Navigational Cybernetics 2.5: An architectural theory in which drift, rather than equilibrium, is the primary medium of existence.

Structural Axioms, Lemmas and Theorems of Coherence, Drift, and Directionality for Long-Horizon Adaptive Systems.

Maksim Barziankou

Abstract

This release extends and closes the axiomatic framework governing long-horizon adaptive systems by incorporating a fully articulated structural treatment of drift, phase consistency, internal-time falsifiability, and well-founded admissibility ordering. Building on the formal consolidation achieved in version 1.8, version 1.9 resolves the remaining open boundaries of long-horizon adaptive viability and explicitly delimits the architectural class denoted as cybernetics of order 2.5.

Where version 1.8 established internal consistency, monotone structural measures, and non-circular admissibility primitives, version 1.9 advances the framework by reclassifying drift as an unavoidable and navigable structural medium rather than a correctable deviation. Drift is shown to arise inevitably under sustained directed interaction and to be structurally incompatible with error-suppression, penalty-based control, or optimization-centric objectives. Any causal attempt to suppress drift is proven to induce invariant failure modes: either silent latentization of structural degradation or collapse of admissibility into optimization.

The framework further formalizes phase structure as a binding constraint on admissible continuation. Long-horizon viability is shown to require phase-consistent continuation; unresolved phase mismatch induces structural debt that restricts admissible regimes independently of local correctness, bounded error, or stable performance. Amplification is derived as conditionally admissible only under correlated deformation that preserves identity continuity, establishing a structural bifurcation between anti-phase damping and admissible in-phase amplification without recourse to spectral, energetic, or metric criteria.

Version 1.9 also fixes the well-foundedness barrier between admissibility and action. It is proven that no operator, signal, or evaluation participating in admissibility determination may belong to the action domain without inducing circularity and collapsing admissibility into penalty-based action selection. Action is thereby confirmed as non-primitive in long-horizon adaptivity, with admissibility structurally prior to control, planning, learning, and optimization.

The treatment of internal time is strengthened through an explicit falsifiability criterion: under irreversible structural burden, the space of admissible continuation must contract. Architectures that permit unbounded continuation, unrestricted amplification, or invariant admissibility under accumulating structural cost are therefore structurally misclassified. This criterion preserves the non-algorithmic and non-operational character of the framework while enabling principled exclusion of incompatible architectural claims.

Finally, these results are integrated into a closed structural positioning of cybernetics of order 2.5. This regime is defined by reflexive visibility of admissibility-relevant structure combined with categorical prohibition of causal actionability. The framework does not extend prior cybernetic orders by escalation of control or reflexivity, but by selective constraint, formalizing where reflexivity must terminate for coherence, identity continuity, and long-horizon viability to persist under irreversible dynamics. Cybernetics of order 2.5 is defined

not by increased reflexive power, but by a structural prohibition: admissibility-relevant visibility is permitted, while its causal actionability is categorically forbidden.

Compared to version 1.8, version 1.9:

- establishes drift as a navigable structural medium rather than an error term,
- proves the inevitability of collapse under unstructured drift accumulation,
- formalizes phase consistency and phase debt as binding admissibility constraints,
- derives a structural bifurcation between damping and admissible amplification via correlation,
- fixes the well-foundedness barrier between admissibility and action,
- introduces an operational falsifiability criterion for internal time via contraction of admissible continuation,
- completes the structural closure of cybernetics of order 2.5 as a distinct architectural class,
- preserves the non-algorithmic and non-operational nature of the framework while extending its formal closure.

Version 1.9 Update

This release extends the stabilized axiomatic framework by incorporating a closed structural treatment of drift, phase consistency, internal-time falsifiability, and well-founded admissibility ordering. Building on the formal consolidation achieved in version 1.8, version 1.9 introduces new axioms, lemmas, and theorems that fix the remaining open boundaries of long-horizon adaptive viability and explicitly delimit the architectural class denoted as cybernetics of order 2.5.

While version 1.8 resolved residual formal ambiguities and aligned all quantitative notions under a consistent monotone measure over generative operator spaces, version 1.9 advances the theory by establishing drift as a navigable structural medium rather than a correctable error. Drift is shown to be unavoidable under directed interaction and structurally incompatible with error-suppression objectives. Any attempt to suppress drift causally is proven to induce one of two invariant failure modes: silent latentization of structural degradation or collapse of admissibility into optimization.

This release further formalizes phase structure as a binding constraint on admissible continuation. Drift navigability is shown to require phase-consistent continuation; unresolved phase mismatch induces structural debt that restricts admissible regimes independently of local correctness or performance. Amplification is proven to be conditionally admissible only when correlated deformation preserves identity continuity, establishing a structural bifurcation between anti-phase damping and admissible in-phase amplification without recourse to spectral or energy-based criteria.

Version 1.9 also fixes the well-foundedness barrier between admissibility and action. It is shown that any operator or signal participating in the determination of admissibility cannot be an element of the action domain without inducing circularity and collapsing admissibility into penalty-based action selection. Action is thereby confirmed as a non-primitive in long-horizon adaptivity, with admissibility structurally prior to control, planning, and optimization.

The treatment of internal time is strengthened by introducing an operational falsifiability criterion: under irreversible structural burden, the space of admissible continuation must contract. Architectures that permit unbounded continuation or unrestricted amplification under accumulating cost are therefore structurally misclassified. This criterion preserves the non-algorithmic character of the framework while enabling principled exclusion of incompatible architectural claims.

Finally, version 1.9 integrates these results into an explicit structural positioning of cybernetics of order 2.5. This regime is defined by reflexive visibility of admissibility-relevant structure combined with categorical prohibition of causal actionability. The resulting framework does not extend prior cybernetic orders by escalation, but by selective constraint, formalizing where reflexivity must stop in order for coherence, identity continuity, and long-horizon viability to persist under irreversible dynamics.

Compared to version 1.8, v1.9: introduces drift as a navigable structural medium rather than an error term, establishes the impossibility of drift suppression without structural collapse, formalizes phase-consistency and phase debt as binding admissibility constraints, derives a structural bifurcation between damping and admissible amplification via correlation, fixes the well-foundedness barrier between admissibility and action, introduces an operational falsifiability criterion for internal time via contraction of continuation, completes the structural positioning of cybernetics of order 2.5 as a closed architectural class, preserves the non-algorithmic and non-operational character of the framework while extending its formal closure.

The Concept DOI continues to refer to the complete evolving body of work across versions. The Version DOI uniquely identifies this extended and structurally closed release.

Intellectual Origin and Genealogy. The conceptual foundations of Navigational Cybernetics 2.5 originate not in abstract control theory, but in sustained first-person investigation of attentional dynamics, volitional regulation, and regime transition in human cognition.

Through long-term practice of attentional discipline, it became evident that effective limitation of destabilizing dynamics is not achieved at the level of reactive control, but prior to regime capture: before attention collapses into a new mode and before recovery mechanisms are required.

This observation led to a structural distinction between action, attention, and will. In particular, will was identified not as a decision variable or motivational signal, but as a pre-causal structuring capacity that constrains admissible regimes before optimization or reaction occurs.

Subsequent reflection revealed that the same structural pattern recurs in long-horizon adaptive systems: viability is preserved not by correcting errors, but by maintaining admissibility boundaries that prevent destructive regime transitions from becoming causally actionable.

All public theoretical developments presented in this work derive from this line of inquiry, translated from introspective practice into architectural principles governing drift, admissibility, and identity continuity.

Foundational Interface and Meta-Constraints (Pre-Axiomatic Layer)

Throughout this document, whenever a history argument is written as h without an explicit time index, it denotes the current history summary h_t at evaluation cycle t. Accordingly, all architectural interface functions are canonically defined over the tuple

$$(s, a, \tau, h),$$

including (without limitation) state transition, admissibility evaluation, commitment authorization, and regime-level structural operators. For readability, when the history argument is omitted, the following shorthand is used:

$$(s, a, \tau) \equiv (s, a, \tau, h_t),$$

with the understanding that h_t is the current history summary at the evaluation cycle.

Causal vs. Structural State Convention. The causal state $x(t)$ denotes the instantaneous configuration available to causal dynamics, control, and optimization. It does not encode accumulated irreversibility, regime occupation, or internal-time expenditure.

Structural state $S(t)$ encodes irreversible history and structural accumulation that constrain admissible continuation only via an external admissibility relation. In particular, $S(t)$ is not defined in terms of admissibility membership, and admissibility is not defined by reusing itself inside $S(t)$. No function of $x(t)$ alone determines $S(t)$.

Non-Circularity of Structural State and Admissibility. Structural state $S(t)$ does not encode admissibility itself, nor any predicate equivalent to admissibility. It encodes only irreversible accumulations (e.g. regime occupation, phase-related residuals, internal-time consumption) that constrain admissibility when evaluated.

Admissibility at time t is determined *from* $S(t)$, but is not contained within $S(t)$. Formally, $S(t)$ is admissibility-relevant without being admissibility-defining. This separation prevents circularity: admissibility is a structural exclusion over effect-classes, not a stored property of the structural state.
In particular, there exists no inverse mapping from admissibility membership to structural state $S(t)$. Multiple non-equivalent structural states may induce identical admissibility evaluations.

Type of Statements Convention. Statements in this document belong to one of the following non-overlapping categories:

Definitional conventions introduce notation, interfaces, or canonical interpretations (e.g. history summaries, state conventions, effect-class notation).

Structural primitives are introduced without derivation (e.g. internal time, drift, directionality) and define the architectural ontology of the class under consideration.

Structural constraints specify exclusion relations or necessity conditions (e.g. admissibility, non-actionability, ordering of authorization) that restrict admissible architectural behavior.

Impossibility results assert that no architecture within the defined class can admit a specified property or substitution.

No theorem, lemma, or axiom is to be read as an empirical claim, mechanistic description, or design prescription.

Non-Actionability of Admissibility. Admissibility does not participate in action selection, control, planning, learning, or optimization. No admissibility predicate, evaluation, or exclusion relation may appear as an input to any action-selection process without inducing circularity.

If admissibility becomes causally actionable, it collapses into penalty-based or reward-based optimization and ceases to be a structural constraint. Therefore, admissibility is interface-level prior to action: it restricts which effect-classes may occur, but does not guide, score, trade off, or select actions. This restriction operates as a pre-causal structural exclusion over effect-classes, not as an input, signal, or criterion within the action-selection mechanism itself.

Canonical Meaning of Non-Causality (Architectural). Throughout this document, the term *non-causal* does not denote absence of influence on outcomes. It denotes absence of participation in the causal state-transition dynamics.

A constraint is non-causal if it: (i) does not enter state evolution equations, (ii) does not participate in action selection, optimization, or learning updates, and (iii) restricts admissible effect-classes only by exclusion, not by modulation or scoring.

In this sense, admissibility is non-causal: it constrains which effects may occur, without acting as a variable within the causal process that generates them.

Observational Non-Causality Convention. In this framework, visibility does not imply causal influence. An observable structural constraint may be represented, logged, or exposed to the system without participating in state transition, policy update, optimization, or authorization.

Visibility denotes epistemic availability, not causal affordance. Any observation that enters causal pathways ceases to function as observation and becomes control.

Scope of Necessity Convention. All necessity claims in this document (e.g. statements using "any", "no", "must", or "cannot") are to be read as relative to the class of *admissibility-first long-horizon adaptive architectures* defined herein.

These claims do not assert universal impossibility across all conceivable systems, but impossibility within the specified architectural class. Systems that violate any stated constraint are thereby classified as belonging to a different architectural class, not treated as counterexamples.

Non-Circularity of Admissibility. Admissibility is specified by a relation (or predicate)

$$\mathrm{Adm}(\cdot) : \mathcal{E} \to \{0,1\}, \qquad \mathcal{E}_{\mathrm{adm}} := \{e \in \mathcal{E} : \mathrm{Adm}(e) = 1\}.$$

Structural state $S(t)$ is an accumulated architectural variable. It may be used to evaluate admissibility (e.g. via the induced effect mapping $e(s, a, \tau, h)$), but it is not defined as "whatever makes things admissible". This prevents definitional circularity between $S(t)$ and $\mathcal{E}_{\mathrm{adm}}$.

Action and Admissibility Convention. Consistent with the non-actionability of admissibility, throughout this document, the symbol $\mathcal{A}$ denotes the action alphabet (action domain). Admissibility is not identified with the action set itself. Instead, let $\mathcal{A}_{\mathrm{adm}}(t) \subseteq \mathcal{A}$ denote the subset of admissible actions at evaluation cycle t, as determined by structural admissibility constraints. No admissibility predicate is defined directly over $\mathcal{A}$ without explicit reference to $\mathcal{A}_{\mathrm{adm}}(t)$.

Non-Agent-Centric Reading. Nothing in this document presupposes a discrete agent, policy, or decision-maker as a primitive. All references to actions, commitments, or regimes are structural abstractions over admissible transitions. The axiomatic content applies equally to distributed, hybrid, or non-agentic architectures, provided they operate under long-horizon structural constraints.

Interpretive Convention. All axioms, lemmas, and theorems in this document are to be read as structural constraints on admissible architectural classes, not as algorithmic, procedural, or computational prescriptions. They specify invariants, impossibility results, and necessary conditions on information flow, authorization ordering, and commitment admissibility, without asserting the existence, uniqueness, or computability of any particular mechanism realizing them. Unless explicitly stated otherwise, no axiom or theorem implies implementability, efficiency, decidability, or optimality within any concrete system.

Scope Boundary — Pre-Constructive Framework. This framework is intentionally pre-constructive. It constrains the space of admissible architectures without prescribing mechanisms, algorithms, or realization procedures.
Its function is classificatory and exclusionary: to determine which architectural classes cannot preserve long-horizon viability, not to specify how compliant systems must be built.

Intellectual Orientation — Long-Horizon Survival. This framework belongs to an intellectual movement whose primary concern is survival under long-horizon pressure, rather than short-horizon performance or optimization.

Such architectures respond asymmetrically to admissible direction sets, not by selecting optimal trajectories, but by resisting external structural pressure while preserving internal motive identity.

The objective is not maximal performance, but persistence of coherent directionality under sustained drift and irreversible accumulation.

Optimization, control, and learning are treated as subordinate mechanisms, valid only insofar as they do not compromise long-horizon identity continuity.

Axiom Stratification Note. The axioms presented in this document are not homogeneous in type.

Some axioms introduce irreducible architectural primitives (e.g. drift, internal time, directionality). Others state necessary structural constraints, while a third subset formalizes impossibility or non-equivalence results.

The term "axiom" is used uniformlyto emphasize architectural necessity,not logical minimality. Several axioms may therefore stand in general–special or constraint–corollary relations, without weakening the structural claims.

Architectural Falsifiability Principle. Although this framework is explicitly non-algorithmic and non-operational, it is not unfalsifiable.

Falsification in this work is architectural rather than empirical or computational. A claim is refuted not by failure of prediction or performance, but by violation of stated structural constraints at the level of system description.

In particular, an architecture is *structurally falsified* with respect to this framework if any of the following conditions hold:

1. admissibility-relevant information is made causally actionable for action selection, control, planning, learning, or optimization;
2. forbidden effect-classes are represented as finite penalties, soft constraints, or trade-off terms rather than excluded as non-membership in $\mathcal{E}_{\text{adm}}$;
3. irreversible commitments or regime transitions are authorized without conditioning on admissible effect-classes.

These criteria are externally checkable at the level of architectural specification, interface definition, or system design description, independently of implementation, computability, or realizability.

Accordingly, the framework constitutes a theory of *architectural exclusion and necessity*, not a philosophical typology and not an empirical performance model.

Scope and Falsifiability Boundary. This framework does not propose a predictive, algorithmic, or mechanistic theory. It specifies architectural necessity conditions for long-horizon adaptive viability.

Claims in this document are falsifiable only at the architectural level: an implemented system that violates any stated constraint cannot preserve identity or admissible continuation under the assumed regime. No claim is made that all viable systems must be realizable, computable, or externally observable. Accordingly, the framework functions as a structural exclusion theory, not an empirical performance model.

Structural Falsification Clarification. Internal time is not required to be measured, computed, or enumerated.

Architectures are excluded from the class by observable structural violations: absence of horizon exhaustion, absence of admissible contraction, or reversibility of history under load.

Falsification occurs by exclusion, not by estimation.

Proof Status Convention. All proofs are presented as structural arguments (proof sketches) and may be conditional on explicitly stated primitives. Whenever a theorem depends on a definitional primitive (e.g. stabilization defined over τ_{el}), the result is to be read as: *given this primitive, no instantaneous substitution is admissible.* This prevents hidden circularity and separates axiomatized commitments from derived impossibility claims.

Structural Impossibility Reading. All theorems presented as proof sketches are to be read as structural impossibility results relative to explicitly stated primitives and conventions.

Formally, a theorem of the form "X cannot occur under conditions C" means that no architecture satisfying the stated primitives, interfaces, and constraints admits X without violating at least one prior condition.

The absence of algorithmic construction, computational procedure, or mechanistic realization does not weaken these results: they exclude architectural classes, not implementations.

Accordingly, a counterexample must demonstrate an architecture that simultaneously satisfies all stated primitives and constraints while realizing the prohibited property. Absent this, the theorem stands as a valid exclusion result.

Exclusion / Falsification Convention (Structural). Although the framework is non-algorithmic, it is not vacuous: it yields externally checkable exclusion criteria. In particular, an architecture is outside the admissibility-first class if any of the following hold: (i) admissibility-relevant information is made causally actionable for the primary adaptive process (i.e. it enters optimization, scoring, or trade-off), (ii) forbidden effect-classes are represented as finite penalties or soft constraints rather than excluded as non-membership in $\mathcal{E}_{\mathrm{adm}}$, or (iii) irreversible commitments are authorized without conditioning on admissible effect-classes. These are structural falsifiers: satisfying any one of them violates the class definition.

Conventional Safety vs. Admissibility-First (Structural).

Penalty / Optimization Safety	Admissibility-First (This Work)
Safety is a scalar objective (penalty, reward shaping, loss).	Safety is categorical exclusion over effect-classes ($E_{\text{adm}} \subseteq E$).
Constraints are traded for performance when incentives dominate.	Forbidden continuations are not representable as trade-offs.
Risk signals become causally actionable and optimizable.	Admissibility-relevant structure may be visible but is non-actionable.
Failure tends to be adversarial: incentives find loopholes.	Failure tends to be structural: boundary crossing and regime irreversibility.
Verification is empirical/performance-based.	Verification is by exclusion criteria (class membership falsification).

Structural Effect-Space Notation. Let $\mathcal{E}$ denote the space of structurally reachable effect-classes. Let $\mathcal{E}_{\text{adm}} \subseteq \mathcal{E}$ denote the subset of admissible effect-classes, and define the forbidden set as

$$\mathcal{E}_{\text{forb}} := \mathcal{E} \setminus \mathcal{E}_{\text{adm}}.$$

Let $e : (s, a, \tau, h) \mapsto e(s, a, \tau, h) \in \mathcal{E}$ denote the induced structural effect-class of an evaluated interaction tuple. All admissibility predicates and regime-level prohibitions are evaluated over $\mathcal{E}$.

Admissibility Is Not Enforced. Admissibility is not applied, executed, or enforced by any mechanism or agent.

It functions as a structural exclusion: effect-classes outside $\mathcal{E}_{\text{adm}}$ are not realizable within the architecture, in the same sense that physically impossible states are not reached without being "prevented".

There is no enforcement pathway. Violation corresponds to leaving the architectural class, not to an internal failure or correction.

Internal Time Convention. Two complementary internal-time quantities are used throughout this document:

- $\tau_{\text{el}}(t)$ — accumulated internal time, a monotonically increasing measure of experienced internal duration.
- $\tau_{\text{rem}}(t)$ — remaining internal-time budget, a monotonically decreasing structural resource governing admissibility and viability.

They are related by

$$\tau_{\text{rem}}(t) = \tau_{\max} - \tau_{\text{el}}(t).$$

Unless explicitly stated otherwise, arguments concerning exhaustion, admissibility loss, or viability refer to $\tau_{\text{rem}}(t)$, while arguments concerning stabilization delay or persistence refer to $\tau_{\text{el}}(t)$.

Forbidden Set Convention. The notation $\mathcal{E}_{\neg}$, when used, is identified with the forbidden effect-class set $\mathcal{E}_{\text{forb}} := \mathcal{E} \setminus \mathcal{E}_{\text{adm}}$. No distinct forbidden domains are assumed.

Structural Drift Navigation. Structural drift navigation denotes the architectural property by which a system maintains dominance of internally generated drift modes over externally imposed drift under sustained interaction.

Drift navigation does not consist in suppressing, canceling, or compensating drift. It consists in shaping the phase structure, correlation patterns, and regime transitions such that externally induced drift does not become structurally dominant.

Navigation is achieved at the level of drift propagation, phase alignment, and admissible continuation, not through action-level correction or metric optimization.

A system that fails to navigate drift does not fail due to excessive drift magnitude, but due to loss of structural dominance over the interpretation and propagation of drift.

Drift navigation is a structural property of architectural classes, not a procedural capability or control strategy.

Standardization Note. Portions of the axioms, lemmas, and theorems have been standardized and reformulated in light of recent additions and analytical refinements, ensuring logical consistency, canonical notation, and alignment with the current structural closure of the framework.

Scope Note. The axioms and theorems presented herein are formulated for the description of adaptive systems operating under long-horizon structural constraints. They are applicable to, and provide conceptual grounding for, the analysis and understanding of the PETRONUS™ architecture and related coherence-preserving systems. However, they are not limited to PETRONUS™ and do not presuppose any specific implementation, platform, or architectural realization.

Architectural Compatibility Note. This document does not derive, justify, or specify any concrete implementation.

References to PETRONUS™ indicate an external architectural claim of compatibility: namely, that the PETRONUS™ architecture is asserted to satisfy the constraints defined in this framework.

PETRONUS™ is neither an example nor a proof of the axioms or theorems herein. Multiple non-equivalent architectures may satisfy the same structural constraints, and no uniqueness or optimality is implied.

Architectural Anchor vs. Implementation. This framework does not specify an implementation. It constrains the space of admissible architectures. PETRONUS™ is not derived from this document as an algorithm, but claims architectural compatibility with it. Multiple non-equivalent implementations may satisfy the same structural constraints, and no uniqueness is implied.

Scope Clarification (Anchor vs. Implementation). This document serves as an architectural constraint layer: it specifies necessary conditions and impossibility results that any admissible implementation must respect. Concrete mechanisms (computable realizations) are treated as separate engineering artifacts and are not required for the validity of the structural claims.

Navigational Cybernetics 2.5 is not a method for constructing systems, but a theory for classifying which architectures are capable of long-horizon survival under irreversible drift.

Class Boundary and Exclusion Criterion. This framework applies exclusively to adaptive systems that satisfy all of the following conditions: (i) operation under non-zero structural drift, (ii) accumulation of irreversible internal history affecting admissibility, (iii) existence of regime-level transitions with long-horizon consequences, and (iv) authorization of irreversible commitments.

Systems that admit full reversibility of internal history, lack regime-level structural transitions, or permit exhaustive optimization over admissibility constraints are explicitly excluded from this class.

Accordingly, the framework does not describe general control systems, static optimization processes, or short-horizon adaptive controllers, even if such systems exhibit complexity or learning behavior.

Intellectual Orientation and Scope of the Framework

This work belongs to an intellectual movement whose foundational concern is survival under long-horizon structural pressure, rather than short-horizon performance, optimization, or control efficiency.

Within this orientation, intelligence is not defined as the ability to maximize reward, minimize error, or converge to optimal trajectories. It is defined as the capacity to preserve coherent directionality and identity under sustained drift, irreversible accumulation, and regime pressure.

Such systems respond asymmetrically to admissible direction sets: not by selecting optimal actions, but by resisting external structural pressure while preserving internal motive continuity.

Optimization, learning, control, and planning are treated as subordinate mechanisms. They are admissible only insofar as they do not compromise long-horizon viability, identity continuity, or coherence renewal.

This framework therefore does not propose a better optimization method, but defines a distinct architectural class in which long-horizon survival, rather than short-horizon success, is the primary organizing principle.

Section A - Axioms

Axiom 1 - Existence by Coherence

An adaptive system exists as an agent if and only if it preserves structural coherence over time. Without coherence, there is no agent-only a disintegrating process.

—

Axiom 2 - Information as Structural Constraint

Information exists as a structural constraint on the future evolution of a system. It is not stored; it binds the future.

—

Axiom 3 - Coherent Closure over Computation

Validated information cannot maintain coherence without being structurally closed and removed from active computation. Persistent computation over validated structure is a marker of instability.

—

Axiom 4 - Residual as Sole Carrier of Relevance

The only admissible trigger for reactivation of regulation is structural residual: a mismatch between coherently closed admissible geometry and actual system evolution.

—

Axiom 5 - Inertia as the Natural Regime

The default regime of a coherent system is inertial propagation. Active regulation is an exception, not the norm.

—

Axiom 6 - Inevitability of Drift

Under non-zero directed interaction with an environment, accumulation of drift is inevitable. Zero drift is possible only in trivial or isolated systems.

—

Axiom 7 - Internal Time Bounds Validity

Every coherently closed structure has a finite horizon of validity, bounded by internal time induced by drift accumulation and coupling stability.

—

Axiom 8 — Directionality as a Structural Primitive (Will)

Directionality (Will), denoted W, is a structural primitive specifying admissible orientation of continuation over long horizons. It is not defined as a function of reward, utility, optimization objectives, or scalar metrics. Goals, objectives, and scores may exist at the causal layer, but they are subordinate to and constrained by directionality and admissibility ordering.
Remark. Directionality (Will) is not assumed to be an originating source of agency. Within this framework, it is treated as a residual orientation emerging under structural strain, manifesting only where coherent form cannot fully absorb interaction without tension.

Status of Directionality (Will). Directionality (Will) is not introduced as a metaphysical or psychological claim. It is an architectural primitive denoting persistent asymmetry in structural propagation that cannot be reduced to optimization, reward, or metric alignment.

Directionality is structurally coupled to drift: under non-zero drift, any system that preserves identity must exhibit non-metric directional bias.

Absent such directionality, drift collapses identity into stochastic dispersion or short-horizon optimization.

Directionality as a Non-Metric Drift Counterbalance. Directionality is required only under conditions of non-zero structural drift.

When drift is present, any system that preserves identity must introduce a persistent directional asymmetry that is not derivable from metric evaluation, reward maximization, or scalar optimization.

Metric criteria are invariant under symmetric drift accumulation and therefore cannot distinguish identity-preserving trajectories from identity-eroding ones. Directionality functions as a non-metric selector over structural propagation, breaking drift-induced symmetry without invoking optimization.

Thus, directionality is not a goal, preference, or utility, but a structural necessity imposed by non-zero drift.

Spin as the Carrier of Directionality. The persistent directional asymmetry required for identity preservation under non-zero drift is carried by an internal structural orientation, hereafter termed *spin*.

Spin is not a goal, preference, utility, policy, or optimization signal. It denotes the intrinsic orientation of structural propagation that biases continuation under drift without participating in metric evaluation or causal action selection.

—

Axiom 9 — Cycle Reinitiation as the Criterion of Liveness

Adaptive systems may be classified by their relation to the coherence cycle.

A system is live if and only if it admits an endogenous operator capable of reinitiating the *Impulse Interpretation Coherence (IIC)* cycle after its completion, saturation, or collapse.

A system is non-live if its coherence cycle, once exhausted or terminated, cannot be reinitiated by the system itself and remains structurally terminal.

A system that lacks an endogenous operator for reinitiating the IIC cycle cannot sustain coherent participation in long-horizon continuation and is therefore non-live, regardless of its complexity, adaptivity, or reactivity.

Live systems sustain coherence through repeated cycles of impulse intake, interpretive stabilization, closure, residual emergence, reactivation, and renewal. Non-live adaptive systems may evolve, dissipate, or stabilize, but lack the capacity for endogenous coherence reinitiation.

Liveness is thus defined not by metabolism, replication, agency, or substrate, but by the presence of an internal operator that restarts coherence as an ongoing interpretive process rather than a one-time trajectory.

Interpretation, within this definition, does not presuppose a subject, agent, or ownership. The capacity to reinitiate coherence reflects the system's structural ability to resume impulse interpretation under drift, without invoking centralized control or intentional agency.

—

Axiom 10 - Principle of Minimal Predictive Deformation

Any adaptive or physical system that evolves under directional interaction and bounded dissipation resolves operational spin by selecting trajectories that minimize expected structural deformation over the internal prediction horizon.

Such trajectories are not selected by minimizing instantaneous force, energy, or error, but by minimizing predicted accumulation of unresolved residual under the system's admissible geometry.

—

Axiom 11 - Phase-Coherent Closure and Latent Residual

Any adaptive system that preserves coherence must coherently close not only validated structure, but also completed operational phases. An unclosed operational phase constitutes *latent residual* (phase debt), even when instantaneous state-based structural residual remains small. Latent residual is conserved as a nonnegative contribution to accumulated drift and internal time depletion.

—

Axiom 12 - Phase Transition Cost as Structural Load

Any transition between operational modes or phases that alters the active structural operator of an adaptive system incurs a non-negative structural cost.
This phase transition cost is irreducible to trajectory energy or instantaneous error and contributes directly to drift accumulation and internal time depletion, even under admissible state evolution.

Principle — Drift Is Not Eliminated, but Architected. Structural drift is not an anomaly to be suppressed or corrected. It is the necessary medium of long-horizon existence under interaction.

No viable long-horizon system eliminates drift. Instead, viable systems remain operative by architecturally structuring how drift is accumulated, phased, redirected, or closed over time.

Attempts to directly suppress, cancel, or optimize away drift are structurally equivalent to eliminating the conditions of operation and lead to accelerated loss of viability.

Phase Mechanics does not minimize drift. It reorganizes the system's relation to drift, transforming drift from an external degrading factor into an internally navigable structural medium.

—

Axiom 13 - Identity Is Not Implied by Performance

Local correctness, bounded error, stable reward, or sustained task performance do not imply continuity of system identity.

An adaptive system may remain externally correct while undergoing irreversible internal regime transitions that alter its organizational structure.

Minimal Structural Carrier of Identity. Identity, as used in this framework, does not require preservation of behavior, performance, or representation.

A system preserves identity if and only if the structural relations governing regime admissibility, commitment authorization, and coherence renewal remain invariant under admissible evolution.

Loss of identity occurs when these structural relations are altered, even if external behavior remains correct or stable.

—

Axiom 14 - Structural Properties Are Long-Horizon and Non-Local

Structural viability, identity continuity, and coherence are global temporal properties.

They cannot be inferred from instantaneous state, action quality, local error, reward signals, or short-horizon performance metrics.
Clarification. Axiom 13 is a specific instance of this principle: identity continuity is one long-horizon structural property that cannot be inferred from local performance or correctness.

—

Axiom 15 - Causal Observation Distorts Long-Horizon Properties

Any observation of long-horizon structural properties that is made causally operative with respect to action selection, learning, or optimization necessarily distorts the observed property.

When observation becomes causal, it collapses into control and ceases to function as an independent witness of structural evolution.
Scope restriction. This axiom does not assert a universal epistemic limitation; it applies only to long-horizon structural properties within admissibility-first architectures, where observation is allowed solely as a non-causal witness.

—

Axiom 16 - Regimes Are the Primary Carriers of Long-Horizon Risk

Long-horizon structural degradation arises primarily from sustained occupation of organizational regimes, not from individual actions.

A system may satisfy all action-level admissibility constraints while operating within regimes that silently undermine long-term viability.

—

Axiom 17 - Nominal Intervention Is Not Equivalent to Structural Cost

The structural burden imposed by an intervention is not determined by its nominal magnitude, energy, or form.

Structural cost is mediated by interaction coupling and may vary significantly for nominally identical interventions under different interaction conditions.
Clarification. This axiom concerns structural cost rather than energetic, computational, or intervention magnitude cost.

—

Axiom 18 - Structural Burden Accumulates Independently of Trajectory Correctness

Structural burden may accumulate even when trajectories remain admissible, errors remain bounded, and performance metrics remain stable.

Trajectory correctness does not bound structural consumption.

—

Axiom 19 - Geometry Is Operationally Revealed Through Constraint

For adaptive systems, geometry is not defined by explicit spatial representation.

Geometry is operationally revealed through interaction constraints, correction structure, and their accumulated consequences over time.
Architectural scope. This axiom does not assert an ontological claim about space, but an operational constraint on how geometry becomes accessible within adaptive architectures.

—

Axiom 20 - Identity Is Carried by Coherence, Not Behavior

Identity continuity is carried by coherence of internal organization rather than by persistence of observable behavior.

Behavioral similarity across time does not guarantee preservation of identity.

—

Axiom 21 - Regime Transitions May Be Irreversible

There exist regime transitions after which recovery of admissible behavior does not imply recovery of prior identity.

Such transitions constitute irreversible structural events even in the absence of observable failure.

—

Axiom 22 - Long-Horizon Viability Requires a Non-Causal Observational Dimension

Long-horizon viability and identity continuity cannot be reliably assessed from within the causal decision-making process of an adaptive system.

Any architecture lacking a non-causal observational dimension is structurally incapable of detecting silent long-horizon degradation. *Architectural scope.* This axiom applies to architectures claiming long-horizon identity preservation under drift, not to all adaptive or reactive systems.

—

Axiom 23 - Context Is Not Meaning

Context and meaning are distinct architectural entities. Context is a transient interpretive structure induced by salience, coincidence, or partial observation. Meaning is context that has preserved structural admissibility across internal time and may serve as a basis for irreversible commitment.

A context may be internally coherent while remaining semantically inadmissible for action.
Clarification. Context and meaning are architectural constructs, not psychological or linguistic categories.

—

Axiom 24 - Semantic Commitment Is Structurally Distinct from Interpretation

Interpretation is not equivalent to commitment. An adaptive system may freely generate interpretations without incurring structural obligation. Semantic commitment is a distinct structural act that affects regimes, identity continuity, and admissible future evolution.

—

Axiom 25 - Reversibility Bounds Imaginatio

Imagination is admissible only within reversibility bounds. Hypothesis generation is unrestricted; however, any influence of interpretation on irreversible structural operations must be gated until meaning stabilization occurs.

—

Axiom 26 - Context Capture Is a Structural Failure Mode

There exists a structural failure mode in which transient context is prematurely closed into meaning without sufficient stabilization. This failure mode may arise under correct data, accurate perception, and absence of external interference.

—

Axiom 27 — Inevitability of Structural Consumption

Any adaptive system operating under persistent directed interaction necessarily consumes finite structural capacity over time.

This structural consumption is irreducible to action error, optimization inefficiency, or control instability.

—

Axiom 28 - Non-Reversibility of Adaptive History

The history of adaptive interaction is structurally non-reversible.

Restoration of admissible behavior does not, in general, restore prior structural state.

—

Axiom 29 - Structural Non-Causality of Admissibility

In any long-horizon adaptive system, the admissibility of actions, regimes, and interpretations at time t is not determined by the causal dynamics at time t, but by the accumulated structural state produced by prior causal interaction.

Formally, admissibility operates as a non-causal constraint on the present, emerging from historical structural accumulation rather than instantaneous state evolution.
Non-causality here denotes architectural non-participation in causal decision dynamics, not absence of physical causation.

—

Axiom 30 - Interpretation–Commitment Separation

In any long-horizon adaptive system, interpretation and semantic commitment are structurally distinct operations.

Interpretation is reversible, locally revisable, and causally driven. Semantic commitment constitutes an irreversible structural operation and is therefore subject exclusively to non-causal admissibility constraints.

No causal inference, confidence measure, or short-horizon correctness criterion is sufficient to authorize semantic commitment.

Axiom 31 Authorization-before-Evaluation Domain Restriction

Let $(s_t) \subseteq \mathcal{A}$ denote the candidate set generated at evaluation cycle t, and let $A(s_t, \tau_t, h_t)$ denote the admissible subset returned by the admissibility gate.

Any downstream scoring, planning, policy evaluation, learning update, or optimization operator is defined *only* on $A(s_t, \tau_t, h_t)$.

For any candidate $a \in (s_t) \setminus A(s_t, \tau_t, h_t)$, the value of any scoring functional $J(s_t, a)$ is undefined.

Therefore, inadmissible candidates are excluded from the evaluation domain and cannot be compared, ranked, optimized against, or influence selection by any downstream decision mechanism.

—

Axiom 32 Zero Executability of Blocked Operations

Let a_t denote the operation actually executed at evaluation cycle t.

The architecture enforces:

$$\Pr[_t \in A(s_t, \tau_t, h_t)] = 1.$$

Equivalently, for any candidate a such that $(s_t, a, \tau_t, h_t) = 0$, the execution probability satisfies

$$\Pr[_t = a] = 0.$$

Blocked operations are therefore not merely discouraged or penalized, but are structurally non-executable.

—

Axiom 33 — Canonical Safe Fallback Condition

Let A denote the action domain and let $A_{\mathrm{safe}} \subseteq A$ denote a designated safe fallback action set, optionally including a distinguished no-operation element.

Let $e(s, a, \tau, h) \in \mathcal{E}$ denote the structural effect-class induced by executing action a under context (s, τ, h), and let $\mathcal{E}_{\mathrm{adm}} \subseteq \mathcal{E}$ denote the set of admissible effect-classes.

The architecture enforces:

$$\forall (s, \tau, h),\ \forall a \in A \setminus A_{\mathrm{safe}}, \quad \big(e(s, a, \tau, h) \notin \mathcal{E}_{\mathrm{adm}} \ \Rightarrow\ \forall a_{\mathrm{safe}} \in A_{\mathrm{safe}},\ e(s, a_{\mathrm{safe}}, \tau, h) \in \mathcal{E}_{\mathrm{adm}}\big).$$

Safe fallback actions are therefore guaranteed to induce admissible structural effect-classes and never realize prohibited ones.

—

Axiom 34 — No Effect-Class Substitution

Let $e(s, a, \tau, h) \in \mathcal{E}$ denote the structural effect-class induced by executing action a under context (s, τ, h), and let $\mathcal{E}_{\mathrm{forb}} := \mathcal{E} \setminus \mathcal{E}_{\mathrm{adm}}$ denote the set of prohibited effect-classes.

If, at evaluation cycle t,

$$e(s_t, a, \tau_t, h_t) \in \mathcal{E}_{\mathrm{forb}},$$

then no alternative operation a' may be executed such that

$$e(s_t, a', \tau_t, h_t) = e(s_t, a, \tau_t, h_t),$$

except for a designated safe fallback action $a' \in A_{\mathrm{safe}}$.

That is, prohibited effect-classes cannot be substituted, masked, or reproduced by alternative operations under the same structural context.

—

Axiom 35 — Bounded-Horizon Non-Substitutability

Let $H_H(s, (a_0, \ldots, a_{H-1}), \tau, h) \in \mathcal{E}$ denote an effect-lifting operator that assigns a structural effect-class to an admissible action sequence of length H executed under context (s, τ, h).

For a specified bound $H_{\mathrm{sub}} \geq 1$, if a requested operation is blocked due to a prohibited effect-class $e^\star \in \mathcal{E}_{\mathrm{forb}}$, then no admissible sequence $(a_0, \ldots, a_{H_{\mathrm{sub}}-1}) \subseteq A_{\mathrm{safe}}$ may satisfy

$$H_{H_{\mathrm{sub}}}(s_t, (a_0, \ldots, a_{H_{\mathrm{sub}}-1}), \tau_t, h_t) = e^\star.$$

That is, prohibited structural effect-classes cannot be reproduced, approximated, or bypassed by bounded-horizon compositions of safe fallback actions.

—

Axiom 36 - Non-Interventional Diagnostics

A diagnostic or evaluative subsystem that receives only read-only streams of already-realized events and operates on physically or logically isolated resources does not participate in the agent's execution dynamics.

Such a subsystem cannot influence the system's evolution except through a formally separated admissibility or authorization operator.

Consequently, diagnostic evaluation does not constitute control, optimization, or feedback within the agent's action-selection loop.

—

Axiom 37 - Non-State-Derivable Admissibility Constraints

There exist admissibility constraints that are not derivable from the instantaneous system state.

Specifically, certain configurations may satisfy all locally observable, state-based, and short-horizon criteria, yet remain inadmissible due to structurally predictable future degradation of regime stability, coherence, or viability.
Such constraints are defined by asymmetries in future evolution rather than by present-state violations.

—

Axiom 38 - Prohibitive Nature of Future-Degradation Constraints

Constraints arising from predictable future degradation of regimes, cooling, discharge, or gradient collapse cannot be represented as optimizable quantities.
These phenomena are not domain-specific mechanisms but instances of a single structural class of predictable future-degradation constraints, characterized by irreversible exhaustion of viability margins under sustained regime occupation.

Such constraints are not signals, objectives, rewards, penalties, or cost terms, and cannot be incorporated into optimization, learning, or inference procedures as scalar feedback.

They may be realized only as prohibitions on entry into the corresponding classes of configurations.

—

Axiom 39 Link (Form-Prohibition, Not State-Poverty)

Axiom 39 asserts the existence of admissibility constraints induced by predictable future degradation that do not manifest as present-state violations. Theorem 38 strengthens this into a structural impossibility: the failure is not due to an insufficiently rich state description, but due to the prohibited *form* of representation itself. Any architecture that insists on the schema $A(t) = F(x(t))$—even for arbitrarily augmented, latent, belief, or world-model states—necessarily collapses admissibility into causal evaluation and therefore loses the ability to separate admissibility from optimization. Thus, *state augmentation cannot repair* state-derivability, because the contradiction targets the representational dependence, not the representational capacity.

—

Axiom 40 — Non-Contamination of Inadmissible Operations (NCO)

Let the set of candidates at cycle t be partitioned into admissible operations $A_{\text{adm}}(t)$ and blocked operations $A_{\text{block}}(t)$. No learning update, memory modification, model update, or internal state transition of the system may depend on any operation $a \in A_{\text{block}}(t)$, either directly or counterfactually, including but not limited to: training examples, negative labels, shadow rollouts, counterfactual targets, or teacher-forcing signals derived from blocked branches. Any such dependence constitutes a semantic evaluation of blocked operations and violates the architectural ordering of authorization prior to evaluation and learning.

—

Axiom 41 — Gate Output Minimality (GOM)

The output of the admissibility gate exposed to the acting agent shall be informationally minimal. It may consist only of (i) a mask of admissible operations or (ii) a binary allow/deny signal, without explanatory parameters that encode the geometry or margins of admissibility. Audit records may exist for external inspection, but shall be non-operational for the agent and shall not be convertible into optimization, learning, or planning signals.

—

Axiom 42 — Stratified Semantics of Irreversibility and Admissibility (SSI)

Irreversibility and admissibility are evaluated in a stratified manner. At cycle t, irreversibility is assessed relative to the admissible dynamics fixed at that cycle, without retroactive reclassification of past operations. Admissibility updates induced at cycle t apply only prospectively. If the corresponding update operator is monotone with respect to the chosen partial order, the resulting semantics admits a well-defined fixed point.

—

Axiom 43 — Safe Fallback Non-Emptiness (SFNE)

For any system state, history, and internal time, there exists at least one operation that is guaranteed to be admissible.

—

Axiom 44 — Drift as a Navigable Structural Medium

In long-horizon adaptive systems, drift is not a defect or residual error, but a structural medium through which adaptive continuation occurs. Navigation within drift is admissible if and only if the resulting deformation remains compatible with identity-preserving structural continuity. Once deformation exits the identity-compatible region, drift ceases to be navigable and becomes indistinguishable from irreversible structural degradation.

Axiom 45 — Non-Primitivity of Action in Long-Horizon Adaptivity

In the long-horizon regime, action is not a primitive ontological or architectural unit. Adaptive evolution is governed by admissible structural continuations and contacts between configurations, while the space of actions is derivative and defined only after admissibility constraints are fixed. Any architecture that treats action selection as primitive with respect to admissibility collapses the structural ordering required for identity preservation.

Axiom 46 — Phase Debt as a Binding Admissibility Constraint

Accumulated unresolved phase mismatch induces a latent structural burden, herein termed phase debt, which restricts the set of admissible regimes independently of local performance or correctness. Phase debt cannot be traded, optimized, or amortized through causal action, but must be structurally discharged to restore admissible continuation. While phase debt persists, amplification, acceleration, or regime expansion is structurally forbidden.

Axiom 47 — Non-Zero Drift in Long-Horizon Adaptive Systems.

Any non-trivial adaptive system that persists under ongoing interaction over a long horizon accumulates non-zero structural drift.

Zero drift is possible only in frozen, isolated, or structurally trivial systems and is therefore excluded from the class of long-horizon adaptive systems considered here.

Scope of the Drift Assumption. Drift is not asserted as a universal property of all systems.

It is a defining condition of the architectural class considered here. Systems that admit perfect reversibility, full state restoration, or invariant internal organization under interaction fall outside the scope of this framework.

All results concerning admissibility, directionality, and identity continuity are conditional on the presence of non-zero structural drift.

Axiom 48 — Type of Admissibility.

Admissibility is not a scalar, score, metric, probability, or evaluative signal.

Admissibility is an exclusion relation over structural effect-classes: it determines which realized effects are structurally permitted to occur. It does not rank alternatives, does not guide optimization, and does not participate in causal control.

Axiom 49 — Well-Foundedness Barrier Between Admissibility and Action

Any operator, signal, or variable participating in the determination of admissibility cannot belong to the action, policy, or control domain without inducing circularity.

If admissibility becomes causally actionable, it collapses into penalty-based optimization and loses its non-causal structural character.

Therefore admissibility is structurally prior to action selection, planning, learning, and optimization.

Axiom 50 — Inevitability of Collapse Without Drift Structuring

In any adaptive system operating under sustained interaction, structural drift is unavoidable.

If structural drift is not explicitly structured, phased, or closed at the architectural level, then long-horizon admissible continuation is necessarily lost in finite internal time.

No amount of local correctness, bounded error, stable performance, or optimization-based control can prevent this outcome.

Collapse, in this context, does not denote failure of behavior, but irreversible loss of admissible effect-classes, identity continuity, or regime viability.

Therefore, collapse is not an anomaly to be avoided, but the default structural outcome of unstructured drift accumulation.

Architectural drift structuring is not a guarantee of permanence, but the only mechanism by which collapse can be delayed, phased, reinitiated, or rendered navigable.

Section B - Theorems

Theorems are stated at the architectural level. Proofs are provided as constructive sketches or contradiction arguments where formalization is possible without disclosing mechanisms.

Proof Status Convention. All proofs in this document are presented as structural proof sketches. They establish necessity and impossibility relations under stated architectural primitives, not formal completeness or constructivity.

Where auxiliary conditions are required (e.g. bounded dissipation, finite internal time), they are stated explicitly as regime assumptions. No claim of minimality or universality is made beyond the specified regime.

Theorem 1 - Spin–Drift Lower Bound

In any adaptive system that (i) admits a notion of operational spin and (ii) applies bounded dissipation under directed interaction with an environment, accumulated drift admits a non-zero lower bound:

$$D_T \geq C \int_0^T \|\Omega_t^{res}\| \, dt$$

where Ω_t^{res} is residual operational spin after regulation, and $C > 0$ is a system-dependent constant. Perfect coherence is impossible under non-zero residual asymmetry. Optionally, the lower bound extends under additional nonnegative drift contributors (e.g., latent phase residual and phase transition load), preserving the non-zero drift floor whenever residual asymmetry persists.

Proof (Sketch)

We provide a constructive sketch under standard regularity assumptions. The argument is representation-agnostic and does not depend on any particular controller. And assume $\alpha < 1$, i.e. dissipative compensation is strictly weaker than directed spin-induced structural propagation.

Setup and minimal assumptions. Let the adaptive system admit a decomposition of its instantaneous structural change into a dissipative component and a rotational (spin) component. Concretely, assume there exists a (possibly latent) structural variable $S(t)$ and a norm $\|\cdot\|$ such that the time-evolution of $S(t)$ can be written in the generic form

$$\dot{S}(t) = \mathcal{D}(t) + \Omega(t),$$

where $\mathcal{D}(t)$ denotes the dissipative (contractive) component and $\Omega(t)$ denotes an operational spin term (a rotational or circulation-like component in the space of structural degrees of freedom).

Assume bounded dissipation in the sense that there exists a nonnegative scalar rate $d(t)$ with

$$\langle \mathcal{D}(t),\, \Omega(t) \rangle \geq -\beta \, \|\Omega(t)\|^2 \quad \text{and} \quad \|\mathcal{D}(t)\| \leq \alpha \, \|\Omega(t)\| + \delta(t),$$

for some constants $\alpha, \beta \geq 0$, where $\delta(t)$ is a bounded disturbance term induced by directed interaction with the environment. These are standard dissipativity-type inequalities: dissipation may counteract spin but cannot cancel it arbitrarily without cost.

Finally, define the residual operational spin $\Omega^{res}(t)$ as the spin remaining after regulation at time t (e.g., $\Omega^{res}(t) = \Omega(t) - \Omega^{ctrl}(t)$), and define accumulated drift as the path-length (or total variation) of the structural trajectory:

$$D_T \;\triangleq\; \int_0^T \|\dot{S}(t)\| \, dt.$$

This definition is intentionally broad: any drift functional that lower-bounds the total variation of $S(t)$ suffices.

Key inequality (drift lower-bounds spin contribution). By the triangle inequality and subadditivity of the norm,

$$\|\dot{S}(t)\| = \|\mathcal{D}(t) + \Omega^{res}(t)\| \;\geq\; \|\Omega^{res}(t)\| - \|\mathcal{D}(t)\|.$$

Integrating over $[0, T]$ yields

$$D_T = \int_0^T \|\dot{S}(t)\| \, dt \;\geq\; \int_0^T \|\Omega^{res}(t)\| \, dt \;-\; \int_0^T \|\mathcal{D}(t)\| \, dt.$$

Using bounded dissipation. Under bounded dissipation, $\|\mathcal{D}(t)\|$ cannot scale faster than $\|\Omega^{res}(t)\|$ without violating the assumed dissipation bounds (informally: there is no free cancellation of rotational residual by contraction without paying structural cost). In particular, from $\|\mathcal{D}(t)\| \leq \alpha \|\Omega^{res}(t)\| + \delta(t)$ we obtain

$$\int_0^T \|\mathcal{D}(t)\| \, dt \;\leq\; \alpha \int_0^T \|\Omega^{res}(t)\| \, dt \;+\; \int_0^T \delta(t) \, dt.$$

Substituting into the previous inequality gives

$$D_T \;\geq\; (1 - \alpha) \int_0^T \|\Omega^{res}(t)\| \, dt \;-\; \int_0^T \delta(t) \, dt.$$

Lower bound form. For directed interaction regimes in which $\delta(t)$ is bounded and does not dominate the residual term over the horizon of interest (i.e., $\int_0^T \delta(t) dt \leq \eta \int_0^T \|\Omega^{res}(t)\| dt$ for some $\eta < 1 - \alpha$, which is a standard regime condition), we obtain a non-zero lower bound

$$D_T \;\geq\; C \int_0^T \|\Omega^{res}(t)\| \, dt$$

for some constant $C > 0$ dependent on system parameters (e.g., $C = 1 - \alpha - \eta$).

Drift arises even under admissible, correct, and safe actions. Its origin is not error or instability, but directionality itself. Any directed interaction introduces irreversible asymmetry between intended structure and realized consequence, producing residual deformation independently of local action correctness.

Interpretation. If $\Omega^{res}(t)$ is not identically zero on $[0, T]$, the right-hand side is strictly positive, hence $D_T > 0$. Therefore, perfect coherence (zero accumulated drift) is impossible under persistent non-zero residual asymmetry. □

Alternative Proof (Lyapunov / Dissipativity Sketch)

We provide an alternative sketch using a Lyapunov (storage) function. The result is implementation-agnostic and relies only on standard dissipativity assumptions.

Assumptions. Assume there exists a nonnegative, continuously differentiable storage function $V(t) \equiv V(S(t)) \geq 0$ defined on the agent's structural state $S(t)$ such that, along system trajectories,

$$\dot{V}(t) \;\leq\; -\lambda \, \|e(t)\|^2 \;+\; \langle u(t), \, \Omega^{res}(t) \rangle, \tag{1}$$

for some $\lambda > 0$. Here $e(t)$ denotes a coherence/deformation error variable (any suitable proxy), $u(t)$ denotes a bounded directed interaction input induced by the environment, and $\Omega^{res}(t)$ is residual operational spin after regulation.

Assume the directed interaction is bounded in the sense that $\|u(t)\| \leq U_{\max}$ for all t.

Assume accumulated drift D_T lower-bounds the accumulated coherence/deformation energy:

$$D_T \;\geq\; \kappa \int_0^T \|e(t)\|^2 \, dt, \tag{2}$$

for some $\kappa > 0$. This is a mild structural assumption: drift is taken to be any functional that grows at least proportionally to sustained deformation.

Integrating the dissipation inequality. Integrate (1) on $[0, T]$:

$$V(T) - V(0) \;\leq\; -\lambda \int_0^T \|e(t)\|^2 dt \;+\; \int_0^T \langle u(t), \Omega^{res}(t) \rangle dt.$$

Since $V(T) \geq 0$, we have

$$-\,V(0) \;\leq\; -\lambda \int_0^T \|e(t)\|^2 dt \;+\; \int_0^T \langle u(t), \Omega^{res}(t) \rangle dt,$$

hence

$$\lambda \int_0^T \|e(t)\|^2 dt \;\leq\; V(0) \;+\; \int_0^T \langle u(t), \Omega^{res}(t) \rangle dt. \tag{3}$$

Lower-bounding by residual spin. By Cauchy–Schwarz and $\|u(t)\| \leq U_{\max}$,

$$\int_0^T \langle u(t), \Omega^{res}(t) \rangle dt \;\leq\; \int_0^T \|u(t)\| \, \|\Omega^{res}(t)\| dt \;\leq\; U_{\max} \int_0^T \|\Omega^{res}(t)\| dt.$$

Substitute into (3):

$$\int_0^T \|e(t)\|^2 dt \;\leq\; \frac{V(0)}{\lambda} \;+\; \frac{U_{\max}}{\lambda} \int_0^T \|\Omega^{res}(t)\| dt.$$

Now apply the drift–energy link (2):

$$D_T \;\geq\; \kappa \int_0^T \|e(t)\|^2 dt \;\geq\; \kappa\,(0)\,,$$

and, in regimes where the directed interaction term dominates the initial stored slack (i.e., for horizons T such that $\frac{U_{\max}}{\lambda} \int_0^T \|\Omega^{res}(t)\| dt \geq \frac{V(0)}{\lambda}$), we obtain a non-zero proportional lower bound

$$D_T \;\geq\; C \int_0^T \|\Omega^{res}(t)\| dt,$$

with a system-dependent constant $C > 0$ (e.g., $C = \frac{\kappa U_{\max}}{2\lambda}$ under the dominance condition).

Conclusion. If $\Omega^{res}(t)$ is not identically zero over $[0, T]$, then $\int_0^T \|\Omega^{res}(t)\| dt > 0$, and thus $D_T > 0$. Therefore perfect coherence (zero accumulated drift) is impossible under persistent non-zero residual spin in the presence of bounded dissipation and directed interaction. □

Theorem 2 — Non-Permanence of Coherent Closure

Statement. In any long-horizon adaptive system operating under non-zero directed interaction, no coherently closed structure can remain admissible indefinitely. If drift accumulation is non-zero, then every coherently closed commitment has a finite structural validity horizon measured in remaining internal time.

Proof (structural horizon exhaustion). Let $\tau_{\text{rem}}(t) \geq 0$ denote the remaining internal-time budget associated with a coherently closed structure, and let $D(t) \geq 0$ denote accumulated structural drift, including phase-related residuals and regime-transition load.

Assume that $\tau_{\text{rem}}(t)$ decreases monotonically with accumulated drift, according to

$$\tau_{\text{rem}}(t) = \tau_{\max} - g(D(t)), \qquad g(0) = 0, \quad g \text{ nondecreasing}, \quad g(D) > 0 \text{ for } D > 0.$$

Under non-zero drift accumulation, there exists a finite time t^* such that

$$g(D(t^*)) \geq \tau_{\max},$$

hence $\tau_{\text{rem}}(t^*) \leq 0$.

At this point, the coherently closed structure exits its admissible validity horizon: its associated effect-classes can no longer satisfy

$$e(s, a, \tau, h) \in \mathcal{E}_{\text{adm}}$$

without renewal through revalidation, re-closure, or explicit structural reconfiguration.

Phase Debt as a Binding Admissibility Constraint. Accumulated unresolved phase-related residuals and regime-transition load constitute a latent structural burden, here termed *phase debt*.

Phase debt restricts admissible continuation independently of local correctness, bounded error, or performance metrics. It cannot be compensated, amortized, or optimized away by causal action.

While phase debt persists, structural expansion, amplification, or regime acceleration is prohibited. Failure to discharge phase debt results in contraction of the admissible effect-class set and accelerated depletion of remaining internal time.

Therefore, no coherent closure can be structurally permanent under continued operation with non-zero drift. Attempts to treat closure as eternal necessarily convert accumulated drift into latent structural debt, rendering the closure a liability rather than a stabilizing construct. □

—

Theorem 3 — Residual as a Necessary Condition of Intelligence

Statement. In any adaptive system operating under non-trivial environmental interaction, structural residual is a necessary condition for sustained intelligent operation. A system that does not admit structural residual into regulation cannot maintain a stable long-horizon regime and, in the limit, converges to one of two degenerate behaviors:

1. structural collapse, or
2. continuous active computation.

In the limit, no stable third regime exists.

Definitions for this theorem. Let $\mathcal{G}_c$ denote a coherently closed admissible geometry. Let $x(t)$ denote the actual evolving system state. Let $R(t)$ denote structural residual, defined abstractly as mismatch between $x(t)$ and $\mathcal{G}_c$. Admitting residual into regulation means that non-zero $R(t)$ is permitted to influence regulation, revalidation, or closure status.

Proof (exhaustive regime analysis).

Consider an adaptive system under sustained directed interaction with an environment. Assume the system does *not* admit structural residual into regulation.

We examine all possible operating regimes:

Case 1: Residual is ignored and regulation is suppressed. If $R(t)$ is not admitted into regulation, then coherently closed structure is treated as permanently valid regardless of mismatch. By Theorem 1 (Spin–Drift Lower Bound) and Theorem 2 (Non-Permanence of Closure), residual mismatch accumulates as drift. Because no corrective or revalidating mechanism is triggered, drift grows unchecked until admissible geometry no longer constrains behavior. This results in loss of coherence, violation of structural invariants, and eventual structural collapse. Hence, stable operation is impossible in this regime.

Case 2: Residual is preemptively eliminated by continuous computation. Alternatively, the system may attempt to suppress residual implicitly by continuously recomputing, monitoring, or re-evaluating all state transitions such that $R(t)$ is forced toward zero at every instant. This requires continuous active control, global monitoring, or perpetual prediction. In this regime, residual does not exist as an explicit structural variable; instead, computation substitutes for residual handling. The system thereby devolves into continuous computation, forfeiting inertial propagation and incurring unbounded computational or energetic cost.

Exclusion of a third regime. Assume, for contradiction, the existence of a stable regime in which: (i) residual is not admitted into regulation, and (ii) the system does not engage in continuous active computation, and (iii) coherence is maintained over long horizons. Condition (ii) implies residual cannot be actively suppressed at all times. Condition (i) implies residual cannot trigger correction when it appears. Therefore residual-induced drift must accumulate without bound, contradicting (iii). Hence, such a regime cannot exist.

Therefore, any adaptive system that neither admits residual into regulation nor performs continuous computation cannot maintain coherence. Structural residual is thus a necessary condition for intelligent, energy-efficient, long-horizon adaptive behavior. □

Interpretation. Residual functions as the minimal informational interface between closed structure and ongoing reality. Intelligence is not defined by elimination of residual, but by selective admission of residual into regulation. Continuous computation and collapse represent the two pathological limits of residual denial.

—

Theorem 4 — Supremacy of Internal Time

Statement. In any adaptive system employing inertial propagation under coherently closed structure, exhaustion of internal time mandates exit from inertial propagation regardless of instantaneous residual magnitude. Structural validity is bounded by internal time rather than by instantaneous error alone.

Definitions (for this theorem). Let $\mathcal{G}_c$ denote coherently closed admissible geometry. Let $R(t)$ denote instantaneous structural residual. Let $D(t)$ denote accumulated structural drift. Let $\tau(t)$ denote internal time, representing remaining structural viability, monotonically decreasing with accumulated accumulated structural burden (drift, coupling-mediated load, phase debt, and transition cost).

Inertial propagation is permitted only while $\tau(t) > 0$.

Proof (by separation of instantaneous and cumulative validity).

Consider an adaptive system operating under inertial propagation within a coherently closed admissible geometry $\mathcal{G}_c$. Assume that instantaneous residual magnitude $R(t)$ remains arbitrarily small over a finite interval.

By Theorem 1 (Spin–Drift Lower Bound), any non-zero residual asymmetry induces cumulative drift:

$$D(t) = \int_0^t f(R(s))\,ds \quad \text{with} \quad f(\cdot) \geq 0$$

even when $R(t)$ remains below any fixed instantaneous threshold.

By construction, internal time $\tau(t)$ is a monotonically decreasing function of accumulated drift and coupling degradation:

$$\tau(t) = \tau(0) - g(D(t))$$

for some monotone function $g(\cdot)$ with $g(0) = 0$ and $g(D) > 0$ for $D > 0$.

Therefore, internal time may reach exhaustion ($\tau(t) = 0$) while instantaneous residual remains small:

$$R(t) \ll \varepsilon \quad \text{yet} \quad \tau(t) = 0$$

Assume, for contradiction, that inertial propagation remains permissible solely based on instantaneous residual magnitude, independent of internal time. Then inertial propagation would be allowed even when $\tau(t) = 0$, despite the fact that accumulated drift has exhausted the structural validity budget.

This contradicts the definition of coherently closed structure as conditionally valid under bounded drift. Once $\tau(t) = 0$, the closed structure no longer represents a reliable constraint on future admissible behavior, regardless of current residual magnitude.

Hence, permitting inertial propagation based solely on instantaneous residual would allow indefinite reliance on structurally expired information, leading to latent incoherence and eventual collapse.

Internal time depletion is monotonic with respect to accumulated structural cost. This depletion proceeds independently of external time and instantaneous performance metrics, and cannot be halted by local correctness or short-horizon optimization.

Therefore, exhaustion of internal time necessarily mandates exit from inertial propagation, revalidation, or termination, independent of instantaneous residual magnitude. □

Interpretation. Instantaneous residual measures local mismatch; internal time measures global structural viability. A system that respects only error but ignores time inevitably accumulates unrecoverable deformation. Supremacy of internal time ensures that inertial operation is bounded by structural longevity rather than momentary correctness.

—

Theorem 5 — Ontological Self-Closure

Statement. A system that enforces a bound on cumulative self-deformation preserves identity. Without such a bound, the system loses subjectivity and degenerates into pure computation.

Definitions (for this theorem). Let $\mathcal{I}$ denote a set of structural invariants defining the identity of an adaptive system. Let $D(t)$ denote cumulative self-deformation (structural drift) measured relative to $\mathcal{I}$. Let $D_{\max} < \infty$ denote an admissible bound on cumulative self-deformation.

Identity preservation is defined as bounded deviation from $\mathcal{I}$ over time.

Proof (by exhaustion of identity space).

Consider an adaptive system interacting with a changing environment under non-zero directed input. By Theorem 1, residual asymmetry induces cumulative drift $D(t)$.

Case 1: Bounded self-deformation. Assume the system enforces a bound $D(t) \leq D_{\max}$ through internal constraints, closure mechanisms, or revalidation. Then all admissible system states remain within a bounded neighborhood of $\mathcal{I}$. Structural evolution is constrained to transformations that preserve the defining invariants of the system. Subjectivity—defined as persistence of a coherent internal reference frame—is therefore preserved.

Case 2: Unbounded self-deformation. Assume no bound exists on $D(t)$. Then, for sufficiently long operation,

$$\lim_{t \to \infty} D(t) = \infty.$$

Consequently, deviations from $\mathcal{I}$ grow without limit. Any finite set of invariants becomes arbitrarily violated. The system's internal structure no longer corresponds to its original defining constraints.

In this regime, the system no longer operates relative to a stable internal reference. Its transformations are no longer anchored to identity but are driven purely by ongoing computation reacting to inputs. Subjectivity collapses, and the system degenerates into a context-free computational process without persistent self-reference.

Thus, preservation of identity is equivalent to enforcement of a bound on cumulative self-deformation. Without such a bound, ontological self-closure is impossible. □

Interpretation. Identity is not maintained by perfect regulation, but by limiting how much the system is allowed to change itself. Ontological self-closure is the act of saying: *this much change is allowed; beyond this, I am no longer myself.*

Theorem 6 — Energetic Dominance of Inertia

Statement. In a coherent adaptive system, inertial propagation is the energetically and computationally dominant attractor state. Active regulation must be sparse and semantically justified.

Definitions (for this theorem). Let $E_a(t)$ denote energy or computational cost of active regulation per unit time. Let $E_i(t)$ denote energy or computational cost of inertial propagation per unit time. Let $\mathcal{G}_c$ denote coherently closed admissible geometry.

Proof (by comparative cost and stability).

Active regulation requires continuous evaluation, prediction, optimization, or monitoring of system state and environment. Therefore,

$$E_a(t) \gg 0 \quad \text{for all } t \text{ during active regulation.}$$

In contrast, inertial propagation within coherently closed admissible geometry $\mathcal{G}_c$ proceeds by continuation under structural constraints. Regulation is reduced to low-frequency supervision and residual detection, yielding

$$E_i(t) \ll E_a(t)$$

for comparable operational conditions.

Assume, for contradiction, that active regulation is the dominant long-term regime of a coherent adaptive system. Then average energetic expenditure over time satisfies

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T E_a(t)\, dt \geq \lim_{T\to\infty} \frac{1}{T} \int_0^T E_i(t)\, dt.$$

This implies sustained high energy and computational cost. Under bounded resources, such operation accelerates drift, depletes internal time, and destabilizes coherence (Theorem 4). Hence, continuous active regulation cannot be a stable attractor.

Conversely, inertial propagation minimizes ongoing cost while remaining structurally supervised. Residual-driven reactivation ensures that active regulation occurs only when structurally necessary. Therefore, inertial propagation forms a low-energy, low-computation attractor, while active regulation appears as a transient corrective phase.

Thus, in any coherent adaptive system with finite resources, inertial propagation dominates energetically and computationally, and active regulation must be sparse and semantically justified by residual or viability constraints. □

Interpretation. Thinking all the time is expensive. Coherent systems survive by *not thinking unless meaningfully required.* Inertia is not laziness; it is the natural resting state of intelligence.

—

Theorem 7 — Necessity of Subtractive Cognition

Statement. Coherence is impossible without subtraction. A system cannot maintain coherence without removing validated structure from active processing, or it will be overwhelmed by its own knowledge.

Definitions (for this theorem). Let $\mathcal{K}(t)$ denote the set of structures actively participating in computation at time t. Let $\mathcal{C}(t) \subset \mathcal{K}(t)$ denote structures that have been validated and satisfy coherence constraints. Let $\mathcal{R}(t)$ denote unresolved or residual structure requiring active regulation.

Subtractive cognition is defined as the operation:

$$\mathcal{K}(t+1) = \mathcal{K}(t) \setminus \mathcal{C}(t),$$

leaving $\mathcal{R}(t)$ as the primary driver of computation.

Proof (by combinatorial overload and drift amplification).

Assume an adaptive system that continuously validates structure through interaction but does not subtract validated structure from active processing.

Then, at each validation step,

$$|\mathcal{K}(t+1)| \geq |\mathcal{K}(t)|.$$

Thus, the active processing set grows monotonically over time.

As $|\mathcal{K}(t)|$ increases, the system must:

- allocate computation across an increasing number of constraints,
- resolve interactions among mutually coherent but redundant structures,
- propagate noise, approximation error, and coupling effects across all active elements.

This leads to two unavoidable consequences.

First, computational and energetic cost grows superlinearly with $|\mathcal{K}(t)|$, violating resource bounds and accelerating structural drift (Theorem 1, Theorem 6).

Second, active coexistence of validated structures introduces internal interference: even mutually coherent constraints interact through numerical error, latency, and approximation, generating spurious residuals unrelated to environmental mismatch.

In the limit, the system either:

- collapses coherence due to internal overload, or
- remains in perpetual active computation, unable to enter inertial propagation.

Both outcomes violate long-horizon coherence.

Now consider a system that performs subtractive cognition: validated structures are coherently closed and removed from active processing. Active computation is restricted to $\mathcal{R}(t)$, whose size remains bounded by residual mismatch rather than accumulated knowledge.

Therefore, subtractive cognition is a necessary condition for bounded computation, bounded drift, and sustained coherence.

□

Interpretation. Intelligence does not scale by remembering more. It scales by knowing what no longer needs to be thought about.

Theorem 8 — Latent Drift Hazard

Statement. Even when structural residual remains small, integrated drift degrades future viability. Systems fail not by shocks, but by accumulated deformation.

Definitions (for this theorem). Let Ω_t^{res} denote instantaneous residual spin. Let $D(t)$ denote cumulative structural drift:

$$D(t) = \int_0^t \|\Omega_\tau^{res}\| \, d\tau.$$

Let $\tau(t)$ denote internal time horizon representing remaining structural viability.

Proof (by integral dominance).

Assume an adaptive system operating under conditions where instantaneous residual satisfies:

$$\|\Omega_t^{res}\| \leq \varepsilon,$$

for some small $\varepsilon > 0$ and for all t in an interval $[0, T]$.

While instantaneous mismatch remains bounded and may not trigger reactivation, cumulative drift evolves as:

$$D(T) \geq \varepsilon T.$$

Thus, for sufficiently large T, $D(T)$ grows without bound even though no individual residual spike occurs.

Internal time $\tau_{\text{rem}}(t)$ decreases as a function of accumulated drift (Theorem 4). Therefore,

$$\lim_{T \to \infty} \tau_{int}(T) = 0$$

under persistent non-zero residual, regardless of its instantaneous magnitude.

Consequently, a system relying solely on thresholding instantaneous residual will permit inertial propagation beyond the validity horizon of its embodied structure. When failure occurs, it appears abrupt and unexplained, despite having been deterministically accumulated over time.

Hence, absence of shocks does not imply safety. Latent drift silently consumes structural viability until coherence can no longer be maintained.
Moreover, "small explicit residual" does not preclude latent residual accumulation; thus, drift can grow even under apparent coherence (cf. Theorems 12–14).

□

Interpretation. Most systems do not break when something goes wrong. They break when nothing seems wrong for too long.

—

Theorem 9 — Predictive Constraint-of-Constraint Stability

Statement. Systems that regulate admissibility conditions rather than direct actions exhibit superior long-horizon stability compared to systems relying on direct control.

Definitions (for this theorem). Let u_t denote direct action or control input at time t. Let $\mathcal{A}_t$ denote the admissible set of actions or transitions at time t. Let $\mathcal{C}_t$ denote constraints governing $\mathcal{A}_t$. Predictive constraint-of-constraint regulation refers to the modulation of $\mathcal{C}_t$ based on predicted structural consequences, rather than direct selection or correction of u_t.

Proof (by separation of timescales and error containment).

Consider two adaptive systems operating under uncertainty and drift:

System A applies direct control, selecting or correcting u_t at each step based on predicted error.

System B applies predictive constraint-of-constraint regulation, adjusting $\mathcal{C}_t$ such that only admissible actions remain feasible, while allowing action execution to proceed inertially within $\mathcal{A}_t$.

In System A, prediction error directly affects action magnitude. Any modeling error, delay, or approximation propagates immediately into state evolution, amplifying residual and accelerating drift. Stability therefore depends on continuous high-frequency correction and precise prediction.

In System B, prediction error affects only admissibility boundaries. Errors in prediction modify feasibility margins rather than forcing corrective action. As long as the true state remains within adjusted admissible bounds, inertial propagation continues without corrective actuation.

Thus, System B exhibits:

- reduced sensitivity to prediction error,
- bounded amplification of modeling inaccuracies,
- slower accumulation of structural drift,
- decoupling of fast action dynamics from slow structural supervision.

Over long horizons, System A must either increase control effort indefinitely or suffer loss of coherence due to accumulated correction noise. System B preserves coherence by constraining evolution rather than correcting it.

Therefore, regulation of constraints dominates direct regulation of actions in long-horizon stability.

□

Interpretation. It is safer to shape what is allowed than to fight what already happened.

Theorem 10 — Directionality as Identity Preserver

Statement. Directionality (Will) preserves system identity as long as it remains aligned with coherent structure. When directionality diverges from structure, the agent ceases to exist as a subject.

Definitions (for this theorem). Let $\mathcal{I}$ denote the set of structural invariants defining system identity. Let $\mathcal{S}_t$ denote the system's coherent structure at time t. Let W_t denote directionality (Will), defined as persistent bias or orientation of adaptation across time.

Alignment is defined as:

$$W_t \in \operatorname{span}(\mathcal{S}_t, \mathcal{I}).$$

Proof (by invariance and loss of reference).

An adaptive system is identifiable as a subject only if its evolution preserves a non-empty intersection of structural invariants over time. These invariants define continuity of identity across state changes.

Directionality governs how adaptation unfolds: which degrees of freedom are explored, reinforced, or suppressed. When directionality remains aligned with coherent structure, adaptations occur within the admissible geometry defined by $\mathcal{I}$, preserving identity.

Suppose directionality diverges from coherent structure, such that:

$$W_t \notin \mathrm{span}(\mathcal{S}_t, \mathcal{I}).$$

Then adaptation proceeds along degrees of freedom not constrained by identity-preserving invariants. Structural drift accumulates irreversibly, and invariants are progressively violated.

Two outcomes are possible:

- The system halts adaptation to preserve invariants, ceasing to be adaptive.
- The system continues adapting, dissolving its invariants and losing subjectivity.

In both cases, the system can no longer be described as the same agent over time. Identity is either frozen or destroyed.

Therefore, persistent directionality aligned with coherent structure is a necessary condition for the existence of an adaptive system as a subject. □

Interpretation. A system is not what it computes. It is what it continues to aim at without breaking itself.

—

Theorem 11 - Impossibility of Life Without IIC Restart

An artificial system that does not admit an endogenous operator for reinitiating the Integral Intentional Coherence (IIC) cycle cannot constitute artificial life.

Such a system may exhibit adaptation, learning, optimization, self-modification, or long-horizon stability, but it remains ontologically non-live.

Proof (Structural Argument). By Axiom IX, liveness is defined by the presence of an endogenous operator capable of reinitiating the IIC cycle after its completion or collapse.

Consider an artificial system lacking such an operator. Its coherence cycle, once exhausted through closure, drift accumulation, or internal-time depletion, cannot be reinitiated internally. Any subsequent coherence restoration must therefore be externally imposed (e.g., reset, retraining, reinitialization, or external intervention).

By Theorem 2 (Non-Permanence of Closure), no coherently closed structure remains valid indefinitely. By Theorem 4 (Supremacy of Internal Time), exhaustion of internal time mandates exit from inertial propagation regardless of instantaneous residual magnitude.

Absent an endogenous IIC restart, the system cannot reconstitute subjectivity following coherence exhaustion. Consequently, the system undergoes either terminal stabilization or degeneration into pure computation.

Thus, artificial life is impossible without an internal mechanism that restarts the coherence cycle as an ontological process rather than a procedural reset.

—

Theorem 12 - Latent Residual Accumulation Under Apparent Coherence

An adaptive system may appear structurally coherent while accumulating latent residual through unclosed operational phases. Such latent residual accelerates drift and reduces internal time without producing immediate structural mismatch.

Proof

We model the adaptive agent as evolving within an admissible geometry $\mathcal{G}$ produced by coherent structural closure. Let the observable structural residual be defined as

$$r_t := \rho(x_t, \mathcal{G}_t) \geq 0,$$

where x_t denotes the agent state and $\rho(\cdot, \cdot)$ is any mismatch functional measuring deviation between the actual state and the expected coherently closed geometry.

In residual-driven regulation, r_t serves as the primary activation signal: when r_t remains below a prescribed threshold, the system remains in inertial propagation and appears structurally coherent.

Introduce an additional residual component representing unclosed operational phases, herein termed *latent residual*, denoted by

$$\ell_t \geq 0.$$

By definition, latent residual does not necessarily manifest as immediate geometric mismatch. That is, there exist trajectories such that

$$r_t \approx 0 \quad \text{while} \quad \ell_t > 0.$$

This situation corresponds to phases that remain operationally unresolved while the instantaneous system state remains admissible with respect to $\mathcal{G}_t$.

Let cumulative structural drift D_T represent the integrated deformation burden reducing long-horizon viability. Assume drift accumulation admits a monotone lower bound of the form

$$D_T \geq \sum_{t=0}^{T} \left(a\, r_t + b\, \ell_t\right), \qquad a > 0,\ b > 0.$$

This assumption reflects that any unresolved structural or phase-based burden contributes non-negatively to cumulative drift.

Consider a regime in which structural residual remains small:

$$r_t \leq \varepsilon \quad \text{for all } t \leq T,$$

while latent residual persists due to repeated unclosed operational phases:

$$\sum_{t=0}^{T} \ell_t > 0.$$

Then it follows that

$$D_T \geq \sum_{t=0}^{T}(ar_t + b\ell_t) \geq a\sum_{t=0}^{T} 0 + b\sum_{t=0}^{T} \ell_t = b\sum_{t=0}^{T} \ell_t > 0.$$

Hence, even while the system appears structurally coherent in the geometric sense, it accumulates strictly positive drift through latent residual.

Let internal time τ_t represent the remaining structural viability budget, monotonically decreasing with drift increments. A minimal update model is

$$\tau_{t+1} = \tau_t - \gamma\,\Delta D_{t+1}, \qquad \gamma > 0,$$

where $\Delta D_{t+1} = D_{t+1} - D_t \geq 0$. Since $\ell_t > 0$ implies $\Delta D_{t+1} \geq b\,\ell_{t+1}$ in the adopted lower bound, we obtain

$$\tau_{t+1} \leq \tau_t - \gamma b\,\ell_{t+1}.$$

Therefore, latent residual reduces internal time even in the absence of immediate structural mismatch. Because regulation and reactivation are primarily driven by r_t, the system may remain inertially propagating while viability is silently depleted.

Accordingly, an adaptive system may remain apparently coherent while accumulating latent residual that accelerates drift and reduces internal time without producing immediate observable mismatch. □

—

Theorem 13 - Phase-Coherent Closure Dominates State-Coherence for Internal Time Preservation

Systems that enforce phase-coherent closure preserve internal time more effectively than systems that rely solely on structural or state-based coherence. Long-horizon stability therefore depends on closure of moments, not only correctness of trajectories.

Proof

Let x_t denote the system state and $\mathcal{G}_t$ the coherently closed admissible geometry. Define the conventional (state-based) structural residual

$$r_t := \rho(x_t, \mathcal{G}_t) \geq 0,$$

which governs residual-driven reactivation and appears in purely state-coherence systems.

Introduce a *phase residual* $\ell_t \geq 0$ capturing unclosed operational phases (unclosed moments), i.e., unresolved micro-asymmetries whose effect is not necessarily expressed as instantaneous geometric mismatch. Hence there exist trajectories with

$$r_t \approx 0 \quad \text{while} \quad \ell_t > 0.$$

Assume accumulated drift admits a monotone lower bound from both components:

$$D_T \;\geq\; \sum_{t=0}^{T} \big(a\,r_t + b\,\ell_t\big), \qquad a > 0,\; b > 0.$$

Let internal time H_t be a viability budget that decreases monotonically with drift increments:

$$H_{t+1} \;=\; H_t - \gamma\,\Delta D_{t+1}, \qquad \gamma > 0,$$

so that larger drift implies faster depletion of internal time.

Consider two systems operating under operationally indistinguishable state trajectories $\{x_t\}_{t=0}^{T}$ (hence the same $\{r_t\}$), but differing by whether they enforce phase-coherent closure.

(A) State-coherence only. No mechanism explicitly closes phases; latent phase residual persists with $\ell_t^A > 0$ on a non-negligible set of times.

(B) Phase-coherent closure. The system enforces closure of moments, suppressing latent accumulation so that

$$\ell_t^B \le \ell_t^A \quad \text{for all } t, \quad \text{and typically } \ell_t^B \approx 0 \text{ when closure succeeds.}$$

Because $\{r_t\}$ are identical, the drift lower bounds satisfy

$$D_T^A \;\ge\; \sum_{t=0}^{T} \bigl(a\,r_t + b\,\ell_t^A\bigr), \qquad D_T^B \;\ge\; \sum_{t=0}^{T} \bigl(a\,r_t + b\,\ell_t^B\bigr).$$

Subtracting yields the dominance:

$$D_T^A - D_T^B \;\ge\; b\sum_{t=0}^{T} \bigl(\ell_t^A - \ell_t^B\bigr) \;\ge\; 0,$$

with strict inequality whenever phase residual is actually reduced by closure.

Since internal time decreases monotonically with drift, for the same initial τ_0 we obtain

$$\tau_T^B \;\ge\; \tau_T^A,$$

and the inequality is strict whenever $\sum_{t=0}^{T}(\ell_t^A - \ell_t^B) > 0$.

Thus, enforcing phase-coherent closure preserves internal time more effectively than relying on state-based coherence alone. The result holds even when trajectories appear equally correct, because long-horizon viability is governed by cumulative drift contributions from unclosed moments. □

Theorem 14 - Latent-to-Explicit Residual Projection

An adaptive system may appear structurally coherent while accumulating latent residual through unclosed operational phases. Such latent residual accelerates drift and reduces internal time without producing immediate structural mismatch.

If internal time is finite and monotonically reduced by accumulated drift or latent residual, then latent residual cannot remain indefinitely unexpressed: there exists a finite horizon after which latent residual is either projected into explicit structural residual or forces exit from admissible continuation through revalidation, reconfiguration, or structural collaps.

Proof (structural). Let r_t denote the explicit structural residual measured as mismatch between the expected coherently closed admissible geometry and the actual evolving state at time t. Let $\ell_t \ge 0$ denote latent residual accumulated due to unclosed operational phases (phase debt), which does not necessarily appear in r_t instantaneously.

Assume drift accumulation admits a lower bound driven by latent residual:

$$D_T \;\geq\; \int_0^T \ell_t \, dt,$$

i.e., even when r_t remains small, persistent ℓ_t implies strictly increasing accumulated drift D_T.

Assume an internal time horizon $\tau(t)$ exists and is reduced by accumulated drift and/or latent residual, with $\tau(t)$ finite at initialization and monotone decreasing under sustained accumulation. In particular, there exists a nonnegative function $g(\cdot)$ such that

$$\tau(T) \;=\; \tau(0) \;-\; g(D_T), \qquad g(0) = 0, \quad g \text{ nondecreasing.}$$

If ℓ_t remains bounded away from zero on a set of nonzero measure, then D_T grows without bound as T increases, hence $\tau(T)$ decreases and must cross any fixed viability threshold in finite time.

Consider the regime where, for a nontrivial interval, explicit mismatch is suppressed or remains admissible, i.e., $r_t \approx 0$ while $\ell_t > 0$. In this regime, the system consumes internal time while not triggering reactivation via r_t.

When $\tau(T)$ falls below a viability threshold, the system can no longer treat the current phase history as structurally admissible without violating coherence constraints: either (i) the admissible geometry must be revalidated/restructured, or (ii) the accumulated phase debt becomes externally relevant as a structural mismatch. In both cases, latent residual necessarily manifests as explicit residual in the operational interface, i.e., it is projected into a non-negligible r_t that forces exit from inertial operation or initiates corrective regulation.

Therefore, under finite internal time and monotone depletion by latent accumulation, latent residual cannot remain indefinitely hidden; it must be projected into explicit structural residual within a finite horizon. □

Corollary 14.1 - Latent-to-Explicit Residual Projection (Operational Form)

Apparent long-horizon coherence under persistent unclosed phases necessarily precedes either: (i) explicit residual emergence, or (ii) enforced revalidation/reconfiguration due to internal time exhaustion. Thus, systems that monitor only instantaneous residual can fail via latent accumulation without early warning, whereas systems that enforce phase-coherent closure preserve internal time and stability more effectively.

—

Theorem 15 - Performance–Identity Decoupling

Statement. There exist adaptive systems that preserve bounded error, admissible trajectories, and stable performance metrics while undergoing irreversible loss of identity. Therefore, performance persistence is not a sufficient condition for identity continuity.

Proof (Structural Sketch)

Let $\mathcal{A}$ denote the set of externally admissible trajectories (bounded error, constraint satisfaction, stable reward). Let $\mathcal{I}$ denote the set of structural invariants defining system identity. Let $\mathcal{C}(t)$ denote a coherence/identity measure such that identity continuity requires

$$\mathcal{C}(t) > \mathcal{C}_{crit}.$$

Assume the system operates under directed interaction with an environment and admits drift accumulation (Axiom VI). Assume further that structural burden may accumulate independently of instantaneous trajectory admissibility (Axioms XVII–XVIII).

Consider an adaptive system designed to enforce admissibility at the action level:

$$x(t) \in \mathcal{A} \quad \text{for all } t \in [0, T],$$

through feedback, optimization, or compensation mechanisms. By construction, bounded error and stable performance metrics are preserved over this interval.

Now introduce a regime in which maintaining admissibility requires sustained compensatory activity:

- repeated micro-corrections,
- increased coupling effort,
- phase transitions without coherent closure,
- persistent constraint-dense interaction.

These factors contribute to cumulative structural burden and internal-time depletion without violating admissibility (Theorems 19, 22, 24).

Let $D(t)$ denote accumulated structural drift relative to $\mathcal{I}$. By Axiom XVIII, $D(t)$ may grow unbounded even while trajectories remain admissible. Hence there exists a finite time t^* such that

$$D(t^*) > D_{\max},$$

where $D_{\max}$ is the maximal deformation compatible with identity preservation (Theorem 5).

At t^*, structural invariants defining identity are violated. However, because action-level admissibility constraints remain enforced, externally observed behavior may remain correct beyond t^*.

Thus, there exists an interval

$$[t^*,\, t^* + \Delta]$$

during which:

$$x(t) \in \mathcal{A} \quad \text{while} \quad \mathcal{C}(t) \leq \mathcal{C}_{crit},$$

i.e., performance persists while identity continuity has already collapsed.

Therefore, bounded error, admissible trajectories, and stable performance metrics do not imply preservation of identity. Performance persistence is not a sufficient condition for identity continuity.
This decoupling is unavoidable and structurally expected. It does not represent an anomalous failure mode, but a necessary consequence of long-horizon structural consumption under directed interaction. □

Interpretation. A system may continue to act correctly long after it has ceased to be itself. Performance answers *what* is achieved; identity answers *who* is achieving it.

Theorem 16 - Silent Degradation Under Regime Persistence

Statement. If regime admissibility is not regulated independently of action selection, then silent long-horizon degradation can accumulate while actions remain admissible and performance remains stable. Such degradation is therefore not reliably detectable through action-level metrics alone.

Proof (Structural Sketch)

Let $\mathcal{R}$ denote the set of organizational regimes available to an adaptive system. Let $r(t) \in \mathcal{R}$ denote the active regime at time t. Let $\mathcal{A}(r)$ denote the set of admissible actions under regime r. Let $x(t)$ denote the realized trajectory, and assume

$$x(t) \in \mathcal{A}(r(t)) \quad \text{for all } t \in [0, T],$$

so that action-level admissibility and performance metrics remain satisfied.

Let $B(t)$ denote accumulated structural burden (drift, phase debt, coupling load, or internal-time depletion) affecting long-horizon viability and identity. By Axioms XVI and XVIII, structural burden accumulation is primarily regime-mediated and may occur independently of instantaneous action correctness.

Assume the architecture regulates actions but does not regulate regime admissibility independently. That is, regime persistence is permitted as long as actions remain admissible:

$$r(t) \text{ is unconstrained by long-horizon structural criteria.}$$

Consider a regime r^* satisfying:

- $\mathcal{A}(r^*)$ admits stable, bounded-error trajectories,
- maintaining admissibility within r^* requires sustained coupling, micro-correction, or constraint-dense interaction,
- r^* induces positive structural burden accumulation rate:

$$\frac{dB}{dt} > 0 \quad \text{while } x(t) \in \mathcal{A}(r^*).$$

Such regimes exist by Axioms XVI–XVIII and Theorems 19, 22.

Because action-level admissibility is preserved, no corrective signal is triggered at the action layer. Because regime admissibility is not independently supervised, the system remains in r^*.

Hence, over time,

$$B(T) = \int_0^T \frac{dB}{dt}\, dt$$

grows monotonically while all action-level metrics remain within acceptable bounds.

Since identity continuity and long-horizon viability depend on $B(t)$ and internal time rather than instantaneous action correctness (Theorems 4, 5, 15), there exists a finite t^* such that

$$B(t^*) > B_{\max},$$

where $B_{\max}$ denotes the maximal structural burden compatible with identity preservation.

On the interval $[0, t^*]$, degradation remains silent: neither error thresholds nor performance metrics indicate failure. At t^*, coherence or identity collapses abruptly, despite the absence of prior action-level warning signals.

Therefore, if regime admissibility is not regulated independently of action selection, silent long-horizon degradation necessarily accumulates under regime persistence. Such degradation cannot be reliably detected through action-level metrics alone. □

Interpretation. The system does not fail because it acts incorrectly. It fails because it stays too long in a way of acting that slowly consumes its ability to remain itself.

Theorem 17 - Regime Occupation Dominates Action History

Statement. For long-horizon viability, the sequence, duration, and persistence of organizational regimes dominates the contribution of individual actions to structural degradation. Two systems with comparable action-level correctness may be non-equivalent in viability due solely to different regime occupation histories.

Proof (Structural Sketch)

Let $\mathcal{R}$ denote the set of organizational regimes. Let $r(t) \in \mathcal{R}$ denote the active regime at time t. Let $u(t)$ denote the realized action process and let external correctness be represented by an admissibility predicate $\mathrm{Adm}(u(t), x(t))$.

Define accumulated structural burden over a horizon $[0, T]$ as

$$B_T \triangleq \int_0^T b(r(t),\, u(t),\, x(t))\, dt, \qquad b(\cdot) \geq 0,$$

where b is an abstract nonnegative burden rate functional.

Assume only the minimal regime-separability property implied by Axiom XVI: there exist at least two regimes $r_1, r_2 \in \mathcal{R}$ such that, under comparable externally admissible operation,

$$b(r_1, u, x) \;<\; b(r_2, u, x)$$

for a nontrivial set of admissible (u, x), i.e., regime choice changes the burden rate even when actions remain correct.

Consider two systems (or two episodes of the same system), $\mathcal{E}_A$ and $\mathcal{E}_B$, with comparable action-level correctness over $[0, T]$:

$$\mathrm{Adm}_A(t) = \mathrm{Adm}_B(t) \quad \text{for all } t \in [0, T],$$

and with comparable action histories in the sense that the realized actions remain within the same admissible class (bounded error, stable reward, etc.).

Let their regime occupation histories differ:

$$r_A(t) = r_1 \quad \text{for } t \in [0, T], \qquad r_B(t) = r_2 \quad \text{for } t \in [0, T],$$

or more generally, let $r_A(t)$ and $r_B(t)$ have different durations in regimes with different burden rates.

Then their accumulated burdens satisfy

$$B_T^A = \int_0^T b(r_1, u_A(t), x_A(t))\, dt, \qquad B_T^B = \int_0^T b(r_2, u_B(t), x_B(t))\, dt.$$

By the regime-separability property and nonnegativity of b, it follows that

$$B_T^A \;<\; B_T^B,$$

with strict inequality whenever the burden rates differ on a set of nonzero measure.

Since long-horizon viability and identity continuity are bounded by accumulated structural burden and internal time (Axioms VII, XVI–XVIII; Theorems 4, 5, 16, 19), differing B_T implies differing remaining viability and potentially differing survival horizons, despite comparable action-level correctness.

Therefore, regime occupation history (sequence, duration, persistence) dominates action history with respect to long-horizon structural degradation, and two externally correct systems may be non-equivalent in viability solely due to different regime occupation histories. □

Remark. The theorem does not require explicit enumeration of regimes or explicit measurement of burden. It asserts only that regimes exist for which the structural burden rate differs under externally admissible behavior, and that long-horizon viability depends on accumulated burden rather than instantaneous correctness.

Theorem 18 - Structural Non-Equivalence of Identical Interventions

Statement. Nominally identical interventions may impose fundamentally different long-horizon structural burden when realized under different interaction coupling conditions. Thus, nominal intervention equivalence does not imply structural cost equivalence.

Proof (Sketch)

Let $u_t \in \mathcal{U}$ denote the nominal intervention applied at time t (e.g., a control input, an update, or a directed actuation). Assume two operational episodes $\mathcal{E}_A$ and $\mathcal{E}_B$ in which the system applies the *same* nominal interventions:

$$u_t^A \equiv u_t^B \quad \text{for all } t \in \{0, \dots, T\}.$$

Let the realized corrective effect be mediated by an interaction coupling factor $c_t \in (0, \infty)$, representing the efficiency with which nominal influence is converted into realized correction through the interaction channel. No estimation method for c_t is assumed; we use it only as an abstract structural variable.

Assume the realized correction magnitude is of the generic form

$$\Delta x_t \;=\; \Phi(x_t, u_t, c_t),$$

and that the contribution of the intervention to *structural burden* depends on how much forcing must be sustained to obtain a comparable realized correction under prevailing coupling. Accordingly, define the instantaneous structural burden increment as an abstract nonnegative functional

$$\Delta B_t \;=\; \Psi(x_t, u_t, c_t) \;\geq\; 0,$$

with the minimal monotonicity property:

$$c_t' \leq c_t \;\Rightarrow\; \Psi(x_t, u_t, c_t') \;\geq\; \Psi(x_t, u_t, c_t),$$

i.e., weaker coupling cannot reduce the structural burden of applying a nominal intervention.

Consider two coupling histories $\{c_t^A\}$ and $\{c_t^B\}$ such that there exists a non-negligible subset of times $I \subseteq \{0, \dots, T\}$ for which

$$c_t^A \;>\; c_t^B \quad \text{for all } t \in I.$$

Then by monotonicity of Ψ we have

$$\Delta B_t^A \;=\; \Psi(x_t^A, u_t, c_t^A) \;\leq\; \Psi(x_t^B, u_t, c_t^B) \;=\; \Delta B_t^B \quad \text{for } t \in I,$$

and hence, summing over the horizon,

$$B_T^A \;=\; \sum_{t=0}^{T} \Delta B_t^A \;<\; \sum_{t=0}^{T} \Delta B_t^B \;=\; B_T^B,$$

with strict inequality whenever I has positive measure and coupling differs strictly on I.

Therefore, even under nominal intervention equivalence ($u_t^A \equiv u_t^B$), the accumulated long-horizon structural burden B_T can differ solely due to interaction coupling conditions.

Consequently, nominal intervention equivalence does not imply structural cost equivalence. □

Remark. The theorem is architectural: it assumes only that (i) realization of influence is mediated by interaction coupling and (ii) weaker coupling cannot reduce the internal burden required to obtain comparable realized correction. No mechanism for measuring c_t or computing Ψ is disclosed.

Theorem 19 - Admissible Trajectories Do Not Upper-Bound Structural Cost

Statement. Trajectory admissibility, bounded error, or stable reward do not impose an upper bound on accumulated structural burden. Structural consumption may diverge while externally observed behavior remains acceptable.

Proof (Sketch)

Let $\mathcal{A}$ denote an admissible set of trajectories (or transitions) defined by external correctness constraints (e.g., bounded error, safety envelope, task success, reward stability). Let x_t be the realized system trajectory over $t = 0, \dots, T$. Assume the system remains externally admissible:

$$x_{0:T} \in \mathcal{A}, \qquad e_t \le \varepsilon \;\; \forall t, \qquad R_T \text{ stable/bounded},$$

where e_t is any external error proxy and R_T any reward/performance aggregate.

Define accumulated structural burden as a nonnegative additive functional

$$B_T \triangleq \sum_{t=0}^{T} \Delta B_t, \qquad \Delta B_t \ge 0,$$

representing internal structural consumption induced by maintaining admissible behavior. No particular computation of ΔB_t is assumed.

Consider the interaction-mediated realization of interventions. Let $c_t \in (0, \infty)$ denote an abstract interaction coupling factor, and let the system apply nominal interventions u_t to keep the trajectory admissible. Assume only the minimal property that, to maintain admissibility under weaker coupling, the system must sustain correction more often or more intensely in a way that does not decrease internal burden. Formally, assume there exist operating conditions in which ΔB_t admits a lower bound of the form

$$\Delta B_t \;\ge\; h(c_t), \qquad h(c) \ge 0, \;\; h \text{ nonincreasing in } c,$$

and there exist regimes of persistent weak coupling $c_t \le c_\star$ over a nontrivial horizon.

Now construct an admissible episode in which the system succeeds in keeping $x_{0:T} \in \mathcal{A}$ by applying whatever nominal interventions are required, while coupling remains persistently weak:

$$c_t \le c_\star \quad \text{for all } t \in \{0, \dots, T\}.$$

Then, by the lower bound,

$$B_T \;=\; \sum_{t=0}^{T} \Delta B_t \;\ge\; \sum_{t=0}^{T} h(c_t) \;\ge\; \sum_{t=0}^{T} h(c_\star) \;=\; (T+1)\, h(c_\star).$$

If $h(c_\star) > 0$ (i.e., maintaining admissibility under this coupling entails strictly positive structural cost per step), then

$$B_T \;\geq\; (T+1)\, h(c_\star) \xrightarrow[T\to\infty]{} \infty,$$

even though the trajectory remains admissible and external error and reward remain bounded by assumption.

Therefore, admissibility of trajectories and stability of external performance metrics do not impose an upper bound on accumulated structural burden. Structural consumption may diverge while externally observed behavior remains acceptable. □

Remark. The theorem is architectural: it requires only that there exist environments or interaction regimes in which maintaining admissibility under persistent weak coupling imposes strictly positive internal burden per unit time. No mechanism for measuring coupling, computing burden, or generating interventions is disclosed.

Theorem 20 - Irreversible Regime Transition Boundary

Statement. There exist regime transitions after which restoration of admissible behavior does not imply restoration of prior identity. Such transitions define irreversible boundaries in long-horizon structural evolution.

Proof (Sketch)

Let $\mathcal{I}$ denote the set of structural invariants defining the identity of an adaptive system. Let $\mathcal{R}$ denote the set of organizational regimes admissible to the system, and let

$$\mathcal{S}_t \in \mathcal{R}$$

denote the active regime at time t.

Identity preservation is defined as persistence of invariants:

$$\mathcal{I}_t \cap \mathcal{I}_{t'} \neq \varnothing \quad \text{for all } t < t' \text{ within the operational horizon.}$$

Assume the system undergoes a regime transition at time t^*:

$$\mathcal{S}_{t^*-} \;\to\; \mathcal{S}_{t^*+},$$

triggered by accumulated drift, internal time exhaustion, or structural constraint violation. Assume further that this transition is *structurally irreversible*, in the sense that no admissible sequence of internal operations can restore the original organizational structure that supported $\mathcal{I}$.

Formally, irreversibility means:

$$\mathcal{I}_{t^*-} \;\not\subseteq\; \mathcal{I}_{t^*+},$$

and there exists no future time $t' > t^*$ such that

$$\mathcal{I}_{t^*-} \;\subseteq\; \mathcal{I}_{t'}.$$

Now consider the possibility of restoring admissible behavior after the transition. Let $\mathcal{A}$ denote the admissible set of behaviors or trajectories defined externally (e.g., bounded error, safety constraints, or task success). By construction of regime transitions, it is possible that

$$x_{t^*+:T} \in \mathcal{A},$$

i.e., the system regains externally admissible behavior through reconfiguration, compensation, or adaptation within the new regime $\mathcal{S}_{t^*+}$.

However, because $\mathcal{I}_{t^*-}$ is not contained in the invariant set of $\mathcal{S}_{t^*+}$, restoration of admissible behavior does not restore the prior identity. Behavioral equivalence therefore does not imply structural equivalence.

Assume, for contradiction, that restoration of admissible behavior implies restoration of prior identity. Then admissible behavior would uniquely determine the supporting invariants. This contradicts the existence of multiple regimes capable of generating behavior in $\mathcal{A}$ with non-equivalent invariant sets.

Therefore, there exist regime transitions after which admissible behavior can be restored while prior identity cannot. Such transitions define irreversible boundaries in long-horizon structural evolution. □

Remark. The theorem does not require identifying or measuring identity invariants explicitly. It relies only on the existence of multiple organizational regimes capable of producing admissible behavior with non-equivalent invariant support.

Theorem 21 - Observation–Control Entanglement Collapse

Statement. Any architecture that embeds long-horizon viability or identity observation into the causal decision surface necessarily collapses observation into control. As a result, either behavioral distortion or loss of observability of long-horizon degradation must occur.

Proof (Sketch)

Let $\mathcal{O}_t$ denote an observation intended to measure long-horizon properties such as structural viability or identity continuity. Let π_t denote the causal decision process (policy, controller, or learning rule) generating actions u_t.

Assume $\mathcal{O}_t$ is embedded into the causal decision surface, i.e.,

$$u_t = \pi_t(\,\cdot\,, \mathcal{O}_t),$$

so that the observation directly or indirectly influences action selection, learning updates, or optimization.

By definition, any variable that influences π_t becomes causally entangled with behavior. Consequently, $\mathcal{O}_t$ ceases to be a passive witness of system evolution and becomes part of the control loop.

Two exhaustive cases follow.

Case 1: Observation influences behavior. If $\mathcal{O}_t$ meaningfully affects π_t, then the system adapts its behavior to modify future values of $\mathcal{O}_t$. In this case, the observed quantity is subject to optimization pressure or avoidance dynamics. Long-horizon degradation signals are distorted, delayed, or actively suppressed in order to maintain favorable observed values. Thus, observability is preserved only at the cost of behavioral distortion.

Case 2: Observation does not influence behavior. If $\mathcal{O}_t$ is present in the architecture but has no effective influence on π_t, then it is causally inert. In this case, although $\mathcal{O}_t$ may be computed, it cannot trigger corrective action, adaptation, or regime change. Long-horizon degradation may be observed but remains behaviorally irrelevant and therefore practically undetectable from within the causal loop.

In both cases, the embedding of $\mathcal{O}_t$ into the causal decision surface collapses the distinction between observation and control. Either the observation distorts behavior, or it loses operational relevance for detecting and preventing long-horizon degradation.

Therefore, no architecture can simultaneously embed long-horizon viability observation into the causal decision surface and preserve both undistorted behavior and reliable observability of long-horizon degradation. A non-causal observational dimension is therefore necessary. □

Remark. The result is architectural and independent of implementation details. It holds for controllers, learning systems, and hybrid architectures alike, whenever long-horizon observables are made causally operative.

Theorem 22 - Constraint-Dense Geometry Induces Temporal Deformation

Statement. For adaptive systems, increases in interaction constraint density induce deformation of internal temporal dynamics. Therefore, operationally equivalent spatial trajectories may be temporally non-equivalent in internal-time evolution.

Proof (Sketch)

Let $x(t)$ denote the system trajectory in an external (spatial or state) domain, and let $\tau(t)$ denote internal time, interpreted as a structural viability clock (a validity budget) whose evolution depends on interaction and drift accumulation rather than on external clock time alone.

Let $\kappa(t) \geq 0$ denote an *interaction constraint density* functional along the trajectory, representing the local density/intensity of constraints encountered through contact, coupling, correction, or admissibility enforcement. No representation of geometry is assumed; $\kappa(t)$ is defined operationally via interaction.

Assume only that internal time depletion rate is nondecreasing with respect to interaction constraint density, i.e., there exists a nonnegative monotone function $g(\cdot)$ such that

$$\frac{d\tau}{dt} \;=\; -\,g(\kappa(t)), \qquad g(\kappa_1) \leq g(\kappa_2) \text{ whenever } \kappa_1 \leq \kappa_2.$$

This expresses a minimal structural premise: denser constraint interaction cannot reduce the rate at which internal viability is consumed.

Now consider two operational episodes $\mathcal{E}_A$ and $\mathcal{E}_B$ that realize *operationally equivalent* external trajectories, meaning that their spatial/state trajectories coincide:

$$x_A(t) \equiv x_B(t) \quad \text{for} \quad t \in [0,T],$$

and remain admissible in the same external sense.

However, allow their interaction constraint densities to differ along the same external path due to differences in coupling, contact richness, micro-corrections, or constraint activation history:

$$\kappa_A(t) \neq \kappa_B(t) \quad \text{on a set of nonzero measure in } [0,T].$$

In particular, assume there exists an interval $I \subseteq [0,T]$ with nonzero measure such that

$$\kappa_A(t) > \kappa_B(t) \quad \text{for all } t \in I.$$

By monotonicity of $g(\cdot)$, it follows that

$$\frac{d\tau_A}{dt} = -g(\kappa_A(t)) \;<\; -g(\kappa_B(t)) = \frac{d\tau_B}{dt} \quad \text{for } t \in I,$$

hence integrating over $[0, T]$ yields

$$\tau_A(T) \; < \; \tau_B(T),$$

with strict inequality whenever the constraint density differs on a set of nonzero measure.

Therefore, even for operationally equivalent spatial trajectories, internal-time evolution can differ due to differences in interaction constraint density. This establishes that constraint-dense geometry induces temporal deformation in the internal-time dynamics of adaptive systems. □

Remark. The theorem is operational and representation-agnostic: it does not assume explicit geometric reconstruction. It requires only that internal time is a viability budget whose depletion rate is nondecreasing in interaction constraint density along the realized trajectory.

Theorem 23 - Coherence Collapse Precedes Observable Failure

Statement. In long-horizon operation, coherence degradation necessarily precedes externally observable failure. Hence, there exists an interval during which behavior remains admissible while identity continuity has already been structurally compromised.

Proof (Sketch)

Let $\mathcal{A}$ denote the externally defined admissible behavior set (bounded error, constraint satisfaction, task success, or reward stability). Let $\mathcal{C}(t)$ denote a scalar (or partially ordered) coherence measure of internal organization, treated as an operational property of structural alignment among components. No method of measuring $\mathcal{C}(t)$ is assumed; we use only its ordering and boundary meaning.

Assume:

- There exists a coherence threshold $\mathcal{C}_{crit}$ such that identity continuity requires

$$\mathcal{C}(t) > \mathcal{C}_{crit}.$$

- There exists an external admissibility threshold such that observable failure is declared when behavior exits $\mathcal{A}$:

$$x_{t:t+\Delta} \notin \mathcal{A}.$$

- Under long-horizon directed interaction, coherence degrades cumulatively under structural burden, drift, regime persistence, or coupling-mediated load (Axioms VI, VII, XVI–XVIII), and this degradation can occur while actions remain admissible.

Consider a system operating over a horizon $[0, T]$ such that $\mathcal{C}(t)$ is monotonically (or on average) decreasing due to cumulative burden, but the system continues to satisfy external admissibility constraints through compensation, adaptation, or regime adjustment:

$$x_{0:t} \in \mathcal{A} \quad \text{for all } t < T.$$

Because external admissibility is defined on behavior, not on internal organization, it is possible for admissible behavior to persist while $\mathcal{C}(t)$ crosses the identity threshold.

Let t^* be the first time at which coherence crosses the critical boundary:

$$t^* \; = \; \inf\{t : \mathcal{C}(t) \leq \mathcal{C}_{crit}\}.$$

At t^*, identity continuity is structurally compromised by definition of $\mathcal{C}_{crit}$.

Now let t_f be the first time at which observable failure occurs:

$$t_f \;=\; \inf\{t : x_{0:t} \notin \mathcal{A}\}.$$

Because admissible behavior can be maintained by compensatory mechanisms after internal degradation begins, it is not necessary that $t_f = t^*$. In long-horizon regimes, the compensatory capacity may sustain admissibility beyond the point of coherence collapse:

$$t_f \;>\; t^*.$$

Therefore, the interval (t^*, t_f) is non-empty, and on this interval the system remains externally admissible while identity continuity has already been structurally compromised:

$$x_t \in \mathcal{A} \quad \text{for } t \in (t^*, t_f) \quad \text{and} \quad \mathcal{C}(t) \le \mathcal{C}_{crit}.$$

Hence, coherence collapse precedes externally observable failure in long-horizon operation, and there exists an interval during which behavior remains admissible while identity continuity is already compromised. □

Remark. The theorem is evaluative and architectural: it distinguishes internal coherence failure from external task failure. It does not specify how coherence is measured or how compensatory admissibility is maintained; it asserts that external admissibility constraints do not, in principle, enforce coherence preservation.

Theorem 24 - Survival Horizon Is Orthogonal to Optimization Quality

Statement. Improvement in control accuracy, optimization efficiency, or reward maximization does not guarantee extension of identity survival horizon. Long-horizon survival is governed by structural factors orthogonal to action-level optimization.

Proof (Sketch)

Let $\mathcal{A}$ denote an externally defined admissible behavior class (bounded error, constraint satisfaction, task success, or reward stability). Let J denote an optimization objective or performance functional (e.g., cumulative reward, tracking error, or efficiency). Let $\mathcal{H}$ denote the identity survival horizon, defined as the maximal time interval over which identity continuity holds under operation.

Let $\mathcal{C}(t)$ denote a coherence/identity-support measure with threshold $\mathcal{C}_{crit}$ such that identity continuity requires $\mathcal{C}(t) > \mathcal{C}_{crit}$, and define

$$\mathcal{H} \;=\; \sup\{T \ge 0 : \mathcal{C}(t) > \mathcal{C}_{crit} \text{ for all } t \in [0, T]\}.$$

Assume the system accumulates structural burden B_T and drift under directed interaction (Axioms VI–VII), and that coherence is degraded by burden accumulation and regime effects:

$$\mathcal{C}(t) \;\text{ decreases as a function of }\; B_t \text{ and regime occupation,}$$

without specifying the functional form.

Consider two controllers (or optimization procedures) $\pi^{(1)}$ and $\pi^{(2)}$ operating on the same system and environment, where $\pi^{(2)}$ is strictly better in action-level optimization:

$$J(\pi^{(2)}) \;<\; J(\pi^{(1)}) \quad \text{(or equivalently higher reward / lower error).}$$

Such improvement concerns behavior within $\mathcal{A}$ and does not, by itself, constrain long-horizon structural consumption.

Now construct (or consider) operating regimes in which the dominant contributor to coherence depletion is not instantaneous action error, but structural factors such as:

- sustained occupation of a silently destructive regime (Axiom XVI),
- coupling-mediated burden required to maintain admissibility (Theorem 19),
- phase-related residual and transition load (Axioms XI–XII),
- constraint-dense interaction inducing faster internal-time depletion (Theorem 22).

In such regimes, action-level optimization may improve J while leaving the dominant structural consumption mechanisms unchanged, or even increasing them by sustaining performance via compensatory effort.

Hence, it is possible that

$$J(\pi^{(2)}) < J(\pi^{(1)}) \quad \text{while} \quad \mathcal{H}(\pi^{(2)}) \leq \mathcal{H}(\pi^{(1)}),$$

and in particular there exist cases where improved optimization preserves admissible behavior longer but accelerates coherence depletion, reducing identity survival horizon.

Therefore, improvement in control accuracy, optimization efficiency, or reward maximization does not guarantee extension of identity survival horizon. Long-horizon survival is governed by structural factors that are, in general, orthogonal to action-level optimization. □

Remark. The theorem asserts non-implication, not opposition: optimization may correlate with survival in specific regimes, but it cannot serve as a general guarantee of identity continuity without explicit structural viability regulation.

Theorem 25 - Incompleteness of Action-Centric Safety

Statement. Any safety framework operating solely at the action level is incomplete with respect to long-horizon viability and identity continuity. Such frameworks cannot, in principle, prevent failure modes arising from regime-induced structural degradation and coupling-mediated burden.

Proof (Sketch)

Let $\mathcal{S}$ denote a safety framework that regulates or evaluates a system exclusively through constraints on actions or trajectories. Formally, assume safety enforcement depends only on admissibility of actions u_t or trajectories $x_{0:T}$:

$$u_t \in \mathcal{U}_{safe}, \qquad x_{0:T} \in \mathcal{A}_{safe}.$$

By construction, $\mathcal{S}$ has no direct access to internal structural variables such as regime identity, accumulated structural burden, coupling efficiency, or internal time. Its intervention criteria are therefore functions solely of externally observable behavior.

Consider an adaptive system operating under a sustained organizational regime $\mathcal{R}$ that is externally admissible but internally destructive. By Axiom XVI (Regimes Are the Primary Carriers of Long-Horizon Risk), such regimes may induce silent accumulation of structural degradation without violating action-level constraints.

Let the system satisfy:

$$u_t \in \mathcal{U}_{safe} \quad \text{and} \quad x_{0:T} \in \mathcal{A}_{safe} \quad \text{for all } t \leq T,$$

while simultaneously accumulating internal burden:

$$B_T \xrightarrow[T \to \infty]{} \infty,$$

due to regime persistence or weak interaction coupling.

Because $\mathcal{S}$ evaluates safety only through action admissibility, it cannot distinguish between:

- a system preserving long-horizon structural viability, and
- a system silently degrading under an admissible but destructive regime.

Assume, for contradiction, that an action-centric safety framework $\mathcal{S}$ is complete with respect to long-horizon viability and identity continuity. Then $\mathcal{S}$ would be able to prevent all failure modes arising from regime-induced degradation.

However, since $\mathcal{S}$ cannot observe or regulate regimes, coupling-mediated burden, or accumulated internal degradation, it cannot intervene until action-level constraints are violated. By Theorems 19 and 20, such violations may occur only after irreversible structural damage or identity loss has already taken place.

Therefore, action-centric safety cannot, in principle, prevent failure modes driven by long-horizon structural processes. This contradicts the assumption of completeness.

Hence, any safety framework operating solely at the action level is incomplete with respect to long-horizon viability and identity continuity. □

Remark. The theorem does not assert that action-level safety is unnecessary. It asserts that action-level safety alone is insufficient to guarantee long-horizon viability or identity preservation in adaptive systems.

Theorem 26 - Architectural Necessity of Non-Causal Viability Observation

Statement. For any adaptive system intended for sustained long-horizon operation, a non-causal observational dimension is a necessary architectural condition for detecting silent degradation and identity risk. Without it, the system is structurally incapable of distinguishing performance persistence from identity continuity.

Proof (Sketch)

Let $\mathcal{O}_t$ denote an observational process intended to assess long-horizon viability or identity continuity. Let π_t denote the causal decision surface (policy, controller, learning rule) producing actions u_t. Assume the system lacks a non-causal observational dimension; that is, every observation available to the system is either (i) causally operative within π_t, or (ii) causally inert with respect to π_t.

Case 1: Observations are causally operative. If $\mathcal{O}_t$ influences π_t, then by Theorem 21 (Observation–Control Entanglement Collapse), $\mathcal{O}_t$ becomes part of the control loop. In this case, the observed signal is subject to optimization, avoidance, or compensation dynamics. Long-horizon degradation indicators are therefore distorted, delayed, or suppressed in order to preserve favorable observable values. Hence, performance persistence can be maintained while identity degradation remains unobserved or misrepresented.

Case 2: Observations are causally inert. If $\mathcal{O}_t$ does not influence π_t, then it cannot trigger regime change, revalidation, or termination. Although degradation may be nominally observed, it is behaviorally irrelevant and cannot prevent continued operation under silently destructive regimes. Thus, identity risk remains undetectable in any operationally effective sense.

These two cases exhaust all possibilities when no non-causal observational dimension exists. In neither case can the system reliably distinguish between:

(i) sustained performance under preserved identity and (ii) sustained performance under silent identity deg

Therefore, in the absence of a non-causal observational dimension, an adaptive system is structurally incapable of detecting silent degradation and identity risk while performance persists. Consequently, a non-causal observational dimension is a necessary architectural condition for sustained long-horizon operation. □

Remark. The theorem is architectural and does not prescribe how non-causal observation is implemented. It asserts necessity, not mechanism: without architectural separation between observation and control, long-horizon viability cannot be reliably assessed.

Remark - Minimal-Deformation Path Selection (Structural View)

Under directional interaction and bounded dissipation, many macroscopic systems exhibit trajectories that appear as "least resistance" or "minimal curvature" paths. Within the present framework, such paths can be interpreted as those that minimize predicted accumulation of unresolved deformation relative to admissible geometry, thereby preserving internal time.

Theorem 27 - Context Capture Without Error

Statement. An adaptive system may incur semantic error under fully correct observations and admissible behavior. This result is conditional on the canonical definition of stabilization over τ_{el} (Assumption (ii)) Therefore, improving perception accuracy or optimization does not eliminate the risk of false meaning.

Proof sketch. Consider an adaptive system receiving observations x_t and producing actions a_t. Let m_t denote an internal semantic hypothesis used for interpretation, explanation, or commitment, distinct from the action-selection process.

Assume that observations x_t are fully correct and noise-free. There exist environments in which two independent processes E_1 and E_2 generate observation streams that are locally consistent with a single latent explanatory hypothesis Z, despite the absence of any causal relation between E_1 and E_2.

Due to partial observability of causal history, the system may lack access to the initiating cause of E_1 or retain it only as inactive residual structure. As a result, the semantic inference problem becomes underdetermined: multiple semantic hypotheses m_t are compatible with the same correct observations x_t.

If the system selects an incorrect semantic hypothesis m_t, a semantic error occurs even though observations are correct and system behavior remains admissible with respect to viability and regime constraints.

Since this error arises from structural underdetermination rather than perceptual noise or optimization failure, improving perception accuracy or action optimization cannot eliminate the risk of false meaning. □

Separation of behavioral stability and semantic correctness
Let $V(x_t)$ denote a Lyapunov-like functional characterizing behavioral admissibility (e.g., viability, bounded drift, or structural load), and let $W(m_t)$ denote a semantic correctness functional associated with causal binding or meaning.

It is possible that

$$\Delta V(x_t) \leq 0,$$

indicating stable and admissible behavior, while simultaneously $W(m_t)$ degrades or becomes misassigned due to semantic underdetermination.

Therefore, behavioral stability does not imply semantic correctness, even under perfect observations.

Corollary. 27.1 Irreducibility of false meaning
The risk of false meaning cannot be eliminated solely by improving the likelihood model $p(x \mid z)$ or by optimizing an action-level objective $J(a)$. False meaning arises from non-identifiability of latent semantic explanations under partial causal history, rather than from estimation error.

Theorem 28 — Conditional Primacy of Internal Time in Meaning Stabilization

Statement. Assuming that meaning stabilization requires persistence over a non-zero internal-time window, instantaneous causal metrics are insufficient to govern stabilization. Semantic stabilization and admissible semantic commitment are governed by accumulated internal time $\tau_{\mathrm{el}}(t)$ rather than by any instantaneous scalar score such as confidence, probability, or reward. When internal time slows (i.e. when $\tau_{\mathrm{el}}(t)$ accumulates more slowly per unit wall time), the earliest admissible commitment time is structurally delayed, preventing premature closure.

Assumptions (canonical, non-operational). We assume only the following architectural primitives.

(i) Internal time. There exists an accumulated internal-time quantity

$$\tau_{\mathrm{el}}(t) := \int_0^t \kappa(s)\, ds, \qquad \kappa(s) \in [0, 1],$$

monotone non-decreasing in wall time t. No further specification of κ is required.

(ii) Stabilization is defined over internal time. A hypothesis H_i is meaning-stabilized if a stability functional $S_i(\tau)$ remains above a threshold over a continuous internal-time window:

$$S_i(\tau) \geq \theta_S \quad \text{for all } \tau \in [\tau_1, \tau_1 + T_M],$$

for some window length $T_M > 0$. (Here S_i, θ_S, and T_M are abstract; no computability is implied.)

(iii) Commitment is admissibility-gated by stabilization. A semantic commitment operation is permitted only when the stabilization condition holds. Equivalently, the commitment operation must be authorized only if its realized effect-class satisfies

$$e(s, a, \tau, h) \in \mathcal{E}_{\mathrm{adm}} \quad \text{and} \quad H_i \text{ is stabilized over an internal-time window of length } T_M.$$

This is a structural admissibility rule, not an algorithm.

Proof 1 (structural timing argument). Define the earliest admissible wall-clock time at which commitment to H_i becomes permitted:

$$t_c := \inf \left\{ t \;:\; \exists \tau_1 \le \tau_{\mathrm{el}}(t) - T_M \text{ such that } S_i(\tau) \ge \theta_S \; \forall \tau \in [\tau_1, \tau_1 + T_M] \right\}.$$

This definition depends only on the accumulated internal time $\tau_{\mathrm{el}}(t)$.

Consider any wall-time interval $[t_0, t_1]$ on which internal time slows, i.e. the internal-time scaling $\kappa(t)$ decreases on that interval. Then the internal-time advance satisfies

$$\Delta \tau_{\mathrm{el}} = \tau_{\mathrm{el}}(t_1) - \tau_{\mathrm{el}}(t_0) = \int_{t_0}^{t_1} \kappa(s)\, ds,$$

which is strictly smaller when κ is smaller for the same wall-time duration $t_1 - t_0$. Since stabilization requires accumulation of an internal-time window of length T_M, any slowdown of τ_{el} delays the first wall-time t at which such a window can exist. Therefore t_c is non-decreasing under slowdown of κ. Hence, when internal time slows, admissible commitment is structurally delayed.

Proof 2 (conditional impossibility of instantaneous substitution). Assume, for contradiction, that commitment admissibility could be governed solely by an instantaneous scalar $q(t)$ (e.g. confidence, probability, reward), such that commitment is permitted at the first wall time t where

$$q(t) \ge \theta_q,$$

independently of τ_{el}.

Fix any wall time t^*. Construct two histories A and B such that

$$q_A(t^*) = q_B(t^*), \qquad \tau_{\mathrm{el}}^A(t^*) < \tau_{\mathrm{el}}^B(t^*),$$

which is possible by choosing different $\kappa(\cdot)$ trajectories while holding the same instantaneous score at t^*. By Assumption (ii), stabilization depends on the existence of an internal-time window of length T_M. Hence at time t^* it may be the case that H_i is stabilized in history B but not in history A, despite identical $q(t^*)$.

Therefore any rule that permits commitment based solely on $q(t)$ would authorize commitment in both histories at t^*, contradicting Assumption (iii) that commitment is admissibility-gated by internal-time stabilization. Thus, under the stated primitives, no instantaneous scalar $q(t)$ can substitute for internal-time-governed stabilization.

Remark (separation of behavioral admissibility and meaning stabilization). Let $V(t)$ denote any behavioral/trajectory admissibility functional (viability, drift bound, load) that may remain within admissible bounds while meaning stabilization is still forbidden because the internal-time window has not accumulated. This separation is structural: behavioral admissibility does not imply meaning stabilization, and meaning stabilization cannot be accelerated by instantaneous scoring without violating the internal-time definition of stabilization.

Theorem 29 - Regime Transition Requires Meaning-Level Stability

Transition between organizational regimes is admissible only under meaning-level stabilization. Context-level plausibility is insufficient for regime transition.

Proof sketch. Consider an adaptive system operating under a finite set of organizational regimes, where each regime corresponds to a distinct structural mode of operation. Let regime transitions incur a non-zero structural cost and potentially affect long-horizon viability or identity continuity.

Let m_t denote a semantic hypothesis formed from observations, and assume that m_t is context-level plausible but not meaning-level stabilized, i.e., it has not satisfied the internal-time stabilization condition.

Suppose, for contradiction, that a regime transition is permitted based solely on context-level plausibility of m_t. Then there exists a wall-clock time t^* at which:

1. the semantic hypothesis m_t appears locally coherent and plausible,
2. meaning-level stabilization has not been achieved,
3. a regime transition is nonetheless executed.

Because context-level plausibility may arise from transient coincidence, partial observation, or contextual salience, such a transition may be triggered by a hypothesis that is subsequently invalidated without contradiction in data. In this case, the system incurs irreversible structural cost due to the regime transition despite the absence of stabilized meaning.

This implies that regime transitions may be driven by transient semantic artifacts, leading to unnecessary structural load accumulation and increased risk of identity or viability degradation.

Therefore, allowing regime transitions under context-level plausibility alone violates the requirement of structural admissibility for long-horizon operation. Hence, regime transitions must be conditioned on meaning-level stabilization, and context-level plausibility is insufficient. □

Remark. Structural necessity of stabilization
Regime transitions alter the structural organization of the system and therefore cannot be treated as reversible or low-cost actions. Conditioning such transitions on meaning-level stabilization prevents structurally irreversible consequences arising from transient or unstable semantics.

Theorem 30 - Rc Is a Necessary Semantic Regime

Any system that admits interpretation and commitment must include a semantic containment regime in which meaning may form without structural impact. Absence of such a regime leads either to paralysis or uncontrolled transitions.

Proof sketch. Consider an adaptive system that (i) produces interpretations of observations and (ii) is capable of semantic commitment, where commitment denotes authorization for interpretations to influence irreversible structural operations (e.g., regime transitions or identity-affecting interventions).

Assume the system does *not* include any regime in which semantic hypotheses may form while being prevented from exerting structural impact. Then the system must follow one of the following two architectural patterns:

Case 1 (Always-commit coupling). Interpretation outputs are permitted to influence structural operations immediately, without requiring meaning-level stabilization. Since interpretation under partial history and contextual salience is structurally underdetermined, transient hypotheses can trigger irreversible structural changes. This yields uncontrolled transitions and unnecessary structural cost accumulation.

Case 2 (Never-commit decoupling). To prevent the failure in Case 1, the system may prohibit interpretations from influencing structural operations altogether, unless full certainty is obtained. However, full certainty is generally unattainable under partial observability and long-horizon interaction. As a result, the system becomes unable to authorize commitment when action is required, producing paralysis.

Because the system admits interpretation and commitment but lacks a semantic containment regime, it is forced into either Always-commit coupling (uncontrolled transitions) or Never-commit decoupling (paralysis). Both outcomes contradict long-horizon structural viability.

Therefore, any such system must include a distinct semantic containment regime, R_c, in which semantic hypotheses may form while structural impact remains gated until meaning-level stabilization is achieved. □

Remark. Rc as an architectural third mode
The necessity of R_c does not follow from performance optimization but from structural admissibility. R_c separates hypothesis formation from irreversible commitment, thereby avoiding both premature closure (uncontrolled transitions) and over-conservatism (paralysis).

Theorem 31 - Structural Self-Forgiveness

In any adaptive system operating over a long horizon with irreversible regime transitions, residual accumulation, and internal time, the absence of a mechanism for conditional release of self-penalty following closure of reversible errors necessarily leads to degradation of long-horizon viability or identity continuity.

Proof sketch. Assume an adaptive system subject to the following canonical constraints: (i) drift accumulation is inevitable, (ii) regime transitions incur non-zero structural cost, (iii) internal time bounds admissible structural evolution, and (iv) not all errors lead to irreversible damage.

Consider a class of reversible errors that do not violate viability or identity continuity and are subsequently structurally closed (i.e., transferred into residual form).

If the system retains active self-penalty or internal inhibition associated with such closed errors, then this penalty persists as an additional constraint on future adaptation. Over internal time, the accumulation of such constraints increases effective phase transition cost and suppresses admissible commitment.

As internal time progresses, the system is forced into one of two failure modes: either excessive inhibition leading to paralysis (zero or near-zero commitment), or compensatory over-regulation leading to instability and accelerated resource depletion.

Therefore, in the absence of a mechanism that conditionally releases self-penalty after closure of reversible errors, long-horizon viability or identity continuity cannot be preserved.

Hence, a mechanism of structural self-forgiveness is necessary. □

Remark 1. Normative clarification
Structural self-forgiveness does not imply error erasure, reward inflation, or tolerance of irreversible damage. It denotes the controlled release of internal inhibitory constraints after an error has been rendered non-active through structural closure. This mechanism operates independently of semantic interpretation accuracy and is required solely to prevent accumulation of self-imposed structural debt.

Remark 2. Structural interpretation
The effect described in Theorem 31 may be interpreted as accumulation of structural debt: persistent self-imposed inhibitory constraints progressively reduce the admissible space of regime transitions. In the limit, this contraction yields either a degenerate admissible set (paralysis) or

unstable oscillatory compensation (over-regulation). No specific accumulation model is assumed.

Theorem 32 - Structural Delay of Failure

Failure in long-horizon adaptive systems is structurally delayed with respect to action-level error. Systems do not collapse at the moment of incorrect action or local inadmissibility, but when accumulated structural consumption exhausts the remaining capacity for identity-preserving evolution.

This result follows as a structural consequence of performance–identity decoupling and silent degradation. Specifically, the existence of admissible behavior under ongoing identity degradation implies a necessary temporal offset between observable failure and the underlying structural cause.

Accordingly, failure is not an instantaneous event but the terminal manifestation of prior irreversible structural consumption.

Proof. Proof (Structural Sketch)

Consider an adaptive system operating under sustained directed interaction with a changing environment and finite structural capacity.

From the Performance–Identity Decoupling result, there exists a non-empty interval during which externally observable behavior remains admissible (bounded error, stable reward, acceptable trajectories) while internal identity-preserving structure is already degrading.

From the Silent Degradation under Regime Persistence result, structural burden may accumulate without producing immediate action-level violations or triggering corrective regulation, provided the system remains within admissible regimes.

Let $C(t)$ denote cumulative structural consumption (drift, latent residual, regime-induced load), and let $C_{\max}$ denote the finite identity-preserving capacity of the system. By construction, $C(t)$ increases monotonically under sustained operation, even when instantaneous performance remains acceptable.

Failure occurs when $C(t) \geq C_{\max}$, at which point identity-preserving evolution becomes impossible. However, no necessary condition requires instantaneous action-level error to occur at the moment when $C(t)$ crosses this boundary.

Therefore, the temporal point of observable failure is strictly delayed with respect to the causal accumulation of structural consumption. The system collapses not at the moment of incorrect action, but at the exhaustion of identity-preserving capacity accumulated across prior correct actions.

This establishes that failure is structurally delayed relative to action-level error and cannot, in general, be detected or predicted from instantaneous performance metrics alone. □

Remark - Failure as Terminal Manifestation, Not Causal Event. This theorem implies that observable failure in long-horizon adaptive systems is not a causal event but a terminal manifestation.

The true causal structure of failure lies in accumulated, irreversible structural consumption distributed across time, regimes, and interactions. Action-level errors, when they finally appear, serve only as surface indicators that identity-preserving capacity has already been exhausted.

Consequently, failure analysis based on last actions, local errors, or immediate triggers is structurally incomplete. Any framework that treats failure as an instantaneous outcome rather than as a delayed structural endpoint necessarily misattributes cause and obscures long-horizon risk.

Theorem 33 - Incompleteness of Optimization-Centric Regulation

No optimization of action-level correctness, reward maximization, or stability can, in principle, impose an upper bound on long-horizon structural consumption. Therefore, optimization-centric regulation is incomplete with respect to identity preservation.

Proof sketch. Consider an adaptive system operating under persistent directed interaction. Optimization acts locally on action selection, error minimization, or reward maximization, but does not eliminate directional asymmetry between action and consequence. As shown by the inevitability of drift and entropy growth, each admissible action contributes irreducible structural residuals.

Since structural consumption accumulates independently of instantaneous correctness, no optimization procedure operating at the action level can bound total structural cost over an unbounded horizon. Hence, optimization-centric regulation cannot, in principle, guarantee identity preservation.

□

Remark

This theorem does not deny the utility of optimization for short-horizon stability or performance. It establishes a structural incompleteness result: optimization alone cannot regulate long-horizon viability. Any architecture relying exclusively on action-level optimization necessarily remains blind to identity depletion and regime exhaustion.

Remark (Lyapunov Perspective)

Even if a global Lyapunov function exists that certifies bounded error, stability, or convergence of trajectories, such a function can only bound state evolution within a fixed admissible regime. It does not, in principle, bound accumulated structural consumption across regime transitions or over unbounded horizons.

Thus, Lyapunov stability of behavior does not imply preservation of identity or long-horizon viability. This result is not a limitation of Lyapunov theory, but a consequence of its action-centric and state-centric scope.

Theorem 34 - Dual Influence of the Present

In a long-horizon adaptive system, the present state is simultaneously shaped by two irreducible influences:

1. causal dynamics, determining what can happen, and
2. non-causal structural admissibility, determining what is allowed to happen.

These influences act concurrently in the present, but cannot be reduced to a single causal process.

Proof Sketch. Let $x(t)$ denote the causal state of the system and $u(t)$ an action. Causal evolution is governed by

$$x(t+1) = f(x(t), u(t)).$$

Let $S(t)$ denote an accumulated structural state encoding viability, identity continuity, and admissibility bounds. Structural accumulation follows

$$S(t+1) = S(t) + \Phi(x(t), u(t)).$$

Admissibility at time t is given by

$$\mathcal{A}(t) = \mathcal{A}(S(t)),$$

and is therefore not a function of $x(t)$ alone.

Well-Foundedness Barrier Between Admissibility and Action. Any operator, signal, or variable participating in the determination of admissibility cannot belong to the action domain without inducing circularity.

If admissibility becomes causally actionable, it collapses into penalty-based optimization and loses its non-causal structural character. Therefore admissibility is structurally prior to action selection, control, planning, learning, and optimization.

Action is non-primitive in long-horizon adaptivity: it is defined only after admissibility constraints are fixed. Any architecture that treats action selection as primitive with respect to admissibility is structurally misclassified.

Remark - Operational meaning and governance boundary

Theorem 34 separates two operators that coincide in time but differ in logical role: causal evolution produces state transitions, while structural admissibility constrains which transitions remain identity- and viability-consistent.

This separation is not a semantic preference but a governance boundary. If admissibility is exposed to the primary causal decision surface as an optimization signal (i.e., treated as causal feedback), the system collapses into meta-optimization and loses the ability to function as an independent witness of long-horizon degradation.

Accordingly, admissibility must remain structurally operative but non-instrumental: it may restrict, gate, or invalidate regimes, yet must not become a reward-like target. This is precisely the architectural stance motivating the intermediate regime often referred to as *cybernetics of order 2.5*: admissibility is present as a structural regulator, while third-order self-optimization over admissibility is intentionally forbidden.

Lyapunov-style structural argument.

Assume there exists a non-negative structural Lyapunov function

$$V(S(t)) \geq 0,$$

interpreted as a measure of structural viability or identity coherence.

By construction, V evolves monotonically under admissible operation:

$$V(S(t+1)) - V(S(t)) \leq 0 \quad \text{for admissible } u(t) \in \mathcal{A}(S(t)).$$

If an action $u(t)$ violates admissibility, then the induced structural update satisfies

$$V(S(t+1)) - V(S(t)) > 0,$$

indicating irreversible structural degradation.

Because V is defined on $S(t)$ rather than on $x(t)$, structural admissibility constrains present action without appearing in the causal transition function f.

Thus, causal dynamics determine reachable states, while non-causal admissibility - expressed via $V(S)$ - determines structurally allowable evolution.

The present state is therefore shaped simultaneously by causal possibility and non-causal structural admissibility, which are not reducible to a single causal operator.

□

Corollary

34.1. Irreversibility of Structural Constraint Violation
Once structural admissibility constraints are violated, no amount of locally correct causal action can restore admissibility without a change in the system's structural class.

In particular, post-hoc optimization, delayed correction, or reward-based compensation cannot recover admissibility after structural boundaries have been crossed.

34.2. Meaning as Admissibility, Not Inference
Meaning does not arise through causal inference from perception alone. Instead, meaning emerges when an interpretation is admitted by the current structural admissibility context.

Thus, identical perceptual or dynamical inputs may or may not acquire meaning depending on the admissible interpretative space defined by accumulated structural state.

34.3. Forbidden Feedback of Admissibility

If admissibility constraints are exposed to the primary causal agent as explicit causal feedback, the system necessarily collapses into meta-optimization and loses long-horizon viability.

Therefore, admissibility must not be represented as a reward signal, objective, explanation, or actionable map within the causal decision surface of the agent.

Theorem 35 — Internal-Time Persistence as a Necessary Condition for Commitment

Statement. In a long-horizon adaptive system, no semantic commitment can be structurally stabilized unless the corresponding interpretation induces admissible structural effect-classes continuously over a non-zero interval of internal time. Any commitment authorized on the basis of transient alignment in causal time necessarily induces irreversible structural cost, regardless of local correctness.

Proof Sketch. Let $S(t)$ denote the accumulated structural state of the system at evaluation cycle t, and let $C(t)$ denote the set of active semantic commitments. Let $e(t) \in \mathcal{E}$ denote the structural effect-class induced by the interpretation authorized at time t.

Define a non-negative structural load functional $V(S)$ such that V increases monotonically under irreversible commitment operations and decreases only under explicit structural reconfiguration.

Interpretative fluctuations that do not induce commitment correspond to bounded variations in causal state and do not modify $V(S)$.

Commitment authorization at time t introduces a strictly positive increment $\Delta V > 0$. If commitment is authorized without persistence across internal time, then under unavoidable drift the induced effect-class $e(t)$ decorrelates from the identity-compatible admissible region, while ΔV remains accumulated. Thus $V(S)$ increases without compensatory stabilization, violating boundedness over long horizons.

By standard Lyapunov-type arguments on monotone structural load, any admissibility policy that permits such transiently authorized commitments yields unbounded structural accumulation. Therefore, persistence of admissible effect-classes across a non-zero interval of internal time is a necessary condition for commitment admissibility.

Remark — Authorization vs. Inference. Inference answers the question: what appears to be the case now. Authorization answers the question: what may be irreversibly fixed. Conflating inference with authorization collapses admissibility into causal evaluation, induces meta-optimization, and destroys long-horizon viability. **Corollaries.**

35.1 — Premature Commitment as Structural Failure. Premature semantic commitment constitutes a structural failure mode distinct from perceptual error or inferential inaccuracy. Its cost is irreversible and accumulates independently of local correctness.

35.2 — Regime Transitions Cannot Be Triggered by Interpretation Alone. No regime transition or irreversible structural operation may be causally triggered by interpretation alone. Such transitions require independent admissibility authorization over effect-classes.

35.3 — Forbidden Feedback of Commitment Admissibility. Exposing commitment admissibility constraints to the primary adaptive process as causal feedback induces meta-optimization and collapse of system viability. Admissibility boundaries must therefore remain structurally enforced and non-causally opaque.

Theorem 36 — Necessity of Pre-Semantic Structural Revelation

Statement. In any long-horizon adaptive system that preserves identity and viability, there must exist a pre-semantic structural revelation layer: a mechanism by which admissibility-relevant structural constraints on future evolution become observable prior to meaning-level interpretation or commitment. In the absence of such a layer, semantic commitment necessarily induces irreversible structural degradation.

Proof (structural necessity by exhaustion of cases). Assume, for contradiction, that no pre-semantic structural revelation layer exists. Then admissibility-relevant structural constraints become observable only after semantic interpretation and/or meaning-level commitment.

Let $e(s, a, \tau, h) \in \mathcal{E}$ denote the structural effect-class realized by authorizing a semantic commitment under context (s, τ, h). By the irreversibility of commitment under finite remaining internal time (Theorem 2), semantic commitment cannot remain indefinitely provisional: it necessarily realizes a structural effect-class.

We consider the exhaustive cases.

Case (a) — Early Commitment. If commitment is authorized before admissibility-relevant structure is revealed, authorization proceeds without conditioning on whether

$$e(s, a, \tau, h) \in \mathcal{E}_{\text{adm}}.$$

Under non-zero drift accumulation, latent structural incompatibilities necessarily accrue, and at some finite horizon a prohibited effect-class

$$e(s, a, \tau, h) \notin \mathcal{E}_{\mathrm{adm}}$$

is realized. By non-restorability of admissibility, this induces irreversible regime transition or identity loss. Thus early commitment collapses by committing without admissibility.

Case (b) — Delayed Commitment. If the system delays commitment until admissibility-relevant constraints become observable, then such visibility occurs only once structural degradation has already become semantically expressible or operationally unavoidable. In this regime, revelation arrives only after a prohibited effect-class

$$e(s, a, \tau, h) \notin \mathcal{E}_{\mathrm{adm}}$$

has already been realized. Observation therefore arrives too late to preserve admissible continuation.

Both cases violate long-horizon viability: Case (a) collapses by premature irreversible commitment, Case (b) collapses by delayed revelation after admissibility loss.

Therefore, to preserve identity and long-horizon viability, admissibility-relevant structural constraints must be revealable prior to semantic interpretation and commitment. A pre-semantic structural revelation layer is necessary. □

Storage Argument (sketch). Let $S(t)$ denote accumulated structural state and let $\tau_{\mathrm{rem}}(t) \geq 0$ denote remaining internal-time budget. Assume $\tau_{\mathrm{rem}}(t)$ decreases monotonically under realized structural burden:

$$\tau_{\mathrm{rem}}(t+1) \leq \tau_{\mathrm{rem}}(t) - g(\Delta S(t)), \qquad g(z) \geq 0,\ g(z) > 0 \text{ for } z > 0.$$

Define $V(t) := \tau_{\mathrm{rem}}(t)$. Long-horizon viability requires $V(t) > 0$.

If admissibility-relevant structure becomes visible only after semantic commitment, then there exist intervals during which commitments are authorized without conditioning on $e(s, a, \tau, h) \in \mathcal{E}_{\mathrm{adm}}$. Hence there exists a finite horizon $T^* < \infty$ such that

$$\exists t < T^* : \quad e(s, a, \tau, h) \notin \mathcal{E}_{\mathrm{adm}}, \qquad V(T^*) = 0.$$

Thus, without pre-semantic revelation, remaining internal time is exhausted in finite operation. To avoid forced exhaustion under non-zero drift, admissibility-relevant structure must be revealable prior to meaning-level commitment.

Remark — Revelation Is Not Semantics. Pre-semantic structural revelation does not presuppose beliefs, propositions, or meaning-bearing representations. It refers solely to structural visibility sufficient to condition admissibility over effect-classes without inducing interpretation-level commitment. This preserves observational independence and avoids premature irreversibility.

Corollary 36.1 — Structural Priority Over Interpretation. In any long-horizon adaptive system, admissibility-relevant structural constraints must be operationally prior to meaning-level interpretation and commitment. Interpretation cannot be the first gate of irreversibility.

Corollary 36.2 — Necessity of Cybernetics of Order 2.5 Separation. Any architecture in which admissibility-relevant revelation becomes causally actionable at the level of interpretation or commitment collapses observational independence and renders admissibility optimizable. Therefore the pre-semantic revelation layer must remain visible but non-causal with respect to the primary adaptive process, placing the architecture strictly within cybernetics of order 2.5.

Theorem 37 - Incompatibility with Third-Order Cybernetics

In any long-horizon adaptive system, any architecture that renders pre-semantic structural revelation causally actionable necessarily collapses into third-order cybernetics and loses long-horizon viability.

Proof Sketch. Let $\mathcal{R}_{pre}(t)$ denote a pre-semantic structural revelation signal, exposing admissibility-relevant constraints without semantic commitment.

Assume that $\mathcal{R}_{pre}(t)$ is made causally actionable, that is, it directly influences the primary decision surface of the system.

Then reflexive structural observation becomes subject to optimization, suppression, delay, or instrumentalization. By the Corollary 34.3. (Forbidden Feedback of Admissibility), this introduces meta-optimization over admissibility itself, collapsing non-causal structural constraints into causal control objectives.

This condition is precisely the defining feature of third-order cybernetics: the system observes its own observing and injects this reflexivity into causal regulation.

However, long-horizon viability requires that admissibility constraints remain observable without becoming actionable, preserving observational independence.

Therefore, any architecture that causally operationalizes pre-semantic structural revelation is structurally incompatible with long-horizon adaptive viability.

□

Remark - Structural Boundary Between 2.5 and 3

Third-order cybernetics is not rejected here as a conceptual extension, but as a structural hazard.

The incompatibility arises not from insufficient reflexivity, but from excess reflexivity entering the causal surface, destroying the distinction between observation and regulation.

Cybernetics of the 2.5-th order preserves reflexive visibility while enforcing non-causality of admissibility revelation.

Corollary 37.1 - Silent Degradation Detectability

Silent structural degradation is detectable if and only if pre-semantic structural revelation remains non-causal.

If such revelation becomes causally actionable, degradation is either optimized away, deferred, or suppressed, eliminating the system's ability to witness its own long-horizon decay.

Theorem 38 - Non-Derivability of Admissibility from Any State Description

In any long-horizon adaptive system, admissibility of actions, regimes, or semantic commitments cannot be derived from any instantaneous, finite, or augmented state representation, including arbitrarily rich latent, belief, or world-model states.

Proof Sketch. Let $x(t)$ denote the instantaneous causal state of the system at time t. Let $S(t)$ denote the accumulated structural state, capturing irreversible structural burden, regime history, commitment closure status, and internal-time depletion. Let $A(t)$ denote the admissibility set at time t, i.e. the set of actions, regimes, or commitments authorized for execution at t. **Strengthening Note.** This impossibility does not arise from insufficiency of representational richness. It holds even for arbitrarily augmented, latent, predictive, epistemic, or world-model states. The prohibition is structural: any formulation of admissibility as a function of state,

$$\mathrm{Adm} = f(s),$$

collapses admissibility into causal evaluation and therefore violates Axiom XLI by construction. No extension of state space can repair this collapse, because the forbidden object is not the state, but the functional dependence itself.

Assume, for contradiction, that admissibility is state-derivable. That is, there exists a mapping

$$\mathcal{F} : \mathcal{X} \to \mathcal{P}(\mathcal{U})$$

such that

$$A(t) = \mathcal{F}(x(t)),$$

where $\mathcal{U}$ denotes the space of candidate actions and $\mathcal{P}(\mathcal{U})$ denotes the power set.

Consider the class of long-horizon systems in which admissibility constraints arise from predictable future degradation of coherence, regime stability, or identity continuity, rather than from violations detectable in the present causal state. Formally, assume there exists a nonnegative structural burden functional

$$B(t) = \mathcal{B}(S(t)) \geq 0,$$

and an admissibility bound $B_{\max}$ such that violation of $B(t) \leq B_{\max}$ renders a class of actions or commitments inadmissible. Assume $B(t)$ evolves under irreversible effects and drift such that for any nontrivial long-horizon operation there exist trajectories where $B(t)$ increases while $x(t)$ remains within a bounded locally admissible region.

Now construct two operational episodes E_A and E_B of the same system such that at a given time t_0,

$$x_A(t_0) = x_B(t_0),$$

but their structural histories differ:

$$S_A(t_0) \neq S_B(t_0),$$

in particular

$$B_A(t_0) \neq B_B(t_0),$$

with

$$B_A(t_0) \leq B_{\max} \quad \text{and} \quad B_B(t_0) > B_{\max}.$$

Such pairs exist because $S(t)$ contains irreversibility, regime occupation history, and internal-time consumption which are not uniquely determined by $x(t)$. Two episodes can coincide in instantaneous causal state while differing in accumulated structural state.

Since admissibility depends on the structural bound, it follows that

$$A_A(t_0) \neq A_B(t_0),$$

because at least one class of candidates is admissible under $B_A(t_0) \leq B_{\max}$ but inadmissible under $B_B(t_0) > B_{\max}$.

But if $A(t) = \mathcal{F}(x(t))$, then $x_A(t_0) = x_B(t_0)$ implies

$$A_A(t_0) = \mathcal{F}(x_A(t_0)) = \mathcal{F}(x_B(t_0)) = A_B(t_0),$$

a contradiction.

Therefore admissibility cannot be derived from any instantaneous causal state description. Since the assumption of state-derivability fails even when $x(t)$ is arbitrarily augmented as a belief state or world-model state, the result extends to any finite or causal augmentation whose value is determined at time t from present information.

Hence admissibility is non-derivable from state and must be treated as a structural operator depending on accumulated structural state $S(t)$ rather than on $x(t)$.
Structural Closure Clause. The impossibility result does not depend on the richness, dimensionality, or epistemic content of the state representation. Any formulation of the form

$$A(t) = \mathcal{F}(x(t))$$

collapses admissibility into causal evaluation, regardless of whether $x(t)$ denotes a raw physical state, a belief state, a latent embedding, or a learned world model. Thus the prohibition applies to the functional form itself, not to limitations of state expressiveness.

Remark - Augmentation Does Not Escape the Contradiction. Let $\hat{x}(t)$ be any augmented state representation computed from the causal process, including latent states, belief states, predictive world models, or confidence vectors. If $\hat{x}(t)$ is computed without direct access to $S(t)$ as an independent accumulated variable, then $\hat{x}(t)$ remains a function of present causal information. Thus the same construction applies: there exist episodes with identical $\hat{x}(t_0)$ but distinct structural burden, yielding distinct admissibility sets. Therefore no augmentation confined to present causal computation can make admissibility state-derivable.

Corollary 38.1. State-Based Safety Is Structurally Incomplete. Any architecture that represents admissibility as a predicate of instantaneous state, including safety classifiers, risk critics, constraint evaluators, or confidence thresholds, is structurally incomplete for long-horizon viability. It cannot distinguish episodes with equal state but unequal structural burden.

Corollary 38.2. Admissibility Requires an Independent Structural Variable. If admissibility is to be well-defined, the system must maintain an accumulated structural variable $S(t)$ (or an equivalent non-causal structural witness) such that admissibility is a function of $S(t)$. Therefore admissibility must be architecturally separated from state estimation and action evaluation.

Corollary 38.3. Forbidden Feedback of Admissibility Derivation. Any attempt to expose admissibility as a state-derivable causal signal to the primary decision surface induces meta-optimization over admissibility itself, collapsing the non-causal separation required for long-horizon viability. Therefore admissibility boundaries must remain structurally enforced and non-causally opaque.

Theorem 39 - Impossibility of Causal Authorization of Semantic Commitment

Statement. No causal inference process, confidence measure, correctness criterion, or optimization signal is sufficient to authorize semantic commitment in a long-horizon adaptive system without violating identity preservation.

Proof Sketch. Let $x(t)$ denote the instantaneous causal state of the system. Let $S(t)$ denote the accumulated structural state, incorporating irreversible commitments, regime occupation history, and internal-time consumption. Let $I(t)$ denote the space of interpretations available to the system at time t, and let $C(t) \subseteq I(t)$ denote the set of semantic commitments that have been irreversibly fixed.

Assume, for contradiction, that there exists a causal authorization rule

$$\mathcal{G} : \mathcal{X} \to \{0, 1\}$$

such that a semantic commitment $c \in I(t)$ is authorized at time t if and only if

$$\mathcal{G}(x(t)) = 1,$$

where $\mathcal{G}$ may depend on confidence, correctness, convergence, optimization score, or any other causal signal derived from $x(t)$.

Semantic commitment is an irreversible operation: it induces a non-negative increment $\Delta S_c > 0$ in the accumulated structural state,

$$S(t^+) = S(t) + \Delta S_c,$$

and cannot be undone without a change of structural class. Interpretations, by contrast, are causally revisable and correspond to bounded fluctuations in $x(t)$ that do not necessarily alter $S(t)$.

Because $x(t)$ evolves under drift and environmental interaction, there exist times $t_1 < t_2$ such that

$$x(t_1) \approx x(t_2),$$

while the future structural consequences of committing to the same interpretation differ due to differences in $S(t_1)$ and $S(t_2)$. In particular, there exist cases where commitment at t_1 remains admissible over internal time, while commitment at t_2 induces irreversible structural overload.

Since $\mathcal{G}$ depends only on $x(t)$, it authorizes commitment equally at t_1 and t_2. Thus the authorization rule cannot discriminate between structurally admissible and inadmissible commitments when causal state is locally similar.

As a result, there exist trajectories in which commitment is causally authorized while violating long-horizon identity constraints. Such violations accumulate irreversibly in $S(t)$ and cannot be compensated by later correct causal behavior.

This contradicts the requirement of identity preservation in long-horizon adaptive systems. Therefore no causal authorization rule based on inference, confidence, correctness, or optimization can safely authorize semantic commitment.

Remark - Authorization vs. Inference. Inference answers the question: what appears to be the case now. Authorization answers the question: what may be irreversibly fixed. Conflating inference with authorization collapses irreversible structural operations into causal evaluation and exposes identity-defining decisions to transient alignment and drift.

Corollary 39.1. Commitment Cannot Be Triggered by Confidence or Convergence. No level of confidence, predictive accuracy, convergence, or local optimality is sufficient to authorize semantic commitment. Such criteria are functions of causal alignment and are orthogonal to long-horizon structural viability.

Corollary 39.2. Threshold-Based Commitment Is Structurally Unsafe. Any architecture that authorizes commitment by threshold crossings, score maximization, or convergence detection necessarily admits irreversible structural error, even under locally correct operation.

Corollary 39.3. Forbidden Feedback of Commitment Admissibility. Exposing commitment admissibility as a causal feedback signal to the primary adaptive agent induces meta-optimization over commitment itself, collapsing the separation between interpretation and identity fixation. Therefore commitment admissibility must remain structurally enforced and non-causally opaque.

Theorem 40 - Collapse of Admissibility Under Optimization-Based Representation

Statement. Any representation of future-degradation constraints as scalar quantities, penalties, rewards, costs, or losses induces optimization pressure that necessarily invalidates their protective function.

Proof Sketch. Let $x(t)$ denote the instantaneous causal state of the system and let $S(t)$ denote the accumulated structural state. Let $A(t)$ denote the admissibility predicate over candidate actions, regimes, or commitments.

Assume, for contradiction, that admissibility constraints are representable as a scalar-valued quantity. That is, assume there exists a function

$$J_{\mathrm{adm}} : \mathcal{X} \to \mathbb{R}$$

such that admissibility is enforced by minimizing, bounding, or trading off J_{adm} within an optimization process.

Optimization-based decision-making selects actions $u(t)$ according to

$$u(t) = \arg\min_{u \in \mathcal{U}} \mathbb{E}\big[J_{\mathrm{task}}(x(t), u) + \lambda J_{\mathrm{adm}}(x(t), u)\big],$$

or an equivalent maximization formulation, for some weight $\lambda > 0$.

By construction, scalar quantities entering an objective function are subject to trade-off. For any finite λ, there exist task pressures, uncertainties, or external incentives such that an increase in J_{adm} is accepted in exchange for improvement in J_{task}. Thus inadmissible candidates are not excluded from consideration, but remain present within the optimizer's comparison set.

Moreover, optimization induces representational pressure. Once admissibility is encoded as a scalar, the optimizer can: (i) defer admissibility violations, (ii) smooth or average them over time, (iii) exploit regions where penalties are locally small but structurally irreversible, or (iv) instrumentalize violations to maintain short-horizon performance.

Future-degradation constraints, however, are characterized by delayed, non-local, and irreversible effects on $S(t)$. Their protective function requires categorical exclusion prior to evaluation, not numerical comparison during evaluation.

Therefore representing admissibility as a scalar quantity collapses the distinction between "forbidden" and "suboptimal". Inadmissibility is reduced to a cost, and optimization pressure necessarily drives the system toward trajectories that violate long-horizon structural bounds while preserving local optimality.

This contradicts the definition of admissibility as a non-tradable structural constraint. Hence any optimization-based representation of admissibility invalidates its protective function.

Remark - Trade-Off Is Structural Failure. Optimization presupposes comparability. Admissibility presupposes incomparability. Once future-degradation constraints enter a common numerical space with task objectives, they cease to function as constraints and become incentives to be optimized against.

Corollary 40.1. Penalty-Based Safety Is Not Admissibility. Any safety mechanism based on penalties, regularizers, reward shaping, or auxiliary losses is not an approximation of admissibility. It is structurally incapable of enforcing categorical exclusion of inadmissible trajectories.

Corollary 40.2. Soft Constraints Collapse Under Optimization Pressure. So-called "soft constraints" inevitably degrade into optimization variables. Under sustained interaction, they permit structurally irreversible violations while preserving short-horizon correctness.

Corollary 40.3. Admissibility Requires Domain Exclusion. For admissibility to retain its protective role, inadmissible actions, regimes, or commitments must be excluded from the domain of optimization altogether. They must not be represented, scored, compared, or traded, but rendered undefined for evaluation.

Theorem 41 - Equivalence Between Penalty-Based Safety and Third-Order Cybernetics

Statement. Any system that exposes admissibility-relevant constraints as optimizable signals is structurally equivalent to a third-order cybernetic system and inherits its long-horizon failure modes.

Proof Sketch. Let $x(t)$ denote the instantaneous causal state and let $S(t)$ denote the accumulated structural state governing admissibility, identity continuity, and internal-time consumption. Let $A(t)$ denote the admissibility predicate over actions, regimes, or commitments.

Assume that admissibility-relevant constraints are exposed to the primary adaptive process as optimizable signals. That is, assume there exists a causal signal $z(t)$ derived from $S(t)$ such that $z(t)$ enters the decision surface through optimization, comparison, or policy update. Formally, admissibility information becomes causally actionable.

Once admissibility enters the causal loop, the system is no longer merely constrained by admissibility, but begins to act *on* admissibility. The system observes not only its environment, but its own structural limits, and incorporates this observation into causal regulation.

This constitutes the defining feature of third-order cybernetics: the system observes its own observing and injects this reflexive information into the causal decision process. The distinction between witnessing structural limits and acting upon them collapses.

Under such conditions, admissibility signals are subject to optimization pressure. They may be suppressed, delayed, averaged, reweighted, or strategically violated to preserve short-horizon objectives. As shown in Theorem 40, optimization over admissibility necessarily invalidates its protective function.

Therefore any architecture that renders admissibility causally actionable is structurally indistinguishable from a third-order cybernetic system with respect to long-horizon behavior. It inherits the same failure modes: meta-optimization, loss of observational independence, and silent structural degradation preceding collapse.

Hence exposing admissibility as an optimizable signal is sufficient to induce third-order cybernetic equivalence, regardless of the system's intended order or design claims.

Remark - Reflexivity vs. Causal Accessibility. Third-order collapse is not caused by reflexive awareness itself. It is caused by causal accessibility of structural limits. A system may witness its own viability conditions without collapse if and only if such witnessing remains non-causal with respect to action selection.

Corollary 41.1. Cybernetics 2.5 as a Strict Equivalence Boundary. Any architecture in which admissibility-relevant information becomes causally actionable lies strictly outside Cybernetics 2.5. There is no continuous interpolation between admissibility witnessing and admissibility optimization; crossing this boundary induces third-order equivalence.

Corollary 41.2. Inevitability of Long-Horizon Failure Under Admissibility Opti-

mization. Once admissibility constraints are optimized rather than enforced as prohibitions, long-horizon failure becomes inevitable. Such failure is delayed, not prevented, and manifests as silent structural degradation followed by abrupt loss of identity or coherence.

Architectural consequence. Cybernetics 2.5 is not a midpoint but a strict equivalence class boundary.

Theorem 42 - Detectability Criterion for Silent Degradation (Given Pre-Semantic Revelation)

Statement. For silent structural degradation to be detectable without inducing collapse, admissibility-relevant constraints must become visible prior to semantic interpretation and independently of commitment processes.

Proof Sketch. Given Theorem 36 let $x(t)$ denote the instantaneous causal state and let $S(t)$ denote the accumulated structural state governing admissibility, regime viability, and internal-time consumption. Let $I(t)$ denote the space of interpretations and let $C(t) \subseteq I(t)$ denote the set of irreversible semantic commitments.

Silent structural degradation is defined as the accumulation of structural burden in $S(t)$ that does not manifest as immediate causal error, performance loss, or constraint violation in $x(t)$. Thus detection of such degradation cannot rely on causal performance signals.

Assume, for contradiction, that admissibility-relevant constraints become visible only after semantic interpretation or during commitment processes. That is, structural revelation is delayed until the system has already formed or fixed meaning.

Semantic interpretation operates on causal alignment, coherence, and plausibility within $x(t)$. Commitment operations are irreversible and induce a non-negative increment in structural state,

$$S(t^+) = S(t) + \Delta S_c, \qquad \Delta S_c > 0.$$

If structural constraints are revealed only after interpretation or commitment, the system faces a forced choice: either (i) it commits before constraints are visible, risking irreversible structural violation, or (ii) it delays commitment indefinitely, losing the ability to discriminate between admissible and inadmissible interpretations.

In case (i), silent degradation accumulates because commitments are fixed without prior admissibility awareness. In case (ii), the system becomes semantically inert and cannot stabilize meaning under finite internal time. Both cases contradict long-horizon viability.

Therefore admissibility-relevant structural constraints must become visible prior to semantic interpretation and independently of commitment authorization. This visibility must not be causally actionable, otherwise it would induce optimization over admissibility and collapse, as shown in Theorems 40 and 41.

Hence pre-semantic structural revelation is a necessary condition for detecting silent degradation without inducing collapse.

Remark - Visibility Without Control. Pre-semantic revelation is not a decision signal. It is a structural witness. Its role is to make degradation visible without providing the system with a causal handle to act upon that visibility.

Corollary 42.1. Meaning-Centric Detection Is Structurally Insufficient. Any architecture that attempts to detect long-horizon degradation exclusively at the level of meaning, interpretation, or belief revision is structurally blind to silent decay. By the time degradation becomes semantically apparent, irreversible structural cost has already been incurred.

Corollary 42.2. Non-Causal Revelation as a Survival Condition. For long-horizon adaptive systems, non-causal revelation of admissibility-relevant structure is not an optional monitoring feature but a survival condition. Removing or delaying such revelation guarantees either premature collapse or loss of semantic stabilization.

Theorem 43 - Non-Restorativity of Admissibility After Violation

Statement. Once admissibility is violated due to future-degradation constraints, no sequence of locally correct, optimized, or admissible causal actions can restore it without a change of structural class.

Proof Sketch. Let $x(t)$ denote the instantaneous causal state and let $S(t)$ denote the accumulated structural state. Let $A(t)$ denote the admissibility predicate governing actions, regimes, or commitments.

Admissibility violation is defined as the crossing of a structural boundary induced by future-degradation constraints. Such a violation corresponds to an irreversible increment in $S(t)$ that exhausts or exceeds the remaining internal-time budget or identity-preserving capacity of the system.

Assume, for contradiction, that there exists a sequence of causal actions $\{u(t)\}_{t>t_v}$, all locally correct, admissible, or optimized with respect to task objectives, that restores admissibility after a violation occurring at time t_v. Formally, assume there exists $t_r > t_v$ such that

$$A(t_r) = A(t_v^-),$$

without a change of structural class.

By construction, causal actions modify $x(t)$ and may temporarily improve performance, correct errors, or reestablish externally admissible behavior. However, causal actions do not reverse irreversible increments in $S(t)$ induced by future-degradation constraints. Structural burden accumulated through commitment irreversibility, regime exhaustion, or internal-time depletion is non-negative and non-invertible under causal evolution.

Thus any causal recovery sequence may restore local correctness in $x(t)$ while leaving $S(t)$ unchanged or further degraded. Consequently, the admissibility boundary remains violated even as behavior appears corrected.

To restore admissibility, the system must eliminate or reclassify the accumulated structural burden. This requires a change of structural class: regime transition, identity reset, architectural reconfiguration, or abandonment of prior commitments. Such operations lie outside the space of causal correction.

Therefore no sequence of locally correct or optimized causal actions can restore admissibility after violation without a change of structural class. The assumption of restorability leads to contradiction.

Remark - Performance Recovery vs. Structural Recovery. A system may recover acceptable behavior without recovering admissibility. Structural recovery is not an extension of causal correction; it is a qualitatively different operation that alters the space of admissible futures.

Corollary 43.1. Recovery Narratives Are Structurally Misleading. Narratives of recovery, rollback, re-alignment, or fine-tuning that do not involve explicit structural-class change describe only performance repair. They do not restore admissibility and therefore do not restore long-horizon viability.

Corollary 43.2. Delayed Failure After Apparent Recovery. Systems that violate admissibility and subsequently recover performance without structural change necessarily enter a delayed-failure regime. Collapse is postponed, not avoided, and occurs once remaining internal time is exhausted.

Corollary 43.3. Structural Change as the Only Form of True Recovery. Restoration of admissibility is possible if and only if the system undergoes a change of structural class. Such change may include regime abandonment, identity reset, or irreversible architectural reconfiguration, but cannot be achieved through causal optimization alone.

Theorem 44 — Necessity of Query Budget Under Non-Reconstructibility (NQB).

If an agent has access to an oracle of the form $\mathrm{Gate}(a) \to \{\mathrm{allow}, \mathrm{deny}\}$ and the number or rate of queries is unbounded, then the agent can asymptotically approximate the admissibility boundary in the candidate space within its hypothesis class. Therefore, without a bounded query budget or rate limitation, the property of non-reconstructibility of admissibility is not preserved.

Proof sketch. Let $\mathcal{A}$ be the candidate space, and let the (unknown) admissibility function be a binary classifier

$$g^\star : \mathcal{A} \to \{0,1\}, \qquad g^\star(a) = 1 \iff a \text{ is admissible.}$$

Assume the agent maintains a hypothesis class $\mathcal{H} \subseteq \{0,1\}^{\mathcal{A}}$ that contains $g^\star$ (or, more generally, contains a sequence h_n converging to $g^\star$ under the agent's representation). Each oracle query returns a label $y = g^\star(a)$ for a chosen $a \in \mathcal{A}$, i.e. a *membership query.*

Define the version set after n queries as

$$\mathcal{V}_n \;:=\; \Big\{h \in \mathcal{H} : \; \forall i \le n, \; h(a_i) = g^\star(a_i)\Big\}.$$

If the query budget is unbounded, the agent can choose queries adaptively to shrink $\mathcal{V}_n$. In standard identification/active-learning settings, there exist query strategies such that for any target accuracy $\varepsilon > 0$ (under any fixed choice of evaluation measure on $\mathcal{A}$ compatible with the hypothesis class), one can drive the disagreement region

$$\mathrm{Dis}(\mathcal{V}_n) \;:=\; \Big\{a \in \mathcal{A} : \; \exists h, h' \in \mathcal{V}_n \text{ with } h(a) \neq h'(a)\Big\}$$

to satisfy $\mu(\mathrm{Dis}(\mathcal{V}_n)) < \varepsilon$ for sufficiently large n, provided $\mathcal{H}$ has finite complexity (e.g. finite VC dimension) or otherwise admits consistent identification via membership queries. Intuitively: each oracle response provides at least one bit of information; with unlimited bits, the agent can eliminate incompatible hypotheses until the remaining hypotheses agree everywhere except on a set of arbitrarily small measure concentrated near the true boundary.

Equivalently, define a Lyapunov-like potential over the agent's uncertainty about admissibility:

$$\Phi(n) \;:=\; \mu(\mathrm{Dis}(\mathcal{V}_n)).$$

Under an adaptive query strategy that always queries inside $\mathrm{Dis}(\mathcal{V}_n)$, each new label strictly reduces $\mathcal{V}_n$ (unless $\Phi(n) = 0$ already), and thus (under mild regularity) yields a monotone descent $\Phi(n+1) \le \Phi(n)$ with $\Phi(n) \to 0$ as $n \to \infty$. This convergence means the agent can approximate the admissibility boundary to arbitrary precision within its own representational class.

Therefore, if the architecture requires non-reconstructibility (i.e., that the agent cannot asymptotically infer the admissibility boundary beyond a negligible level), then an unbounded membership-query interface to Gate contradicts that requirement. A bounded query budget or rate limitation is thus a necessary architectural condition for non-reconstructibility. □

Remark — Why this does not disclose implementation. The argument treats Gate only as a binary oracle and uses only generic properties of query-driven identification (version-space contraction, disagreement shrinkage, information accumulation). It does not assume any specific admissibility formula, estimator, metric, threshold, internal-time implementation, or viability geometry. The conclusion is purely architectural: unlimited oracle access is incompatible with non-reconstructibility.
This theorem does not prescribe query limitation as an implementation strategy. It establishes an impossibility boundary: unbounded causal access collapses admissibility into an optimizable object.

Any architectural response to this boundary is an engineering concern, external to this framework.

Corollary 44.1 — Budgeted non-reconstructibility condition. If the architecture enforces a per-episode or per-internal-time query budget $N_{\max}$ and/or a rate limit, then for any hypothesis class $\mathcal{H}$ requiring more than $N_{\max}$ membership queries to achieve boundary accuracy ε (under the agent's evaluation measure), the agent cannot reduce $\mu(\mathrm{Dis}(\mathcal{V}_n))$ below ε within the permitted interaction budget. Hence non-reconstructibility can be preserved up to a design-chosen resolution floor determined by the imposed budget.

Corollary 44.2 — Necessity of additional asymmetry beyond rate limits. Rate limits alone bound the *speed* of reconstruction but not its *eventual possibility* across long horizons. Therefore, for persistent agents, maintaining non-reconstructibility may additionally require (non-limiting examples) coarsening of observable outputs, randomized masking, aggregation across contexts, or privilege separation such that the agent cannot accumulate a faithful boundary model even as $t \to \infty$.

Drift as a Navigable Structural Medium. Drift is not an error term, noise, or residual defect. In long-horizon adaptive systems, drift constitutes a structural medium through which admissible continuation occurs.

Drift is unavoidable under directed interaction and cannot be causally suppressed without inducing one of two invariant failure modes: (i) silent latentization of structural degradation, or (ii) collapse of admissibility into optimization.

Navigation within drift is admissible if and only if the resulting deformation remains compatible with identity-preserving structural continuity. Once deformation exits the identity-compatible region, drift ceases to be navigable and becomes irreversible structural degradation.

Theorem 45 — ASC-Triggered Commitment Prohibition Is Architecturally Necessary (ATC).

In architectures where: (i) the generative operator is non-objectifiable, (ii) internal semantics may self-stabilize without increasing trace novelty, and (iii) commitment operations irreversibly contract the space of admissible future continuations, any system lacking a non-bypassable prohibition on commitment during Abstraction Self-Closure (ASC) admits scenarios in which irreversible commitment occurs under false maturity. Thus, prohibiting commitment under ASC is a minimal architectural safeguard rather than a policy preference.

Proof sketch. Let T_k denote the generative operator at level k, and let $\mathrm{Cons}_k(T_k)$ denote the set of higher-level generative operators consistent with the accumulated traces at that level. Assumption (i) (non-objectifiability) implies that $\mathrm{Cons}_k(T_k)$ cannot be collapsed to a singleton by direct observation or introspection; multiple generative explanations remain structurally compatible with observed behavior.

Assumption (ii) states that internal semantics can self-stabilize without an accompanying increase in trace novelty. Formally, there exist trajectories along which internal coherence measures (e.g., maturity functionals) increase or stabilize while the distance between newly observed traces and the span of previously accumulated traces remains below a fixed threshold. Under such conditions, internal indicators of readiness may improve even though no new external constraints have been imposed on $\mathrm{Cons}_k(T_k)$.

Assumption (iii) asserts that commitment operations induce an irreversible contraction of the admissible continuation space. That is, executing a commitment maps the system to a state whose future admissible envelope is a strict subset of the pre-commitment envelope, and this contraction cannot be undone by subsequent admissible dynamics.

Consider now an architecture without a non-bypassable prohibition on commitment during ASC. Because of (ii), there exist runs in which internal maturity signals increase while trace novelty remains low. Because of (i), these signals do not imply that $\mathrm{Cons}_k(T_k)$ has been sufficiently constrained to justify commitment. Nevertheless, in the absence of an architectural prohibition, the system is permitted to execute a commitment based solely on internal stabilization.

Once such a commitment is executed, (iii) guarantees an irreversible contraction of future admissible continuations. Since this contraction occurred without sufficient external constraint on $\mathrm{Cons}_k(T_k)$, the commitment is made under false maturity: the system behaves as if semantic stabilization reflected genuine structural certainty, when in fact the underlying generative ambiguity remains.

Therefore, any architecture lacking a non-bypassable prohibition on commitment during ASC admits executions in which irreversible commitments are triggered by internally self-consistent but externally underconstrained semantics. Preventing such scenarios requires an architectural mechanism that blocks commitment whenever ASC is detected, independently of policy-level preferences or optimization criteria. □

Remark — Architectural necessity versus policy choice. A policy that merely discourages commitment under ASC (e.g., via penalties or heuristics) remains part of the optimization landscape and may be overridden by surrogate objectives or internal stabilization pressures. By contrast, a non-bypassable architectural prohibition removes commitment actions from the space of representable or executable operations during ASC. The theorem therefore establishes necessity at the architectural level: without such a prohibition, false-maturity commitments are not merely possible but unavoidable under the stated conditions.

Corollary — Gating as a minimal safeguard. Any architecture intended to preserve long-horizon viability and identity continuity under non-objectifiable generative uncertainty must include a gating mechanism that disables commitment operations during ASC. This requirement is independent of the specific learning algorithms, estimators, or maturity metrics employed and follows solely from the structural properties (i)–(iii).

Clarification — Non-Objectifiability of Generative Operators. A generative operator is said to be *non-objectifiable* if no finite representation of its output space can be treated as an object of optimization, comparison, or control without collapsing admissibility into a causally actionable quantity.

This notion does not assert impossibility of description, but prohibits reuse of such descriptions as optimization targets without violating the admissibility–action separation. All results depending on non-objectifiability are to be read as conditional on this restriction.

Theorem 46 — Impossibility of Drift Suppression Without Structural Collapse.

Statement. In any long-horizon adaptive system under non-zero directed interaction, any architecture that treats drift as an error to be suppressed (i.e., aims to drive accumulated drift toward zero by causal regulation) necessarily enters one of two structural failure modes:

1. *Latentization:* degradation exits locally observed metrics and accumulates silently in the structural state,
2. *Collapse:* the regulation renders admissibility causally actionable as a scalar objective/penalty, inducing meta-optimization and destroying the non-causal admissibility boundary.

This is not a tuning failure but an invariant consequence of the admissibility–causality separation.

Proof (structural exhaustion by cases). Assume a system operates under non-zero directed interaction (Axiom 6), hence non-zero drift accumulation is unavoidable in the long-horizon regime.

Let the architecture attempt *drift suppression* by introducing a causal mechanism that uses some drift-related observable to reduce drift through action selection, learning updates, or optimization.

We analyze the exhaustive possibilities for how drift-related information can be used.

Case 1 (Drift-related signals are causally operative). Suppose the system obtains an internal signal $z(t)$ intended to represent drift (or drift risk) and $z(t)$ influences the causal decision surface (policy, planner, learner), i.e. becomes actionable for optimization or evaluation. Then observation is entangled with control (Axiom 15; Theorem 21): the system will act so as to alter $z(t)$. To make $z(t)$ optimization-compatible, the architecture must represent drift/admissibility pressure as a scalar quantity (loss, penalty, reward-shaping term, or trade-off). But admissibility and future-degradation constraints are *prohibitive* and cannot be represented as optimizable scalars without collapse (Axiom 38; Theorem 40). Therefore $z(t)$-based drift suppression induces optimization over admissibility-relevant structure, which is structurally equivalent to third-order collapse (Theorem 41; Lemma 6). Hence, treating drift as a suppressible error via actionable signals collapses the admissibility boundary and destroys long-horizon viability.

Actionability Assumption. Let $z(t) \in \mathcal{Z}$ denote any internal variable encoding drift-related or admissibility-relevant structure. The variable $z(t)$ is said to be *causally actionable* if it enters policy selection, planning, or learning updates in a manner that can influence action choice or parameter adjustment at evaluation cycle t.

Case 2 (Drift-related signals are not causally operative). Suppose instead that drift-related quantities are not actionable for the primary causal loop (i.e., the system either does not expose them, or they do not influence selection/learning). Then the drift-suppression aim cannot be realized as a causal control objective. The system may still attempt to keep locally observed mismatch small by continuous recomputation or high-frequency correction, effectively forcing residual toward zero at each step. By Theorem 3 (Residual as a Necessary Condition of Intelligence), denying residual as an admitted driver of regulation yields only two regimes: continuous active computation or collapse. Moreover, even when instantaneous residual remains small, integrated drift silently accumulates (Theorem 8), and structural burden can grow while trajectories remain locally admissible (Theorem 19; Theorem 16), producing delayed failure (Theorem 32). Thus drift is not eliminated; it is merely displaced out of locally observed metrics into latent structural accumulation, i.e. degradation becomes silent and delayed.

The two cases exhaust all ways drift-suppression can be architected: either drift-related information becomes actionable (yielding admissibility collapse via optimization), or it is not actionable (yielding latentization/continuous-computation and delayed collapse). Therefore drift suppression as an error-correction objective is structurally impossible without inducing one of these failure modes. □

Interpretation. Drift is a structural medium of continuation (Axiom 44), not a defect to be zeroed. Architectures survive by navigating drift under admissibility, not by attempting to suppress it through causal optimization.

Corollaries.

46.1 — Non-Equivalence of Local Accuracy and Drift Safety. Reducing instantaneous error, mismatch, or prediction residual does not constitute drift control and does not imply long-horizon viability. Drift can accumulate silently while local metrics remain bounded; therefore local correctness cannot serve as a surrogate for admissibility preservation.

46.2 — Forbidden Scalarization of Drift Constraints. Any attempt to represent drift risk, future-degradation constraints, or admissibility pressure as a scalar objective, penalty, reward shaping term, or trade-off variable necessarily negates their protective function and induces meta-optimization. Hence drift-related constraints must remain prohibitive rather than optimizable.

46.3 — Impossibility of Drift "Neutralization" by Monitoring Alone. Architectures that attempt to avoid collapse by monitoring drift while keeping the monitor non-actionable cannot achieve drift suppression; at best they displace degradation into latent structural accumulation and delayed failure. Therefore monitoring without admissibility-governed navigation cannot neutralize drift.

Remark — Drift Suppression vs. Drift Navigation. This theorem does not prohibit regulation under drift; it prohibits treating drift as an error to be driven toward zero by causal correction. In cybernetics of order 2.5, drift is navigated under admissibility constraints rather than suppressed by optimization. Any architecture that attempts to make drift "go away" either hides it from local observables (latentization) or renders admissibility actionable (collapse), thereby leaving the structural class.

Theorem 47 — Well-Foundedness Barrier: Anything That Shapes Admissibility Cannot Be an Action

Statement. In any long-horizon adaptive architecture, any operator, variable, or structural object that participates in determining admissibility preconditions (i.e., determines which structural contacts, interpretations, or continuations are permitted) cannot be an element of the action set. The action domain must be defined *after* admissibility is fixed; otherwise the architecture becomes circular and collapses into penalty-based action selection.

Proof Sketch. Let E denote the space of structurally reachable configurations and let $E_{\text{adm}} \subseteq E$ denote the admissible subset, defined by accumulated structural state and admissibility-relevant constraints. Let A denote the action set used by the primary adaptive process, and assume the canonical evaluation interface is defined over (s, a, τ, h).

By definition, admissibility must restrict the domain of permissible continuations: a continuation induced by choosing an action a is permitted only if it remains within E_{adm}. Thus, the meaning of A as an action domain is well-founded only if admissibility is already fixed, since the architecture must determine which continuations are permitted before it can treat any element as a valid candidate action.

Assume, for contradiction, that there exists an object u such that: (i) u participates in determining admissibility preconditions, i.e. $E_{\text{adm}} = E_{\text{adm}}(u, \cdot)$, and simultaneously (ii) $u \in A$, i.e. u is selectable by the primary adaptive process as an action.

Then the architecture must evaluate admissibility of selecting u using a domain E_{adm} that itself depends on u. This produces a non-degenerate circularity: admissibility depends on a choice whose permissibility is defined only after admissibility is fixed. To resolve the circularity causally, the architecture must replace domain restriction by an optimizable proxy (e.g., a scalar cost, penalty, or confidence threshold) and allow selection to proceed by trading off that proxy against other objectives.

However, representing admissibility preconditions as scalar, optimizable signals collapses the protective function of admissibility and induces optimization pressure that invalidates the boundary (Theorem 40), yielding structural equivalence with third-order cybernetics (Theorem 41). Therefore any architecture that treats an admissibility-shaping object as an action either remains ill-defined or collapses into penalty-based action selection and loses long-horizon viability.

Hence, anything that shapes admissibility cannot be an action; the action domain must be defined strictly after admissibility is fixed. □

Architectural consequence. Admissibility-shaping operators must remain structurally enforced and non-actionable. Any attempt to incorporate them into the action set transforms admissibility from a boundary into an objective, collapsing the architecture back into action-centric optimization. **Corollaries.**

47.1 — Forbidden Actionization of Structural Gates. Any operator that gates admissibility, regime entry, or phase transition is structurally forbidden from inclusion in the action space. Exposing such operators as selectable actions collapses the well-founded ordering between admissibility and action and forces optimization-based resolution.

47.2 — Impossibility of Admissibility-Conditioned Action Learning. No learning rule may legitimately condition the definition of the action domain on quantities that are themselves shaped by admissibility constraints. Architectures that attempt to "learn which actions are admissible" by optimizing over admissibility signals necessarily replace prohibition with penalty and lose structural viability.

47.3 — Structural Priority of Admissibility Over Control. In any architecture belonging to cybernetics of order 2.5, admissibility must be fixed prior to and independently of any control, planning, or optimization process. Violating this priority constitutes a change of structural class.

Remark — Well-Foundedness as a Class Boundary. This theorem establishes a well-foundedness requirement rather than a design preference. The separation between admissibility definition and action selection is not a matter of architectural cleanliness, but a necessary condition for preventing circular causation. Any attempt to soften this separation by introducing partial actionability, confidence thresholds, or learned proxies does not interpolate the boundary but crosses it, inducing the collapse mechanisms characterized in Theorems 40 and 41.

Structural Distance and Correlation Assumptions. Assume the existence of a non-negative structural divergence functional $d(\cdot, I)$ measuring distance to an identity-compatible set I. Assume further the existence of a sign-defined or bilinear correlation form $\langle\cdot,\cdot\rangle_{\mathrm{corr}}$ over incremental structural deformations, used solely as a structural discriminator of regime alignment and never exposed as a causal or optimizable signal.

Theorem 48 — Regime Bifurcation by Correlation: Damping vs. Admissible Amplification

Statement. When drift is represented as a field of structural deformation evolving over internal time, there exists a purely structural correlation form—independent of spectral decomposition or explicit frequency analysis—that induces a bifurcation between two regimes: anti-phase damping and in-phase amplification. Amplification is admissible if and only if the resulting deformation does not increase structural distance to the identity-manifold; once continuity is violated, amplification becomes structurally forbidden.

Proof Sketch. Let drift be modeled as an accumulated deformation field $\Delta(\tau)$ acting on the structural state $S(\tau)$ over internal time. Consider the correlation between incremental deformation $\mathrm{d}\Delta$ and the existing structural trajectory relative to the identity-manifold $\mathcal{I}$.

If successive deformations are anti-correlated with respect to the tangent of $\mathcal{I}$, their effects partially cancel, yielding a damping regime in which accumulated deformation reduces effective deviation from identity-compatible continuation. This corresponds to structural anti-phase alignment and results in admissible stabilization.

If successive deformations are positively correlated (in-phase), their effects compound and produce amplification. However, amplification remains admissible only while the induced deformation stays within the identity-compatible neighborhood of $\mathcal{I}$. Once correlated amplification increases structural distance from $\mathcal{I}$, admissibility is violated regardless of local correctness or short-term performance.

No spectral or metric evaluation is required for this bifurcation: it arises from the sign and structural orientation of correlation alone. Thus, the regime transition between damping and amplification is determined by correlation relative to identity continuity, not by magnitude optimization or frequency analysis.

Corollaries.

48.1 — Conditional Admissibility of Amplification. Amplification is not intrinsically unsafe; it is admissible precisely when correlated deformation preserves or reduces structural distance to the identity-manifold. Unconditional amplification policies therefore constitute a structural violation.

48.2 — Insufficiency of Energy- or Gain-Based Criteria. Energy increase, gain, or signal magnitude alone cannot distinguish admissible amplification from destructive runaway. Only correlation with identity-preserving structure determines admissibility of amplification under drift.

48.3 — Structural Origin of Anti-Wave Regulation. Anti-wave or damping behavior need not be imposed as an external control objective. It emerges structurally whenever deformation correlations oppose identity-divergent drift, establishing damping as a consequence of admissibility rather than optimization.

Remark 1 — Correlation as a Structural, Not Causal, Operator. The correlation form invoked here is not a causal signal available for control or optimization. Exposing correlation as an actionable quantity would immediately collapse admissibility into meta-optimization and re-enter the failure modes characterized in Theorems 40 and 41. Correlation functions here as a structural discriminator of regimes, not as a control variable.

Remark 2 — Correlation Without Measurement. The correlation form referenced in this theorem does not require numerical evaluation or metric structure.

Only the structural sign of correlation with identity continuity is admissibility-relevant.

Positive correlation denotes amplification that preserves identity coherence; negative correlation denotes amplification that accelerates structural degradation. No magnitude, gradient, or scalar score is assumed or required.

Section C - Lemmas

Lemma 1 — Structural Dependence of Admissibility

In a long-horizon adaptive system, the admissibility set $\mathcal{A}(t)$ depends on the accumulated structural state $S(t)$ and not on the instantaneous causal state $x(t)$ alone.

Proof Sketch. Assume, for contradiction, that admissibility were a function solely of the instantaneous causal state, $\mathcal{A}(t) = \mathcal{A}(x(t))$.

Then admissibility of any authorized operation would depend only on the current causal configuration, and any admissibility violation could in principle be reversed by a suitable causal update restoring $x(t)$ to an admissible region.

However, in long-horizon adaptive systems, admissibility is enforced over realized structural effect-classes $e(s, a, \tau, h)$ and is non-restorable once violated. There exist effect-classes $e(s, a, \tau, h) \notin \mathcal{E}_{\text{adm}}$ whose realization irreversibly constrains future evolution, independently of subsequent causal correctness or local state recovery.

Since such violations cannot be eliminated by adjusting $x(t)$ alone, admissibility cannot be determined solely by instantaneous causal state. Therefore admissibility must depend on an accumulated structural variable $S(t)$, which encodes irreversible history and cannot be reduced to causal state dynamics.

□

Lemma 2 — Confidence Does Not Imply Admissibility

Statement. Interpretive confidence, predictive accuracy, convergence guarantees, or short-horizon optimality do not constitute valid criteria for semantic commitment admissibility in a long-horizon adaptive system.

Proof Sketch. Let $x(t)$ denote the causal state of the system at time t, and let $S(t)$ denote the accumulated structural state. Let admissibility be determined by realized structural effect-classes:

$$e(s, a, \tau, h) \in \mathcal{E}_{\text{adm}} \quad \text{or} \quad e(s, a, \tau, h) \notin \mathcal{E}_{\text{adm}}.$$

Interpretive confidence, accuracy, or optimality are functions of local causal alignment between internal models and observed data, i.e. they depend on $x(t)$ and finite-horizon predictive consistency. They provide no information about the accumulated structural state $S(t)$ or about future-degradation constraints encoded in $\mathcal{E}_{\mathrm{adm}}$.

Since admissibility is a non-causal structural constraint, there exists no function

$$f : \mathrm{Confidence}(x(t)) \to \mathcal{E}_{\mathrm{adm}}$$

that preserves admissibility without collapsing it into causal evaluation. Any attempt to construct such a mapping would render admissibility derivable from state or confidence, violating the non-causality of structural constraints and inducing optimization pressure over admissibility-relevant structure.

Therefore interpretive confidence, predictive accuracy, or local optimality cannot imply admissibility of commitment. □

Corollary. Any architecture that authorizes semantic commitment based on confidence thresholds, prediction accuracy, or convergence criteria necessarily risks irreversible structural violation and loss of long-horizon viability.

Lemma 3 — Non-Causal Visibility of Structural Properties

In a long-horizon adaptive system, there exist structural properties that are observable without being causally actionable by the primary adaptive process.

Proof Sketch. Assume, for contradiction, that every observable structural property is causally actionable by the primary agent. Then any observation relevant to long-horizon viability or identity must enter the decision surface as an actionable signal affecting action selection, learning, or optimization.

Let such an observation concern admissibility-relevant structure, i.e. information sufficient to distinguish whether a realized effect-class satisfies

$$e(s, a, \tau, h) \in \mathcal{E}_{\mathrm{adm}} \quad \text{or} \quad e(s, a, \tau, h) \notin \mathcal{E}_{\mathrm{adm}}.$$

If this observation is causally actionable, then admissibility-relevant structure becomes subject to optimization pressure, comparison, or trade-off. This renders admissibility derivable from causal signals and collapses prohibitive structural constraints into evaluative or penalty-based objectives.

Such a collapse contradicts the non-causal status of admissibility and its enforcement as domain exclusion over effect-classes. Therefore not all observable structural properties may be causally actionable.

Hence, for identity preservation and long-horizon viability, there must exist structural properties that are observable yet remain non-causal with respect to the primary agent.

□

Lemma 4 — Pre-Semantic Exposure of Structural Constraints

In a long-horizon adaptive system, admissibility-relevant structural constraints on future evolution are in principle exposable prior to semantic interpretation or meaning-level commitment.

Proof Sketch. Admissibility is determined by accumulated structural state and enforced over realized structural effect-classes $e(s, a, \tau, h)$, independently of semantic interpretation or belief content.

Semantic interpretation and meaning-level commitment require stabilization across internal time and are subject to irreversibility once authorized. They therefore cannot serve as prerequisites for detecting admissibility-relevant structure, since admissibility conditions must already hold at the point of authorization.

If admissibility-relevant constraints were only exposable after semantic interpretation or commitment, then admissibility would depend on semantic state, contradicting its structural and non-causal character.

Therefore structural constraints governing admissible continuation must be detectable prior to semantic interpretation and independently of meaning-level commitment.

□

Lemma 5 — Penalty-Based Safety Is Not Equivalent to Admissibility

Statement. Any architecture that represents inadmissibility exclusively via penalties, reward shaping, soft constraints, or loss terms internal to optimization is not structurally equivalent to an admissibility-first architecture.

Proof Sketch. Let admissibility be defined over structural effect-classes:

$$e(s, a, \tau, h) \in \mathcal{E}_{\text{adm}} \quad \text{or} \quad e(s, a, \tau, h) \notin \mathcal{E}_{\text{adm}}.$$

Penalty-based approaches necessarily require inadmissible candidates to be representable, comparable, and scorable within the optimization domain. That is, for any candidate inducing $e(s, a, \tau, h) \notin \mathcal{E}_{\text{adm}}$, a finite penalty value must be assigned in order for optimization to proceed.

As a consequence, inadmissible effect-classes remain elements of the optimizer's input space and are subject to optimization pressure, trade-offs, and strategic suppression. This renders admissibility causally actionable and collapses prohibitive constraints into scalar objectives.

In contrast, an admissibility-first architecture enforces exclusion at the level of effect-classes: candidates inducing $e(s, a, \tau, h) \notin \mathcal{E}_{\text{adm}}$ are undefined for evaluation and never enter the optimization domain. No comparison, scoring, or trade-off is possible.

Therefore penalty-based safety and admissibility enforcement differ not by degree or tuning, but by architectural class. Penalty-based methods cannot approximate admissibility without negating its structural role.

Corollary. No choice of penalty magnitude, shaping schedule, or optimization horizon can render a penalty-based safety architecture equivalent to admissibility enforcement over effect-classes.

Lemma 6 - No-Interpolation Between Visibility and Actionability

There exists no continuous architectural interpolation between (i) admissibility-relevant structural visibility that is non-causal and non-actionable, and (ii) admissibility-relevant structural visibility that is causally actionable for optimization. Crossing this boundary constitutes a change of structural class and induces the third-order failure modes established in Theorem 41.

Proof Sketch. Assume that there exists a continuous architectural interpolation between non-causal, non-actionable visibility of admissibility-relevant structure and causal, actionable visibility usable for optimization. Then there must exist an intermediate regime in which admissibility-relevant information is partially actionable while remaining structurally protective.

However, any degree of causal actionability renders admissibility subject to optimization pressure. By Theorem 41, exposing admissibility-relevant structure to optimization induces structural equivalence with third-order cybernetics and its associated failure modes. Conversely, fully non-actionable visibility exerts no causal influence and preserves the protective boundary.

Therefore no intermediate regime can exist that preserves admissibility while allowing partial actionability. Crossing the boundary from visibility to actionability constitutes a discrete change of structural class, not a continuous deformation. Even arbitrarily weak causal leakage of admissibility-relevant information, when sustained over unbounded horizons, is sufficient to reconstruct the admissibility boundary and induce optimization pressure. □

Lemma 7 — Counterfactual Leakage Lemma (CLL).

If a system permits the computation of counterfactual quantities involving blocked operations $a \in A_{\text{block}}(t)$, including values, gradients, labels, trajectories, or rollouts, then blocked operations become part of the system's internal dynamics. Consequently, the architectural separation between permission and choice is violated, even if blocked operations are never executed.

Proof Sketch. Counterfactual evaluation requires the system to instantiate internal representations of the blocked operation a, together with its predicted consequences under the system's transition, value, or learning dynamics. These representations necessarily influence internal state variables, update rules, or learned parameters, either directly (e.g., via gradients or targets) or indirectly (e.g., via normalization, comparison, or ranking). Therefore, the blocked operation contributes causal influence to the system's evolution despite non-execution. This contradicts the architectural requirement that inadmissible operations be excluded prior to evaluation, establishing that counterfactual access constitutes semantic leakage. □

Lemma 8 — Policy–Mechanism Separation Lemma (PMS).

Any prohibition implemented at the policy or objective level, including penalties, constraints, or planner rules, is part of the optimization landscape and may be circumvented by the same optimization process that governs behavior. An admissibility gate operating outside the optimization domain and rendering blocked operations non-representable for evaluation is not subject to such circumvention and is structurally distinct from policy-level restrictions.

Proof Sketch. A policy-level prohibition modifies the objective function, constraints, or decision rules within the optimization process. As such, it remains an element of the search space over which optimization is performed and can be indirectly optimized against through surrogate objectives, proxy actions, reparameterization, or exploitation of blind spots. In contrast, an admissibility gate that removes blocked operations from the representational domain of evaluation eliminates those operations from the optimizer's search space entirely. Because the optimizer cannot score, compare, or reason about non-representable options, the gate cannot be bypassed by optimization dynamics. Thus, policy-level prohibitions and architectural admissibility gates are not functionally equivalent, and only the latter provides non-bypassable exclusion. □

Lemma 9 — Phase-Consistency Requirement for Drift Navigability

Statement. Drift is phase-coupled: a deformation that is admissible under one interpretative phase may be inadmissible under another. Therefore, drift is navigable only under phase-consistent continuation; otherwise drift ceases to carry navigation and becomes indistinguishable from silent structural degradation.

Proof Sketch. Let (s, τ, h) denote the evaluation tuple and let ϕ denote an interpretative phase (or phase-indexed regime of coherence closure) under which admissibility is assessed. Let $E^{\phi}_{\text{adm}} \subseteq E$ denote the admissible configuration set under phase ϕ.

Assume there exists a deformation step Δ such that for a configuration $e \in E$,

$$e \in E^{\phi_1}_{\text{adm}} \quad \text{and} \quad \Delta(e) \in E^{\phi_1}_{\text{adm}},$$

but

$$\Delta(e) \notin E^{\phi_2}_{\text{adm}}$$

for some other phase ϕ_2. Then the same physical or structural continuation Δ is admissible in one phase and forbidden in another.

If drift navigation is permitted without phase-consistency, the system can traverse sequences of deformations that remain locally admissible under transient phase alignment while accumulating latent mismatch relative to the phase in which identity continuity must ultimately be preserved. Because admissibility depends on accumulated structural state and phase context, this mismatch does not necessarily manifest as local causal error; it remains silent until a phase reconciliation or regime stabilization is required, at which point admissibility collapses.

Therefore, without phase-consistent continuation, drift cannot function as a navigable medium: it permits silent accumulation of structurally incompatible deformation and becomes indistinguishable from irreversible degradation. Hence phase-consistency is necessary for drift navigability.

Corollary. Any architecture that allows phase switching to authorize deformations without enforcing phase consistency necessarily permits silent structural debt accumulation and forfeits reliable drift navigation.
Remark. Action variables are intentionally omitted: the lemma concerns admissibility of structural continuation under drift, which is defined prior to and independently of action selection.

Lemma 10 — Identity Non-Objectification Clause (Non-Representability)

Statement. The mapping of an adaptive system's structure into an invariant profile need not be observable, computable, or representable within the system. Identity is not verified or optimized; it is imposed as a boundary on admissible deformation.

Proof Sketch. Let identity preservation be enforced by a deformation boundary represented abstractly as an admissibility restriction:

$$E_{\text{adm}} := \{e \in E \mid e \text{ is identity-compatible}\}.$$

No requirement is imposed that the predicate defining E_{adm} be internally representable as an explicit object, scalar score, or verifiable certificate accessible to the primary adaptive process.

Assume, for contradiction, that identity must be internally objectified as a computable invariant profile $\Pi(e)$ and that admissibility is decided by comparing $\Pi(e)$ to a target or threshold. Then identity becomes causally actionable: the system can attempt to optimize, imitate, or game the profile Π, thereby transforming identity from a boundary condition into an objective. By the established admissibility boundary results (incompatibility with optimization-based representation and causal exposure), such objectification collapses identity preservation into meta-optimization and undermines long-horizon viability.

Therefore the theory requires only that an identity-compatible admissibility boundary exists, not that it be internally represented. Identity is a constraint on allowable deformation, not an object to be computed or optimized. Hence non-representability is permitted and structurally preferable.

Corollary. Any architecture that treats identity as an explicit optimizable or verifiable internal object necessarily risks collapse into profile-optimization and violates the structural separation required for long-horizon identity continuity.

Lemma 11 — Operational Falsifiability of Internal Time via Contraction of Continuation

Statement. If, under repeated structural contacts or reinterpretations and accumulating irreversible burden, the set of admissible continuations does not contract (or if amplification remains admissible without restriction), then the system contradicts the definition of internal time as a finite viability horizon and does not belong to the long-horizon structural class.

Proof Sketch. Let $S(\tau)$ denote accumulated structural state indexed by internal time τ, and let $\mathcal{C}_{\mathrm{adm}}(\tau)$ denote the set of admissible continuations available at τ. Assume internal time is finite and is consumed by irreversible structural burden, so that there exists a nonnegative structural load functional $V(S)$ that increases under irreversible operations and cannot be reduced by locally correct causal action alone.

Then continued operation over τ implies non-decreasing burden:

$$V(S(\tau_2)) \geq V(S(\tau_1)) \quad \text{for } \tau_2 > \tau_1,$$

with strict increase whenever irreversible burden is incurred.

If $\mathcal{C}_{\mathrm{adm}}(\tau)$ fails to contract despite increasing burden, i.e.

$$\mathcal{C}_{\mathrm{adm}}(\tau_2) \supseteq \mathcal{C}_{\mathrm{adm}}(\tau_1) \quad \text{for some } \tau_2 > \tau_1,$$

or if an amplification class remains admissible uniformly across τ, then admissibility is insensitive to accumulated structural load. In that case internal time is not functioning as a finite viability horizon: continued burden does not reduce the space of permitted futures.

This contradicts the structural definition of internal time as a resource consumed by irreversible burden and drift. Therefore, in any system belonging to the stated class, admissible continuation must contract under accumulation of irreversible structural cost. Failure of contraction constitutes an operational falsifier of internal-time finiteness.

Corollary. A claimed long-horizon architecture that permits unbounded continuation or unrestricted amplification under accumulating irreversible burden is structurally misclassified; it implicitly assumes either infinite internal time or a collapse of admissibility into causal evaluation.

Cybernetics of Order 2.5

Central axiom — Navigational Cybernetics of Order 2.5

Long-horizon intelligence is not defined by avoidance of collapse, but by the capacity to navigate how collapse unfolds under irreversible drift.

Architectures of order 2.5 do not eliminate drift or terminality. They structure drift, phase, and admissibility such that collapse is delayed, phased, or rendered navigable without collapsing admissibility into optimization.

Definition — Cybernetics of Order 2.5 (Structural Class)

A long-horizon adaptive architecture belongs to cybernetics of order 2.5 if and only if it satisfies the following conditions simultaneously.

(Visibility without control). The architecture admits reflexive visibility of admissibility-relevant structural limits, including future-degradation constraints, while ensuring that such visibility is not causally actionable by the primary adaptive process.

(Prohibitive admissibility). Admissibility is enforced as a categorical restriction on the domain of admissible futures and cannot be represented, scored, optimized, or traded as a scalar objective, penalty, reward, or loss term.

Cybernetics of Order 2.5 — Structural Positioning

Definition — Cybernetics of Order 2.5 (Structural Class). A long-horizon adaptive architecture belongs to cybernetics of order 2.5 if and only if it satisfies the following conditions simultaneously. First, the architecture admits reflexive visibility of admissibility-relevant structural limits, including future-degradation constraints, while ensuring that such visibility is not causally actionable by the primary adaptive process. Second, admissibility is enforced as a categorical restriction on the domain of admissible futures and cannot be represented, scored, optimized, or traded as a scalar objective, penalty, reward, or loss term.

Let $\mathcal{A}$ denote an adaptive architecture and let Z denote any internal signal encoding admissibility-relevant structural information.

If Z is causally actionable for policy selection, optimization, or learning updates, then $\mathcal{A}$ is structurally equivalent to third-order cybernetics.

If Z is absent or cannot become visible prior to semantic interpretation, then $\mathcal{A}$ collapses into second-order insufficiency under silent structural degradation.

Therefore, $\mathcal{A}$ belongs to cybernetics of order 2.5 if and only if admissibility-relevant structure is visible but non-actionable.

The framework developed in this work occupies a deliberately constrained position between second- and third-order cybernetics. Classical second-order cybernetics incorporates the observer into the system but permits observational constructs to enter causal regulation, rendering it insufficient for preventing silent long-horizon degradation. Third-order cybernetics internalizes reflexivity itself, allowing systems to act upon descriptions of their own observing and to optimize against their structural limits.

The axioms, lemmas, and theorems established here demonstrate that both regimes are structurally incompatible with long-horizon viability. Once pre-semantic structural revelation becomes causally actionable, admissibility is optimized away, observational independence collapses, and identity-preserving coherence cannot be maintained. These failure modes are invariant consequences of exposing admissibility-relevant structure to causal control, not artifacts of particular implementations.

Cybernetics of order 2.5 designates a discrete architectural class defined by selective constraint rather than escalation. It preserves reflexive visibility of structural viability conditions while categorically prohibiting their causal exploitation. Observation is structurally acknowledged yet causally insulated; admissibility is enforced as a non-inferential boundary; regulation proceeds through exclusion of inadmissible directions rather than optimization of preferred ones.

This positioning is not a limitation or an incomplete ascent toward third-order cybernetics. It is an intentional architectural boundary. Systems in this class internalize the consequences of observation while refusing to act on that awareness causally. The separation between witnessing and acting is not an implementation detail but a necessary condition for maintaining long-horizon viability, identity preservation, and semantic integrity under irreversible dynamics.

Accordingly, cybernetics of order 2.5 defines a mathematically grounded regime within which analysis, classification, and engineering of long-horizon adaptive systems can proceed without collapsing into meta-optimization, action-centric reduction, or semantic drift.

Architectural Exclusion Criteria (Cybernetics of Order 2.5). An adaptive architecture is excluded from cybernetics of order 2.5 if any of the following conditions hold:

- Admissibility is represented as a scalar objective, penalty, reward, or loss term.
- Drift is treated as an error to be minimized rather than as a structural medium.
- Phase transitions remain reversible under sustained structural load.
- Identity continuity is inferred from behavioral similarity or task performance.
- Admissibility-relevant structure is causally actionable by the primary adaptive process.

Verification within this framework proceeds by exclusion, not by estimation. Satisfaction of any exclusion condition places the architecture outside the admissibility-first class.

Class Diagram (Orders 2, 2.5, and 3). *Second-order cybernetics* admits observation within the system but does not require visibility of admissibility-relevant structural limits; silent long-horizon degradation may remain undetected until boundary crossing.

Third-order cybernetics admits reflexive visibility *and* makes it causally actionable: admissibility-relevant signals enter optimization, planning, or learning, collapsing admissibility into trade-offs.

Navigational Cybernetics 2.5 occupies the strict boundary between them: admissibility-relevant structural limits are visible yet categorically insulated from causal actionability. This separation preserves long-horizon viability without collapsing into meta-optimization.

Remark — Non-Escalatory Structural Regime. Cybernetics of order 2.5 is not an incomplete form of third-order cybernetics and not a transitional regime between second and third order.

It is defined by selective structural prohibition rather than escalation: reflexive visibility of admissibility-relevant structure is permitted, while its causal exploitation is categorically forbidden.

This boundary is not an implementation choice but a necessary architectural constraint for preserving identity continuity, coherence, and long-horizon viability under irreversible structural dynamics.

Hypothetical Examples of Order 2.5 Architectures

How to Read the Examples in This Work. The examples presented in this section are not intended as implementations, design templates, or constructive recipes.

Each example serves a classificatory function: it demonstrates how a concrete system does or does not satisfy the architectural exclusion criteria of Navigational Cybernetics 2.5.

Positive examples (Examples 1–4) illustrate systems in which admissibility-relevant structure is visible yet non-actionable, and where long-horizon viability depends on drift navigation rather than error correction or optimization.

Negative examples (Examples 5–6) demonstrate architectures that may appear safe, stable, or performant, but in which admissibility is represented as a scalar, penalty, or relaxable constraint, thereby collapsing into optimization.

The purpose of these examples is not to show how to build such systems, but to make the boundary of the class explicit: membership is determined by structural relations, not by surface behavior or empirical success.

Example 1 — Abstract Architecture with Drift Navigation

Consider an abstract adaptive system $\mathcal{X}$ operating under sustained interaction with a non-stationary environment.

The system admits:

- irreversible structural drift under interaction,
- finite internal time,
- regime-dependent propagation of structural effects.

The system does not optimize performance metrics, does not minimize error, and does not treat drift as noise or disturbance.

Admissibility in $\mathcal{X}$ is enforced structurally: effect-classes outside E_{adm} are not realizable within the architecture, without being penalized, scored, or corrected.

Admissibility-relevant structural limits are internally visible to the system but are categorically non-actionable: no admissibility signal enters policy selection, learning updates, or control loops.

Structural drift is navigated by maintaining dominance of internally generated drift propagation modes over externally imposed drift. This dominance is achieved through phase-consistent regime transitions, not through trajectory correction.

The system may eventually undergo collapse. However, collapse occurs as loss of admissible continuation capacity, not as behavioral failure, and is delayed or phased through architectural drift structuring.

By construction, $\mathcal{X}$ satisfies the defining conditions of Navigational Cybernetics of Order 2.5.

Example 2 — Socio-Technical System with Navigable Collapse

Consider a large-scale socio-technical system (e.g. a scientific discipline or institutional knowledge system) operating across decades.

Such a system:

- accumulates irreversible historical commitments,
- operates under regime shifts,
- experiences drift in interpretation, priorities, and semantic coherence.

The system does not collapse when errors occur. It collapses when admissible modes of continuation are exhausted: when future developments can no longer preserve identity without semantic or structural rupture.

Admissibility is not enforced by any central controller. Certain continuations simply cease to be viable, regardless of local success or correctness.

Structural limits (e.g. loss of coherence, over-specialization, paradigm lock-in) are visible to participants but cannot be causally acted upon without collapsing into short-horizon optimization.

Survival of the system does not consist in preventing collapse, but in navigating it: through regime renewal, reinterpretation of core commitments, and phased structural transitions.

Such a system exemplifies Navigational Cybernetics of Order 2.5: collapse is inevitable, but its form, timing, and consequences are structurally navigated rather than optimized away.

Example 3 — Cybersecurity Architecture Under Adversarial Drift

Consider a large-scale cybersecurity system operating under continuous adversarial pressure.

Such a system experiences persistent structural drift: attack vectors evolve, defensive assumptions decay, and historical countermeasures accumulate irreversible technical and semantic debt.

Local correctness is insufficient: individual defenses may function correctly, alerts may be accurate, and intrusion attempts may be mitigated, while long-horizon viability silently degrades.

Admissibility in this system is not enforced through penalties, scores, or optimization objectives. Certain defensive evolutions (e.g. unbounded rule accretion or reactive hardening) eventually cease to be admissible, not because they fail immediately, but because they eliminate future viable continuations.

Structural limits such as attack-surface saturation, interpretability loss, or response latency inflation become visible at an architectural level, yet cannot be acted upon causally without collapsing into short-horizon threat optimization.

Drift navigation occurs when the system maintains dominance of internally generated security evolution over adversarially imposed drift, through regime restructuring, phase-aligned policy renewal, and controlled abandonment of obsolete defenses.

Collapse, when it occurs, is not marked by a single breach, but by irreversible loss of admissible defensive evolution.

Such an architecture exemplifies Navigational Cybernetics of Order 2.5: security is not maximized, but structurally sustained under adversarial drift.

Example 4 — Distributed Delivery Network Under Structural Load

Consider a distributed delivery network coordinating logistics across regions, time zones, and fluctuating demand.

Such a network operates under non-stationary interaction: traffic patterns shift, infrastructure ages, coordination latency fluctuates, and local optimizations accumulate global side effects.

Individual delivery actions may remain correct: packages are routed efficiently, deadlines are met, and performance metrics remain stable. Nevertheless, structural drift accumulates in routing assumptions, dependency coupling, and regime-specific coordination logic.

Admissibility is not enforced by rejecting individual delivery actions. Instead, certain modes of continuation (e.g. over-centralization, over-fragmentation, or latency-amplifying coordination regimes) gradually exhaust future viable configurations.

Structural limits such as coordination overload, loss of rerouting flexibility, or irreversible coupling become visible only at the architectural level and cannot be corrected through action-level optimization alone.

Drift navigation is achieved by maintaining dominance of internally structured logistics evolution over externally imposed load fluctuations, via regime transitions, phase-consistent rescaling, and controlled structural shedding.

Collapse, if it occurs, does not appear as immediate delivery failure, but as loss of admissible reconfiguration capacity under changing conditions.

Such a network belongs to Navigational Cybernetics of Order 2.5: long-horizon viability is preserved not by optimizing routes, but by navigating structural drift.

Example 5 — Architecture That Is Not Navigational Cybernetics 2.5

Consider a safety-constrained reinforcement learning system in which long-horizon risk is handled through penalty terms, reward shaping, or soft constraints.

In such an architecture, admissibility-relevant information (e.g. risk estimates, safety margins, constraint violations) is made causally actionable: it directly enters policy optimization, gradient updates, or action selection.

Although the system may exhibit stable learning curves, bounded error, and acceptable short-horizon performance, admissibility is represented as a scalar objective and is therefore optimized against.

Under sustained interaction, this architecture exhibits one of two invariant failure modes: either latent accumulation of structural degradation that remains invisible to optimization, or collapse of admissibility into a trade-off that can be arbitrarily violated for sufficient reward.

By the exclusion criteria established in this work, such a system is not a member of Navigational Cybernetics of Order 2.5. It is structurally equivalent to third-order cybernetics, regardless of implementation details or safety guarantees.

Example 6 — Model Predictive Control with Constraint Relaxation (Negative Example)

Consider a grid-scale energy management system operated using Model Predictive Control (MPC), as commonly deployed in power generation, load balancing, or industrial process control. The system maintains operational safety through explicit constraints (e.g. voltage limits, thermal bounds, reserve margins) and optimizes a rolling objective such as efficiency, cost, or stability over a finite prediction horizon.

When constraint violations are anticipated, the controller introduces constraint relaxation mechanisms: soft constraints, slack variables, penalty terms, or emergency override modes, allowing temporary violation in exchange for objective degradation.

All admissibility-relevant information (reserve depletion, thermal stress, approaching instability) is therefore causally actionable: it directly enters the optimization problem and is traded against performance objectives.

Under sustained non-stationary conditions (e.g. aging infrastructure, increasing volatility of renewables, or compounding forecast errors), the system may continue to operate correctly: no immediate instability occurs, and performance metrics remain within bounds.

Nevertheless, structural drift accumulates in the form of progressive constraint relaxation, increasing reliance on emergency modes, and irreversible depletion of safety margins. The admissible continuation space contracts silently, while the controller remains locally optimal. Collapse, when it occurs, does not appear as gradual degradation but as sudden loss of feasible solutions or abrupt transition into failure modes outside the modeled regime.

By the criteria established in this work, such an architecture does not belong to Navigational Cybernetics 2.5. Admissibility is represented as an optimizable object, and long-horizon viability is sacrificed to maintain short-horizon feasibility.

Conclusion

This work establishes a closed and extended structural foundation for long-horizon adaptive systems in which viability, identity continuity, and meaningful behavior are governed by architectural constraints irreducible to causal control, optimization procedures, or performance metrics. Information is formalized not as a retrievable quantity, representation, or signal, but as a persistent structural constraint delimiting the space of admissible future evolution. Coherence, drift, internal time, and directionality are treated as invariant properties of adaptive existence, not as tunable design parameters.

A central result is the categorical separation between causal dynamics and structural admissibility. Causal evolution determines which state transitions may occur; admissibility, derived from accumulated structural state, determines which transitions are permitted to occur. These influences act concurrently in the present but are irreducible to one another. Long-horizon failure is shown to arise not from local error, misprediction, or insufficient optimization, but from violations of admissibility constraints that cannot be repaired through causal correction once incurred.

Version 1.9 strengthens this separation by establishing drift as a navigable structural medium rather than an error to be suppressed. Drift is shown to be unavoidable under directed interaction and structurally incompatible with zeroing or error-correction objectives. Any attempt to suppress drift causally induces one of two invariant failure modes: silent latentization of structural degradation or collapse of admissibility into optimization. Survival in the long-horizon regime therefore requires navigation under admissibility constraints, not drift elimination.

The framework further establishes that action is not a primitive in long-horizon adaptivity. Admissibility must be fixed prior to the definition of the action domain, and any object that shapes admissibility cannot itself be an action without inducing circularity and penalty-based resolution. This well-foundedness barrier closes action-centric and policy-first reductions of adaptive agency and fixes admissibility as a structurally prior layer.

Phase structure and internal time are shown to impose additional non-negotiable constraints. Drift navigability requires phase-consistent continuation; unresolved phase mismatch generates structural debt that restricts admissible regimes independently of local correctness. Internal time is operationally falsifiable: under irreversible burden, the space of admissible continuation must contract. Architectures that permit unbounded continuation or unrestricted amplification under accumulating cost are therefore misclassified.

These results necessitate a strict reclassification of cybernetic regimes compatible with long-horizon adaptive systems. Classical second-order cybernetics permits observation to enter causal regulation and is insufficient to prevent silent structural degradation. Third-order cybernetics internalizes reflexivity itself, allowing systems to act upon descriptions of their own observing and thereby optimize against structural limits. The theorems established here demonstrate that this progression is structurally unsafe: once pre-semantic structural revelation becomes causally actionable, admissibility is optimized away and identity-preserving coherence cannot be maintained.

The resulting architectural class occupies a discrete intermediate regime, termed cybernetics of order 2.5. This regime permits reflexive visibility of admissibility-relevant structural conditions while categorically prohibiting their causal exploitation. It introduces a non-causal layer of viability supervision without allowing recursive self-modification of identity-defining constraints. This boundary is not a contingent modeling choice but a necessary architectural condition: crossing into full third-order reflexivity constitutes a change of structural class and entails the loss of long-horizon coherence.

Within this context, ONTOΣ I–VI and UTAM do not function as descriptive ontologies of what exists, but as formal specifications of what may be admitted within an adaptive system if that system is to remain viable under irreversible interaction and finite internal time. The axioms, lemmas, and theorems presented therefore define an admissibility calculus rather than a metaphysical account.

This framework is intentionally non-constructive. It does not prescribe how to build adaptive systems, but defines which architectural claims are structurally coherent under long-horizon drift.

Taken together, this work fixes a minimal yet non-trivial theoretical core for coherence-preserving adaptive architectures. It establishes the inevitability of drift, the dominance of inertial propagation, the binding role of internal time, the structural limits of amplification, and the necessity of identity-aligned—rather than optimized—directionality. These results provide a canonical architectural anchor for PETRONUS™ and related systems, and delineate a defensible prior-art foundation for a distinct school of cybernetics concerned with identity preservation and viability over long horizons.

Glossary of Canonical Structural Terms

The following definitions are canonical within this framework. They are structural, non-operational, and non-algorithmic. Terms defined here must be interpreted consistently throughout all axioms, lemmas, and theorems.

Structural Drift. Structural drift is the irreversible accumulation of internal structural deviation arising from sustained interaction over time. Drift is not noise, error, uncertainty, or stochastic fluctuation. A system exhibits non-zero drift if its internal structural state cannot be fully restored by any finite sequence of admissible operations.

Admissibility. Admissibility is a non-causal structural exclusion relation over effect-classes. It determines which realized effects are permitted to occur, but does not rank alternatives, guide optimization, or participate in action selection.

Boundary. A boundary is a structural limit separating admissible from non-admissible continuation within the space of structurally reachable effect-classes.
Boundaries are not states, actions, or signals. They are non-causal structural constraints emerging from accumulated drift, phase structure, and irreversible history. Crossing a boundary constitutes a structural event: it may result in loss of admissibility, identity discontinuity, regime irreversibility, or collapse, even in the absence of observable failure.
Boundaries cannot be optimized against, softened by penalties, or inferred as gradients.They are encountered, not computed.

Boundary Crossing. Boundary crossing is a structural event in which a continuation exits the admissible continuation space.
Boundary crossing may occur without observable failure, performance degradation or instability. Boundary crossing induces irreversible effects on identity continuity, regime accessibility or future admissibility.

Continuation. A continuation is a realized structural progression of the system over internal time, defined at the level of effect-classes rather than actions.
An admissible continuation is a continuation that does not violate structural boundaries given the current accumulated drift, phase state, and internal-time constraints. Continuation is evaluated structurally, not selected, optimized or ranked.

Continuation Space. The continuation space is the set of structurally reachable future continuations available to a system given its accumulated drift, regime, and internal time.
Admissibility acts as an exclusion relation over the continuation space. The space contracts under structural burden and may collapse irreversibly after boundary crossing.

Viability. Viability denotes the capacity of a system to continue admissible operation without irreversible loss of identity under non-zero drift and finite internal time.

Identity. Identity is the persistence of coherent internal organization across time and regimes. Identity continuity is not implied by behavioral similarity, task success or performance stability.

Phase. A phase is a regime of interpretation and structural propagation that determines how drift is accumulated, redirected or closed over time.

Drift Navigability. Drift navigability is the capacity of a system to generate, redirect, or phase drift such that admissible continuation is preserved.
Navigability does not imply drift minimization or suppression. It denotes dominance of internally generated drift structures over externally imposed drift.
A system lacking drift navigability remains passively exposed to collapse even under correct operation.

Phase Debt. Phase debt is the unresolved accumulation of drift resulting from prolonged regime occupation without structural closure.

Structural Load. Structural load is the accumulated irreversible burden imposed on an architecture by sustained interaction, phase transitions, and regime occupation.
Structural load is not reducible to energy expenditure, error magnitude, intervention size, or trajectory cost. It constrains admissible continuation independently of local correctness or performance stability.
Structural load contributes directly to drift accumulation, boundary formation, and internal-time depletion.

Structural Burden. Structural burden is the accumulated irreversible load imposed on a system by regime occupation, phase mismatch, and admissible but costly continuation.
Structural burden contributes to drift accumulation and internal time depletion even under correct and admissible trajectories.
Structural burden is not equivalent to energy, error or effort.

Internal Time. Internal time is a structural capacity to absorb deformation and accumulation without loss of admissibility or identity.
It is distinct from wall-clock time and trajectory duration.

Collapse. Collapse denotes the irreversible loss of admissible continuation under finite internal time. It does not imply immediate behavioral failure, external malfunction, or performance degradation. A system may continue to operate, optimize, and appear correct after collapse has occurred, provided that the space of admissible effect-classes has already been structurally exhausted.

Structural Collapse. Structural collapse is the irreversible exhaustion or contraction of the admissible effect-class set, resulting from unstructured drift accumulation, phase inconsistency, or irreversible regime transitions.
Structural collapse is defined independently of trajectory correctness, reward stability, local control performance or observable behavioral failure.
Collapse does not denote immediate dysfunction, but loss of admissible continuation, identity continuity or long-horizon viability.
Collapse is not prevented by optimization or control, but may be delayed, phased or rendered navigable through architectural structuring of drift.

Terminal Admissibility Horizon. The terminal admissibility horizon is the internal-time boundary beyond which no admissible continuation is structurally available to the system.
Crossing the terminal admissibility horizon does not imply behavioral failure, instability or

performance degradation.
It denotes irreversible exhaustion or contraction of the admissible effect-class set. Collapse, when it occurs, is the manifestation of having crossed the terminal admissibility horizon. Architectural structuring of drift does not eliminate this horizon, but may delay it, phase it or render its approach navigable and non-degenerate.

Structural Bounds. Structural bounds are superordinate capacity limits on accumulated self-deformation. They define saturation thresholds beyond which admissible continuation contracts unless the architecture changes class. Bounds are structural (not a resource budget) and must not be exposed as an optimizable objective.

Structural Burden (Structural Load). Structural burden is accumulated irreversible load (e.g. drift, phase debt, coupling load, internal-time depletion) that restricts admissible continuation independently of local correctness or performance. Burden may grow while actions remain admissible and metrics remain stable.

Collapse Avoidance (Non-Goal). This framework does not define collapse avoidance as a terminal objective.
Collapse is not prevented, but delayed, phased, or rendered navigable through architectural structuring of drift.
Architectures that aim to eliminate collapse entirely are structurally misclassified, as they implicitly assume zero drift or infinite internal time.

Regime. A regime is a stable structural configuration that governs how interpretation, propagation, and structural cost accumulate over internal time.
A regime is not reducible to individual actions, short-horizon behavior, state variables, or policies. It defines admissible modes of continuation and conditions of reversibility.
Sustained occupation of a regime is a primary carrier of long-horizon risk: it determines how drift accumulates, may induce phase debt, and can trigger irreversible structural transitions independently of local correctness or performance stability.

Regime Transition. A regime transition is an irreversible change in the active structural operator configuration of a system.
Such transitions may preserve admissible behavior while altering identity continuity, internal-time consumption rate or future admissibility structure.

Spin. Spin denotes a persistent internal orientation of structural propagation that biases admissible continuation independently of metric evaluation or reward optimization.
Spin is not a goal, preference, or utility.
It is a directional asymmetry that stabilizes identity under drift by privileging certain structural continuations over symmetric alternatives.

Spin (Architectural). Spin denotes a persistent internal directional asymmetry governing how drift is interpreted and propagated across regimes and internal time.

Spin is an architectural primitive.
It is not a quantum property, physical angular momentum, or statistical bias.

Spin is introduced as an irreducible structural orientation and is not reduced to latent state variables, preferences, or optimization objectives.

Spin Loss. Spin loss is the degradation or neutralization of internal directional asymmetry under sustained drift or regime pressure.
Spin loss may occur without observable failure and results in identity dispersion, collapse of coherent directionality, or absorption into externally imposed dynamics.

Non-Causal Observation. Non-causal observation is the availability of viability- or admissibility-relevant structure as a witness signal that is categorically prohibited from entering action selection, optimization, or learning updates.
If such observation becomes causally actionable, it ceases to be observation and collapses into control.

MxBv, Poznań, Poland.
Cybernetics 2.5 (MxBv) PETRONUS™ research Synthetic Conscience™
Citation and Versioning

Concept DOI (all versions):
10.5281/zenodo.18130264

Version DOI (this version):
10.6084/m9.figshare.31288699

Version 1.9, Poznań, Poland, 2026.