Intelligence Under Constraint: Construction, Legibility, and Boundary Conditions

Flyxion

December 16, 2025

Abstract

Contemporary discussions of intelligence, both biological and artificial, are dominated by performance metrics, representational models, and optimization-based framings. These approaches obscure a more fundamental question: what structural conditions must remain invariant for intelligence to persist, generalize, and survive exposure under reuse?

This essay develops a constraint-first account of intelligence grounded in construction history, invariant-preserving abstraction, modularity, and refusal. Intelligence is treated not as a static capacity or a bundle of skills, but as a mode of regulated interaction with an environment, mediated by semi-permeable boundaries. Drawing on biological membranes, cortical synchronization, evolutionary redundancy, operating system privilege separation, and formal mathematical structure, the essay argues that mature intelligence necessarily regulates its own legibility.

On this view, refusal is not an ethical afterthought but a constitutive operation; abstraction is compression under lawful transformation; and responsible publication requires effort-gated legibility rather than secrecy or unrestricted openness. The form of the argument enacts these claims by requiring cumulative discipline rather than extractable instruction.

1 Introduction

There is a persistent expectation, in both technical and public discourse, that intelligence should be immediately legible. Explanations should be compressible, systems should be reproducible, and ideas should be deployable without extended reconstruction. Within artificial intelligence research, this expectation appears as benchmark-driven evaluation, architectural recipes, and the presumption that progress consists primarily in scaling or optimization. In broader intellectual culture, it appears as demands for accessibility, summarizability, and rapid dissemination.

This essay begins from the claim that this expectation is not neutral. It functions as a selection pressure that privileges shallow correctness, discourages cumulative reasoning, and systematically erodes the very properties that make intelligence general rather than brittle. Ideas that survive such environments do so not because they are structurally robust, but because they are easily paraphrased, weakly constrained, or trivially recontextualized.

The central thesis advanced here is that intelligence is best understood not as a set of capabilities or representations, but as *construction under constraint*. To be intelligent is to build structures

over time, to reuse them lawfully, to compress them without destroying invariants, and to refuse extrapolation beyond jurisdiction. These properties are not optional. They are the conditions under which any adaptive system remains coherent when exposed to reuse, transfer, and scrutiny. This perspective aligns with early cybernetic and complexity-theoretic accounts of adaptive systems as processes governed by constraint rather than by static description (Ashby 1956; Weaver 1948).

A secondary but unavoidable consequence follows. If intelligence itself depends on regulated interaction and semi-permeable boundaries, then the communication of intelligence—including the publication of powerful ideas—must obey the same structural logic. Unfiltered disclosure is not inherently virtuous; neither is concealment inherently responsible. What matters is whether the form of disclosure preserves the invariants that give the ideas meaning.

The argument proceeds in a deliberately cumulative manner. Early sections establish ontological commitments regarding events, states, and construction history. Subsequent sections develop abstraction as invariant-preserving compression, modularity as an ontological requirement, and refusal as a first-class operation. Only after these constraints are in place does the essay address general intelligence, legibility, biological and computational boundary conditions, and the ethics of publication.

The structure of the essay is not incidental. Later claims depend on earlier constraints, and no section is intended to stand alone. This is not a stylistic choice but a substantive one: intelligence that can be meaningfully summarized without discipline is not the kind of intelligence under consideration.

2 Events Over States

2.1 The Insufficiency of State-Based Accounts

Most contemporary theories of intelligence, whether biological or artificial, are formulated in terms of states. Neural activations, parameter vectors, symbolic configurations, or world models are treated as the primary bearers of meaning and competence. Learning, on such accounts, consists in moving through a space of states toward configurations that optimize some criterion.

This framing obscures a fundamental asymmetry. A state is always a summary. It collapses history, erases order, and forgets the contingencies by which it was produced. Two systems may occupy indistinguishable states while having arrived there through radically different processes, with radically different implications for reuse, generalization, and failure.

State-based descriptions therefore lack explanatory power with respect to persistence. They can describe what a system *is like* at a moment, but not what it *can survive*. Intelligence, however, is not defined by momentary adequacy. It is defined by the ability to remain coherent across time, exposure, and reuse, a point emphasized early in cybernetic and complexity-theoretic treatments of adaptive systems (Ashby 1956; Weaver 1948).

2.2 Construction History as Primary

To address this limitation, we take *events* rather than states as ontologically prior. An event is an irreversible contribution to construction history: an observation made, a commitment taken, a refusal issued, a structure compressed. States, where they appear, are projections derived from histories under particular queries and scopes.

Definition 2.1. A *construction history* is an ordered, irreversible sequence of events from which any state description is derivable only as a partial projection.

On this view, memory is not storage but replayability. To remember is to be able to reconstruct how a structure came to be, not merely to access its current form. Learning is the reorganization of construction history into forms that are cheaper to reuse without loss of function, consistent with accounts that treat learning as transformation of process rather than accumulation of representations (Ashby 1956).

This shift has immediate consequences. Intelligence can no longer be identified with static representations, nor can learning be reduced to parameter adjustment. What matters is the system's ability to transform its own history into reusable abstractions while preserving the conditions under which those abstractions remain valid.

Remark 2.1. Throughout what follows, any appeal to a "state" should be understood as shorthand for a view compiled from construction history under explicit constraints. No state is treated as ontologically primitive.

2.3 Irreversibility, Time, and Constraint Accumulation

Treating events as primary commits the theory to irreversibility. An event, once incorporated into construction history, cannot be undone without altering the identity of the system that incorporates it. This irreversibility is not an implementation detail but a structural feature: it is what gives time its direction and learning its cost, echoing foundational analyses of irreversibility in complex adaptive systems (Weaver 1948).

In a state-based account, time is often treated as an index labeling successive configurations. In an event-based account, time is instead the accumulation of constraints. Each event narrows the space of future admissible constructions by fixing commitments, excluding alternatives, and conditioning subsequent possibilities. Learning therefore increases not only competence but responsibility: what has been built constrains what may coherently follow.

This perspective clarifies why intelligence cannot be adequately characterized by instantaneous performance. A system that performs well only by discarding its history, resetting its commitments, or erasing prior structure is not intelligent in the relevant sense. It is merely reactive. Intelligence is distinguished by the capacity to carry constraints forward without collapse.

2.4 Why Compression Requires History

Compression plays a central role in most theories of intelligence, but it is often misunderstood as a purely representational operation. On the present account, compression is inseparable from history.

To compress is not merely to shorten a description, but to replace a detailed construction history with a more economical one that can be replayed to the same effect under specified conditions.

Such replacement is only meaningful relative to what is being preserved. A compressed history is not a lossless encoding of all prior detail; it is a commitment to ignore certain distinctions while maintaining others. These commitments are intelligible only when the original history remains, at least in principle, reconstructible, a requirement closely related to information-theoretic treatments of compression as invariance under transformation (Jaynes 2003; Cover and Thomas 2006).

Compression without history degenerates into pattern matching. It may succeed locally, but it cannot support lawful reuse, because the reasons for the pattern's validity have been erased. A genuinely intelligent system therefore compresses its history while retaining the ability to justify that compression relative to the constraints it preserves.

2.5 Event-Based Identity and Generalization

An immediate consequence of the event-based view is that identity becomes path-dependent. Two systems with identical outward behavior may nonetheless differ profoundly in what they can safely generalize, because their construction histories differ. What one system can extend lawfully, another may only approximate dangerously.

Generalization, on this account, is not the application of a rule to new cases, but the transport of a construction across contexts while preserving its enabling conditions. Such transport presupposes knowledge of how the construction was achieved in the first place. Without access to construction history, there is no principled way to determine whether reuse is valid or merely coincidental.

This observation already rules out a wide class of naïve general intelligence proposals. Any approach that treats learned structure as context-free, or that discards the conditions of its acquisition, forfeits the ability to regulate its own extension. What appears as flexibility in the short term becomes fragility under exposure.

2.6 Consequences for the Structure of the Argument

The remainder of this essay builds on the commitments established here. By treating events as ontologically prior to states, we commit to a conception of intelligence that is cumulative, constraint-sensitive, and irreducibly historical. Abstraction, modularity, refusal, and legibility will all be shown to follow from this starting point, not as independent principles but as necessary consequences.

Readers accustomed to state-based or representation-centric accounts may find this shift initially disorienting. That disorientation is itself diagnostic. Any framework that seeks to explain intelligence without accounting for the cost and irreversibility of construction has already abstracted away the very phenomenon it purports to explain.

In the next section, abstraction will be reconsidered under these constraints. Rather than treating abstraction as representational simplification, it will be developed as the identification of invariants under lawful transformation—a move that preserves the primacy of history while enabling reuse (Arnold 1989; Olver 1993).

3 Modularity as Ontology, Not Convenience

3.1 Why Monolithic Intelligence Fails

It is tempting to imagine intelligence as a single, unified faculty: a global workspace, a central optimizer, or a comprehensive world model in which all information is integrated and adjudicated. Such monolithic conceptions are attractive because they promise coherence by fiat. If all distinctions collapse into a single representational space, then conflict, inconsistency, and ambiguity appear tractable.

This promise is illusory. A monolithic system has no principled way to regulate interaction between heterogeneous processes, because all interactions are already internal. As a result, every change propagates everywhere. Local failure becomes global corruption. The system gains expressive power at the cost of fragility, a trade-off long recognized in cybernetic analyses of large adaptive systems (Ashby 1956).

More importantly, monolithic intelligence cannot generalize safely. Without internal boundaries, there is no mechanism to prevent abstractions learned in one regime from being applied illegitimately in another. What appears as flexibility in small domains becomes catastrophic overreach as scope expands, a failure mode familiar from both biological and engineered systems (Weaver 1948).

3.2 Modularity as a Structural Requirement

Modularity is often introduced as an engineering strategy: a way to manage complexity, enable parallel development, or improve maintainability. In the present framework, modularity is not optional and not pragmatic. It is ontological.

Definition 3.1. A system is *modular* if it can be decomposed into subsystems whose internal coherence does not depend on global synchronization, and whose interactions are mediated exclusively by explicit interfaces.

Each module constitutes a locally coherent domain with its own construction history, abstractions, and refusal conditions. Modules may interact, but only through transformations that preserve the invariants of both sides. No module is privileged by default; no global authority resolves conflicts by erasing distinctions.

This separation is what allows abstraction to remain lawful under reuse. By forcing interactions to pass through interfaces, the system makes scope explicit and prevents accidental generalization, consistent with classical principles of separation and constraint in complex systems design (Saltzer and Schroeder 1975).

3.3 Interfaces, Jurisdiction, and Scope

An interface is not merely a channel for data exchange. It is a declaration of jurisdiction. It specifies what kinds of transformations are permitted, what kinds of commitments may be imported, and what kinds of effects may be exported.

Jurisdictional boundaries are the modular analogue of physical membranes. They do not block interaction; they regulate it. By constraining how information and influence cross module boundaries, interfaces preserve local coherence while enabling coordination, paralleling the role of semi-permeable boundaries in biological and cognitive systems (Friston 2010; Friston 2015).

This has a crucial consequence for learning. When a module fails, the failure can be localized. Its construction history can be revised without invalidating unrelated structures. The system learns without unlearning everything else.

3.4 Why Global Workspaces Are Insufficient

Global workspace theories attempt to reconcile modularity with unity by positing a shared representational arena into which local processes broadcast their contents. While such architectures can support coordination, they reintroduce the very fragility modularity is meant to avoid.

Once information enters a global workspace, it becomes available everywhere, regardless of whether the receiving processes share the assumptions under which it was produced. Scope is lost. Refusal becomes difficult, because the system lacks a principled way to prevent uptake, a limitation noted in critiques of globally shared control architectures (Ashby 1956).

In contrast, a genuinely modular system does not require a global workspace to achieve coherence. Coordination emerges from structured interaction among modules, not from universal exposure.

3.5 Modularity and the Accumulation of Constraints

Because modules maintain their own construction histories, they accumulate constraints independently. This independence is essential for long-term adaptation. Different modules may explore different abstractions, maintain different buffers of variation, and refuse different classes of transformation.

The system as a whole benefits from this diversity. It can compose modules when their invariants align and keep them separate when they do not. General intelligence arises not from homogenization, but from the disciplined management of heterogeneity, a theme that recurs across studies of complex adaptive systems (Weaver 1948).

3.6 Transition to Refusal

Modularity alone is insufficient unless the system can decline illegitimate interaction. Interfaces must not only specify what is allowed; they must also make refusal possible when constraints cannot be preserved.

In the next section, refusal will be developed as a first-class cognitive operation. Rather than treating refusal as failure or ignorance, it will be shown to be a necessary condition for coherent modular interaction and, ultimately, for general intelligence.

4 Refusal as a First-Class Cognitive Operation

4.1 Refusal Is Not Failure

Within many accounts of intelligence, refusal is treated implicitly as a defect: an absence of knowledge, an error condition, or a temporary limitation to be overcome by improved optimization. Systems are evaluated by how rarely they fail to produce an answer or an action. Silence, hesitation, or constraint are framed as shortcomings.

This framing is incompatible with any conception of intelligence that is cumulative and general. A system that never refuses has no mechanism to protect the invariants that make its abstractions meaningful. It cannot distinguish between domains where reuse is lawful and domains where it is destructive. Apparent competence in such a system is purchased by borrowing against hidden fragility.

Refusal, properly understood, is not the negation of intelligence. It is one of its essential expressions.

4.2 Refusal as Jurisdictional Enforcement

Every abstraction has a domain of validity, whether acknowledged or not. To apply an abstraction outside that domain is to violate the conditions under which it was constructed. A system that generalizes without checking these conditions will eventually collapse its own coherence.

Refusal is the operation by which a system enforces jurisdiction. It is the recognition that no admissible transformation exists between a current context and the contexts in which a given abstraction is known to hold.

Definition 4.1. A *refusal* is the structurally grounded non-existence of an admissible transformation between a proposed action or inference and the system's preserved invariants.

This definition is deliberately negative. Refusal is not an alternative action chosen from a menu. It is the absence of a lawful path forward.

4.3 Uncertainty, Irreversibility, and Risk

Refusal becomes especially important in the presence of irreversibility. When an action or inference cannot be undone without corrupting construction history, the cost of error increases dramatically. Under such conditions, approximation is not a virtue; it is a liability.

A system that treats uncertainty as a reason to guess rather than to suspend commits itself to irreversible commitments it may later be unable to repair. By contrast, a system that can refuse preserves optionality. It delays commitment until sufficient structure exists to justify action.

This capacity to delay is not passivity. It is a form of active constraint management.

4.4 Refusal and Learning

Refusal plays a central role in learning precisely because learning involves revision. To revise a construction, the system must be able to identify which prior commitments remain valid and which

must be set aside. Without refusal, all commitments are treated as equally binding, and revision becomes destructive rather than selective.

When a system refuses, it marks a boundary in its own competence. That boundary becomes informative. It directs exploration, motivates the acquisition of new constraints, and guides the construction of new abstractions. Refusal is therefore not the end of learning, but its compass.

4.5 Refusal in Modular Systems

In a modular architecture, refusal prevents the collapse of scope. When one module proposes an interaction that another cannot support without violating its invariants, refusal blocks composition. This preserves local coherence and prevents error from propagating globally.

Importantly, refusal need not be symmetric. One module may accept a transformation that another must reject. Such asymmetry is not inconsistency; it reflects differing construction histories and jurisdictions. Coherence is maintained not by forcing agreement, but by respecting boundaries.

4.6 Refusal as a Condition of Generality

Generality is often conflated with permissiveness: the ability to act across many domains with minimal restriction. On the present account, the opposite is true. A system is general to the extent that it can distinguish where its abstractions apply and where they do not.

A system that always acts is not general. It is reckless.

A system that refuses indiscriminately is not general. It is inert.

General intelligence occupies the narrow regime in which refusal is selective, justified, and structurally grounded.

4.7 Transition to Transfer and Generality

Once refusal is recognized as a first-class operation, the problem of general intelligence can be posed more precisely. The question is no longer how to build a system that acts everywhere, but how to build a system that transfers structure lawfully across contexts while refusing illegitimate reuse.

In the next section, general intelligence will be characterized in these terms: not as breadth of competence, but as the regulated transport of invariant structure under constraint.

5 Generality as Lawful Transfer

5.1 Against Breadth as a Criterion of Generality

Generality is commonly identified with breadth: the number of domains, tasks, or environments in which a system performs adequately. On such accounts, intelligence becomes a matter of coverage. A system is more general insofar as it succeeds in more places.

This criterion is misleading. Breadth alone cannot distinguish between lawful reuse and accidental success. A system may perform well across many domains by relying on shallow correlations, heuristics, or brute-force adaptation, while lacking any principled understanding of why those

strategies work or where they fail. Such systems generalize until they do not, and when they fail, they fail catastrophically.

A more demanding criterion is required—one that distinguishes genuine transfer from coincidental overlap.

5.2 Transfer as the Transport of Construction

On the event-based account developed earlier, generalization is not the application of a static rule to new cases, but the transport of a construction from one context to another. This transport is only legitimate if the conditions that made the construction successful in its original domain are preserved, or suitably transformed, in the new one.

Transfer therefore presupposes knowledge of construction history. A system must know not only what worked, but why it worked, and under what constraints. Without this knowledge, reuse becomes guesswork.

Definition 5.1. A *lawful transfer* is a transformation of a construction history from one context to another that preserves the invariants required for its validity.

Lawful transfer is conservative. It does not promise success everywhere. It promises coherence where it applies and refusal where it does not.

5.3 Invariants as the Currency of Generality

Invariants play a central role in distinguishing lawful transfer from overgeneralization. An invariant is a property of a construction that remains stable under a specified family of transformations. These transformations encode what counts as a permissible change of context.

A system that cannot articulate, even implicitly, which invariants it is preserving cannot regulate its own generalization. It may apply an abstraction successfully in familiar regimes, but it has no principled way to detect when it has left those regimes.

Generality, on this view, is proportional not to the number of abstractions a system possesses, but to the number of invariants it can preserve under transformation.

5.4 Why Transfer Requires Modularity and Refusal

Lawful transfer cannot be centralized without collapsing distinctions. If a single global mechanism adjudicates all reuse, then all contexts are implicitly treated as commensurable. This undermines the very notion of invariance.

Instead, transfer must be mediated by modules that maintain local coherence. Each module evaluates proposed transformations relative to its own construction history and invariants. When alignment exists, transfer proceeds. When it does not, refusal blocks propagation.

This distributed evaluation is what allows generality to scale without fragility. The system does not need to know in advance which transfers will succeed. It needs only to enforce refusal when constraints cannot be preserved.

5.5 Failure Modes of Illegitimate Transfer

When transfer is attempted without constraint, several characteristic failure modes arise. Abstractions become overextended, losing specificity until they are vacuous. Systems appear flexible until they encounter edge cases, at which point failure is abrupt and difficult to localize. Responsibility for error diffuses, because no boundary marks where reuse ceased to be valid.

These failures are not incidental. They follow necessarily from treating generality as permissiveness rather than as regulated transport.

5.6 Generality Without Collapse

A system that implements lawful transfer will appear conservative by contemporary standards. It will refuse often. It will decline to extrapolate without sufficient alignment. It will prioritize coherence over coverage.

Paradoxically, such a system is capable of deeper generality. Because it preserves its invariants, it can reuse constructions repeatedly without degradation. Its competence accumulates rather than dissolves.

This is the sense in which general intelligence should be understood: not as the absence of limits, but as the disciplined navigation of them.

5.7 Transition to Locality and Context

Lawful transfer presupposes an account of context. Invariants are preserved relative to transformations, and transformations are defined relative to local conditions. To make this precise, the next section develops a view of intelligence as locally coherent and only conditionally global.

Rather than assuming a single unified worldview, intelligence will be shown to arise from the structured coordination of partial perspectives under explicit gluing conditions.

6 Local Coherence and Conditional Globality

6.1 Why Global Consistency Is the Wrong Ideal

Many theories of intelligence implicitly assume that coherence requires global consistency: a single, unified model of the world in which all facts, beliefs, and abstractions are mutually compatible. In such frameworks, inconsistency is treated as a defect to be eliminated, and intelligence is equated with the capacity to resolve all local discrepancies into a coherent whole.

This ideal is misplaced. In systems that learn, adapt, and operate across heterogeneous environments, global consistency is neither achievable nor desirable. Different contexts impose different constraints, admit different abstractions, and tolerate different approximations. Forcing these into a single globally consistent framework often requires discarding precisely the information that makes local reasoning effective.

Intelligence does not require global consistency. It requires *local coherence* and principled criteria for when local constructions may, or may not, be combined.

6.2 Local Coherence as the Unit of Sense

A locally coherent construction is one whose internal abstractions, constraints, and refusals are mutually compatible within a defined scope. Such a construction may rely on assumptions that do not hold elsewhere, and it may tolerate inconsistencies that would be unacceptable outside its jurisdiction.

Local coherence is not a weakness. It is what allows intelligence to function in environments that are themselves fragmented, non-stationary, and partially observable. By maintaining multiple locally coherent structures rather than a single global one, a system preserves flexibility without sacrificing rigor.

This perspective reframes contradiction. Apparent inconsistencies between local constructions are not necessarily errors. They are signals that different invariants are being preserved under different conditions.

6.3 Context as Boundary, Not Background

Context is often treated as ancillary information that modifies the interpretation of otherwise universal rules. On the present account, context is primary. It defines the boundary conditions under which abstractions are valid and transfers are lawful.

A context is not merely a set of parameters. It is a structured environment with its own admissible transformations and refusal conditions. To move between contexts is to cross a boundary, and not all boundaries are traversable.

This makes explicit why abstraction cannot be context-free. Any abstraction that purports to apply universally without qualification has already erased the conditions of its own validity.

6.4 Gluing and the Limits of Integration

While global consistency is neither required nor attainable, intelligence does require coordination across contexts. Local constructions must sometimes be combined to support action, explanation, or learning at larger scales. The challenge is to determine when such combination is legitimate.

Combination is legitimate only when local constructions agree on the overlaps that matter for the task at hand. Where such agreement exists, local structures may be *glued* together to form a larger coherent whole. Where it does not, separation must be maintained.

This criterion replaces the demand for universal consistency with a demand for conditional compatibility. Integration is not assumed; it is earned.

6.5 Failure Modes of Forced Globality

When systems attempt to impose global coherence prematurely, characteristic failures occur. Local distinctions are flattened, leading to abstractions that are broadly applicable but weakly informative. Conflicts are resolved by erasure rather than reconciliation, obscuring the reasons for disagreement. Errors propagate widely because boundaries that would have contained them have been removed.

These failures are especially damaging in general systems, where errors are reused across domains. What appears as elegance in design becomes brittleness in operation.

6.6 Conditional Globality as an Achievement

On the present account, global structure is not a starting point but an outcome. It emerges only where local constructions align sufficiently to support integration. Even then, such globality remains conditional. It may dissolve as contexts change or new constraints arise.

This view aligns with the broader theme of the essay. Intelligence is not the elimination of boundaries, but their disciplined management. Coherence is not imposed from above; it is negotiated across levels.

6.7 Transition to Legibility and Exposure

Once intelligence is understood as locally coherent and only conditionally global, the problem of legibility takes on a new form. Making a construction legible outside its original context is itself a kind of transfer. It risks erasing boundaries, flattening constraints, and inviting illegitimate reuse.

In the next section, legibility will be treated as a selection pressure acting on intelligent systems. The capacity to regulate what is made visible, when, and under what conditions will be shown to be as essential as the capacity to act or infer.

7 Legibility as a Selection Pressure

7.1 Legibility Is Not Neutral

Legibility is often treated as an unqualified good. To be legible is to be understood, reproduced, and adopted. In scientific, technical, and public discourse alike, there is a presumption that increasing legibility increases value. Systems, theories, and explanations are praised to the extent that they are accessible, transparent, and easily summarized.

This presumption ignores a fundamental asymmetry. Making a structure legible does not merely reveal it; it alters the environment in which it operates. Once legible, a construction becomes subject to extraction, recombination, and reuse by agents that may not share the constraints under which it was built. Legibility therefore functions as an exposure mechanism.

From an evolutionary perspective, exposure is not benign. It creates new selection pressures that act on the system itself.

7.2 Competence Increases Targeting

In biological systems, increased competence often correlates with increased visibility. Bright coloration, complex behavior, or pronounced structure may confer advantages in coordination or reproduction, but they also attract predation. The same trait that signals fitness can become a liability when the environment changes.

This dynamic generalizes. Any system that demonstrates capability creates surface area. It becomes easier to exploit, copy, attack, or repurpose. Intelligence that cannot regulate its own

visibility will be selected against, not because it lacks power, but because it cannot survive its own success.

Legibility is therefore not merely communicative. It is ecological.

7.3 Legibility as an Internal Control Problem

If exposure alters the environment, then the decision to expose must itself be regulated. For an intelligent system, legibility becomes an internal control variable alongside action and inference.

To make something legible is to choose a projection of internal structure into an external context. Different projections preserve different invariants and discard different constraints. A full exposition may support deep collaboration in trusted contexts while becoming dangerous in adversarial or extractive ones. A minimal projection may preserve safety at the cost of utility.

There is no universally correct level of legibility. What matters is that the choice be constrained, justified, and reversible where possible.

7.4 Camouflage and Signaling

Biological systems exhibit two complementary strategies for managing legibility: camouflage and signaling. Camouflage reduces visibility to avoid exploitation, while signaling amplifies specific traits to coordinate with allies or attract mates. Both are adaptive responses to selection pressure.

Crucially, these strategies are not expressions of intent or deception. They are regulatory mechanisms. The same organism may employ both, depending on context.

Intelligence exhibits the same duality. There are contexts in which displaying capability accelerates coordination and learning, and contexts in which concealment preserves coherence. A system that can do only one or the other is brittle.

7.5 Why Unregulated Legibility Is Fragile

When legibility is treated as an unconditional virtue, systems are incentivized to collapse structure into forms that travel easily. Explanations become slogans, abstractions become heuristics, and architectures become recipes. What survives exposure is not what is robust, but what is portable.

This dynamic selects against constraint. Systems that preserve invariants resist summarization and are penalized for opacity. Systems that discard constraints gain adoption at the cost of correctness. Over time, the environment fills with ideas that are easy to reuse and hard to repair.

From the perspective developed in this essay, this is not progress. It is a form of epistemic drift.

7.6 Legibility and the Risk of Illegitimate Transfer

Making a construction legible outside its original context invites transfer. If the conditions of validity are not preserved, such transfer becomes illegitimate. The resulting failures are often attributed to misuse or misunderstanding, but structurally they arise from boundary collapse.

An intelligent system must therefore treat legibility itself as a form of transfer, subject to the same criteria as action or inference. Where invariants cannot be preserved, exposure must be limited or delayed.

7.7 Transition to Boundary Conditions

The regulation of legibility is not an ad hoc defensive maneuver. It follows from the same structural principles that govern biological membranes, neural synchronization, and modular cognition. To make this continuity explicit, the next section examines boundary conditions across substrates, showing how semi-permeable boundaries, buffering, and refusal recur wherever intelligence persists under selection pressure.

8 Boundary Conditions Across Substrates

8.1 Boundaries as Conditions of Persistence

The regulation of legibility developed in the previous section is not a peculiarity of social or epistemic systems. It is a special case of a more general requirement: any system that persists under selection pressure must instantiate boundaries that are neither rigid barriers nor unrestricted channels. These boundaries regulate exchange, preserve internal gradients, and prevent uncontrolled propagation of perturbation.

A boundary, in this sense, is not defined by separation alone. It is defined by *selective permeability*. What matters is not whether interaction occurs, but under what conditions it is admitted, delayed, transformed, or refused.

This requirement recurs across biological, neural, computational, and epistemic domains. Its manifestations differ in material realization, but its structural role is invariant.

8.2 Cell Membranes and Ion Channels

At the cellular level, life depends on membranes that maintain electrochemical gradients. These gradients are not passive features of matter; they are actively regulated through ion channels whose opening and closing is context-sensitive. The cell survives by allowing some exchanges while preventing others, and by doing so at rates that preserve internal coherence.

Total openness would collapse the gradient and destroy the cell's capacity to perform work. Total closure would prevent adaptation and starve the cell of information and resources. The membrane therefore enforces a narrow regime in which selective exchange sustains function.

The lesson is structural rather than biological. Work, communication, and adaptation require differences to be maintained. Boundaries exist to protect those differences.

8.3 Cortical Synchronization and Markov Blankets

In nervous systems, boundaries reappear as patterns of synchronization and desynchronization among neural populations. Cortical regions are not globally synchronized at all times. Instead, they couple transiently through oscillatory alignment, forming functional assemblies that dissolve as conditions change.

These assemblies are often described in terms of Markov blankets: sets of variables that mediate interaction between a subsystem and its environment while shielding internal states from direct perturbation. The blanket does not block influence; it structures it.

Temporal gating plays the role of refusal. Signals that arrive out of phase are ignored or delayed. Information passes only when timing constraints align. Coherence is maintained not by global integration, but by regulated coupling.

8.4 Redundancy and Noncoding Structure in Evolution

At the scale of evolution, boundaries take the form of redundancy and buffering. Genomes contain large regions of noncoding material that do not directly specify phenotypic traits. These regions are often described as waste, but this description is misleading.

Redundant and noncoding sequences provide a buffer against mutation. They allow variation to occur without immediately corrupting essential functions. Over time, such variation may be co-opted into new structures, but only after selection has filtered it.

Evolution thus relies on apparent inefficiency to preserve long-term adaptability. Constraint is not imposed by minimizing variation, but by regulating its exposure to selection.

8.5 Policy Selection and Latent Variation

In cognitive and artificial systems, similar buffering appears as latent policy space. An intelligent system does not act on every possible policy it can generate. Instead, it maintains internal variation while exposing only a small subset of actions to the environment.

Policy selection functions as a boundary. It evaluates potential actions relative to constraints, suppresses those that violate jurisdiction, and delays commitment where uncertainty is high. Much of the system's internal computation never becomes externally visible.

This apparent waste is functional. Without it, exploration would be destructive rather than informative, and learning would collapse into immediate exploitation.

8.6 Boundaries as Constraints on Transfer

Across these substrates, the role of boundaries is consistent. They regulate transfer: of matter, of energy, of information, of influence. Transfer that is too permissive destroys internal structure; transfer that is too restrictive prevents adaptation.

The optimal regime is not fixed. It depends on context, history, and threat. Boundaries must therefore be dynamically regulated rather than statically imposed.

This dynamic regulation is precisely what distinguishes living, learning, and intelligent systems from inert ones.

8.7 Transition to Formalization

The recurrence of semi-permeable boundaries across domains suggests that the phenomenon is not contingent but necessary. To make this necessity precise, the next section introduces a formal characterization of boundaries, refusal, and invariance. Rather than appealing to metaphor, it will articulate the structural conditions under which adaptive systems remain coherent under reuse and exposure.

9 Formalizing Boundaries, Invariants, and Refusal

9.1 Why Formalization Is Necessary

Up to this point, the argument has proceeded by structural analysis across domains. The recurrence of semi-permeable boundaries, buffering, and refusal suggests that these are not contingent design choices but necessary conditions for persistence under selection pressure. However, without formalization, this necessity remains suggestive rather than binding.

Formalization serves a specific role here. It is not introduced to increase precision for its own sake, nor to provide executable specifications. Its function is to make explicit which transformations are admissible, which are not, and why. In doing so, it enforces scope and prevents illegitimate generalization.

The formalism employed is deliberately minimal. It aims to capture invariance and refusal without presupposing any particular substrate, representation, or implementation.

9.2 Objects, Transformations, and Invariants

We begin by treating locally coherent subsystems as abstract entities and interactions between them as transformations. What matters is not the internal constitution of these entities, but the conditions under which interactions preserve coherence.

Definition 9.1. Let an *object* denote a locally coherent construction, characterized by a construction history and a set of preserved invariants.

Definition 9.2. A *transformation* between two objects is admissible if it preserves the invariants required for the coherence of both source and target.

Invariants may encode conservation laws, semantic constraints, timing relations, jurisdictional limits, or other conditions discovered through construction history. They are not assumed to be universal. Each object may preserve a different family of invariants.

9.3 Admissibility and the Non-Existence of Paths

A crucial feature of this framework is that not all pairs of objects admit admissible transformations. The absence of a lawful path is not a failure of the formalism; it is its central expressive feature.

Definition 9.3. A *refusal* occurs when no admissible transformation exists between a proposed interaction and the invariants preserved by an object.

Refusal is thus modeled negatively. It is not a special action or signal. It is the recognition that a transformation cannot be composed without violating coherence.

This treatment avoids a common pitfall. If refusal were modeled as an ordinary operation, it would itself be subject to misuse or overextension. By contrast, modeling refusal as non-existence makes it structurally enforced rather than procedurally optional.

9.4 Boundaries as Restricted Domains of Interaction

Boundaries arise naturally from admissibility constraints. For any object, the set of admissible transformations defines a domain of interaction. Outside this domain, interactions are refused.

This yields a notion of semi-permeability. Some interactions are allowed, others are not, and the distinction depends on preserved invariants rather than on external authority or global rules.

Dynamic regulation enters through the evolution of invariants. As construction histories change, invariants may be strengthened, weakened, or refined. What was once admissible may later be refused, and vice versa. Boundaries are therefore not static walls but evolving constraints.

9.5 Buffering and Redundancy

The formal framework also accommodates buffering. Redundant internal structure corresponds to alternative constructions or latent transformations that are not currently admissible but may become so under future refinement of invariants.

Such redundancy is essential for adaptation. Without it, the system would have no internal degrees of freedom with which to explore new admissible interactions. With it, variation can occur internally without immediate exposure to irreversible external consequences.

Buffering therefore appears not as inefficiency, but as the formal precondition for learning under constraint.

9.6 Implications for General Intelligence

Within this framework, general intelligence is not characterized by the existence of many transformations, but by the system's capacity to regulate admissibility. A system is general to the extent that it can identify which invariants matter in a given context, evaluate proposed transformations against those invariants, refuse interactions that violate coherence, and revise its invariants through construction without collapse.

This definition is deliberately conservative. It privileges coherence over coverage and refusal over indiscriminate action. Yet it is precisely this conservatism that allows competence to accumulate rather than dissolve.

9.7 Transition to the Boundary Lemma

The formal considerations developed here allow the recurring boundary phenomena observed earlier to be stated as a general structural result. In the next section, this result is expressed as a lemma capturing the necessity of semi-permeable boundaries for any system that remains adaptive under selection pressure.

The lemma does not depend on biology, neuroscience, or computation in particular. It follows from the minimal requirements of invariance, interaction, and persistence.

10 The Semi-Permeable Boundary Theorem

10.1 Definitions

Definition 10.1. An adaptive system is a dynamical system $(\mathcal{Q}, \varphi_t)$ with irreversible construction history such that, under persistent environmental perturbation, it maintains a nonempty set of admissible future trajectories preserving its defining invariants.

Definition 10.2. A boundary is a mapping $\mathcal{B} : \mathcal{Q} \to \mathcal{C}$ that regulates admissible interactions between internal and external degrees of freedom by restricting the existence of transformations across contexts.

Definition 10.3. A boundary is *semi-permeable* if it admits a proper, nontrivial subset of interactions whose admissibility depends on context and history.

10.2 Theorem Statement

Theorem 10.1 (Semi-Permeable Boundary Theorem)**.** *Any adaptive system must instantiate semi-permeable boundaries. Systems whose boundaries are either fully permeable or fully impermeable cannot preserve adaptive capacity over time.*

10.3 Proof

Assume first that boundaries are fully permeable. Then for any perturbation, all transformations are admissible. Since invariants are defined by preservation under admissible transformations, this implies that no nontrivial invariants exist. Consequently, the system cannot distinguish lawful from unlawful reuse, and adaptive capacity collapses.

Assume instead that boundaries are fully impermeable. Then no interaction with the environment is admissible. The system cannot respond to perturbation, update construction history, or adjust internal structure. Adaptive capacity again collapses.

The only remaining regime admits some interactions while refusing others, with admissibility varying by context and history. This is precisely semi-permeability. Therefore any adaptive system must instantiate semi-permeable boundaries. □

This lemma does not assert that all adaptive systems look alike. It asserts that any system which remains coherent under reuse and exposure must satisfy these conditions in some material or formal realization.

10.4 Justification by Structural Necessity

The lemma follows from the minimal requirements of interaction and persistence. A system that adapts must exchange information or influence with its environment; otherwise it cannot respond to change. Yet if this exchange is unrestricted, internal distinctions erode faster than they can be reconstructed. The system loses the gradients upon which work, prediction, and learning depend.

Selective permeability resolves this tension. It allows interaction while preserving difference. Dynamic regulation is required because the environment is not static; what is admissible under

one set of conditions may be destructive under another. Buffering is required because adaptation proceeds through variation and selection, which demand internal slack. Constraint preservation is required because invariants are the currency of reuse.

None of these requirements can be relaxed independently. Removing any one collapses the regime in which adaptation is possible.

10.5 Non-Existence as a Positive Condition

A key feature of the lemma is its treatment of refusal. Refusal appears not as a compensatory mechanism added to an otherwise permissive system, but as the structural non-existence of certain interactions.

This is counterintuitive only if one assumes that capability consists in the availability of options. On the present account, capability consists in the availability of *lawful* options. The absence of an unlawful transformation is therefore not a limitation, but a condition of coherence.

Boundaries enforce this absence. They do not need to represent what is forbidden; they simply do not admit it.

10.6 Substrate Independence

The lemma is deliberately substrate-neutral, applying equally to biological systems that maintain metabolic gradients, neural systems that coordinate through oscillatory synchrony, evolutionary systems that buffer variation through redundancy, computational systems that enforce privilege separation, and epistemic systems that regulate disclosure and reuse. In each case, the material realization differs, yet the structural role remains identical. Where boundaries fail, systems either stagnate or disintegrate.

10.7 Transition to Scope and Refusal

The lemma establishes boundaries as a necessary condition for adaptation. It remains to show how these boundaries operate internally in intelligent systems. In particular, how scope is determined, and how refusal functions as a boundary operator rather than as a failure mode.

The next section develops this consequence formally, showing that scope restriction and refusal are not optional safeguards, but constitutive features of any general intelligence.

11 Scope, Refusal, and Boundary Operators

11.1 From Boundaries to Scope

The Semi-Permeable Boundary Lemma establishes that adaptive systems must regulate interaction. For intelligent systems, this regulation appears internally as *scope*. Scope determines where a construction, abstraction, or policy applies, and where it does not.

Scope is not an annotation added after the fact. It is an emergent property of construction history. Each abstraction inherits the conditions under which it was formed, refined, and validated. These conditions delimit the contexts in which reuse remains lawful.

A system that cannot track scope implicitly treats all contexts as equivalent. This is not generality; it is indiscrimination.

11.2 Scope as a Boundary Operator

Within the formal framework introduced earlier, scope functions as a boundary operator on admissible transformations. It restricts the domain in which morphisms may exist by enforcing invariants derived from construction history.

Definition 11.1. The *scope* of a construction is the set of contexts for which admissible transformations preserving its invariants exist.

Outside this set, no lawful transformation exists. Attempted reuse is therefore refused, not because the system lacks capacity, but because the required invariants cannot be preserved.

Scope thus encodes both competence and its limits. To know what one can do is inseparable from knowing where that doing remains valid.

11.3 The Scope–Refusal Corollary

The relationship between boundaries, scope, and refusal can now be stated explicitly.

Corollary 11.1 (Scope–Refusal Corollary)**.** *In any system capable of general intelligence, scope restriction and refusal are necessary boundary operators. A system that lacks mechanisms to determine the scope of its abstractions and to refuse actions, inferences, or disclosures outside that scope cannot preserve its invariants under reuse and therefore cannot remain coherent as it generalizes.*

This corollary follows directly from the Semi-Permeable Boundary Lemma. If boundaries are required for adaptation, and if scope determines where boundaries apply internally, then refusal is the operation by which scope is enforced.

11.4 Refusal as Structural Non-Existence

It is important to emphasize that refusal is not implemented as a special case or exception. In the formal account, refusal corresponds to the non-existence of an admissible transformation. This makes refusal robust. It cannot be overridden by optimization pressure or convenience, because there is nothing to override.

This treatment also explains why refusal often appears unintuitive or frustrating to external observers. From outside the system, refusal looks like an absence of response. From inside, it is the only coherent response available.

11.5 Asymmetry and Partial Applicability

Scope need not be symmetric. A construction may apply in one direction but not in another; a transformation may preserve invariants in one context but violate them in reverse. Such asymmetries are common in biological, cognitive, and epistemic systems.

Recognizing this prevents a common error: assuming that applicability must be mutual or universal. General intelligence does not require symmetry. It requires discrimination.

11.6 Implications for Learning and Revision

Scope and refusal also govern revision. When a construction fails outside its scope, the system must decide whether to expand the scope by discovering new invariants, or to maintain refusal. Both options are legitimate. What matters is that expansion is earned through construction, not assumed by default.

Learning, on this view, is the gradual reshaping of scope through experience. Refusal marks the current boundary; exploration tests whether that boundary can be safely moved.

11.7 Transition to Invariance Across Systems

The corollary clarifies how boundaries operate within intelligent systems. The final step is to show that these same boundary operators recur across biological, neural, computational, and epistemic domains. This recurrence is not accidental. It reflects a deeper invariance.

In the next section, this invariance is stated as a general theorem unifying membranes, Markov blankets, privilege separation, and effort-gated legibility under a single structural principle.

12 The Boundary Invariance Theorem

12.1 Statement of the Theorem

The preceding analysis has shown that boundaries, scope, and refusal are necessary for coherent adaptation within individual systems. What remains is to show that these requirements are not domain-specific, but invariant across substrates. This can now be stated formally.

Theorem 12.1 (Boundary Invariance Theorem)**.** *Any system—biological, neural, computational, or epistemic—that remains adaptive under reuse, exposure, and selection pressure must implement boundary conditions satisfying all of the following: selective permeability, admitting only admissible interactions; buffering, maintaining latent or redundant internal structure; dynamic regulation, with boundary conditions varying by context and history; refusal, enforced as the non-existence of illegitimate transformations; and graded access, distinguishing inspection from modification. These conditions are invariant across substrate and scale. Systems that fail to satisfy them may exhibit short-term performance but cannot sustain general intelligence.*

12.2 Proof by Structural Correspondence

The proof proceeds by demonstrating that each clause of the theorem is realized, in structurally equivalent form, across distinct domains.

12.2.1 Biological and Neural Systems

In biological systems, selective permeability is instantiated by cell membranes and ion channels that regulate the exchange of matter and charge. Buffering appears in metabolic reserves and noncoding genetic material, which allow variation without immediate functional loss. Dynamic regulation is achieved through context-sensitive channel gating and regulatory networks.

In neural systems, selective permeability appears as oscillatory synchronization that admits signals only under phase alignment. Markov blankets mediate interaction between internal and external states, enforcing refusal through desynchronization or inhibition. Graded access arises through layered processing, where sensory input may influence internal states without directly modifying long-term structure.

Failure of these mechanisms leads to loss of coherence, as seen in pathological synchronization or metabolic collapse.

12.2.2 Computational Systems

In computational systems, particularly operating systems, selective permeability is enforced through permission models. Read access allows inspection without modification; write and execute permissions regulate alteration. Privilege separation enforces graded access, requiring explicit escalation for operations that affect global state.

Buffering appears as redundancy, indirection layers, and sandboxed processes. Dynamic regulation occurs through runtime checks and context-dependent permissions. Refusal is enforced structurally: illegal system calls do not exist within unprivileged contexts.

Systems lacking such boundaries are not more powerful. They are less stable.

12.2.3 Epistemic and Intelligent Systems

In epistemic systems, selective permeability is realized through formal constraints, scope conditions, and effort-gated legibility. Buffering appears as redundancy in exposition, delayed conclusions, and non-operational formalism that preserves structure without enabling immediate deployment.

Dynamic regulation is achieved by varying levels of disclosure across contexts. Refusal appears as the absence of recipes, stepwise instructions, or context-free rules. Graded access distinguishes understanding from authority to modify or apply.

When these boundaries fail, ideas propagate without constraint, leading to distortion, misuse, and epistemic collapse.

12.3 Non-Equivalence to Secrecy

It is essential to distinguish boundary enforcement from secrecy. In all cases considered, the boundaries do not conceal structure; they regulate interaction. Inspection remains possible. What is restricted is modification, deployment, or transfer without constraint.

This distinction explains why effort-gated legibility is ethically cleaner than concealment. It preserves openness while preventing illegitimate reuse.

12.4 Consequences for General Intelligence

The theorem implies that general intelligence cannot be equated with maximal openness, maximal capability, or maximal transferability. It is characterized instead by the disciplined placement of boundaries that preserve invariants under reuse.

Any proposal for general intelligence that omits refusal, buffering, or graded access is structurally incomplete. Such systems may scale briefly, but they do not persist.

12.5 Transition to Architectural Implications

The Boundary Invariance Theorem completes the formal core of the argument. What remains is to articulate its consequences for the design and interpretation of general intelligence architectures, and for the ethics of publishing powerful ideas.

In the next section, these consequences are developed explicitly, not as prescriptions, but as architectural implications that follow from the invariants established here.

13 A Blueprint for General Intelligence Under Constraint

13.1 Architectures, Not Recipes

The Boundary Invariance Theorem places a strict limit on what it can mean to specify a general intelligence. Any attempt to present a stepwise procedure, an algorithmic recipe, or an operational checklist immediately violates the very constraints that make general intelligence coherent. What can be specified, instead, is an *architecture*: a set of structural conditions that any admissible realization must satisfy.

An architecture constrains what may exist without dictating how it must be built. It describes invariants, interfaces, and refusal conditions rather than executable steps. This distinction is not rhetorical. It is the only way to speak meaningfully about general intelligence without collapsing it into extractable technique.

13.2 Replayable Construction History

A general intelligence must be grounded in replayable construction history. Its identity cannot be reduced to a static configuration or parameter set. Every competence it exhibits must be traceable to a sequence of irreversible events: observations, abstractions, compressions, and refusals.

States, where they appear, are projections compiled from this history under explicit scope. They are not authoritative. Learning consists in reorganizing history into forms that are cheaper to replay while preserving the invariants that justify reuse.

Any system that discards its history in order to optimize present performance forfeits generality.

13.3 Strong Modularity and Jurisdiction

The architecture must be modular in the strong sense developed earlier. Functional units maintain independent construction histories and preserve local invariants. No privileged global workspace

exists in which all distinctions collapse.

Interaction occurs only through explicit interfaces that declare jurisdiction. These interfaces do not merely transmit information; they enforce scope. A module may inspect the outputs of another without acquiring the authority to modify its internal structure or to reuse its abstractions outside their domain of validity.

This separation is the condition under which learning accumulates rather than interferes with itself.

13.4 Abstraction as Invariant-Preserving Compression

Abstraction operates by replacing detailed construction histories with more economical ones that behave equivalently under admissible transformations. The architecture must therefore track not only abstractions, but the invariants they preserve and the transformations under which they remain valid.

Summarization and extension are the same operation viewed from opposite directions. Both are evaluated by preservation, not novelty. A general intelligence improves over time not by increasing the number of abstractions it holds, but by refining the invariants that govern their reuse.

13.5 Lawful Transfer and Distributed Evaluation

Generality is achieved through lawful transfer. Proposed reuse of a construction is evaluated locally by the modules involved, relative to their own invariants and construction histories. Where alignment exists, transfer proceeds. Where it does not, refusal blocks composition.

There is no centralized authority that forces generalization. Coherence emerges from distributed refusal.

This architecture scales because it fails locally. Errors are contained, diagnoses are possible, and revision does not require global rollback.

13.6 Refusal as a Constitutive Operation

Refusal is not an error condition or a safety override. It is a constitutive operation of the architecture. It appears whenever no admissible transformation exists between a proposed action, inference, or disclosure and the preserved invariants of the system.

Refusal preserves optionality under uncertainty and prevents irreversible commitments from being made without justification. A system that cannot refuse cannot remain coherent as it generalizes.

13.7 Local Coherence and Conditional Globality

The architecture maintains multiple locally coherent constructions that may or may not integrate. Global structure is not assumed. It is achieved only where gluing conditions are satisfied. Even then, such integration remains conditional and revisable.

This prevents the collapse of local constraints into globally inconsistent abstractions. It allows the system to operate effectively across heterogeneous domains without enforcing premature unification.

13.8 Regulation of Legibility

Finally, the architecture treats legibility as an internal control variable. External projections of internal structure are chosen relative to context, threat, and irreversibility. Different projections preserve different invariants.

The system may expose competence where coordination is beneficial and conceal structure where extraction would be destructive. This regulation is not deception. It is boundary maintenance.

13.9 What This Blueprint Excludes

This architectural characterization excludes, by necessity, monolithic objective functions, unrestricted optimization, context-free abstractions, and unconditional disclosure. These exclusions are not normative judgments; rather, they follow directly from the boundary conditions required for persistence under reuse and exposure.

13.10 Transition to Publication and Ethics

The blueprint completes the internal characterization of general intelligence. What remains is to consider its external consequences. If intelligence itself depends on boundaries, refusal, and regulated transfer, then the communication of intelligent structures must obey the same logic.

The next section addresses this implication directly, examining the ethics of publication, openness, and legibility under the constraints established throughout the essay.

14 Publication, Openness, and Epistemic Responsibility

14.1 From Internal Architecture to External Disclosure

The preceding section characterizes general intelligence in terms of internal architecture: construction history, modularity, abstraction, refusal, and regulated transfer. These properties do not terminate at the boundary of the system. They extend outward into the system's interactions with its environment, including how its structures are communicated, taught, or published.

Publication is itself a form of transfer. It transports constructions from one epistemic context into others, often far removed from the conditions under which they were formed. As such, publication is subject to the same constraints as action, inference, and reuse. To treat it as exempt is to introduce a structural inconsistency.

If intelligence depends on boundaries internally, then responsible disclosure must respect boundaries externally.

14.2 Why Openness Is Not a Scalar Virtue

Openness is frequently framed as a scalar good: more transparency is better, more access is better, more reproducibility is better. This framing collapses distinct kinds of interaction into a single dimension and ignores the role of scope.

Inspection, understanding, modification, and deployment are not interchangeable. They require different kinds of access and carry different risks. A system that permits inspection without permitting modification may remain coherent; a system that permits unrestricted modification may not.

Treating openness as unqualified obscures this distinction. It conflates the right to see with the right to act, and the ability to describe with the authority to apply.

14.3 Effort-Gated Legibility

One way to preserve this distinction without resorting to secrecy is through effort-gated legibility. An artifact is effort-gated when understanding it requires reconstructing the constraints that make it valid. The effort is not imposed artificially; it arises from the structure itself.

Formal definitions, cumulative dependencies, and non-operational descriptions serve this role. They allow scrutiny and critique while preventing immediate extraction. Nothing is hidden, but nothing is portable without discipline.

This approach mirrors the graded access observed in other domains. Reading source code does not grant permission to write to the kernel. Observing neural activity does not grant control over cognition. Understanding an abstraction does not automatically authorize its deployment.

14.4 Why Recipes Are Categorically Different

A recipe collapses architecture into procedure. It removes context, suppresses refusal, and presents transfer as unconditional. In doing so, it destroys the invariants that made the original construction coherent.

Publishing recipes for general intelligence is therefore not an act of openness but a category error. It invites reuse without scope, optimization without constraint, and deployment without responsibility.

By contrast, publishing architectures preserves structure while refusing extraction. It allows understanding to scale with effort and competence rather than with reach alone.

14.5 Misuse as a Structural Failure

Misuse is often framed as a moral failing of users rather than as a design failure of artifacts. While intent matters, this framing is incomplete. When misuse is predictable, it is structural.

An artifact that can be easily misapplied without reconstructing its constraints has already failed to enforce its own scope. Responsibility does not lie solely with the user; it lies with the absence of boundaries that would have prevented illegitimate transfer.

Effort-gated legibility addresses this failure by aligning access with competence.

14.6 Refusal at the Level of Publication

Just as intelligent systems must refuse illegitimate actions, authors and institutions must refuse illegitimate forms of disclosure. This refusal is not censorship. It is the preservation of coherence.

Refusal at the level of publication may take many forms: declining to provide stepwise instructions, delaying operational detail, emphasizing formal constraints over application, or requiring cumulative engagement. These are not evasions. They are boundary operators.

14.7 The Ethics of Constraint Preservation

The ethical stance that follows from this framework is neither maximal openness nor protective secrecy. It is constraint preservation. The primary obligation is not to maximize access, but to ensure that what is accessed remains meaningful and survivable under reuse.

This stance may appear conservative in environments that reward immediacy and portability. Yet it is precisely this conservatism that allows knowledge to accumulate rather than degrade.

15 Variational Formulation of General Intelligence

15.1 State Variables and Configuration Space

We model a general intelligence as a dynamical system evolving on a constrained configuration space rather than as a static computational artifact. The system is characterized by a set of generalized coordinates capturing coherence, action tendency, latent variation, and boundary structure.

Let the configuration at time t be given by

$$q(t) = (\Phi(t), \mathbf{v}(t), S(t), \pi(t), \mathcal{B}(t)),$$

where $\Phi(t)$ denotes a scalar coherence or abstraction density, $\mathbf{v}(t)$ denotes a vector-valued policy flow or action tendency, $S(t)$ denotes internal entropy or latent slack, $\pi(t)$ denotes an internal policy distribution, and $\mathcal{B}(t)$ denotes boundary or scope structure mediating interaction.

These variables are not representations of the environment. They are intensities describing internal structure and admissible change.

The system evolves on a constrained manifold $\mathcal{M} \subset \mathcal{Q}$ determined by preserved invariants. Points outside $\mathcal{M}$ correspond to incoherent or illegitimate configurations.

15.2 Action Principle

The evolution of the system is determined by an action functional

$$\mathcal{S}[q] = \int_{t_0}^{t_1} \mathcal{L}(q, \dot{q})\, dt,$$

where admissible trajectories are those that extremize $\mathcal{S}$ subject to the constraints defining $\mathcal{M}$.

General intelligence, on this account, is not defined by the ability to reach arbitrary configurations, but by the existence of stable low-action trajectories that preserve invariants under reuse and exposure.

15.3 Lagrangian Structure

We decompose the Lagrangian as

$$\mathcal{L} = T - V - \Lambda + \Xi,$$

where each term encodes a distinct structural requirement.

15.3.1 Kinetic Term

The kinetic term measures the capacity for change:

$$T = \frac{1}{2} m_\Phi \dot{\Phi}^2 + \frac{1}{2} m_v \|\dot{\mathbf{v}}\|^2 + \frac{1}{2} m_\pi \|\dot{\pi}\|^2.$$

These terms do not encode performance, but responsiveness: the system's ability to adapt its internal structure.

15.3.2 Invariant Potential

Invariant preservation is enforced by a potential term

$$V = \alpha \, d(\Phi, \mathcal{I})^2 + \beta \, d(\mathbf{v}, \mathcal{A}_{\text{adm}})^2,$$

where $\mathcal{I}$ denotes the invariant manifold associated with coherent abstraction, and $\mathcal{A}_{\text{adm}}$ denotes the admissible action set.

Deviation from preserved invariants incurs increasing energetic cost. Illegitimate generalization corresponds to climbing the potential.

15.3.3 Boundary and Legibility Penalty

Boundary integrity is enforced by

$$\Lambda = \gamma \|\nabla \mathcal{B}\|^2 + \delta \, \kappa(\mathcal{B}, E),$$

where κ measures coupling to external observational or extractive fields E.

This term penalizes uncontrolled boundary collapse and excessive exposure. Interaction remains possible, but unregulated legibility is energetically disfavored.

15.3.4 Learning and Compression Term

Learning appears as constrained entropy descent:

$$\Xi = -\eta \frac{d}{dt} \mathrm{KL}\big(\pi \,\|\, \pi_{\text{comp}}\big),$$

where π_{comp} denotes a compressed policy distribution preserving invariant behavior.

This term favors reorganization of internal structure that reduces description length without violating scope.

15.4 Euler–Lagrange Dynamics

Admissible trajectories satisfy the Euler–Lagrange equations

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}_i} - \frac{\partial \mathcal{L}}{\partial q_i} = 0,$$

subject to the constraints defining $\mathcal{M}$.

Constraint forces arise naturally from the geometry of the configuration space and require no explicit safety rules.

15.5 Hamiltonian Formulation

Define conjugate momenta

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i}.$$

The Hamiltonian is given by

$$\mathcal{H}(q, p) = \sum_i p_i \dot{q}_i - \mathcal{L}.$$

In this formulation, refusal corresponds to regions of phase space where

$$\mathcal{H} \to \infty.$$

No admissible trajectory enters such regions. Refusal is therefore structural rather than procedural.

15.6 Refusal as Forbidden Phase Space

Configurations that violate invariants, collapse boundaries, or destroy replayable construction history lie outside $\mathcal{M}$ and are separated by divergent action.

The system does not evaluate these configurations and reject them. They are dynamically inaccessible.

This realizes refusal as a property of the action landscape rather than as an explicit decision.

15.7 Interpretation

Under this formulation, general intelligence is characterized by stable low-action trajectories under reuse, invariant-preserving abstraction as compression, lawful transfer constrained by geometry, refusal encoded as inaccessible phase space, and legibility regulated by boundary energetics. No objective function is maximized, and no global reward is assumed. Coherence is preserved by the structure of the dynamics itself.

16 Field-Theoretic Extension of the Variational Model

16.1 Spatially Extended Intelligence Fields

To model general intelligence as a distributed system rather than a point dynamical process, we extend the variational formulation to fields defined over a spatial domain $\Omega \subset \mathbb{R}^n$.

Let the system be described by fields

$$\Phi(x,t), \quad \mathbf{v}(x,t), \quad S(x,t), \quad \pi(x,t), \quad \mathcal{B}(x,t),$$

where $x \in \Omega$ indexes internal structure, memory locality, or semantic region.

The action functional becomes

$$\mathcal{S} = \int dt \int_\Omega \mathcal{L}(\Phi, \partial_t\Phi, \nabla\Phi, \mathbf{v}, \partial_t\mathbf{v}, \nabla\mathbf{v}, S, \pi, \mathcal{B})\ dx.$$

This formulation treats intelligence as a continuous medium whose coherence depends on the regulated propagation of structure.

16.2 Field Lagrangian Density

A representative Lagrangian density is given by

$$\begin{aligned}\mathcal{L} = &\frac{1}{2}\rho_\Phi(\partial_t\Phi)^2 - \frac{1}{2}c_\Phi\|\nabla\Phi\|^2 \\ &+ \frac{1}{2}\rho_v\|\partial_t\mathbf{v}\|^2 - \frac{1}{2}c_v\|\nabla\mathbf{v}\|^2 \\ &- U(\Phi,\mathbf{v},\mathcal{J}) - \Lambda(\mathcal{B},\nabla\mathcal{B}) + \Xi(\pi,S).\end{aligned}$$

Gradient penalties enforce smoothness and locality, U penalizes deviation from invariant manifolds, Λ enforces boundary integrity and limits information flux, and Ξ encodes constrained entropy descent. Sharp discontinuities correspond to boundary failure and incur unbounded action.

16.3 Euler–Lagrange Field Equations

The resulting field equations take the form

$$\partial_t\left(\frac{\partial\mathcal{L}}{\partial(\partial_t\Phi)}\right) - \nabla\cdot\left(\frac{\partial\mathcal{L}}{\partial(\nabla\Phi)}\right) + \frac{\partial\mathcal{L}}{\partial\Phi} = 0,$$

with analogous equations for $\mathbf{v}$ and the remaining fields.

Information, policy influence, and abstraction propagate as waves or diffusions only where permitted by boundary structure. Illegitimate propagation is dynamically suppressed.

17 Invariants and Conserved Quantities

17.1 Symmetry and Constraint

Invariant structure in the intelligence dynamics corresponds to symmetry of the action functional. Unlike physical symmetries of space and time, the relevant symmetries here act on abstraction, policy, and boundary variables.

Let $\mathcal{G}$ be a continuous group of transformations acting on the fields such that

$$\mathcal{S}[q] = \mathcal{S}[g \cdot q] \quad \forall g \in \mathcal{G}.$$

These symmetries encode lawful reuse: transformations under which abstraction remains valid.

17.2 Noether-Type Theorem

Theorem 17.1 (Invariant Preservation Theorem)**.** *If the action $\mathcal{S}$ is invariant under a continuous group of transformations $\mathcal{G}$ acting on the configuration fields, then there exists a conserved quantity J along all admissible trajectories.*

Violation of the associated invariant corresponds to a breakdown of conservation and induces divergent action.

17.3 Interpretation of Conserved Quantities

The conserved quantities arising here are not physical momenta but structural measures, including abstraction coherence flux, policy consistency current, and boundary integrity charge. Conservation expresses the fact that lawful transfer preserves structure. Attempts to reuse abstractions outside their symmetry class necessarily inject or dissipate conserved quantities and are therefore dynamically forbidden.

17.4 Refusal as Symmetry Breaking

Refusal corresponds to the absence of symmetry extension. When a proposed transformation lies outside $\mathcal{G}$, no conserved current exists, and the action becomes unbounded.

Thus refusal is not an exception to the dynamics; it is the absence of a symmetry under which motion could proceed.

18 CLIO Dynamics as a Slow–Fast Decomposition

18.1 Separation of Timescales

General intelligence requires simultaneous stability and adaptability. This is realized through a separation of timescales between fast policy dynamics and slow structural optimization.

Let:

$$(\Phi, \mathcal{B}) \quad \text{evolve on a slow timescale,}$$

$$(\mathbf{v}, \pi) \quad \text{evolve on a fast timescale.}$$

This induces a foliation of phase space into slow manifolds parameterized by boundary and abstraction fields.

18.2 Fast Policy Flow

On short timescales, policy variables evolve according to

$$\dot{\pi} = -\nabla_{\pi} \mathcal{H}_{\text{fast}}(\pi \mid \Phi, \mathcal{B}),$$

where $\mathcal{H}_{\text{fast}}$ is the Hamiltonian restricted to fixed structural variables.

This corresponds to CLIO-style rapid inference and local policy adjustment.

18.3 Slow Structural Evolution

On longer timescales, abstraction and boundary fields evolve via

$$\partial_t \Phi = -\epsilon \, \frac{\delta \mathcal{H}}{\delta \Phi}, \quad \partial_t \mathcal{B} = -\epsilon \, \frac{\delta \mathcal{H}}{\delta \mathcal{B}},$$

with $0 < \epsilon \ll 1$.

Structural change occurs only after repeated fast dynamics demonstrate invariant preservation or violation.

18.4 Policy Refusal via Manifold Geometry

Policies that require leaving the slow manifold correspond to directions of infinite curvature in the Hamiltonian landscape. These directions are dynamically inaccessible.

Refusal therefore appears as geometric obstruction: no slow-fast trajectory exists that preserves invariants.

18.5 Interpretation

CLIO emerges here not as an algorithm but as a dynamical regime in which fast loops explore policy space locally, slow loops reshape abstraction and boundary geometry, and only invariant-preserving couplings survive averaging. This realizes intelligence as constrained exploration over time rather than immediate optimization.

18.6 Averaging and Emergent CLIO Dynamics

Let the system dynamics be written as

$$\dot{x} = f(x, y), \qquad \dot{y} = \frac{1}{\epsilon} g(x, y),$$

where x denotes slow structural variables $(\Phi, \mathcal{B})$, y denotes fast policy variables $(\pi, \mathbf{v})$, and $0 < \epsilon \ll 1$.

For fixed x, the fast subsystem $\dot{y} = g(x, y)$ admits a unique invariant measure μ_x and is mixing with respect to μ_x.

Theorem 18.1 (CLIO Averaging Theorem)**.** *As $\epsilon \to 0$, trajectories of the full system converge uniformly on compact time intervals to solutions of the averaged system*

$$\dot{x} = \int f(x, y)\, d\mu_x(y).$$

18.7 Interpretation

The averaged dynamics define an effective flow on the slow manifold. Fast policy exploration does not directly determine long-term behavior; only invariant-preserving effects survive averaging.

CLIO therefore arises as a dynamical regime: rapid internal exploration coupled to slow structural consolidation. No explicit algorithmic separation is required. The separation is enforced by timescale geometry.

18.8 Failure Without Separation

If timescale separation fails, averaging breaks down. Fast dynamics leak directly into structural variables, inducing boundary collapse and invariant violation. Such systems exhibit brittle generalization and unstable transfer.

19 Stability of Invariant-Preserving Dynamics

19.1 Motivation

A theory of general intelligence that emphasizes constraint, refusal, and boundary regulation must establish that such systems are not merely coherent but dynamically stable. Without stability, the architecture would describe a fragile ideal rather than a realizable regime.

Stability here does not mean convergence to a fixed point. It means persistence of coherent behavior under perturbation, reuse, and partial exposure.

19.2 Invariant-Preserving Attractors

Let $\mathcal{M} \subset \mathcal{Q}$ denote the manifold of admissible configurations preserving the system's invariants. The Hamiltonian dynamics induce a flow φ_t on $\mathcal{Q}$.

Definition 19.1. An *invariant-preserving attractor* is a compact set $\mathcal{A} \subset \mathcal{M}$ such that trajectories starting in a neighborhood of $\mathcal{A}$ remain in $\mathcal{M}$ for all future time, deviations transverse to $\mathcal{M}$ experience restoring forces induced by the invariant potential, and motion within $\mathcal{A}$ preserves the conserved quantities associated with the system's symmetries.

Such attractors correspond to stable regimes of intelligent behavior rather than static solutions.

19.3 Stability Theorem

Theorem 19.1 (Invariant Stability Theorem)**.** *Assume the following conditions: the invariant potential V is coercive outside $\mathcal{M}$; boundary penalties Λ diverge under boundary collapse; entropy descent Ξ is bounded below; and slow–fast separation holds between policy and structural variables. Then the induced Hamiltonian flow admits invariant-preserving attractors $\mathcal{A} \subset \mathcal{M}$ that are Lyapunov-stable under admissible perturbations.*

19.4 Sketch of Argument

Coercivity of V ensures that trajectories leaving $\mathcal{M}$ incur unbounded energy cost, producing restoring forces normal to the manifold. Divergence of Λ prevents collapse of boundary structure, eliminating destabilizing modes associated with unregulated coupling.

Bounded entropy descent ensures that learning does not induce runaway collapse of internal degrees of freedom. Slow–fast separation ensures that rapid policy fluctuations average out before structural variables evolve, yielding effective stability on long timescales.

Together, these conditions guarantee persistence of coherent trajectories without requiring global optimization or centralized control.

19.5 Basins of Attraction and Convergence Rates

Let $\mathcal{A} \subset \mathcal{M}$ be an invariant-preserving attractor.

Definition 19.2. The basin of attraction $\mathcal{U}(\mathcal{A})$ is the set of initial conditions whose trajectories converge to $\mathcal{A}$ while remaining in $\mathcal{M}$.

If the invariant potential V is strongly coercive transverse to $\mathcal{M}$, then $\mathcal{U}(\mathcal{A})$ contains an open neighborhood of $\mathcal{A}$.

19.6 Exponential Stability

Theorem 19.2. *If, in addition, V is locally strictly convex in directions normal to $\mathcal{M}$, then trajectories converge exponentially to $\mathcal{A}$.*

19.7 Bifurcation Under Boundary Weakening

Let λ parameterize boundary strength in Λ.

As $\lambda \to 0$, invariant-preserving attractors generically lose stability through boundary-collapse bifurcations, resulting in unbounded information flux and loss of coherence.

19.8 Interpretation

Stability is not incidental; it depends quantitatively on boundary strength and invariant enforcement. Robust intelligence occupies a bounded region of parameter space, outside of which collapse

is generic. Stability is achieved not by suppressing change, but by constraining it: the system remains adaptive within $\mathcal{M}$ while being dynamically excluded from incoherent regimes. This explains how intelligence can remain flexible without dissolving itself under generalization pressure.

20 The State-Projection No-Go Theorem

20.1 Motivation

Many prevailing accounts of intelligence, learning, and generalization implicitly or explicitly assume that a system's operative identity can be captured by its instantaneous state. On this view, learning updates a state, generalization operates over states, and admissibility of action or inference is decided by reference to the current state alone.

The framework developed in this work rejects that assumption. Lawful transfer, refusal, and invariant preservation depend essentially on construction history. This section makes that dependence precise by establishing a no-go result: any purely state-based account necessarily discards information required to determine admissible reuse.

20.2 Formal Setup

Let $\mathcal{H}$ denote the space of construction histories, where each history is an ordered, irreversible sequence of events. Let $\mathcal{S}$ denote a space of states.

A *state projection* is a mapping

$$P : \mathcal{H} \to \mathcal{S}$$

that assigns to each history a state intended to summarize all relevant information for future behavior.

Let $\mathcal{A}(h)$ denote the set of admissible transformations (actions, inferences, transfers) available after history $h \in \mathcal{H}$. Admissibility is defined relative to preserved invariants, as developed in previous sections.

20.3 Lawful Transfer Criterion

A transformation τ is *lawful* after history h if and only if applying τ preserves the invariants induced by h. Lawfulness is therefore a property of the pair (h, τ), not of τ alone.

Two histories h_1, h_2 are *lawfully equivalent* if

$$\mathcal{A}(h_1) = \mathcal{A}(h_2).$$

A state projection P is said to be *lawful* if

$$P(h_1) = P(h_2) \;\Rightarrow\; \mathcal{A}(h_1) = \mathcal{A}(h_2).$$

That is, identical states must imply identical admissible transformations.

20.4 Theorem Statement

Theorem 20.1 (State-Projection No-Go Theorem)**.** *There exists no non-injective state projection*

$$P : \mathcal{H} \to \mathcal{S}$$

that is lawful with respect to admissible transformations. In particular, any many-to-one projection necessarily collapses histories with distinct admissibility structures.

20.5 Proof

Let P be any non-injective projection. Then there exist distinct histories $h_1 \neq h_2$ such that

$$P(h_1) = P(h_2).$$

Because histories are sequences of irreversible events, h_1 and h_2 differ by at least one event e whose occurrence or non-occurrence contributes to construction history.

By the definition of invariants, there exists at least one transformation τ whose admissibility depends on whether e occurred. Otherwise, e would be invariant-irrelevant and could be removed from the construction history without consequence, contradicting irreversibility.

Thus,

$$\tau \in \mathcal{A}(h_1) \quad \text{and} \quad \tau \notin \mathcal{A}(h_2),$$

or vice versa.

But since $P(h_1) = P(h_2)$, any decision procedure based solely on state must assign the same admissibility judgment to τ in both cases. Therefore P cannot preserve admissibility.

Hence P is not lawful. Since P was arbitrary, no non-injective lawful state projection exists. □

20.6 Corollaries

Corollary 20.1. *Any system that decides admissibility solely on the basis of instantaneous state cannot correctly implement refusal.*

Corollary 20.2. *Any post hoc safety or filtering mechanism applied after state projection cannot recover admissibility information lost in the projection.*

Corollary 20.3. *Generalization error arising from state-only representations is structural, not contingent on data, training, or optimization quality.*

20.7 Implications for Architecture

The no-go theorem does not imply that states are useless. It implies that states are necessarily partial views: projections compiled from history under specific scope.

Admissibility, refusal, and lawful transfer cannot be functions of state alone. They require access to construction history or to structures that are provably equivalent to it with respect to invariants.

Architectures that treat state as ontologically primary must therefore externalize refusal, imposing it as a rule, filter, or objective. Architectures grounded in construction history enforce refusal structurally, through the non-existence of admissible transformations.

20.8 Relation to Prior Results

This theorem formalizes the claims made earlier regarding events over states, modularity, and refusal. It explains why boundary-regulated, history-sensitive systems admit stable generalization, while state-centric systems exhibit brittle transfer.

The result is independent of substrate, learning rule, or representational format. It follows solely from the irreversibility of construction history and the definition of lawful transfer.

21 Computational Realizability Under Constraint

21.1 Scope of the Question

The preceding sections have developed a theory of general intelligence as constrained dynamics on a manifold of admissible configurations. The formulation has employed idealized objects: infinite barriers, continuous fields, unbounded construction histories, and exact invariants.

This section addresses a natural concern: whether such a theory is merely mathematical, or whether its essential claims survive under finite, approximate, and computationally realizable conditions.

The aim is not to provide an implementation recipe. Rather, it is to show that the theory is robust under approximation, and that its core constraints do not rely on unphysical idealizations.

21.2 Finite Approximations of Infinite Barriers

Infinite action barriers were introduced to model refusal as the non-existence of admissible trajectories. In computational systems, infinite quantities cannot be represented. However, exact infinity is not required for structural effect.

Let $V_\lambda(q)$ be a family of potentials indexed by $\lambda > 0$ such that

$$V_\lambda(q) \to \infty \quad \text{as } q \to \partial\mathcal{M}$$

and

$$V_\lambda(q) \geq \lambda \quad \text{for } q \notin \mathcal{M}.$$

For sufficiently large λ, the qualitative refusal behavior induced by V_λ is equivalent to that of an infinite barrier: trajectories starting in $\mathcal{M}$ remain in $\mathcal{M}$ for all practical timescales.

Thus refusal need not be absolute to be effective. It need only dominate the relevant dynamics.

21.3 Approximate Invariants

Exact invariant preservation is similarly idealized. In finite systems, invariants may only be preserved up to tolerance.

Definition 21.1. An ε-invariant is a property whose deviation under admissible dynamics remains bounded by ε over relevant timescales.

If invariant violation incurs a restoring force whose magnitude dominates noise and discretization error, then ε-invariants suffice to preserve lawful transfer and refusal behavior.

This mirrors robustness results in physical systems, where approximate symmetries still yield approximately conserved quantities sufficient for stability.

21.4 Complexity of Admissibility Checking

A further concern is the computational cost of determining whether a proposed transformation is admissible.

Let the *admissibility decision problem* be defined as:

$$\text{Given } (h, \tau), \text{ decide whether } \tau \in \mathcal{A}(h).$$

In the general case, exact admissibility checking may be computationally expensive or undecidable, particularly when construction histories are large.

However, the theory does not require exact global checking.

Local admissibility approximations that conservatively underestimate $\mathcal{A}(h)$ preserve refusal and stability, at the cost of reduced but coherent capability.

False refusals reduce generality but do not induce collapse. False admissions, by contrast, violate invariants and are structurally catastrophic. The asymmetry aligns naturally with conservative approximation.

21.5 Discretization of Continuous Dynamics

The variational and field-theoretic formulations assume continuous time and space. In computational realizations, dynamics are discretized.

Standard results from numerical analysis apply.

If discretization schemes respect the coercivity of invariant potentials and boundary penalties, then discrete trajectories remain within an ε-neighborhood of admissible continuous trajectories.

Boundary collapse under discretization occurs only when the discretization fails to respect constraint geometry, not because constraints themselves are unrealizable.

21.6 Finite Memory and Bounded History

The theory emphasizes construction history, which may appear to require unbounded memory. In practice, systems have finite storage.

This does not undermine the framework.

Construction history need not be stored verbatim. It need only be represented up to equivalence with respect to admissibility.

If two histories induce identical admissible transformation sets, then they are equivalent for the purposes of refusal and lawful transfer.

Compression of history into sufficient statistics is therefore compatible with the theory, provided the compression preserves admissibility structure.

21.7 What Cannot Be Approximated Away

21.8 Non-Negotiable Features

The preceding results demonstrate that many idealizations are benign. However, certain features are non-negotiable: removal of boundary penalties cannot be compensated by approximation; elimination of construction history cannot be recovered by post hoc filters; and replacement of refusal with weighted preference destroys invariance. These failures are qualitative rather than quantitative. No amount of scaling or tuning can repair them.

21.9 Interpretation

Computational realizability does not require abandoning rigor. It requires understanding which aspects of the theory are structural and which are representational conveniences.

Infinite barriers may be approximated. Exact symmetries may be softened. Continuous fields may be discretized.

But boundaries, refusal, and history-sensitive admissibility cannot be removed without collapsing the regime the theory describes.

The theory therefore characterizes not a fragile ideal, but a robust class of realizable systems whose defining properties survive approximation while resisting illegitimate simplification.

22 Failure Modes and Diagnostic Signatures

22.1 Purpose of Failure Analysis

A theory that characterizes intelligence as a constrained dynamical regime must also characterize how that regime fails. Failure, in this context, does not mean poor performance on tasks. It means loss of coherence: the breakdown of invariant preservation, boundary regulation, or lawful transfer.

This section classifies failure modes implied by the framework and derives diagnostic signatures that distinguish structural collapse from benign limitation.

22.2 Boundary Collapse

Definition 22.1. Boundary collapse occurs when the penalty enforcing semi-permeability becomes insufficient to regulate interaction between internal and external degrees of freedom.

In the variational formulation, this corresponds to the weakening or removal of the boundary term Λ. When Λ loses dominance, information, influence, or policy flow propagates without constraint.

Signature. Boundary collapse manifests as rapid growth of coupling terms, loss of locality in field dynamics, and sensitivity of internal structure to arbitrarily small perturbations.

Interpretation. Boundary collapse manifests empirically as uncontrolled synchronization, runaway feedback, or overfitting to transient context. It is not a failure of learning, but a failure of insulation.

22.3 Invariant Drift

Definition 22.2. Invariant drift occurs when quantities intended to be conserved under lawful transformation vary systematically over time.

In the Hamiltonian picture, this corresponds to violation of the conditions required for Noether-type conservation.

Signature. Invariant drift is characterized by gradual degradation of previously stable abstractions, increasing inconsistency in admissibility judgments, and dependence of reuse validity on irrelevant contextual variation.

Interpretation. Invariant drift indicates that compression or learning is occurring faster than constraint enforcement. The system remains active but loses the ability to distinguish lawful from unlawful transfer.

22.4 Illegitimate Transfer

Definition 22.3. Illegitimate transfer occurs when a construction is reused outside the scope in which its invariants are preserved.

This failure mode is structurally prohibited in the ideal theory but appears under approximation or boundary weakening.

Signature. Illegitimate transfer presents as apparent generalization followed by abrupt breakdown, success in narrow contexts accompanied by catastrophic errors elsewhere, and retrospective inability to explain failure using preserved abstractions.

Interpretation. Illegitimate transfer is not overgeneralization in the usual sense. It is a violation of admissibility structure caused by missing refusal.

22.5 Refusal Suppression

Definition 22.4. Refusal suppression occurs when refusal is replaced by graded preference, penalty weighting, or optimization pressure.

This corresponds to flattening infinite or dominant barriers into finite costs.

Signature. Refusal suppression manifests in rare but catastrophic boundary crossings, increasing reliance on external correction or filtering, and the emergence of brittle safety mechanisms.

Interpretation. Refusal suppression transforms structural non-existence into probabilistic discouragement, thereby eliminating the asymmetry that protects coherence.

22.6 Timescale Entanglement

Definition 22.5. Timescale entanglement occurs when fast policy dynamics directly perturb slow structural variables without averaging.

Signature. Timescale entanglement is characterized by oscillatory or chaotic structural evolution, dependence of long-term behavior on short-term fluctuations, and loss of stable attractors.

Interpretation. This failure mode corresponds to the breakdown of the CLIO regime, wherein the system remains responsive but loses memory and structure.

22.7 History Erasure

Definition 22.6. History erasure occurs when construction history is discarded or replaced by state-only summaries that are not admissibility-equivalent.

Signature. History erasure leads to an inability to justify refusals or constraints, inconsistent behavior across identical states, and reliance on ad hoc patches to restore safety.

Interpretation. History erasure directly violates the State-Projection No-Go Theorem. It produces systems that appear competent until lawful transfer is required.

22.8 Diagnostic Hierarchy

The failure modes described above admit a partial ordering by severity:

$$\text{Invariant Drift} \prec \text{Illegitimate Transfer} \prec \text{Boundary Collapse}.$$

Early-stage failures may be corrected by strengthening constraints. Late-stage failures indicate loss of regime and require architectural revision.

22.9 Self-Diagnosis and Legibility

A crucial consequence of the framework is that failure should be detectable internally. Diagnostic quantities correspond to deviations in conserved currents, boundary flux, and invariant measures.

Systems that cannot diagnose their own boundary collapse are already outside the regime of general intelligence described here.

22.10 Summary

Failure, in this framework, is not mysterious. Each mode corresponds to violation of a specific structural requirement. The diagnostic signatures follow directly from the dynamics.

This completes the characterization of general intelligence as a constrained, stable, and diagnosable dynamical regime. Systems that persist do so because they cannot fail silently.

23 Comparison to Existing Formalisms

23.1 Relation to Variational Inference

Variational inference frames intelligence as minimization of a divergence between internal models and external data. While formally similar in its use of variational principles, the present framework differs in its treatment of constraints.

In variational inference, constraints typically appear as regularizers. Here, constraints define the admissible manifold itself. Violations are not penalized softly; they are dynamically inaccessible.

As a result, refusal is structural rather than statistical, and generalization is lawful rather than approximate.

23.2 Relation to the Free Energy Principle

The Free Energy Principle (FEP) characterizes adaptive systems as minimizing variational free energy under a generative model. Markov blankets play a central role in mediating internal and external states.

The present framework agrees with the necessity of boundaries but diverges in emphasis. Boundaries are not introduced to support inference, but to preserve invariants under reuse. Entropy descent occurs only where it preserves abstraction coherence, not universally.

Where FEP emphasizes prediction, this framework emphasizes survivability of structure.

23.3 Relation to Control-Theoretic AGI

Control-theoretic approaches typically assume a predefined objective or reward function. Intelligence is identified with optimal control under uncertainty.

In contrast, the present framework does not assume a global objective. It replaces optimality with admissibility. Actions are selected not because they maximize reward, but because they preserve coherence.

Refusal emerges naturally where no admissible control exists.

23.4 Relation to Foundation Model Paradigms

Large foundation models emphasize scale, transfer, and emergent capability through parameterization. Their success depends on weak constraints and broad generalization.

The present framework explains both their power and their fragility. In the absence of strong boundary enforcement, generalization proceeds by collapse of invariants rather than by lawful transfer. Stability must be imposed externally rather than emerging from dynamics.

This comparison is descriptive rather than critical. It identifies structural differences rather than proposing replacements.

23.5 Summary of Distinctions

Across these comparisons, the distinguishing features of the present framework are invariants as first-class objects, refusal as phase-space exclusion, boundaries as dynamical fields, generality as symmetry-preserving transfer, and stability through constraint rather than optimization. These features are not additional mechanisms; rather, they are consequences of treating intelligence as a constrained dynamical system rather than as an objective-maximizing machine.

23.6 Final Thoughts

The argument has now come full circle. Beginning from the primacy of events and construction history, it has developed a view of intelligence as boundary-regulated adaptation. The same principles that govern cells, brains, and operating systems govern ideas and their dissemination.

In the concluding section, these threads are drawn together to restate the central claim: intelligence that cannot regulate its own boundaries, internally or externally, does not remain intelligent for long.

24 Intelligence That Persists Regulates Itself

The argument of this essay has proceeded by constraint rather than by accumulation. Beginning from the primacy of events over states, it has treated intelligence as an irreversible process of construction whose coherence depends on the preservation of invariants under reuse. Abstraction has been redefined as invariant-preserving compression; modularity as an ontological condition rather than an engineering convenience; refusal as a constitutive operation rather than a failure mode; and generality as lawful transfer rather than breadth of application.

From these commitments, the necessity of boundaries has followed without appeal to external ethics or precautionary principles. Semi-permeable boundaries appear wherever adaptive systems persist: in biological membranes, neural synchronization, evolutionary buffering, computational privilege separation, and epistemic practice. These boundaries regulate interaction, preserve gradients, and prevent collapse under exposure. They are not obstacles to intelligence. They are its enabling conditions.

Legibility, viewed through this lens, is no longer an unqualified good. It is a form of exposure that alters the environment in which intelligence operates. Unregulated legibility selects for portability over robustness, extraction over coherence, and immediacy over survivability. Intelligence that cannot regulate its own visibility becomes a target of its own success.

This does not license secrecy, obscurantism, or authority-based restriction. On the contrary, the framework developed here distinguishes sharply between concealment and constraint. Inspec-

tion without modification, understanding without authorization, and critique without deployment are not limitations on openness; they are its disciplined forms. Effort-gated legibility preserves falsifiability while preventing illegitimate reuse.

The architectural characterization of general intelligence that emerges is therefore conservative by design. It privileges coherence over coverage, refusal over indiscriminate action, and cumulative understanding over extractable technique. Such conservatism is often mistaken for timidity. In fact, it is what allows intelligence to scale without dissolving itself.

A final analogy may serve to close the loop. An operating system that grants all processes root access does not become more powerful; it becomes unstable. A nervous system that synchronizes all regions at once does not become more intelligent; it seizes. A genome without buffering cannot evolve. An epistemic system without boundaries cannot accumulate knowledge.

The same is true of intelligence itself.

Systems that cannot refuse do not remain general. Systems that cannot regulate their own legibility do not remain coherent. Systems that cannot preserve their invariants under reuse do not remain intelligent.

What persists is not maximal capability, but disciplined construction under constraint.

25 Conclusion

This work has developed a structural theory of general intelligence grounded in constrained dynamics rather than task performance or objective maximization. Intelligence has been treated as an irreversible process of construction whose persistence depends on the preservation of invariants under reuse, transfer, and exposure.

Beginning from the primacy of events over states, the analysis established that adaptive systems require semi-permeable boundaries to regulate interaction. These boundaries arise not as safeguards but as necessary conditions for maintaining coherence. Refusal was shown to be a structural property of such systems, corresponding to the non-existence of admissible transformations rather than to error or intervention.

A variational formulation was introduced to formalize these principles. Within this framework, abstraction appears as invariant-preserving compression, learning as constrained entropy descent, and generality as lawful transfer along symmetry-preserving directions. Boundary structure was incorporated directly into the action, making illegitimate configurations dynamically inaccessible.

Extending the model to a field-theoretic setting demonstrated how intelligence can be distributed across space or semantic domains while preserving locality and coherence. Conserved quantities associated with invariants were derived via symmetry considerations, establishing a Noether-type correspondence between lawful reuse and structural conservation.

The resulting dynamics admit stable invariant-preserving attractors under mild conditions. Stability is achieved not through global optimization or centralized control, but through the geometry of the admissible manifold and the energetic cost of boundary violation. Fast policy dynamics and slow structural evolution were shown to coexist via a natural separation of timescales, yielding adaptive behavior without collapse.

Comparison to existing formalisms clarified the distinctiveness of this approach. Unlike optimization-centric or inference-based models, the present framework treats refusal, boundary regulation, and invariant preservation as foundational rather than auxiliary. Capability is constrained by coherence, not expanded at its expense.

Taken together, these results support a conception of general intelligence as a physically realizable regime of constrained dynamics. Such systems remain adaptive because they cannot act, infer, or disclose beyond what their structure permits. Generality emerges not from breadth alone, but from the disciplined preservation of invariants across change.

This view suggests that intelligence, in any substrate, persists only where boundaries are maintained, refusal is enforced by geometry, and learning proceeds through lawful compression rather than unconstrained optimization.

A Configuration Manifolds, Constraints, and Refusal

Definition A.1 (Configuration manifold)**.** Let $\mathcal{Q}$ be a smooth finite-dimensional manifold (or Fréchet manifold in the field case) with local coordinates

$$q = (\Phi, \mathbf{v}, S, \pi, \mathcal{B}).$$

Definition A.2 (Invariant constraints and admissible manifold)**.** Let $F : \mathcal{Q} \to \mathbb{R}^k$ be smooth and define the admissible manifold

$$\mathcal{M} := F^{-1}(0).$$

Assume 0 is a regular value of F, so $\mathcal{M}$ is an embedded submanifold with tangent space

$$T_q\mathcal{M} = \ker DF(q).$$

Definition A.3 (Admissible variations)**.** A variation $\delta q(t)$ along a curve $q(t) \in \mathcal{M}$ is admissible if

$$DF(q(t))\,\delta q(t) = 0 \quad \text{for all } t.$$

Definition A.4 (Refusal set)**.** Let $\mathcal{H} : T^*\mathcal{Q} \to \mathbb{R} \cup \{+\infty\}$. Define the refusal set

$$\mathcal{R} := \{(q,p) \in T^*\mathcal{Q} : \ \mathcal{H}(q,p) = +\infty\}.$$

A state is refused iff it lies in $\mathcal{R}$.

Lemma A.1 (Normal coercivity implies manifold confinement)**.** *Assume there exists a neighborhood U of $\mathcal{M}$ and constants $c_1, c_2 > 0$ such that*

$$V(q) \ \geq \ c_1 \operatorname{dist}(q, \mathcal{M})^2 - c_2 \quad \textit{for all } q \in U,$$

and that $\mathcal{H}$ satisfies $\mathcal{H}(q,p) \geq V(q)$ on T^*U. Then for any energy sublevel set

$$\Sigma_E := \{(q,p) \in T^*U : \ \mathcal{H}(q,p) \leq E\},$$

there exists $r(E) > 0$ such that $\Sigma_E \subset \{(q,p) : \mathrm{dist}(q, \mathcal{M}) \le r(E)\}$.

Proof. For $(q,p) \in \Sigma_E$, one has $E \ge \mathcal{H}(q,p) \ge V(q) \ge c_1 \mathrm{dist}(q,\mathcal{M})^2 - c_2$, hence $\mathrm{dist}(q,\mathcal{M})^2 \le (E + c_2)/c_1$. Take $r(E) = \sqrt{(E+c_2)/c_1}$. □

B Field-Theoretic Euler–Lagrange Equations

Definition B.1 (Field configuration space)**.** Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with smooth boundary. Let the field variables be

$$\Phi : \Omega \times \mathbb{R} \to \mathbb{R}, \quad \mathbf{v} : \Omega \times \mathbb{R} \to \mathbb{R}^n, \quad S : \Omega \times \mathbb{R} \to \mathbb{R}, \quad \pi : \Omega \times \mathbb{R} \to \Delta, \quad \mathcal{B} : \Omega \times \mathbb{R} \to \mathbb{R}^m,$$

where Δ is a (finite-dimensional) simplex or a suitable function space.

Definition B.2 (Action functional)**.** Let $\mathcal{L}$ be a Lagrangian density depending smoothly on $(q, \partial_t q, \nabla q)$. Define

$$\mathcal{S}[q] = \int_{t_0}^{t_1} \int_{\Omega} \mathcal{L}(q, \partial_t q, \nabla q)\, dx\, dt.$$

Theorem B.1 (Euler–Lagrange field equations)**.** *Assume variations δq vanish on $\partial\Omega \times [t_0, t_1]$ and at times t_0, t_1. Then a stationary point of $\mathcal{S}$ satisfies, componentwise for each field q^α,*

$$\partial_t \left(\frac{\partial \mathcal{L}}{\partial(\partial_t q^\alpha)} \right) + \sum_{i=1}^{n} \partial_{x_i} \left(\frac{\partial \mathcal{L}}{\partial(\partial_{x_i} q^\alpha)} \right) - \frac{\partial \mathcal{L}}{\partial q^\alpha} \;=\; 0.$$

Proof. Compute the first variation:

$$\delta \mathcal{S} = \iint \left(\frac{\partial \mathcal{L}}{\partial q^\alpha} \delta q^\alpha + \frac{\partial \mathcal{L}}{\partial(\partial_t q^\alpha)} \partial_t(\delta q^\alpha) + \frac{\partial \mathcal{L}}{\partial(\partial_{x_i} q^\alpha)} \partial_{x_i}(\delta q^\alpha) \right) dx\, dt,$$

(sum over α and i). Integrate by parts in t and x_i and use boundary/endpoint vanishing to remove boundary terms:

$$\delta \mathcal{S} = \iint \left(\frac{\partial \mathcal{L}}{\partial q^\alpha} - \partial_t \left(\frac{\partial \mathcal{L}}{\partial(\partial_t q^\alpha)} \right) - \partial_{x_i} \left(\frac{\partial \mathcal{L}}{\partial(\partial_{x_i} q^\alpha)} \right) \right) \delta q^\alpha\, dx\, dt.$$

Since δq^α are arbitrary, the coefficient must vanish. □

C Hamiltonian Structure and Refusal Barriers

Definition C.1 (Legendre transform)**.** Let $\mathcal{L}(q, \dot{q})$ be C^2 and strictly convex in $\dot{q}$. Define momenta $p_i = \partial \mathcal{L} / \partial \dot{q}_i$ and Hamiltonian

$$\mathcal{H}(q,p) = \sup_{\dot{q}} \left(\sum_i p_i \dot{q}_i - \mathcal{L}(q, \dot{q}) \right).$$

Theorem C.1 (Hamilton's equations)**.** *If $\mathcal{H}$ is finite and C^2 on an open set $U \subset T^*\mathcal{Q}$, then the induced flow satisfies*

$$\dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i}.$$

Proof. On U, strict convexity gives the smooth inverse $\dot{q} = \dot{q}(q,p)$. Then $\mathcal{H}(q,p) = p \cdot \dot{q} - \mathcal{L}(q,\dot{q})$. Differentiate:

$$d\mathcal{H} = \dot{q} \cdot dp + p \cdot d\dot{q} - \frac{\partial \mathcal{L}}{\partial q} \cdot dq - \frac{\partial \mathcal{L}}{\partial \dot{q}} \cdot d\dot{q} = \dot{q} \cdot dp - \frac{\partial \mathcal{L}}{\partial q} \cdot dq,$$

since $p = \partial \mathcal{L}/\partial \dot{q}$ cancels the $d\dot{q}$ terms. Thus $\partial \mathcal{H}/\partial p = \dot{q}$ and $\partial \mathcal{H}/\partial q = -\partial \mathcal{L}/\partial q$. With Euler–Lagrange $d/dt(\partial \mathcal{L}/\partial \dot{q}) = \partial \mathcal{L}/\partial q$ we obtain $\dot{p} = -\partial \mathcal{H}/\partial q$. □

Lemma C.1 (Infinite barrier is forward-invariant)**.** *Let $\mathcal{H}$ be lower semicontinuous and define $\mathcal{R} = \{\mathcal{H} = +\infty\}$. If a trajectory $(q(t), p(t))$ satisfies $\mathcal{H}(q(t), p(t)) < \infty$ for some $t = t_*$ and energy is conserved on $\{\mathcal{H} < \infty\}$, then $(q(t), p(t)) \notin \mathcal{R}$ for all t in its maximal interval of existence.*

Proof. Energy conservation gives $\mathcal{H}(q(t), p(t)) = \mathcal{H}(q(t_*), p(t_*)) < \infty$ whenever the solution remains in $\{\mathcal{H} < \infty\}$. If it entered $\mathcal{R}$ at some time, $\mathcal{H}$ would be $+\infty$, contradicting conservation. □

D Noether Currents for Field Symmetries

Definition D.1 (Infinitesimal symmetry)**.** Let q^α be fields and $\mathcal{L}(q, \partial_\mu q)$ with $\mu = 0, \dots, n$ ($x^0 = t$). An infinitesimal symmetry is a variation

$$\delta q^\alpha = \varepsilon \, X^\alpha(q, x)$$

such that the induced Lagrangian variation is a total divergence:

$$\delta \mathcal{L} = \varepsilon \, \partial_\mu K^\mu$$

for some K^μ.

Theorem D.1 (Noether current)**.** *Assume q satisfies the Euler–Lagrange equations. Then the current*

$$J^\mu \;=\; \frac{\partial \mathcal{L}}{\partial(\partial_\mu q^\alpha)} \, X^\alpha \;-\; K^\mu$$

is conserved:

$$\partial_\mu J^\mu = 0.$$

Proof. Compute

$$\delta \mathcal{L} = \frac{\partial \mathcal{L}}{\partial q^\alpha} \delta q^\alpha + \frac{\partial \mathcal{L}}{\partial(\partial_\mu q^\alpha)} \partial_\mu(\delta q^\alpha).$$

Rewrite the second term:

$$\frac{\partial \mathcal{L}}{\partial(\partial_\mu q^\alpha)} \partial_\mu(\delta q^\alpha) = \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu q^\alpha)} \delta q^\alpha \right) - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu q^\alpha)} \right) \delta q^\alpha.$$

Hence

$$\delta\mathcal{L} = \left(\frac{\partial \mathcal{L}}{\partial q^\alpha} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu q^\alpha)} \right) \right) \delta q^\alpha + \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu q^\alpha)} \delta q^\alpha \right).$$

On-shell the Euler–Lagrange bracket vanishes, so

$$\delta\mathcal{L} = \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu q^\alpha)} \delta q^\alpha \right).$$

By symmetry, $\delta\mathcal{L} = \partial_\mu K^\mu$, therefore

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu q^\alpha)} \delta q^\alpha - K^\mu \right) = 0.$$

Divide by ε and substitute $\delta q^\alpha = \varepsilon X^\alpha$. □

E Slow–Fast Decomposition and CLIO Regimes

Definition E.1 (Slow–fast Hamiltonian system)**.** Let $(x, y) \in \mathbb{R}^m \times \mathbb{R}^\ell$ with $0 < \varepsilon \ll 1$ and Hamiltonian

$$\mathcal{H}_\varepsilon(x, y, p_x, p_y) = \mathcal{H}_0(x, p_x) \; + \; \mathcal{H}_{\text{fast}}(x, y, p_y) \; + \; \varepsilon\, \mathcal{H}_1(x, y, p_x, p_y),$$

with canonical symplectic form $\omega = dx \wedge dp_x + dy \wedge dp_y$.

Definition E.2 (Critical manifold)**.** Assume for each fixed (x, p_x) the fast subsystem has an equilibrium set

$$\mathcal{C}_0 := \{(x, y, p_x, p_y) : \; \partial_y \mathcal{H}_{\text{fast}} = 0, \; \partial_{p_y} \mathcal{H}_{\text{fast}} = 0\}.$$

Theorem E.1 (Persistence of normally hyperbolic slow manifolds (Fenichel-type))**.** *Assume $\mathcal{C}_0$ is a compact normally hyperbolic invariant manifold for the $\varepsilon = 0$ fast flow and that $\mathcal{H}_\varepsilon$ is C^r, $r \geq 2$. Then for sufficiently small $\varepsilon > 0$ there exists a locally invariant manifold $\mathcal{C}_\varepsilon$ C^{r-1}-close to $\mathcal{C}_0$ and the reduced flow on $\mathcal{C}_\varepsilon$ is C^{r-1} conjugate to the slow drift induced by $\mathcal{H}_0 + \varepsilon\mathcal{H}_1$.*

Proof. Standard Fenichel theory for normally hyperbolic invariant manifolds applies to the fast-slow vector field generated by $\mathcal{H}_\varepsilon$. Normal hyperbolicity provides exponential contraction/expansion transverse to $\mathcal{C}_0$, yielding persistence and smoothness of $\mathcal{C}_\varepsilon$ for small ε and conjugacy of reduced dynamics. □

Lemma E.1 (Geometric refusal via infinite curvature)**.** *Let $\mathcal{H}$ be C^2 on an open set $U \subset T^*\mathcal{Q}$ and extend $\mathcal{H}$ by $+\infty$ outside U. If a direction ξ at $(q, p) \in U$ approaches ∂U such that $\mathcal{H}(q + \tau\xi_q, p + \tau\xi_p) \to +\infty$ as $\tau \uparrow \tau_*$, then no finite-energy trajectory can cross ∂U along ξ.*

Proof. Along any finite-energy trajectory, $\mathcal{H}$ is conserved and finite. Approaching ∂U along ξ forces $\mathcal{H} \to +\infty$, contradicting conservation. Hence crossing is impossible. □

F Lyapunov Stability for Invariant-Preserving Attractors

Definition F.1 (Lyapunov function)**.** A continuous function $W : U \to \mathbb{R}_{\geq 0}$ on a neighborhood U of a compact set $\mathcal{A}$ is a Lyapunov function if:

$$W^{-1}(0) = \mathcal{A}, \qquad \dot{W} \leq 0 \text{ along trajectories in } U.$$

Theorem F.1 (Lyapunov stability from coercive invariant potentials)**.** *Let $\mathcal{H} = T + V + \Lambda - \Xi$ on $T^*\mathcal{Q}$, with $T \geq 0$, Ξ bounded above, and suppose there exists a compact $\mathcal{A} \subset \mathcal{M}$ and constants $c, C > 0$ such that in a neighborhood U of $\mathcal{A}$:*

$$V(q) + \Lambda(q) \;\geq\; c\,\mathrm{dist}(q, \mathcal{M})^2, \qquad V(q) + \Lambda(q) \;\to +\infty \text{ as } q \to \partial U.$$

Then $\mathcal{A}$ is Lyapunov-stable for the Hamiltonian flow restricted to the energy sublevel set $\{\mathcal{H} \leq E\}$ for any E with $\mathcal{A} \subset \{\mathcal{H} \leq E\} \subset U$.

Proof. Fix such E and define $W(q,p) = V(q) + \Lambda(q)$ on $\{\mathcal{H} \leq E\}$. Energy conservation implies trajectories remain in $\{\mathcal{H} \leq E\} \subset U$. By the lower bound, W controls $\mathrm{dist}(q, \mathcal{M})^2$. Since $\mathcal{A} \subset \mathcal{M}$, $W = 0$ on $\mathcal{A}$ and $W > 0$ off $\mathcal{M}$. Let $\epsilon > 0$ be given. Choose $\delta > 0$ such that $W < \delta \Rightarrow \mathrm{dist}(q, \mathcal{M}) < \epsilon$. If initial data satisfy $W(q(0), p(0)) < \delta$, then by energy confinement and the barrier condition at ∂U, the trajectory cannot exit the set $\{W < \delta'\}$ for a sufficiently small $\delta' \geq \delta$ contained in U, hence $\mathrm{dist}(q(t), \mathcal{M}) < \epsilon$ for all t. □

A Worked Examples of Constrained Intelligence Dynamics

A.1 Purpose and Scope

This appendix presents minimal mathematical examples illustrating the structural phenomena developed in the main text. The examples are not intended as implementations or blueprints. Their role is to demonstrate, in low-dimensional and analytically tractable settings, how invariants, refusal, boundaries, and stability arise naturally from constrained dynamics.

All examples are deliberately simple. They are sufficient because the properties under study—refusal as non-existence, boundary-regulated interaction, slow–fast separation, and invariant-preserving stability—are structural rather than dimensional.

A.2 One-Dimensional Refusal Barrier

Consider a one-dimensional configuration variable $q \in \mathbb{R}$ evolving under a Hamiltonian

$$\mathcal{H}(q,p) = \frac{1}{2}p^2 + V(q),$$

where

$$V(q) = \frac{1}{1-q^2}.$$

The admissible manifold is

$$\mathcal{M} = (-1, 1).$$

As $q \to \pm 1$, the potential diverges, and the Hamiltonian becomes unbounded. Consequently, no finite-energy trajectory reaches or crosses the boundary of $\mathcal{M}$.

Refusal is realized here as geometric exclusion. Configurations outside $\mathcal{M}$ are not evaluated or rejected; they are dynamically inaccessible. The boundary enforces admissibility without invoking decision logic.

This example illustrates the minimal mechanism by which refusal can be encoded as phase-space geometry.

A.3 Two-Dimensional Boundary Collapse

Let $(x, y) \in \mathbb{R}^2$, where x denotes a slow structural variable and y a fast policy variable. Consider the Hamiltonian

$$\mathcal{H}_\lambda(x, y, p_x, p_y) = \frac{1}{2}p_x^2 + \frac{1}{2}p_y^2 + U(x) + \lambda W(x, y),$$

where $U(x)$ is coercive and $W(x, y)$ penalizes boundary violation.

For $\lambda > 0$, the coupling term constrains y relative to x, inducing an invariant-preserving attractor. As $\lambda \to 0$, the constraint weakens, and trajectories exhibit unbounded excursions in y, destabilizing x through backreaction.

This bifurcation corresponds to boundary collapse. It demonstrates that stability depends quantitatively on boundary strength, not merely on the existence of a boundary term.

A.4 Slow–Fast Averaging and CLIO Dynamics

Let the system evolve according to

$$\dot{x} = f(x, y), \qquad \dot{y} = \frac{1}{\epsilon}g(x, y),$$

with $0 < \epsilon \ll 1$.

Assume that for fixed x, the fast subsystem admits a unique invariant measure μ_x and is mixing. The averaged dynamics are

$$\dot{x} = \int f(x, y)\, d\mu_x(y).$$

Fast policy exploration occurs in y without directly altering x. Structural change in x reflects only invariant-preserving averages of fast dynamics. This realizes CLIO as a dynamical regime rather than an algorithmic loop.

When mixing fails or timescale separation collapses, averaging breaks down and fast dynamics directly perturb structure, producing unstable generalization.

A.5 Information Flux Across a Boundary

Model boundary permeability by a scalar parameter $\kappa \geq 0$ governing coupling strength between internal and external variables. Let information flux $I(\kappa)$ be a monotone increasing function of κ.

For small κ, flux is insufficient to support adaptation. For large κ, internal gradients collapse. Optimal operation occurs at intermediate κ, where information transfer is sufficient but constrained.

This reproduces, in abstract form, the same regime observed in biological membranes, neural synchronization, and policy buffering. Boundaries are neither maximally open nor maximally closed.

A.6 Redundancy and Latent Slack

Introduce an auxiliary variable s representing latent slack or redundancy, with dynamics

$$\dot{s} = -\alpha s + \eta,$$

where η is bounded noise.

The presence of s buffers perturbations to primary variables, absorbing fluctuations without immediate structural impact. Removing s increases sensitivity and induces instability.

This mirrors the role of noncoding genetic material and latent policy variation. Apparent inefficiency is a condition for robust selection and learning.

A.7 Summary of Structural Lessons

Across these examples, the same conclusions recur: refusal is realized as inaccessible regions of phase space; boundaries regulate transfer without concealment; stability depends on constraint strength and geometry; slow–fast separation enables learning without collapse; and redundancy functions as adaptive slack rather than waste. These properties do not depend on dimensionality, substrate, or implementation. They arise wherever intelligence persists under constraint.

References

Arnold, V. I. (1989). *Mathematical Methods of Classical Mechanics*. 2nd ed. Springer.

Ashby, W. Ross (1956). *An Introduction to Cybernetics*. Chapman & Hall.

Cover, Thomas M. and Joy A. Thomas (2006). *Elements of Information Theory*. 2nd ed. Hoboken, NJ: Wiley-Interscience. ISBN: 9780471241959.

Friston, Karl (2010). "The Free-Energy Principle: A Unified Brain Theory?" In: *Nature Reviews Neuroscience* 11, pp. 127–138.

— (2015). "Life as We Know It". In: *Journal of the Royal Society Interface*.

Jaynes, Edwin T. (2003). *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press. ISBN: 9780521592710.

Olver, Peter J. (1993). *Applications of Lie Groups to Differential Equations*. Springer.

Saltzer, Jerome H. and Michael D. Schroeder (1975). "The Protection of Information in Computer Systems". In: *Proceedings of the IEEE* 63.9, pp. 1278–1308. DOI: 10.1109/PROC.1975.9939.

Weaver, Warren (1948). "Science and Complexity". In: *American Scientist* 36, pp. 536–544.