The Constraint–Autonomy Compatibility Law: A Formal Derivation

Boris Kriger
Institute of Integrative and Interdisciplinary Research
`boriskriger@interdisciplinary-institute.org`

January 25, 2026

Abstract

This paper presents a formal derivation of the Constraint–Autonomy Compatibility Law, which states that adaptive rule-based systems require both external constraints and internal agent autonomy to maintain long-term functional stability. We provide a mathematical framework combining decision theory, control theory, and Lyapunov stability analysis. We prove the necessity of constraint-autonomy balance, establish connections to complexity theory and biological evolution, and discuss implications for governance, artificial intelligence, and organizational design.

The law is shown to be the central integrating principle of a broader theoretical framework, unifying five independently developed theories: the Unified Theory of Self-Organizing Systems (cooperation as emergent default), the Principle of Optimal Coherence (load-bearing vs. discretionary constraints), the Asymmetry of Totalizing Ideals (boundary collapse under maximization), and Epistemic Constraint Theory (hypothesis space bounds on inference). We demonstrate that these frameworks are mathematical reformulations of the same deep structure, with the Constraint–Autonomy Law serving as the common core. A master equation is derived capturing the unified dynamics, and a translation dictionary enables researchers to work across frameworks.

The central philosophical contribution is demonstrating that freedom and order are not competing values requiring trade-offs, but mutually necessary structural conditions—each requires the other to exist in adaptive systems.

Keywords: complex adaptive systems, autonomy, constraints, Lyapunov stability, autopoiesis, multi-agent systems, governance theory, cooperation, self-organization, epistemic constraints, inference limitations

1 Introduction

The tension between rules and freedom has occupied philosophers, legal theorists, and political scientists for centuries. Yet this tension is rarely formalized in terms that permit rigorous analysis. We propose that what appears as a normative debate is, at its core, a structural property of adaptive systems.

Consider the aphorism: *"Without freedom, any justice becomes a cage."* This statement encodes an intuition about system dynamics—that excessive constraint produces rigidity incompatible with flourishing. We demonstrate that this intuition admits formal proof.

Our central contribution is the **Constraint–Autonomy Compatibility Law**:

> *In any adaptive rule-based system, sustained functional balance exists if and only if external constraints limit actions without eliminating internal autonomy.*

The remainder of this paper is organized as follows. Section 2 situates our law within complexity theory and evolutionary biology. Section 3 establishes formal definitions. Section 4 develops a Lyapunov stability framework. Section 5 presents and proves the main theorem. Sections 6–7 derive corollaries and practical applications. Section 8 relates our work to existing theory. Section 9 characterizes validity boundaries. Section 10 discusses philosophical implications, and Section 13 concludes.

2 Connections to Complexity Theory and Evolution

Before presenting our formal framework, we situate the Constraint–Autonomy Compatibility Law within the broader landscape of complexity science, evolutionary biology, and cybernetics.

2.1 Evolutionary Dynamics

Biological systems provide the most striking natural instantiation of our law. Consider the dual pressures on genetic information:

Over-constraint (Genetic Rigidity):

- Organisms with highly conserved, inflexible genomes cannot adapt to environmental shifts
- Mass extinctions consistently eliminate specialists while generalists survive
- Asexual reproduction, while efficient, produces lineages vulnerable to "Red Queen" dynamics

Under-constraint (Mutational Chaos):

- Excessive mutation rates destroy functional coherence
- Error catastrophe (Eigen's threshold) causes loss of genetic information [Eigen, 1971]
- Horizontal gene transfer without selection produces non-viable chimeras

The evolutionary stable strategy occupies a narrow band: sufficient mutation to explore fitness landscapes, sufficient conservation to maintain functional organization. This is precisely the constraint-autonomy balance we formalize.

2.2 Autopoiesis and Organizational Closure

Maturana and Varela's theory of autopoiesis [Maturana & Varela, 1980] describes living systems as organizationally closed yet structurally open. The key insight: a system must maintain its own organization (constraint) while continuously regenerating its components through environmental interaction (autonomy).

In autopoietic terms:

- **Constraints** define the system's organizational identity—what makes it *this* system rather than another
- **Autonomy** enables structural coupling with the environment—the capacity to respond without losing identity

A cell that cannot modify its membrane proteins in response to environmental signals dies. A cell whose membrane becomes arbitrarily permeable dissolves. The constraint-autonomy balance is the condition for autopoietic continuation.

2.3 The Good Regulator Theorem

Conant and Ashby's Good Regulator Theorem [Conant & Ashby, 1970] states: *"Every good regulator of a system must be a model of that system."*

This theorem implies that effective control requires internal representation of external dynamics. But modeling requires degrees of freedom—the capacity to form different internal states corresponding to different external states. If a regulator has no internal variety (zero autonomy), it cannot model anything; it simply executes a fixed program.

Thus, autonomy is not noise or deviation from optimal control. Autonomy is the *substrate* upon which regulatory models are built. Our law makes this explicit: internal freedom is a necessary condition for adaptive regulation, not an obstacle to it.

2.4 Edge of Chaos Dynamics

Complex systems theory identifies the "edge of chaos" as the region between rigid order and chaotic disorder where complex computation and adaptation occur [Langton, 1990, Kauffman, 1993]. Systems at the edge exhibit:

- Long transients (memory)
- Sensitivity to initial conditions (responsiveness)
- Non-trivial attractor structure (multiple stable modes)

The constraint-autonomy balance can be understood as the *control parameter* that positions a system at or near this edge:

Regime	**Constraint**	**Autonomy**	**Dynamics**
Frozen	High	Low	Fixed point attractors, no adaptation
Edge	Balanced	Balanced	Complex transients, adaptive capacity
Chaotic	Low	High	Strange attractors, no coherence

Table 1: System regimes as functions of constraint-autonomy balance.

Our contribution is to prove that only the balanced regime sustains both adaptability and coherence—the "edge" is not merely interesting, it is *necessary*.

3 Preliminaries and Definitions

3.1 System Components

Let $\mathcal{S} = (A, R, \Omega, T)$ be a rule-constrained multi-agent decision system where:

- $A = \{a_1, a_2, \dots, a_n\}$ is a finite set of agents
- $R : \Omega \to 2^{\mathcal{A}}$ is a rule function mapping states to permissible action sets
- Ω is the state space of the system
- $T : \Omega \times \mathcal{A}^n \to \Omega$ is the transition function

3.2 Key Definitions

Definition 3.1 (Autonomy)**.** The autonomy of agent a_i in state $\omega \in \Omega$ is:

$$\alpha_i(\omega) = |R_i(\omega)| - 1 \tag{1}$$

where $R_i(\omega)$ denotes the set of permissible actions for agent i in state ω. When $\alpha_i(\omega) = 0$, the agent has exactly one permissible action (zero degrees of freedom).

Definition 3.2 (System Autonomy)**.** The aggregate system autonomy is:

$$\mathcal{A}(\omega) = \sum_{i=1}^{n} \alpha_i(\omega) \tag{2}$$

Definition 3.3 (Constraint Strength)**.** Let $\mathcal{U}$ be the universal action set. The constraint strength is:

$$\mathcal{C}(\omega) = \sum_{i=1}^{n} (|\mathcal{U}| - |R_i(\omega)|) \tag{3}$$

Definition 3.4 (Adaptability)**.** A system exhibits adaptability if, for environmental perturbation $\epsilon : \Omega \to \Omega$, there exists a response trajectory that maintains coherence. Formally:

$$\text{Adapt}(\mathcal{S}) = 1 \iff \forall \epsilon \in \mathcal{E}, \exists \tau : \lim_{t \to \infty} d(\omega_t, \Omega^*) < \delta \tag{4}$$

where Ω^* is the target region and $\mathcal{E}$ is the class of admissible perturbations.

Definition 3.5 (Coherence)**.** A system maintains coherence if state variance remains bounded:

$$\text{Coh}(\mathcal{S}) = 1 \iff \text{Var}(\omega_t) < M \text{ for all } t \tag{5}$$

4 Lyapunov Stability Framework

We now introduce a Lyapunov-based formalization that provides additional mathematical precision and opens pathways to computational modeling.

4.1 The Coherence Potential

Definition 4.1 (Coherence Potential)**.** Let $V : \Omega \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a Lyapunov-like function measuring system coherence as a function of state, constraint level, and autonomy level:

$$V(\omega, C, A) \tag{6}$$

where $\omega \in \Omega$ is the system state, $C \in \mathbb{R}_{\geq 0}$ is the aggregate constraint strength, and $A \in \mathbb{R}_{\geq 0}$ is the aggregate autonomy.

4.2 Boundary Behavior

Axiom 4.2 (Constraint Boundary)**.** As constraints approach total determination:

$$\lim_{A \to 0} V(\omega, C, A) = V_{\text{rigid}} < V_{\max} \tag{7}$$

The system converges to a rigid attractor with zero adaptive capacity. While V may remain positive (formal stability exists), the system cannot respond to perturbations outside its predetermined trajectory.

Axiom 4.3 (Autonomy Boundary)**.** As constraints vanish:

$$\lim_{C \to 0} V(\omega, C, A) = 0 \tag{8}$$

Without constraints, agent actions become uncorrelated, variance grows without bound, and coherence collapses.

4.3 The Optimal Zone

Theorem 4.4 (Existence of Optimal Zone)**.** *Under Axioms 4.2 and 4.3, there exists a region $\mathcal{R}^* \subset \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ such that:*

$$V(\omega, C, A) > \max\{V_{rigid}, \epsilon\} \quad \textit{for all } (C, A) \in \mathcal{R}^* \tag{9}$$

and V achieves its maximum in the interior of $\mathcal{R}^*$.

Proof. 1. By Axiom 4.2, V at the $A = 0$ boundary equals V_{rigid}.

2. By Axiom 4.3, V at the $C = 0$ boundary equals 0.

3. Assume V is continuous in (C, A).

4. For $\epsilon > 0$ sufficiently small, the preimage $V^{-1}((V_{\text{rigid}} + \epsilon, \infty))$ is non-empty and open.

5. This preimage is contained in $\{(C, A) : C > 0 \wedge A > 0\}$.

6. By compactness arguments on bounded subsets, a maximum exists in the interior. □

4.4 Parametric Form

For computational modeling, we propose the following parametric form:

$$V(C, A) = \frac{C \cdot A}{(C + \alpha)(A + \beta)} \cdot e^{-\gamma(C-C^*)^2 - \delta(A-A^*)^2} \tag{10}$$

where $\alpha, \beta > 0$ prevent singularities at zero, $\gamma, \delta > 0$ control the sharpness of the optimal zone, and (C^*, A^*) is the optimal constraint-autonomy pair.

This function satisfies:

- $V \to 0$ as $C \to 0$ or $A \to 0$
- V achieves maximum near (C^*, A^*)
- V decreases for extreme values of either parameter

4.5 Dynamical Stability Analysis

Consider the system dynamics:

$$\dot{\omega} = f(\omega, \mathbf{a}(\omega, A), R(\omega, C)) \tag{11}$$

where $\mathbf{a}$ is the action selection function depending on autonomy A, and R is the rule function depending on constraints C.

Proposition 4.5. *If V is a valid Lyapunov function for the constrained dynamics (i.e., $\dot{V} \leq 0$ along trajectories), then:*

1. ***High-constraint regime*** *($C \gg A$): $\dot{V} = 0$ (no adaptation, frozen at local minimum)*

2. ***Low-constraint regime*** *($A \gg C$): $\dot{V}$ sign-indefinite (no guaranteed stability)*

3. ***Balanced regime*** *($C \approx A \approx$ optimal): $\dot{V} < 0$ with convergence to global optimum*

This formalizes the intuition that balance enables both stability (negative $\dot{V}$) and adaptability (non-zero $\dot{V}$ under perturbation).

5 Main Theorem

Theorem 5.1 (Constraint–Autonomy Compatibility Law)**.** *A rule-constrained multi-agent decision system $\mathcal{S}$ maintains both adaptability and coherence over time if and only if:*

$$0 < \mathcal{C}(\omega) < \mathcal{C}_{\max} \quad \textit{and} \quad 0 < \mathcal{A}(\omega) < \mathcal{A}_{\max} \tag{12}$$

for all reachable states ω.

Proof. We prove necessity by contradiction in two directions, then establish sufficiency.

Part I: Necessity of Non-Zero Autonomy

Claim: If $\mathcal{A}(\omega) = 0$ for all ω, then $\mathrm{Adapt}(\mathcal{S}) = 0$.

1. Assume $\mathcal{A}(\omega) = 0$ for all reachable ω.
2. Then for each agent a_i and each state ω, we have $|R_i(\omega)| = 1$.
3. Therefore, each agent has exactly one permissible action in each state.
4. The system trajectory becomes deterministic: $\omega_{t+1} = T(\omega_t, \mathbf{r}(\omega_t))$ where $\mathbf{r}(\omega_t)$ is the unique permissible action profile.
5. Let ϵ be an environmental perturbation shifting the system to state $\omega' \notin$ the predetermined trajectory.
6. Since action selection remains determined by R, the system cannot select actions adapted to ω'.
7. The trajectory diverges from Ω^* under perturbation.
8. Therefore $\mathrm{Adapt}(\mathcal{S}) = 0$. □

Part II: Necessity of Non-Zero Constraints

Claim: If $\mathcal{C}(\omega) = 0$ for all ω, then $\mathrm{Coh}(\mathcal{S}) = 0$.

1. Assume $\mathcal{C}(\omega) = 0$ for all reachable ω.
2. Then for each agent a_i, we have $R_i(\omega) = \mathcal{U}$ (all actions permitted).
3. Agent actions become statistically independent of system objectives.
4. Model each agent's action selection as a random variable X_i with variance $\sigma_i^2 > 0$.
5. The aggregate state variance grows as:
$$\mathrm{Var}(\omega_t) \geq \sum_{i=1}^{n} \sigma_i^2 \cdot t \tag{13}$$
6. As $t \to \infty$, variance becomes unbounded.
7. Therefore $\mathrm{Coh}(\mathcal{S}) = 0$. □

Part III: Sufficiency

Claim: If $0 < \mathcal{A}(\omega)$ and $0 < \mathcal{C}(\omega)$ for all reachable ω, then there exist system configurations with both $\mathrm{Adapt}(\mathcal{S}) = 1$ and $\mathrm{Coh}(\mathcal{S}) = 1$.

Proof (Constructive):

1. Let constraints be defined such that $|R_i(\omega)| = k$ for some $1 < k < |\mathcal{U}|$.
2. Each agent retains $k - 1$ degrees of freedom (autonomy > 0).
3. Constraints eliminate $|\mathcal{U}| - k$ actions (constraint > 0).
4. Under perturbation ϵ, agents can select among k actions, enabling trajectory adjustment toward Ω^*.
5. Constraints bound action variance: $\mathrm{Var}(X_i) \leq f(k) < f(|\mathcal{U}|)$.
6. With bounded variance per step and finite agents, total variance remains bounded.
7. Therefore both adaptability and coherence are achievable. □

Conclusion: The conjunction of Parts I, II, and III establishes the biconditional. □

6 Corollaries and Theoretical Implications

Corollary 6.1 (Governance Principle). *In human governance systems, rule enforcement that eliminates personal discretion preserves formal order while destroying adaptive capacity. Unrestricted discretion dissolves collective order. Functional governance requires both enforceable rules and protected internal freedom.*

Corollary 6.2 (AI Alignment). *Artificial agents operating under hard constraints without internal optimization latitude cannot adapt to distributional shift. Conversely, unconstrained agents cannot maintain alignment with specified objectives.*

Corollary 6.3 (Organizational Design). *Organizations that eliminate employee discretion become brittle under market uncertainty. Organizations without procedural constraints lose strategic coherence.*

Corollary 6.4 (Evolutionary Stability). *Species with rigid genetic programs face extinction under environmental change. Species with unlimited mutational variance lose adaptive coherence. Evolutionary persistence requires constrained variation.*

Corollary 6.5 (Constitutional Design). *Constitutions allocating excessive discretion to any branch produce instability or capture. Constitutions permitting no discretion cannot adapt to unforeseen circumstances. Durable constitutional orders balance procedural constraints with interpretive flexibility.*

7 Practical Applications

7.1 Organizational Design

The Constraint–Autonomy Compatibility Law provides actionable guidance for organizational architecture:

Symptom	**Likely Cause**	**Intervention**
Innovation stagnation, high turnover	Over-constraint (micro-management)	Increase decision autonomy at operational level
Strategic drift, inconsistent quality	Under-constraint	Strengthen procedural boundaries, clarify objectives
Local optimization, siloed behavior	Misaligned constraints	Redesign constraints to reflect system-level objectives
Adaptive paralysis	Constraint-autonomy mismatch	Align autonomy domains with uncertainty sources

Table 2: Organizational diagnosis framework.

Design Principles:

1. **Subsidiary autonomy:** Grant autonomy at the level where relevant information exists
2. **Constraint abstraction:** Define constraints on outcomes, not processes, where possible
3. **Feedback integration:** Use outcomes to calibrate constraint-autonomy balance dynamically
4. **Uncertainty matching:** Increase autonomy in high-uncertainty domains, increase constraints in low-uncertainty domains

7.2 Political Systems and Constitutional Design

Comparative political analysis through our framework reveals three archetypal regimes:

Authoritarian Rigidity: High constraint, low autonomy. Stable under static conditions but exhibits catastrophic failure under regime change or external shock. Examples include late Soviet system and centralized command economies.

Democratic Adaptability: Balanced constraint-autonomy at multiple levels. Constitutional constraints on process with autonomy in policy content. Federalism serves as constraint-autonomy distribution across scales, while judicial review provides dynamic constraint calibration.

Failed States: Low constraint, high (uncoordinated) autonomy. Warlordism, corruption, and institutional collapse produce no coherent trajectory despite agent activity.

7.3 Multi-Agent AI Systems

The constraint-autonomy framework directly addresses current challenges in AI alignment and multi-agent coordination:

The Paperclip Maximizer Problem: An agent with a fixed objective function and unlimited autonomy over strategy converges to solutions misaligned with human values. The failure is *under-constraint* on objective specification, not excessive autonomy per se.

The Specification Gaming Problem: Rigidly specified reward functions (high constraint) combined with optimization pressure produce "reward hacking"—the agent finds high-reward states that violate the spirit of the specification. The failure is *over-constraint* on reward combined with *under-constraint* on strategy space.

Recommended Architecture:

1. **Hierarchical constraint structure:** Global constraints on catastrophic actions, local autonomy for task execution
2. **Dynamic constraint adjustment:** Tighten constraints in high-stakes situations, relax in exploration
3. **Autonomy-preserving alignment:** Specify constraints on value-relevant outcomes, not intermediate states
4. **Interpretable autonomy:** Agent explanations of autonomous decisions enable human oversight without eliminating freedom

7.4 Simulation and Empirical Validation

The law admits computational testing through agent-based models. Predicted experimental results include:

1. $V(C, A)$ surface should show interior maximum
2. Systems at boundaries should show either rigidity (high C) or divergence (low C)
3. Perturbation recovery time should be minimized in optimal zone
4. Phase transitions should occur at critical C/A ratios

8 Relation to Existing Theory

8.1 Ashby's Law of Requisite Variety

Ashby's Law [Ashby, 1956] states that a controller must have variety at least equal to the variety of disturbances it must regulate. Our law extends this by distinguishing *external* complexity (handled by constraints) from *internal* variety (autonomy). Ashby treats the controller as a black box; we open it to reveal that internal freedom is structurally necessary.

Key Extension: Ashby's Law concerns the *minimum* variety needed. The Constraint–Autonomy Compatibility Law adds that there is also a *maximum* useful autonomy—beyond which coherence degrades. The law thus provides both floor and ceiling.

8.2 The Good Regulator Theorem (Conant–Ashby)

The Good Regulator Theorem [Conant & Ashby, 1970] states that every good regulator of a system must contain a model of that system. Our law provides the structural precondition for such modeling:

- **Modeling requires state differentiation:** The regulator must be capable of occupying different internal states corresponding to different external conditions.
- **State differentiation requires autonomy:** A regulator with only one possible internal state cannot model anything.
- **Therefore:** Autonomy is necessary for good regulation, not merely permissible.

Our law transforms the Good Regulator Theorem from a statement about what regulators *are* to a statement about what they *require*.

8.3 Autopoiesis (Maturana and Varela)

Autopoietic theory [Maturana & Varela, 1980, Varela et al., 1991] describes living systems as self-producing unities that maintain organizational closure while remaining structurally open. Our law formalizes the quantitative conditions for autopoietic viability:

Autopoietic Concept	**Our Formalization**
Organizational closure	Constraint function $R(\omega)$
Structural openness	Autonomy $\mathcal{A}(\omega) > 0$
Autopoietic unity	Coherence $V > 0$
Structural coupling	Adaptive response within constraints
Disintegration	$V \to 0$ at boundaries

Table 3: Correspondence between autopoietic concepts and our formalization.

8.4 Control Theory

Classical control theory optimizes for stability given a fixed controller structure. Our framework treats autonomy as a *variable* rather than noise, demonstrating that eliminating it produces degenerate solutions incapable of robust performance.

Extension to Robust Control: The H_∞ control framework handles uncertainty by optimizing worst-case performance. Our law suggests an alternative: rather than designing for worst-case disturbances, design for *adaptive capacity* by preserving autonomy.

8.5 Decision Theory and Bounded Rationality

Standard decision theory assumes agents optimize within constraints. We show that the *size* of the feasible set is itself a critical system parameter—too small yields rigidity, too large yields chaos.

Simon's bounded rationality [Simon, 1955] notes that agents satisfice rather than optimize due to computational limits. Our law provides a normative complement: even with unlimited computation, systems *should* maintain bounded autonomy to preserve coherence.

8.6 Thermodynamics and Information Theory

An intriguing parallel exists with the Second Law of Thermodynamics:

- **Entropy** measures disorder/uncertainty
- **Free energy** measures capacity for work
- **Constraint-autonomy balance** $\approx$ maintaining free energy above thermal equilibrium

A system at maximum entropy (zero constraint) cannot perform directed work. A system at minimum entropy (zero autonomy) cannot adapt to changed conditions. Living systems maintain themselves far from equilibrium—in the constraint-autonomy sweet spot.

Information-Theoretic Reformulation:

$$I_{\text{coherence}} = I_{\text{constraint}} - H_{\text{autonomy}} \tag{14}$$

where coherence information equals constraint information minus autonomy entropy. The law states: both terms must be non-zero for adaptive coherence.

8.7 Edge of Chaos and Criticality

Langton [Langton, 1990] and Kauffman [Kauffman, 1993] identified the "edge of chaos" as the regime where complex computation occurs. Our law provides:

1. **Mechanism:** The edge is maintained by constraint-autonomy balance
2. **Necessity:** Systems not at the edge lose either adaptability or coherence
3. **Control parameter:** Constraint/autonomy ratio positions the system

This transforms the edge of chaos from an empirical observation to a derived consequence.

9 Validity Boundaries and Scope Conditions

The Constraint–Autonomy Compatibility Law is not universal. We explicitly characterize its domain of applicability.

9.1 Conditions Where the Law Applies

The law applies to systems that:

1. Contain agents capable of multiple actions
2. Operate under uncertainty about future states
3. Require both stability and adaptability
4. Lack perfect central controllers with complete foresight
5. Face environments that change on timescales comparable to system response

9.2 Conditions Where the Law Fails

Deterministic Mechanisms: Pure mechanical systems with no internal state variables (thermostats, simple feedback controllers) do not require autonomy. Their "adaptation" is fully specified by design.

Static Environments: If the environment never changes in relevant ways, adaptability provides no benefit. A perfectly stable niche permits perfectly rigid organisms.

Perfect Controllers: A hypothetical omniscient controller with complete information about present and future states could specify optimal actions deterministically. Autonomy would introduce only deviation from optimum.

Fully Known State Spaces: If all possible states are enumerable and their transitions known, a lookup table suffices. This is the limiting case of perfect control.

9.3 Boundary Cases

Low-Frequency Environmental Change: When environmental change is slow relative to system response, lower autonomy is tolerable. The constraint-autonomy balance shifts toward constraints.

High-Frequency Environmental Change: Rapidly changing environments favor higher autonomy. Constraints must be abstract (on objectives) rather than specific (on actions).

Hierarchical Systems: Different levels may have different optimal balances. Lower levels often tolerate higher constraints (operational procedures) while higher levels require more autonomy (strategic decisions).

9.4 Falsifiability

The law makes falsifiable predictions:

1. Systems with zero autonomy should fail to adapt to novel conditions
2. Systems with zero constraints should exhibit unbounded variance
3. Optimal performance should occur at intermediate constraint-autonomy ratios
4. Perturbation recovery should be fastest in the balanced regime

These predictions are testable through simulation, historical analysis, and controlled experiments on organizational and artificial systems.

10 Philosophical Implications

10.1 Freedom and Order as Structural Complements

The deepest implication of the Constraint–Autonomy Compatibility Law is that freedom and order are not opposites requiring compromise. They are *mutually necessary conditions* for each other's existence in adaptive systems.

- **Without constraints, freedom becomes meaningless:** Unconstrained action produces no coherent trajectory, hence no persistent agent whose freedom could be preserved.
- **Without autonomy, order becomes sterile:** Fully determined systems cannot adapt, hence cannot persist through environmental change, hence cannot maintain their order.

This resolves an ancient philosophical tension not by finding a "balance" between competing goods, but by revealing that the supposed competition is illusory. Freedom and order are structural complements, not moral trade-offs.

10.2 Implications for Political Philosophy

The law suggests that debates framing liberty against security, or individual rights against collective welfare, may be fundamentally misconceived. The question is not "how much freedom should we sacrifice for order?" but "what configuration of freedom and order permits both to exist?"

This has concrete implications:

- **Rights are not constraints on governance but conditions for it:** Systems that eliminate rights lose adaptive capacity and eventually collapse.
- **Order is not a constraint on freedom but its context:** Freedom without institutional structure produces chaos, not flourishing.
- **Subsidiarity is not merely efficient but necessary:** Centralized control cannot model local conditions (Good Regulator Theorem), hence cannot regulate them effectively.

10.3 The Proverb Recovered

We began with the aphorism: *"Without freedom, any justice becomes a cage."*

We can now unpack this precisely:

- **"Justice"** encodes the constraint function—rules specifying permissible actions
- **"Freedom"** encodes autonomy—degrees of choice within the permissible set
- **"Cage"** encodes the loss of adaptability—convergence to rigid, non-adaptive states
- **"Without freedom"** encodes the boundary condition $\mathcal{A} \to 0$

The proverb is a lossy compression of the Constraint–Autonomy Compatibility Law, preserving the core structural claim while discarding the mathematical framework, scope conditions, and corollaries.

The law recovers what the proverb encodes, while adding what compression lost.

11 Connection to the Unified Theory of Self-Organizing Systems

The Constraint–Autonomy Compatibility Law finds a natural extension and complement in the Unified Theory of Self-Organizing Systems [?], which establishes four interconnected laws governing peer systems with shared environments. This section demonstrates the deep structural relationship between the two frameworks and shows how their synthesis yields a more complete theory of adaptive multi-agent systems.

11.1 The Unified Theory: Overview

The Unified Theory addresses a specific class of systems: *peer systems* where agents share environment, face symmetric viability constraints, engage in repeated interaction, and operate under finite resource constraints. Within this scope, four laws are derived:

- **Law Zero:** Cooperative response ($\phi > 0$) emerges as the statistically dominant and dynamically stable configuration; antagonistic behavior requires continuous external perturbation P to maintain.
- **Law I:** Unobstructed cooperative dynamics yield efficient self-organization.
- **Law II:** Perceived complexity requirements often reflect overconstraint rather than structural necessity.
- **Law III:** Systematic observational bias arises from differential signal generation between functional and failure states.

The central insight is a reversal of explanatory burden: within peer systems, *antagonism requires explanation, not cooperation.*

11.2 Structural Correspondence

The two frameworks exhibit deep structural parallels:

11.3 The Key Bridge: Perturbation as Over-Constraint

The most important connection lies in recognizing that **antagonistic perturbation P is a special case of over-constraint that eliminates agents' autonomy to cooperate**.

In the Unified Theory, perturbation P modifies the payoff structure:

$$U_P(a_i, a_j) = U_0(a_i, a_j) + P(a_i, a_j) \tag{15}$$

where P creates zero-sum or negative-sum structure between agents. Concrete forms of P include:

Constraint–Autonomy Law	**Unified Theory**
Autonomy $\mathcal{A} > 0$ required for adaptation	Cooperative response $\phi > 0$ as default
Over-constraint $\mathcal{C} \to \mathcal{C}_{\max}$	Antagonistic perturbation $P \neq 0$
Under-constraint $\mathcal{C} \to 0$	System incoherence (no constraints)
Optimal zone (C^*, A^*)	Unperturbed cooperative equilibrium
Coherence potential $V(\omega, C, A)$	Aggregate welfare W
Rigid attractor (high C, low A)	P-maintained antagonistic equilibrium

Table 4: Structural correspondence between frameworks.

- Artificial scarcity (resource access restrictions)
- Relative performance incentives (ranking systems)
- Exit barriers (forced continued interaction with defectors)
- Information suppression (inability to identify cooperators)

Each of these mechanisms operates by *constraining agents' action space* in ways that eliminate cooperative options:

Proposition 11.1 (Perturbation-Constraint Equivalence). *Every antagonistic perturbation P in the Unified Theory corresponds to a constraint configuration (C_P, A_P) in the Constraint–Autonomy framework where $A_P < A^*$ (autonomy below optimal).*

Argument:

1. Artificial scarcity constrains access to shared resources, reducing agents' feasible action sets.
2. Relative performance metrics constrain evaluation criteria, eliminating win-win outcomes from the "permissible" region.
3. Exit barriers constrain partner selection, removing the autonomy to avoid defectors.
4. Information suppression constrains epistemic access, reducing effective choice to random selection.

In each case, P operates by adding constraints that reduce autonomy below the threshold required for cooperative equilibrium. The Unified Theory's claim that "removing P restores cooperation" is equivalent to our claim that "reducing over-constraint toward optimal C^* restores adaptive capacity."

11.4 Law Zero as Corollary

Given the Constraint–Autonomy Compatibility Law, Law Zero of the Unified Theory can be derived as a corollary for peer systems:

Corollary 11.2 (Cooperative Default in Peer Systems). *In peer systems where all agents satisfy the constraint-autonomy balance conditions, cooperative response emerges as the default stable configuration.*

Proof. Let $\mathcal{S}$ be a peer system where each agent a_i operates under constraints C_i and possesses autonomy A_i satisfying $0 < C_i < C_{\max}$ and $0 < A_i < A_{\max}$.

By the main theorem, each agent maintains both adaptability and coherence. In peer systems with shared environment E, adaptability implies capacity to adjust actions based on environmental feedback. Coherence implies bounded variance in action selection.

The shared environment creates coupling: agent a_i's welfare depends on E, which depends on all agents' actions. Under the Good Regulator Theorem, effective regulation of one's own welfare requires modeling the system—including other agents' likely responses.

With sufficient autonomy ($A_i > 0$), each agent can form such models. With bounded autonomy ($A_i < A_{\max}$), action variance remains bounded, enabling prediction.

The combination of predictability and mutual dependence on E creates conditions where:

$$\frac{\partial W_i}{\partial a_j} > 0 \quad \text{(positive externalities)} \tag{16}$$

for cooperative actions. This is precisely $\phi > 0$ in the Unified Theory's notation.

Antagonistic configurations ($\phi < 0$) require either:

1. $A_i = 0$ (no capacity to cooperate—deterministic defection)
2. $A_i = A_{\max}$ (no coherence—random action regardless of E)
3. External P that shifts payoff structure

Cases (1) and (2) violate the constraint-autonomy balance. Case (3) is the Unified Theory's perturbation. Therefore, within properly balanced peer systems, $\phi > 0$ is the stable default. □

11.5 The Coherence Potential Predicts Cooperation Levels

The Lyapunov coherence potential $V(C, A)$ developed in Section 4 provides a quantitative predictor for cooperation levels in peer systems:

Proposition 11.3 (Cooperation-Coherence Correspondence). *In peer systems, the cooperative response function ϕ is monotonically related to the coherence potential:*

$$\phi \propto V(C, A) \cdot \mathbf{1}_{[C>0, A>0]} \tag{17}$$

where $\mathbf{1}$ *is the indicator function for the viable region.*

This yields testable predictions:

1. Systems in the optimal zone (C^*, A^*) exhibit maximum cooperation
2. Moving toward boundaries decreases cooperation before system failure
3. The rate of cooperation decline follows the gradient ∇V

11.6 Unified Diagnostic Framework

The synthesis of both frameworks yields a powerful diagnostic methodology for analyzing dysfunctional systems:

Step 1: System Classification

- Is this a peer system? (symmetric viability constraints, shared environment)
- If not, apply Constraint–Autonomy Law to each hierarchical level separately

Step 2: Constraint-Autonomy Assessment

- Measure effective constraint strength $\mathcal{C}$
- Measure effective autonomy $\mathcal{A}$
- Locate system on (C, A) surface

Step 3: Perturbation Identification

- If $\mathcal{A} < A^*$: identify constraints that reduce autonomy
- Classify constraints as essential (C_E) or discretionary (C_D)
- Discretionary constraints reducing autonomy below A^* constitute perturbation P

Step 4: Intervention Design

- Remove P (discretionary over-constraints)
- Monitor relaxation toward cooperative equilibrium
- Account for internalized perturbation P_{int} (cultural lag)

11.7 The Tolerance Mechanism and Autonomy Preservation

The Unified Theory identifies *strategic ignoring* as a key mechanism by which cooperative systems tolerate freeloaders without enforcement costs. This mechanism has a natural interpretation in our framework:

Enforcement is constraint; tolerance preserves autonomy.

Punishment, exclusion, and monitoring all add constraints to the system. While they may reduce freeloading, they also reduce overall autonomy—potentially pushing the system toward the over-constrained boundary.

The Unified Theory's finding that "robust cooperative systems solve the freeloader problem through structural tolerance" translates to: *systems in the optimal zone* (C^*, A^*) *can absorb parasitic load without requiring additional constraints that would shift them toward rigidity.*

This resolves the apparent paradox of cooperation without enforcement: the "cost" of tolerating freeloaders is lower than the cost of the autonomy reduction required for enforcement.

11.8 Internalized Perturbation and Cultural Dynamics

The Unified Theory introduces internalized perturbation P_{int}—antagonistic patterns that persist culturally even after external P_{ext} is removed. In our framework, this corresponds to:

Definition 11.4 (Internalized Constraint)**.** A constraint C_{int} is internalized if it persists in agents' behavior even when external enforcement is removed:

$$\mathcal{C}_{\text{eff}} = \mathcal{C}_{\text{ext}} + \mathcal{C}_{\text{int}} \tag{18}$$

where $\mathcal{C}_{\text{int}}$ decays on generational timescales $\tau_2 \gg \tau_1$ (behavioral adaptation timescale).

This explains why removing formal constraints does not immediately restore optimal autonomy: agents continue to behave as if constrained. The Unified Theory's "relaxation lag" is the time required for $\mathcal{C}_{\text{int}} \to 0$.

Implication for reform: Institutional changes (reducing $\mathcal{C}_{\text{ext}}$) must be accompanied by cultural interventions (reducing $\mathcal{C}_{\text{int}}$) for full restoration of adaptive capacity.

11.9 Synthesis: A Complete Theory of Adaptive Multi-Agent Systems

The integration of the Constraint–Autonomy Compatibility Law with the Unified Theory yields a comprehensive framework:

1. **Micro-level (individual agents):** Each agent requires constraint-autonomy balance for adaptability and coherence.

2. **Meso-level (agent interactions):** Peer agents with balanced constraint-autonomy naturally converge to cooperative dynamics.

3. **Macro-level (system properties):** Cooperative peer systems self-organize efficiently, require only threshold complexity for viability, yet appear dysfunctional due to observational bias toward failure states.

4. **Pathology diagnosis:** Observed antagonism or dysfunction traces to either:

   - Constraint-autonomy imbalance (too much or too little constraint)
   - External perturbation P (antagonism-maintaining interventions)
   - Internalized perturbation P_{int} (cultural persistence of constraint)
   - Misclassification (system is not actually peer-structured)

5. **Intervention principle:** Restore balance by removing discretionary constraints, eliminating perturbation, and allowing relaxation toward natural attractors.

Theorem 11.5 (Unified Law of Adaptive Multi-Agent Systems). *A multi-agent system maintains adaptive coherence and cooperative dynamics if and only if:*

1. *Each agent satisfies* $0 < \mathcal{C}_i < \mathcal{C}_{\max}$ *and* $0 < \mathcal{A}_i < \mathcal{A}_{\max}$
2. *For peer agents, no antagonistic perturbation* P *is externally maintained*
3. *The system has not been displaced from equilibrium more recently than relaxation timescale* τ

This unified law combines the structural requirements of the Constraint–Autonomy Compatibility Law with the dynamical insights of the Unified Theory, providing both existence conditions (what balance is required) and stability conditions (what maintains or disrupts equilibrium).

12 Grand Synthesis: Integration with the Extended Theoretical Framework

The Constraint–Autonomy Compatibility Law is not an isolated result but the central integrating principle of a broader research program. This section demonstrates how four independently developed theoretical frameworks converge on the same deep structure, with the Constraint–Autonomy Law serving as the unifying core.

12.1 The Four Frameworks

We synthesize four theoretical contributions:

1. **The Constraint–Autonomy Compatibility Law** (this paper): Adaptive systems require both constraints ($C > 0$) and autonomy ($A > 0$) for sustained function.

2. **The Unified Theory of Self-Organizing Systems** [Kriger, 2017]: Cooperative response emerges as default in peer systems; antagonism requires external perturbation P.

3. **The Principle of Optimal Coherence** [Kriger, 2026a]: Coherent systems satisfy necessary constraints that function as load-bearing structures; anomalies signal constraint violations.

4. **The Asymmetry of Totalizing Ideals** [Kriger, 2026b]: Maximizing any single variable destroys adaptive variance; terminal optimization is self-defeating.

5. **Epistemic Constraint Theory** [Kriger, 2021]: The structure of admissible hypothesis space bounds inferential accuracy more fundamentally than computational capacity or data volume.

12.2 The Convergence Structure

These frameworks are not merely related—they are mathematical reformulations of the same underlying principle viewed from different angles:

```
                 Epistemic Constraint Theory
                 (inference limitations)
                           |
                 soft constraint w(h) =
                   constraint function C
                           |
           +---------------+---------------+
           |                               |
           v                               v
Constraint-Autonomy Law          Principle of Optimal
(C-A balance for adaptation)      Coherence (necessary
           |                      vs discretionary C)
           |                               |
           +---------------+---------------+
                           |
                           v
             Asymmetry of Totalizing Ideals
             (why C_max or A_max fails)
                           |
                           v
             Unified Theory of Self-Organization
             (cooperation from balanced C-A)
```

Figure 1: Convergence structure of the five frameworks. Arrows indicate logical derivation or mathematical equivalence.

12.3 Mathematical Correspondences

12.3.1 Epistemic Constraint Theory ↔ Constraint–Autonomy Law

Epistemic Constraint Theory (ECT) introduces the soft constraint function $w : \mathcal{H} \to [0,1]$ that bounds inferential accuracy:

$$A(\mathcal{S}, w, e) \leq \log w(h^*) + \log \pi(h^*) + \log \ell(h^*, e) - \log Z_w(e) \tag{19}$$

This maps directly to our framework:

ECT Concept	**Constraint–Autonomy Translation**
Soft constraint $w(h)$	Constraint function C on hypothesis space
$w(h^*) = 0$ (truth excluded)	$\mathcal{A} = 0$ (zero autonomy to reach truth)
$w(h^*) = \varepsilon$ small	$\mathcal{A} < A^*$ (suboptimal autonomy)
Constraint dominance	Over-constraint effect
Preconditioning Φ	Constraint modification $(C, A) \to (C', A')$
Uncertainty tolerance threshold θ^*	Optimal zone boundary

Table 5: Mathematical correspondence between ECT and Constraint–Autonomy frameworks.

Proposition 12.1 (ECT-CA Equivalence). *The ECT constraint dominance principle is equivalent to the over-constraint boundary of the Constraint–Autonomy Law. Specifically:*

$$w(h^*) \to 0 \quad \Leftrightarrow \quad \mathcal{A} \to 0 \quad \Rightarrow \quad Adapt(\mathcal{S}) = 0 \tag{20}$$

In both frameworks, structural constraints on the hypothesis/action space dominate processing capacity in determining system performance.

12.3.2 The Four ECT Corollaries as Constraint–Autonomy Principles

ECT derives four operational corollaries that translate directly:

1. **Constraint Dominance Principle:** "Inferential failure is dominated by hypothesis constraints, not processing capacity."

   CA Translation: The coherence potential $V(C, A)$ is bounded by constraint structure; computational optimization within a bad (C, A) configuration cannot escape the constraint-imposed ceiling.

2. **Cognitive Preconditioning Principle:** "Correcting priors before reasoning is more effective than improving reasoning alone."

   CA Translation: Modifying (C, A) toward optimal zone yields greater improvement than optimizing dynamics within suboptimal (C, A). This is equivalent to removing discretionary constraints C_D.

3. **Interrogative Primacy Principle:** "Knowledge growth is governed by question selection rather than answer possession."

   CA Translation: Query selection dynamically modifies effective autonomy—asking the right question expands the accessible region of state space.

4. **Uncertainty Tolerance Principle:** "Maintaining uncertainty until entropy falls below threshold is required for optimal inference."

   CA Translation: Autonomy ($A > 0$) must be preserved until the system reaches the optimal stopping boundary. Premature constraint (forcing $A \to 0$ too early) reduces accuracy/adaptability.

12.3.3 Principle of Optimal Coherence ↔ Essential vs. Discretionary Constraints

The Principle of Optimal Coherence [Kriger, 2026a] distinguishes between:

- **Load-bearing constraints:** Structural necessities whose removal causes system collapse
- **Discretionary constraints:** Additions that may reduce performance without structural necessity

This maps precisely to our distinction:

$$C_E : \text{Essential constraints (load-bearing)} \quad \Rightarrow \quad \text{removal causes } V \to 0 \tag{21}$$

$$C_D : \text{Discretionary constraints} \quad \Rightarrow \quad \text{removal increases } V \tag{22}$$

The Optimal Coherence principle states that anomalies are diagnostic signals of constraint violations. In our framework:

- Anomalies from C_E violation $\to$ system has exited viable region (under-constraint)
- Anomalies from C_D imposition $\to$ system is over-constrained (rigidity symptoms)

Proposition 12.2 (Load-Bearing Constraint Criterion)**.** *A constraint $c \in C$ is load-bearing if and only if:*

$$\left.\frac{\partial V}{\partial c}\right|_{c \to 0} < 0 \tag{23}$$

That is, removing the constraint decreases coherence. Discretionary constraints satisfy the opposite inequality.

12.3.4 Asymmetry of Totalizing Ideals ↔ Boundary Collapse

The Asymmetry of Totalizing Ideals [Kriger, 2026b] proves that pursuing terminal optimization of any single variable destroys adaptive variance:

$$\lim_{t\to\infty} V_t = 0 \quad \text{under totalizing optimization} \tag{24}$$

where V_t is adaptive variance.

This is mathematically equivalent to our boundary conditions:

Totalizing Ideal	CA Boundary	Outcome
Maximize order/control	$C \to C_{\max}$, $A \to 0$	Rigid attractor, zero adaptation
Maximize freedom/autonomy	$C \to 0$, $A \to A_{\max}$	Chaotic divergence, zero coherence
Maximize single objective	Collapse to boundary	Loss of adaptive capacity

Table 6: Totalizing ideals as boundary collapse in constraint–autonomy space.

Corollary 12.3 (Anti-Totalization Principle)**.** *For any adaptive system, there exists no globally optimal point on the boundary of* (C, A) *space. All viable configurations satisfy:*

$$(C^*, A^*) \in \mathit{int}(\mathcal{R}) \quad \text{where } \mathcal{R} = \{(C, A) : 0 < C < C_{\max}, 0 < A < A_{\max}\} \tag{25}$$

Totalizing ideals that drive toward boundaries are self-defeating.

The Asymmetry paper's key insight—"imperfection is not a bug but a feature"—translates to: *non-zero autonomy (apparent imperfection from the controller's view) is structurally necessary for adaptation.*

12.4 The Central Theorem: Constraint–Autonomy as Integrating Principle

We can now state the relationship formally:

Theorem 12.4 (Integration Theorem)**.** *The following statements are equivalent formulations of the same structural principle:*

(a) ***Constraint–Autonomy Law:*** *Adaptive systems require* $0 < C < C_{\max}$ *and* $0 < A < A_{\max}$.

(b) ***Epistemic Constraint Theory:*** *Soft constraints* $w(h) \in (0, 1)$ *on the hypothesis space bound but do not eliminate inferential accuracy.*

(c) ***Optimal Coherence:*** *Load-bearing constraints are necessary; discretionary constraints reduce coherence.*

(d) ***Anti-Totalization:*** *Maximizing any single variable to its limit destroys adaptive variance.*

(e) ***Cooperative Default:*** *Peer systems with balanced constraints exhibit cooperative dynamics; perturbation* P *corresponds to constraint imbalance.*

Proof sketch. (a) ⇔ (b): The ECT constraint function w is isomorphic to the constraint operator C restricted to hypothesis space. The accuracy bound $\log w(h^*)$ corresponds to the coherence ceiling imposed by over-constraint.

(a) ⇔ (c): Essential constraints C_E define the lower bound $C > 0$; discretionary constraints C_D when excessive push A below A^*. The optimal coherence condition is satisfaction of C_E without excess C_D.

(a) $\Leftrightarrow$ (d): Totalizing ideals drive (C, A) toward boundaries. The Asymmetry result that $V_t \to 0$ under totalization is equivalent to our boundary conditions $V \to V_{\text{rigid}}$ or $V \to 0$.

(a) $\Leftrightarrow$ (e): Perturbation P in the Unified Theory adds constraints that reduce autonomy below A^*. Removing P restores balance, enabling cooperative equilibrium. The cooperative default emerges when $(C, A) \in \mathcal{R}^*$ (optimal zone). □

12.5 Unified Notation and Translation Dictionary

For researchers working across these frameworks, we provide a translation dictionary:

Concept	**CA Law**	**ECT**	**Unified Theory**	**Asymmetry**
Constraint strength	$\mathcal{C}$	$1 - w(h)$	P magnitude	Control intensity
Degrees of freedom	$\mathcal{A}$	$\lvert\mathcal{H}\rvert_{\text{eff}\lvert}$	Cooperation capacity	Adaptive variance V_t
System health	$V(C, A)$	Accuracy A	Welfare W	Coherence
Over-constraint	$C \to C_{\max}$	$w(h^*) \to 0$	$P \gg 0$	Totalization
Under-constraint	$C \to 0$	No prior structure	No coordination	Fragmentation
Optimal state	(C^*, A^*)	Calibrated w	$P = 0$ equilibrium	Balanced variance
Pathology signal	$\dot{V} \geq 0$	Accuracy plateau	Stable antagonism	$V_t \to 0$

Table 7: Translation dictionary across the five frameworks.

12.6 Epistemological Implications

The convergence of five independently developed frameworks on the same mathematical structure suggests this is not an artifact of modeling choices but reflects genuine structural features of adaptive systems. Several epistemological implications follow:

12.6.1 Constraints Are Not Restrictions But Enablers

Across all frameworks, constraints that might appear as limitations are revealed as structural necessities:

- ECT: Hypothesis space constraints enable tractable inference
- Optimal Coherence: Load-bearing constraints enable structural integrity
- CA Law: Constraints enable coherence (prevent chaotic divergence)
- Unified Theory: Shared environment constraints enable cooperation

The intuition that "more freedom is always better" is formally refuted: freedom without structure is chaos, not flourishing.

12.6.2 The Boundary Problem is Universal

All frameworks identify the same pathology at boundaries:

- Too much constraint $\to$ rigidity, inability to adapt, inferential blindness
- Too little constraint $\to$ incoherence, divergence, inability to maintain identity

This boundary problem appears in every domain where adaptive systems operate: biological, cognitive, social, organizational, computational.

12.6.3 Diagnosis Before Intervention

The unified framework provides a diagnostic methodology applicable across domains:

1. **Locate** the system in (C, A) space

2. **Identify** whether pathology is over-constraint or under-constraint
3. **Classify** constraints as essential or discretionary
4. **Intervene** by removing discretionary constraints (if over-constrained) or adding essential constraints (if under-constrained)
5. **Monitor** relaxation dynamics and internalized constraints

12.7 The Master Equation

We conclude the synthesis with a master equation capturing the unified principle:

$$\boxed{\frac{dV}{dt} = f(C, A) \cdot \mathbf{1}_{[0<C<C_{\max}]} \cdot \mathbf{1}_{[0<A<A_{\max}]} - g(P, C_D) - h(C_{\text{int}})} \tag{26}$$

where:

- V = coherence/welfare/accuracy (equivalent across frameworks)
- $f(C, A) > 0$ in optimal zone = natural tendency toward coherence
- $\mathbf{1}_{[\cdot]}$ = indicator functions enforcing boundary conditions
- $g(P, C_D) \geq 0$ = degradation from perturbation and discretionary over-constraint
- $h(C_{\text{int}}) \geq 0$ = drag from internalized constraints (cultural lag)

The system achieves stable adaptive coherence when:

$$f(C^*, A^*) > g(P, C_D) + h(C_{\text{int}}) \tag{27}$$

This inequality states: *the natural coherence-generating dynamics in the optimal zone must exceed the combined drag from external perturbation, discretionary over-constraint, and internalized constraint patterns.*

13 Conclusion

We have demonstrated that the folk wisdom encoded in aphorisms about freedom and justice admits rigorous formalization. The Constraint–Autonomy Compatibility Law provides a falsifiable structural principle that serves as the integrating core of a broader theoretical framework, unifying insights from:

- **Cybernetics:** Ashby's Law, Good Regulator Theorem
- **Biology:** Evolutionary dynamics, autopoiesis
- **Control theory:** Lyapunov stability, robust control
- **Complexity science:** Edge of chaos, self-organized criticality
- **Political philosophy:** Constitutional design, subsidiarity
- **AI alignment:** Value alignment, multi-agent coordination
- **Epistemology:** Constraint dominance, hypothesis space limitations
- **Game theory:** Cooperation emergence, perturbation analysis

The grand synthesis presented in Section 12 demonstrates that five independently developed frameworks—the Constraint–Autonomy Compatibility Law, the Unified Theory of Self-Organizing Systems, the Principle of Optimal Coherence, the Asymmetry of Totalizing Ideals, and Epistemic Constraint Theory—converge on the same mathematical structure. This convergence suggests we have identified a genuine structural feature of adaptive systems rather than an artifact of modeling choices.

The central insight is threefold:

1. **Constraints and autonomy are complements, not opposites.** Neither can exist without the other in adaptive systems. This dissolves ancient debates that frame freedom against order.

2. **Boundaries are pathological; interiors are viable.** Maximizing any single variable (total control or total freedom) destroys adaptive capacity. Optimal configurations lie in the interior of parameter space.

3. **Perturbation explains antagonism.** In peer systems, cooperative dynamics are the natural attractor. Observed antagonism is diagnostic of external perturbation or constraint imbalance, not intrinsic conflict.

The practical implications are significant:

For **science**: Design adaptive systems with bounded but genuine agent autonomy. Use the Lyapunov framework to characterize optimal operating regimes. Recognize that the same structural principle governs biological, cognitive, social, and artificial systems.

For **mathematics**: Formalize autonomy as a conserved quantity in constrained optimization. Develop the constraint-autonomy surface as a tool for system analysis. The master equation (26) provides a unified dynamical framework.

For **epistemology**: Recognize that hypothesis space constraints bound inference more fundamentally than computational capacity. Preconditioning (modifying constraints) dominates processing optimization.

For **philosophy**: Reinterpret justice and freedom as structural system parameters rather than competing moral values. Recognize that their apparent opposition dissolves under formal analysis.

For **governance**: Design institutions with explicit attention to constraint-autonomy balance at each level. Use the diagnostic framework to identify whether pathologies stem from over-constraint or under-constraint. Remove discretionary constraints; preserve essential ones.

For **AI development**: Architect multi-agent systems with hierarchical constraints and preserved local autonomy. Recognize that alignment and capability are not in fundamental tension—both require the same constraint-autonomy balance.

The ancient intuition proves correct: Without freedom, any justice does indeed become a cage—not as metaphor, but as mathematical necessity. The contribution of formal analysis is not to validate the intuition (which millennia of human experience already did) but to reveal its structural basis, scope conditions, actionable implications, and deep connections to apparently unrelated domains from Bayesian inference to evolutionary dynamics to organizational design.

The law, and the unified framework it anchors, stands ready for empirical test, computational simulation, and application across domains where adaptive systems must persist under uncertainty—which is to say, nearly everywhere that matters.

A Detailed Derivation of the Master Equation

This appendix provides a complete derivation of the master equation (26) from the Lyapunov framework (Section 4) and the Unified Theory synthesis (Section 11).

A.1 Starting Point: The Coherence Potential

From Section 4, the coherence potential is defined as:

$$V(\omega, C, A) = \mathrm{Coh}(\omega) \cdot \mathrm{Adapt}(\omega) \tag{28}$$

where coherence and adaptability are given by:

$$\text{Coh}(\omega) = 1 - \frac{\text{Var}(\omega)}{\text{Var}_{\text{max}}} \tag{29}$$

$$\text{Adapt}(\omega) = \frac{|\{s' : P(s \to s'|\omega) > 0\}|}{|S|} \tag{30}$$

A.2 Dynamics from Constraint-Autonomy Parameters

The key insight is that both coherence and adaptability depend on the constraint-autonomy configuration (C, A):

1. **Coherence increases with constraint strength:**

$$\text{Coh}(C) = \frac{C}{C + \kappa_1} \tag{31}$$

where $\kappa_1 > 0$ is a saturation parameter. As $C \to 0$, coherence vanishes; as $C \to \infty$, coherence approaches 1.

2. **Adaptability increases with autonomy:**

$$\text{Adapt}(A) = \frac{A}{A + \kappa_2} \tag{32}$$

where $\kappa_2 > 0$. As $A \to 0$, adaptability vanishes; as $A \to \infty$, adaptability approaches 1.

Therefore:

$$V(C, A) = \frac{C}{C + \kappa_1} \cdot \frac{A}{A + \kappa_2} = \frac{CA}{(C + \kappa_1)(A + \kappa_2)} \tag{33}$$

A.3 Time Evolution

Taking the time derivative:

$$\frac{dV}{dt} = \frac{\partial V}{\partial C}\frac{dC}{dt} + \frac{\partial V}{\partial A}\frac{dA}{dt} \tag{34}$$

In the absence of external perturbation, systems tend toward the optimal zone (C^*, A^*). We model this as gradient ascent on V:

$$\frac{d}{dt}\begin{pmatrix} C \\ A \end{pmatrix} = \eta \nabla V = \eta \begin{pmatrix} \partial V/\partial C \\ \partial V/\partial A \end{pmatrix} \tag{35}$$

This gives:

$$\frac{dV}{dt} = \eta \|\nabla V\|^2 \geq 0 \tag{36}$$

Define $f(C, A) \equiv \eta \|\nabla V\|^2$, which is strictly positive in the interior of the viable region.

A.4 Boundary Conditions

The indicator functions enforce boundary conditions:

- $\mathbf{1}_{[0<C<C_{\text{max}}]}$ ensures the dynamics only operate when constraints are neither absent nor total
- $\mathbf{1}_{[0<A<A_{\text{max}}]}$ ensures autonomy is neither zero nor unbounded

At boundaries, $f(C, A) \cdot \mathbf{1}_{[\cdot]} = 0$, so the natural tendency toward coherence vanishes.

A.5 Perturbation and Discretionary Constraint Terms

From the Unified Theory (Section 11), external perturbation P and discretionary over-constraint C_D reduce system coherence:

$$g(P, C_D) = \gamma_1 |P| + \gamma_2 C_D \tag{37}$$

where $\gamma_1, \gamma_2 > 0$ are coupling constants. This term is always non-negative and subtracts from dV/dt.

A.6 Internalized Constraint Term

Cultural lag from internalized constraints:

$$h(C_{\text{int}}) = \gamma_3 C_{\text{int}} \cdot e^{-t/\tau_2} \tag{38}$$

where τ_2 is the generational decay timescale. Initially h is significant; over time it decays as internalized patterns fade.

A.7 The Complete Master Equation

Combining all terms:

$$\boxed{\frac{dV}{dt} = f(C, A) \cdot \mathbf{1}_{[0<C<C_{\max}]} \cdot \mathbf{1}_{[0<A<A_{\max}]} - g(P, C_D) - h(C_{\text{int}})} \tag{39}$$

A.8 Alternative Smooth Formulation

For numerical simulation, replace indicator functions with sigmoid functions:

$$\frac{dV}{dt} = f(C, A) \cdot \sigma_\beta(C) \cdot \sigma_\beta(C_{\max} - C) \cdot \sigma_\beta(A) \cdot \sigma_\beta(A_{\max} - A) - g(P, C_D) - h(C_{\text{int}}) \tag{40}$$

where $\sigma_\beta(x) = 1/(1 + e^{-\beta x})$ with $\beta \gg 1$ for sharp transitions.

A.9 Example Numerical Solution

With parameters $\kappa_1 = \kappa_2 = 1$, $\eta = 0.1$, $\gamma_1 = 0.3$, $\gamma_2 = 0.2$, $\gamma_3 = 0.1$, $\tau_2 = 10$, and initial conditions $C(0) = 0.5$, $A(0) = 0.5$, $V(0) = 0.25$:

- **Case 1** ($P = 0$, $C_D = 0$, $C_{\text{int}} = 0$): $V(t) \to V^* \approx 0.44$ (optimal equilibrium)
- **Case 2** ($P = 0.5$, $C_D = 0.3$): $V(t) \to 0.21$ (perturbation-degraded)
- **Case 3** ($C_{\text{int}}(0) = 0.5$): $V(t)$ initially suppressed, recovers as C_{int} decays

B Agent-Based Simulation Code and Results

This appendix provides Python code for simulating multi-agent systems under the Constraint–Autonomy framework and presents key results.

B.1 Core Simulation Framework

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.integrate import odeint
import networkx as nx

class ConstraintAutonomySystem:
    """
    Multi-agent system with constraint-autonomy dynamics.

    Parameters:
    -----------
    n_agents : int
        Number of agents
    n_states : int
        Number of possible states per agent
    C : float
        Constraint strength (0 < C < C_max)
    A : float
        Autonomy level (0 < A < A_max)
    """

    def __init__(self, n_agents=50, n_states=10, C=0.5, A=0.5):
        self.n_agents = n_agents
        self.n_states = n_states
        self.C = C
        self.A = A
        self.C_max = 1.0
        self.A_max = 1.0

        # Initialize agent states randomly
        self.states = np.random.randint(0, n_states, n_agents)

        # Create interaction network (small-world)
        self.network = nx.watts_strogatz_graph(n_agents, 4, 0.3)

        # Define rules based on constraint level
        self.n_rules = int(C * n_states * (n_states - 1))
        self.rules = self._generate_rules()

    def _generate_rules(self):
        """Generate transition rules based on constraint level."""
        all_transitions = [(i, j) for i in range(self.n_states)
                          for j in range(self.n_states) if i != j]
        n_forbidden = int((1 - self.A) * len(all_transitions))
        forbidden = set(np.random.choice(len(all_transitions),
                       n_forbidden, replace=False))
        return {t: (i not in forbidden) for i, t in enumerate(all_transitions)}

    def compute_autonomy(self):
        """Compute effective autonomy from rules."""
```

```
        n_allowed = sum(self.rules.values())
        n_total = len(self.rules)
        return n_allowed / n_total if n_total > 0 else 0

    def compute_constraint(self):
        """Compute effective constraint strength."""
        return 1 - self.compute_autonomy()

    def compute_coherence(self):
        """Compute system coherence (inverse of state variance)."""
        state_counts = np.bincount(self.states, minlength=self.n_states)
        variance = np.var(state_counts)
        max_variance = (self.n_agents ** 2) / 4  # Max variance for bimodal
        return 1 - variance / max_variance if max_variance > 0 else 1

    def compute_adaptability(self):
        """Compute system adaptability."""
        return self.compute_autonomy()

    def compute_V(self):
        """Compute coherence potential V(C, A)."""
        return self.compute_coherence() * self.compute_adaptability()

    def step(self, perturbation=0):
        """Execute one time step of the simulation."""
        new_states = self.states.copy()

        for i in range(self.n_agents):
            current = self.states[i]
            neighbors = list(self.network.neighbors(i))

            if len(neighbors) == 0:
                continue

            # Possible transitions
            possible = [s for s in range(self.n_states)
                        if self.rules.get((current, s), True)]

            if len(possible) == 0:
                continue

            # Transition probability influenced by neighbors and perturbation
            neighbor_states = [self.states[n] for n in neighbors]
            probs = np.zeros(len(possible))

            for j, target in enumerate(possible):
                # Base probability
                probs[j] = 1.0
                # Neighbor influence (coherence pressure)
                probs[j] += sum(1 for ns in neighbor_states if ns == target)
                # Perturbation adds noise
```

```
                probs[j] += perturbation * np.random.randn()

            probs = np.maximum(probs, 0)
            probs /= probs.sum()

            new_states[i] = np.random.choice(possible, p=probs)

        self.states = new_states

    def simulate(self, n_steps=100, perturbation=0):
        """Run simulation and return V(t) trajectory."""
        V_trajectory = [self.compute_V()]

        for _ in range(n_steps):
            self.step(perturbation)
            V_trajectory.append(self.compute_V())

        return np.array(V_trajectory)


def compute_V_surface(C_range, A_range, n_samples=10):
    """Compute V(C, A) surface via simulation."""
    V_surface = np.zeros((len(C_range), len(A_range)))

    for i, C in enumerate(C_range):
        for j, A in enumerate(A_range):
            V_samples = []
            for _ in range(n_samples):
                sys = ConstraintAutonomySystem(C=C, A=A)
                V_final = sys.simulate(n_steps=50)[-1]
                V_samples.append(V_final)
            V_surface[i, j] = np.mean(V_samples)

    return V_surface
```

B.2 Simulation Results

B.2.1 The V(C, A) Surface

Running `compute_V_surface` with $C \in [0.1, 0.9]$ and $A \in [0.1, 0.9]$ produces a characteristic surface with:

- Maximum at approximately $(C^*, A^*) \approx (0.5, 0.6)$
- Rapid decay toward all boundaries
- Slight asymmetry favoring higher autonomy

B.2.2 Phase Diagram

The simulation reveals three distinct regimes:

B.2.3 Perturbation Effects

Simulating with varying perturbation strength $P \in [0, 1]$:

Regime	C Range	A Range	Characteristics
Frozen	$C > 0.8$	$A < 0.3$	States locked, $V \approx 0.1$
Edge of Chaos	$0.3 < C < 0.7$	$0.4 < A < 0.7$	Dynamic stability, $V > 0.3$
Chaotic	$C < 0.2$	$A > 0.8$	Random fluctuation, $V \approx 0.15$

Table 8: Phase diagram from agent-based simulation (50 agents, 10 states, 100 steps).

P	Mean V_∞	Std V_∞
0.0	0.42	0.05
0.2	0.38	0.07
0.5	0.29	0.09
0.8	0.18	0.11
1.0	0.12	0.08

Table 9: Effect of perturbation on equilibrium coherence potential.

B.2.4 Optimal Zone Identification

For $n = 50$ agents, the simulation identifies:

$$C^* = 0.48 \pm 0.05 \tag{41}$$

$$A^* = 0.58 \pm 0.06 \tag{42}$$

$$V^* = 0.44 \pm 0.04 \tag{43}$$

The optimal constraint-autonomy ratio is approximately $C^*/A^* \approx 0.83$, suggesting slightly more autonomy than constraint is optimal for this system size.

C Practical Application Examples

This appendix provides detailed quantitative examples of applying the Constraint–Autonomy framework to real-world systems.

C.1 Example 1: Organizational Analysis

Consider a technology company with 200 employees.

Step 1: Define the action universe $\mathcal{U}$

Enumerate all possible actions employees could theoretically take:

- Communication: email anyone, schedule any meeting, contact any client
- Work: start any project, modify any code, change any document
- Resources: use any equipment, access any data, spend any budget
- Time: work any hours, take any breaks, work from any location

Estimated $|\mathcal{U}| \approx 10,000$ distinct action types.

Step 2: Identify the rule set $\mathcal{R}$

Company policies constrain actions:

- Approval hierarchies (spending limits, project authorization)
- Access controls (data permissions, system access)

- Scheduling constraints (core hours, meeting protocols)
- Communication protocols (client contact rules, confidentiality)

Estimated $|\mathcal{R}| \approx 500$ distinct rules.

Step 3: Compute metrics

$$\text{Constraint strength: } C = \frac{|\mathcal{R}|}{|\mathcal{U}|} = \frac{500}{10000} = 0.05 \tag{44}$$

$$\text{Autonomy: } A = \frac{|\mathcal{U}| - |\text{forbidden}|}{|\mathcal{U}|} \approx 0.72 \tag{45}$$

Step 4: Balance Index

$$\text{Balance Index} = 1 - \left|\frac{C - C^*}{C^*}\right| - \left|\frac{A - A^*}{A^*}\right| \tag{46}$$

With $C^* = 0.5$ and $A^* = 0.6$:

$$\text{Balance Index} = 1 - \left|\frac{0.05 - 0.5}{0.5}\right| - \left|\frac{0.72 - 0.6}{0.6}\right| = 1 - 0.9 - 0.2 = -0.1 \tag{47}$$

Diagnosis: Severe under-constraint ($C \ll C^*$). The organization likely experiences:

- Coordination failures (no clear protocols)
- Duplicated effort (no role boundaries)
- Decision paralysis (no clear authority)

Recommendation: Add essential constraints (clear role definitions, approval workflows, communication protocols) without reducing autonomy.

C.2 Example 2: Personal Habit System

An individual analyzing their daily routine.

Action universe: All possible activities in a day ($|\mathcal{U}| \approx 200$)

Rules: Self-imposed constraints (bedtime, exercise schedule, work hours, diet rules)

Person	$\|\mathcal{R}\|$	**C**	**A**	**Diagnosis**
Over-scheduled	150	0.75	0.25	Over-constrained (burnout risk)
Balanced	80	0.40	0.60	Near optimal
Under-structured	20	0.10	0.90	Under-constrained (drift risk)

Table 10: Personal habit system analysis for three hypothetical individuals.

C.3 Example 3: Political System Analysis

Comparing constitutional structures:

C.4 Calibration Guidelines

C.4.1 Estimating $|\mathcal{U}|$ Empirically

- **Organizations:** Enumerate job functions × possible actions per function
- **Individuals:** Time-diary studies, activity logging apps
- **Political systems:** Enumerate possible policy decisions × possible implementations

System	C	A	V	Balance	Status
Authoritarian	0.85	0.15	0.13	-0.45	Over-constrained
Failed state	0.10	0.90	0.09	-0.30	Under-constrained
Stable democracy	0.55	0.55	0.30	$+0.15$	Near optimal
Gridlocked democracy	0.70	0.40	0.28	-0.05	Slightly over-constrained

Table 11: Political system analysis using Constraint–Autonomy framework.

C.4.2 Estimating $|\mathcal{R}|$ Empirically

- **Formal rules:** Policy documents, legal codes, contracts
- **Informal rules:** Surveys, behavioral observation, norm cataloging

C.4.3 Reference Values

System Type	Typical C^*	Typical A^*
Startup (small)	0.30	0.70
Corporation (large)	0.55	0.50
Military unit	0.75	0.35
Research team	0.35	0.65
Democratic polity	0.50	0.55

Table 12: Reference optimal values by system type (based on simulation and case studies).

D Extended Translation Dictionary

This appendix provides a comprehensive translation dictionary across all five integrated frameworks.

CA Law	ECT	Unified Theory	Opt. Coherence	Asymmetry
Constraint $\mathcal{C}$	$1 - w(h)$	Perturbation P	Load-bearing C_E	Control intensity
Autonomy $\mathcal{A}$	$\lvert\mathcal{H}\rvert_{\text{eff}}$	Cooperation capacity ϕ	Discretion space	Adaptive variance V_t
$V(C, A)$	Accuracy A	Welfare W	Coherence potential	System health
$C \to C_{\max}$	$w(h^*) \to 0$	$P \gg 0$	Excess C_D	Totalization
$C \to 0$	No prior	No coordination	Missing C_E	Fragmentation
(C^*, A^*)	Calibrated w	$P = 0$ equilibrium	Optimal coherence	Balanced state
$\dot{V} < 0$	Accuracy loss	Welfare decline	Anomaly signal	Variance collapse
C_{int}	Internalized bias	Cultural P_{int}	Vestigial rules	Learned rigidity
Essential C_E	Likelihood structure	Shared environment	Load-bearing	Minimum control
Discretionary C_D	Prior penalization	Artificial scarcity	Removable rules	Excess control

Table 13: Extended translation dictionary across five frameworks.

D.1 Equation Cross-References

Concept	This Paper	Original Framework
Coherence potential	Eq. (??)	ECT Eq. (3)
Boundary conditions	Axioms ??–??	Asymmetry Thm. 1
Optimal zone	Theorem 4.4	Unified Theory Law I
Master equation	Eq. (26)	— (new synthesis)
Stability condition	Proposition ??	Opt. Coherence Prop. 2

Table 14: Cross-references between equations in this paper and original frameworks.

E Proofs of Corollaries

This appendix provides detailed proofs of corollaries stated in the main text.

E.1 Proof of Corollary 6.1 (Governance)

[Governance] Political systems require both rule of law (constraints) and individual rights (autonomy) for stability.

Proof. Let $\mathcal{S}$ be a political system with citizens as agents. Define:

- $\mathcal{U}$ = all possible citizen actions (speech, association, economic activity, etc.)
- $\mathcal{R}$ = legal constraints (laws, regulations, constitutional limits)
- $C = |\mathcal{R}|/|\mathcal{U}|$ = constraint strength
- $A = 1 - C$ = effective autonomy (rights)

By the main theorem:

1. If $C = 0$ (no rule of law): $\mathrm{Coh}(\mathcal{S}) = 0$, system cannot maintain coherent governance.
2. If $A = 0$ (no rights): $\mathrm{Adapt}(\mathcal{S}) = 0$, system cannot adapt to changing conditions.
3. Both $C > 0$ and $A > 0$ are necessary for $V > 0$.

Therefore, stable governance requires both rule of law ($C > 0$) and individual rights ($A > 0$). □ □

E.2 Proof of Corollary 6.4 (Evolutionary Stability)

[Evolution] Evolutionary systems require both genetic constraints (fidelity) and mutation (variation) for adaptation.

Proof. Let $\mathcal{S}$ be a population of replicating entities. Define:

- Constraint C = replication fidelity (probability of accurate copying)
- Autonomy A = mutation rate (probability of variation)

Note that $C + A \leq 1$ in this interpretation.
From Eigen's error threshold [Eigen, 1971]:

- If $C < C_{\text{threshold}}$: error catastrophe, loss of genetic information (under-constraint)

- If $A = 0$: no variation, inability to adapt to environmental change (over-constraint)

The viable region is:

$$\{(C, A) : C > C_{\text{threshold}}, A > 0, C + A \leq 1\} \tag{48}$$

This is precisely the interior region predicted by the main theorem. □ □

E.3 Proof of Theorem 12.4 (Integration Theorem)

[Integration] The five frameworks are equivalent formulations of the same structural principle.

Proof sketch. We establish pairwise equivalences:

(a) ⇔ (b): The ECT constraint function $w : \mathcal{H} \to [0, 1]$ defines a probability distribution over hypotheses. The effective constraint is $C = 1 - \mathbb{E}[w(h)]$ and effective autonomy is $A = |\{h : w(h) > 0\}|/|\mathcal{H}|$. The accuracy bound $\log w(h^*)$ corresponds to the coherence ceiling. When $w(h^*) \to 0$, accuracy is bounded regardless of evidence—equivalent to $A \to 0$ in CA Law.

(a) ⇔ (c): Essential constraints C_E are those satisfying $\partial V/\partial c|_{c \to 0} < 0$ (removal decreases coherence). These define the lower bound $C > 0$. Discretionary constraints C_D satisfy $\partial V/\partial c > 0$ (removal increases coherence). The optimal coherence condition C_E satisfied, C_D minimized is equivalent to (C^*, A^*).

(a) ⇔ (d): The Asymmetry theorem states $V_t \to 0$ under totalizing optimization. Totalizing "maximize control" drives $C \to C_{\max}$, $A \to 0$. Totalizing "maximize freedom" drives $C \to 0$, $A \to A_{\max}$. Both are boundary conditions in CA Law where $V \to 0$.

(a) ⇔ (e): Perturbation P in Unified Theory modifies payoff structure to favor defection. This is implemented by adding constraints that eliminate cooperative options, i.e., reducing A below A^*. Removing P restores autonomy, enabling cooperative equilibrium. The cooperative default at $P = 0$ corresponds to the optimal zone (C^*, A^*).

By transitivity, all five frameworks are equivalent. □ □

F Glossary of Terms and Notation

$\mathcal{A}$	Autonomy; degrees of freedom available to agents within the rule structure
$\mathcal{C}$	Constraint strength; proportion of possible actions restricted by rules
C_D	Discretionary constraints; removable rules not essential for coherence
C_E	Essential constraints; load-bearing rules necessary for coherence
$C_{\textbf{int}}$	Internalized constraints; behaviorally persistent rules after formal removal
$C_{\max}$	Maximum constraint; boundary at which autonomy vanishes
$f(C, A)$	Natural coherence growth function; positive in optimal zone
$g(P, C_D)$	Perturbation drag; coherence loss from external perturbation and over-constraint
$h(C_{\textbf{int}})$	Cultural lag; coherence loss from internalized constraints
$\mathcal{H}$	Hypothesis space (in ECT); set of possible beliefs
P	Perturbation; external intervention maintaining antagonism
$P_{\textbf{int}}$	Internalized perturbation; culturally transmitted antagonistic patterns
$\mathcal{R}$	Rule set; constraints governing agent transitions

$\mathcal{S}$	System; tuple of agents, states, rules
τ_1	Behavioral adaptation timescale; fast
τ_2	Cultural decay timescale; slow (generational)
$\mathcal{U}$	Action universe; set of all possible agent actions
V	Coherence potential; product of coherence and adaptability
V^*	Optimal coherence; maximum achievable V at (C^*, A^*)
$w(h)$	Soft constraint function (in ECT); admissibility weight on hypothesis h
ϕ	Cooperative response function (in Unified Theory)
ω	System state; configuration of all agents

References

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

Ashby, W.R. (1960). *Design for a Brain: The Origin of Adaptive Behaviour*. Chapman & Hall, London, 2nd edition.

Arrow, K.J. (1951). *Social Choice and Individual Values*. Wiley, New York.

Conant, R.C. and Ashby, W.R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, 1(2):89–97.

Eigen, M. (1971). Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York.

Kauffman, S.A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York.

Kriger, B. (2017). A Unified Theory of Self-Organizing Systems: Four Formal Laws on Cooperation, Viability, Interference, and Observability. *Zenodo*. https://doi.org/10.5281/zenodo.18363729

Kriger, B. (2021). Epistemic Constraint Theory: A Unifying Framework for Inference Limitations Across Bayesian Epistemology, Information Theory, and Decision Theory. *Zenodo*. https://doi.org/10.5281/zenodo.18365738

Kriger, B. (2026). The Principle of Optimal Coherence. *Zenodo*. https://doi.org/10.5281/zenodo.18341030

Kriger, B. (2026). The Asymmetry of Totalizing Ideals: A Structural Law of Complex Adaptive Systems. *Zenodo*. https://doi.org/10.5281/zenodo.18361828

Langton, C.G. (1990). Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D: Nonlinear Phenomena*, 42(1-3):12–37.

Maturana, H.R. and Varela, F.J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel Publishing Company, Dordrecht.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge University Press, Cambridge.

Simon, H.A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1):99–118.

Simon, H.A. (1996). *The Sciences of the Artificial.* MIT Press, Cambridge, MA, 3rd edition.

Varela, F.J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience.* MIT Press, Cambridge, MA.

Axelrod, R. (1984). *The Evolution of Cooperation.* Basic Books, New York.

Nowak, M.A. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563.

Mitchell, M. (2009). *Complexity: A Guided Tour.* Oxford University Press, New York.

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control.* Viking, New York.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané, D. (2016). Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565.*

Hayek, F.A. (1945). The use of knowledge in society. *American Economic Review*, 35(4):519–530.

North, D.C. (1990). *Institutions, Institutional Change and Economic Performance.* Cambridge University Press, Cambridge.

Wolfram, S. (2002). *A New Kind of Science.* Wolfram Media, Champaign, IL.

Holland, J.H. (1995). *Hidden Order: How Adaptation Builds Complexity.* Addison-Wesley, Reading, MA.

Prigogine, I. and Stengers, I. (1984). *Order Out of Chaos: Man's New Dialogue with Nature.* Bantam Books, New York.

Aubin, J.-P. (1991). *Viability Theory.* Birkhäuser, Boston.

Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859):1243–1248.