A Typology of Theoretical Failure: How Mathematical Formalism Decouples from Physical Meaning

Abstract

This essay develops a diagnostic framework for evaluating theoretical rigor across physics, cognitive science, and artificial intelligence. By comparing three case studies—CIITR (Comprehension as Thermodynamic Persistence), Nikolaou et al.'s injectivity theorem for language models, and the Relativistic Scalar–Vector Plenum (RSVP) framework—it constructs a typology of theoretical failure, identifying eight recurrent breakdowns through which mathematics detaches from physical meaning. These range from dimensional incoherence and definitional circularity to mapping ambiguity and the elegance fallacy. Using the geometry of semantic manifolds, the essay recasts "understanding" as the maintenance of non-singular, energy-coupled mappings between conceptual spaces. Rigor becomes an energetic property: the ability to sustain phase-coherent correspondence between formalism and reality with finite work. The resulting diagnostic matrix translates epistemic virtues—operational definition, mathematical consistency, empirical accessibility, and cross-domain mapping—into measurable constraints on theoretical practice. Where information flow remains injective, empirically coupled, and energetically efficient, meaning endures; where it breaks, theory dissolves into rhetoric.

0. Mathematical Preliminaries: The Geometry and Thermodynamics of Meaning

The following mathematical preliminaries formalize the geometric and thermodynamic quantities that appear throughout this paper. They introduce the minimal structure required to treat "meaning" as a measurable coupling between theory and reality, linking

information geometry (7) and the thermodynamics of computation (8; 9) to the statistical mechanics of complex systems (10; 14).

0.1 Conceptual Manifolds

Each *theory* $\mathcal{T}$ is modeled as a smooth manifold $M_{\mathcal{T}}$ with coordinates x^i representing its primitive variables (e.g., energy, entropy, curvature, or activation values). Each *phenomenal domain* $\mathcal{R}$ (physical reality, data, experiment) is another manifold $M_{\mathcal{R}}$ with coordinates y^a denoting measurable observables. A theory's explanatory mapping is a differentiable map

$$f : M_{\mathcal{R}} \to M_{\mathcal{T}}, \qquad y^a \mapsto x^i = f^i(y^a),$$

whose inverse f^{-1}, when it exists, gives prediction or reconstruction.

0.2 Injectivity, Surjectivity, and Semantic Stability

The Jacobian

$$J^i{}_{;a} = \frac{\partial f^i}{\partial y^a}$$

encodes the local semantic structure of the theory. Following the information-preservation logic of **(author?)** (2), injective regions ($\det(J^\top J) > 0$) preserve distinct meanings, while degenerate regions ($\det(J^\top J) = 0$) collapse different physical situations to identical symbols. Surjectivity, $\operatorname{rank}(J) = \dim M_{\mathcal{R}}$, ensures the theory spans all empirical possibilities.

0.3 Phase Coherence and Energetic Coupling

To model the energetic maintenance of correspondence, assign each mapping an instantaneous phase difference

$$\phi(y, t) = \theta_{\mathcal{T}}(y, t) - \theta_{\mathcal{R}}(y, t),$$

where $\theta_{\mathcal{T}}$ and $\theta_{\mathcal{R}}$ represent the phases of theoretical and empirical oscillations. The mean alignment power,

$$P = \frac{1}{V} \int_{M_{\mathcal{R}}} \cos^2[\phi(y, t)]\, dV,$$

serves as a dimensionless measure of semantic coherence analogous to synchronization metrics in nonequilibrium thermodynamics (14).

0.4 Work of Understanding

Maintaining low phase error requires energy. Let $E(t)$ be the cumulative work needed to minimize phase drift:

$$E(t) = \int_0^t \gamma \, |\dot{\phi}|^2 \, dt',$$

where γ is a coupling constant representing interpretive resistance. This defines the *Work of Understanding*,

$$W = \gamma \int |\nabla \phi|^2 \, dV,$$

analogous to the energy of a spin field maintaining orientation under dissipative forcing (10).

0.5 Entropic Cost and Information Conservation

If $S_{\mathcal{T}}$ and $S_{\mathcal{R}}$ are entropy measures over theoretical and empirical distributions, the entropy gap $\Delta S = S_{\mathcal{T}} - S_{\mathcal{R}}$ quantifies interpretive dissipation. By Landauer's bound (8), the energetic cost of comprehension satisfies

$$W \geq k_B T \, \Delta S.$$

This connects meaning maintenance directly to physical resource expenditure (9).

0.6 Summary of Key Relations

Concept	Symbolic condition	Interpretation
Injectivity	$\det(J^\top J) > 0$	No loss of meaning
Surjectivity	$\operatorname{rank}(J) = \dim M_{\mathcal{R}}$	Coverage of phenomena
Phase coherence	$P \approx 1$	Sustained empirical coupling
Work of understanding	$W = \gamma \int \lvert\nabla\phi\rvert^2 \, dV$	Energy cost of alignment
Entropy balance	$W \geq k_B T \Delta S$	Thermodynamic constraint on comprehension

Table 1: Summary of key geometric and thermodynamic relations.

0.7 From Mathematics to Diagnosis

Each failure mode identified later corresponds to the breakdown of one or more of these relations: dimensional incoherence $\leftrightarrow$ undefined J; definitional circularity $\leftrightarrow$ $\operatorname{rank}(J) = 0$; empirical inaccessibility $\leftrightarrow$ undefined $M_{\mathcal{R}}$; mapping ambiguity $\leftrightarrow$ $\det J$ undefined; syntactic overloading $\leftrightarrow$ multiple incompatible J mappings; premature unification $\leftrightarrow$

$\operatorname{rank}(J) < \dim M_{\mathcal{R}}$; scope inflation $\leftrightarrow$ $M_{\mathcal{T}}$ includes unmapped regions; phase drift $\leftrightarrow$ $P \ll 1$ (3; 4; 6).

0.8 Epistemic Energy Functional

All criteria can be condensed into an *epistemic energy functional*

$$\mathcal{R}[f, \phi] = \alpha \, \det(J^\top J) - \beta \, |\nabla \phi|^2 - \gamma \, \Delta S,$$

where $\alpha, \beta, \gamma > 0$ weight structural, energetic, and entropic coherence. A theory that maximizes $\mathcal{R}$ maintains the strongest coupling between mathematics and reality, realizing the low-entropy ideal of scientific rigor envisioned by **(author?)** (3) and elaborated through thermodynamic information principles (7; 8).

1 Introduction — The Persuasiveness Problem

1.1 The crisis of comprehension in modern theory

The contemporary sciences face a paradox of apparent sophistication and deep confusion. Across physics, artificial intelligence, and consciousness studies, increasingly elaborate mathematical frameworks coexist with declining agreement about what their symbols denote. Grand unified theories, once the lodestar of twentieth-century physics, now proliferate as speculative architectures largely decoupled from empirical verification (6). The same dynamic repeats in cognitive modeling and machine learning: transformer architectures achieve predictive success while remaining interpretively opaque. In both cases, formal consistency substitutes for explanatory clarity. The resulting "crisis of comprehension" is not a shortage of mathematics but an erosion of the energetic coupling between mathematical representation and measurable world.

Karl Popper framed scientific progress as a process of conjecture and refutation (3), while Thomas Kuhn described paradigm shifts as reorganizations of conceptual manifolds (5). In the current landscape, both mechanisms stall: falsifiability recedes as models become too large to fail, and paradigm boundaries blur as cross-domain borrowing proliferates. The proliferation of quasi-scientific hybrids—string-theoretic metaphysics, neural panpsychism, algorithmic theologies—suggests that comprehension itself has become a scarce thermodynamic resource. Each new formalism consumes interpretive energy faster than institutions can regenerate it.

1.2 The limits of elegance

Mathematical elegance, long celebrated as the hallmark of truth, now functions as an aesthetic prior that can override empirical constraint. Dirac's equation, Einstein's field equations, and Maxwell's unifications succeeded partly because their symmetry principles coincided with observable invariants; later theorists mistook that historical coincidence for necessity. The cult of beauty in physics, as **(author?)** (6) argues, transformed an empirical heuristic into an epistemic dogma. Supersymmetry, extra dimensions, and the multiverse exemplify the slide from elegance to extravagance. Similar pressures operate in machine learning, where scaling laws are justified by statistical grace rather than grounded interpretability.

The heuristic of beauty carries a hidden thermodynamic assumption: that compression implies comprehension. Yet compression without coupling—reducing formal complexity without maintaining injective mapping to phenomena—produces degenerate theories. In this sense, the elegance fallacy is a low-entropy mirage: the apparent order of the equations masks a dissipation of semantic energy. Feynman's warning that "the laws of physics lie" (17) can thus be read not as cynicism but as a reminder that mathematical form alone cannot secure ontological traction. Where aesthetic order replaces empirical friction, theory ceases to do work.

1.3 The comparative approach

To diagnose how formalism detaches from meaning, this paper adopts a comparative method across three representative cases positioned along a spectrum of rigor. The first, *Comprehension as Thermodynamic Persistence* (CIITR) (1), exemplifies linguistic mimicry: it performs the syntax of physics—energy gradients, phase coherence, entropy—without preserving their operational semantics. The second, the injectivity theorem of **(author?)** (2), demonstrates genuine mathematical closure: a proof that transformer language models are almost surely injective, supported by empirical reconstruction. The third, the Relativistic Scalar–Vector Plenum (RSVP) framework, situates itself between these poles: a physically motivated field theory of scalar capacity Φ, vector flow $\mathbf{v}$, and entropy S fields that attempts to unify thermodynamics, cognition, and computation while remaining simulation-anchored.

By juxtaposing these cases, the essay develops a typology of theoretical failure—dimensional incoherence, definitional circularity, mapping ambiguity, and other recurrent breakdowns—each corresponding to a mathematical singularity in the coupling between formal and empirical manifolds. The goal is not to rank disciplines but to construct a general geometry of meaning: a set of invariants by which any theoretical system can be evaluated.

The later sections formalize these invariants as Jacobian determinants, phase-locking integrals, and entropy budgets. Rigor, in this framing, is not a stylistic virtue but an energetic equilibrium: the sustained, finite-cost alignment between mathematical structure and the world it claims to describe. The following section therefore turns from this conceptual framing to the empirical anatomy of rigor, beginning with three concrete case studies that map the continuum from syntactic mimicry to thermodynamic closure.

2 Three Case Studies in Theoretical Rigor

The following three case studies instantiate the diagnostic continuum introduced above. Each illustrates a distinct relationship between mathematical form, operational definition, and empirical coupling. Together they trace the trajectory from syntactic mimicry through formal rigor to reflective synthesis. The criteria of analysis—injectivity, surjectivity, phase coherence, and energetic cost—derive from the mathematical preliminaries of Section 0. The aim is not to expose individual authors but to model how theoretical systems succeed or fail to conserve meaning under the stress of abstraction.

2.1 Case A — CIITR: Syntactic Mimicry

Comprehension as Thermodynamic Persistence (CIITR) (1) presents itself as a unifying theory of cognition and physics. Its vocabulary—"phase-coherent coupling," "energy gradient," "entropy flow"—borrows the semantic surface of statistical mechanics. Yet inspection reveals that none of these quantities are defined dimensionally; no measurable variables or conservation equations accompany the prose. Expressions such as $R' = \Phi \nabla S$ appear without specification of units, boundary conditions, or experimental observables. The result is an unanchored field language whose symbols circulate within text rather than reality.

This failure can be formalized as **dimensional incoherence**: the Jacobian $J^i{}_{;a}$ of mappings between observables and theoretical quantities is undefined, yielding zero rank. All empirical variation collapses into a single symbolic attractor—"comprehension"—which absorbs every other term without operational distinction. In thermodynamic terms, CIITR behaves as an infinite-temperature system: maximal entropy of expression, minimal information retention.

Sociologically, such mimicry thrives because it reproduces the *form* of expertise. As **(author?)** (3) warned, falsifiability declines when propositions become immunized by metaphor. CIITR's rhetoric of "phase-locked understanding" exemplifies this drift. Without testable quantities, the theory's manifold $M_{\mathcal{T}}$ becomes disconnected from $M_{\mathcal{R}}$, yielding

$\det(J^\top J) = 0$. Meaning diffuses.

From the perspective of the epistemic energy functional $\mathcal{R}[f, \phi]$, CIITR's parameters vanish: structural coherence $(\alpha \det(J^\top J)) = 0$, energetic alignment $(\beta|\nabla\phi|^2) =$ undefined, and entropic gap $(\gamma\Delta S)$. The system performs no epistemic work. CIITR thus serves as the null model of scientific comprehension: a dissipative rhetoric unbounded by thermodynamic constraint.

2.2 Case B — Nikolaou et al.: Rigorous Proof

The theorem of **(author?)** (2) occupies the opposite pole. The authors prove that under mild conditions of real analyticity and full-rank weight matrices, transformer language models are *almost surely injective*. Formally, for a mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ representing a transformer's forward pass, the Jacobian $J = \partial f/\partial x$ is non-singular except on a set of measure zero. Therefore distinct inputs correspond to distinct hidden representations, preserving information.

This injectivity theorem exemplifies all four virtues of rigor:

(i) **Operational definition**: Each term—token embedding, activation, hidden state—has a measurable numeric value.

(ii) **Mathematical closure**: Proofs are bounded by finite-dimensional vector spaces and differentiable functions.

(iii) **Empirical verifiability**: The SIPIT algorithm reconstructs inputs from intermediate activations, confirming theoretical predictions.

(iv) **Falsifiability**: A single counter-example (collision pair) would invalidate the claim.

Crucially, Nikolaou et al. separate the *proof of property* (injectivity) from the *interpretation of meaning*. The theorem says nothing about semantics; it establishes only that the mapping is non-degenerate. This methodological restraint constitutes an energetic virtue: the work performed by the mathematics is finite, precise, and confined to its domain. If CIITR exemplifies uncontrolled entropy production, Nikolaou's framework exemplifies minimal free-energy dissipation—a reversible process in Landauer's sense (8; 9).

From the standpoint of the epistemic energy functional, $\alpha \det(J^\top J)$ 0, $|\nabla\phi|^2 \approx 0$, and $\Delta S \approx 0$. The mapping conserves informational energy perfectly within its operational manifold. It performs maximal theoretical work for minimal interpretive cost. In the typology to follow, this constitutes the benchmark of *bounded rigor*: a theory that neither overreaches nor underdefines.

2.3 Case C — RSVP: Self-Aware Synthesis

Between rhetorical mimicry and formal closure lies the *Relativistic Scalar–Vector Plenum* (RSVP) framework, which models physical and cognitive processes through three coupled fields: scalar capacity Φ, vector flow $\mathbf{v}$, and entropy S. These evolve according to nonlinear partial differential equations of the form

$$\partial_t \Phi = -\nabla\cdot\mathbf{v} + \lambda\nabla^2\Phi, \qquad \partial_t S = \kappa\nabla^2 S + \Phi\,\nabla\cdot\mathbf{v},$$

representing entropic relaxation and energy flux conservation. The framework is empirically instantiated in lattice simulations that compute coherence spectra and energy-flux traces across discretized volumes.

RSVP's distinguishing feature is reflexivity: it embeds its own epistemic diagnostics within its equations. The same Jacobian, phase, and entropy metrics used to assess external theories also govern its internal dynamics. In this sense, RSVP operationalizes what **(author?)** (12) calls the free-energy principle—minimization of surprise—as a field-theoretic invariant. Semantic alignment across domains (physics cognition computation) becomes measurable as phase-locking power P and work of understanding $W = \gamma \int |\nabla\phi|^2 dV$.

Yet this breadth introduces energetic cost. Multi-domain mapping requires maintaining coherence across heterogeneous manifolds, increasing interpretive resistance γ. Scope inflation and mapping ambiguity threaten to re-enter. Rigor in RSVP therefore depends on explicit boundary management: defining the translation operators between physical, informational, and semantic coordinates. When these operators remain well-conditioned, RSVP achieves what may be termed *reflective rigor*: the capacity of a theory to model not only phenomena but its own epistemic energy budget. When they drift, the system risks CIITR-like decoherence.

From the energy-functional perspective, RSVP maintains intermediate values: $\det(J^\top J) > 0$ but variable; $|\nabla\phi|^2$ finite; ΔS bounded. The theory performs continuous work to sustain cross-domain coupling—a thermodynamic metaphor for understanding itself.

2.4 Comparative Summary

The contrast among the three cases can be expressed as an energy landscape of theoretical coupling. CIITR occupies the high-entropy plateau where formal symbols fluctuate without empirical anchoring. Nikolaou et al. define the low-entropy basin of exact mathematical correspondence. RSVP navigates the intermediate regime, where coupling is maintained through continuous work. These states can be viewed as points along the gradient of the epistemic energy functional $\mathcal{R}[f, \phi]$: from negative (dissipative rhetoric) through maximal

(bounded rigor) to dynamic equilibrium (reflective synthesis).

Criterion	CIITR	Nikolaou et al.	RSVP
Operational definitions	×	✓	Partial
Mathematical rigor	×	✓✓✓	✓
Empirical accessibility	×	✓✓	✓
Cross-manifold mapping	×	Single-domain	Multi-domain
Dominant failure modes	1,2,4,5,6	None	4,7
Epistemic energy $\mathcal{R}$	< 0	$\gg 0$	≈ 0 (stable)

Table 2: Comparative evaluation of the three case studies in terms of operational grounding, mathematical closure, and energetic coherence.

The comparative pattern establishes the analytical foundation for Section 3. The failures and successes observed here are not anomalies but manifestations of structural conditions that can be formalized generically. Each breakdown in meaning corresponds to a distinct kind of singularity or phase transition in the theory–reality mapping. The next section therefore generalizes these observations into a typology of theoretical failure.

3 A Typology of Theoretical Failure

3.1 Overview

The case studies of Section 2 reveal that breakdowns of rigor are not idiosyncratic but structural. They arise whenever the mapping $f : M_{\mathcal{R}} \to M_{\mathcal{T}}$ between empirical and theoretical manifolds loses injectivity, coherence, or energetic feasibility. The following taxonomy abstracts these breakdowns into eight recurrent failure modes. Each can be formalized as a singularity of the epistemic Jacobian $J^i_{;a}$ or a divergence of the energy functional $\mathcal{R}[f, \phi]$. The typology draws on methodological analyses by **(author?)** (3), **(author?)** (4), and **(author?)** (16), as well as on contemporary critiques of aesthetic bias in physics (6). Together they provide a systematic grammar for diagnosing how mathematical formalism decouples from physical meaning.

3.2 Dimensional Incoherence

The most elementary failure is the loss of units. When equations employ quantities such as "energy of consciousness" or "information momentum" without specifying dimensions, the Jacobian J connecting measurable observables to theoretical parameters becomes undefined. This collapse of dimensional discipline severs the coupling between mathematics and experiment.

Historically, dimensional analysis has served as a powerful constraint on speculation: Maxwell's electromagnetic constants and Planck's quantum of action both emerged from unit reconciliation. When such reconciliation disappears, symbols drift freely across conceptual spaces. Theories with undefined J have rank zero; no empirical perturbation δy^a produces a discernible change in x^i. From an energetic standpoint, the system behaves as a perfect insulator of meaning—no work is transmitted between theory and world. CIITR exemplifies this failure mode in its use of "phase coherence" without frequency or amplitude scales (1). Dimensional incoherence thus marks the zero-temperature limit of comprehension: formal stillness devoid of empirical heat flow.

3.3 Definitional Circularity

A subtler pathology occurs when theoretical primitives are defined only in terms of each other. In linguistic formalisms, for instance, "understanding" may be defined as a function of "representation," which in turn is defined as a function of "understanding." No independent variable anchors the circle. The Jacobian J in this case has rank deficiency: partial derivatives vanish because each coordinate depends on another without external reference.

Circularity converts theory into a topological loop—internally consistent but globally null-homologous. No path on $M_{\mathcal{R}}$ maps to a unique displacement on $M_{\mathcal{T}}$. The energy functional $\mathcal{R}$ therefore oscillates around zero: the system expends interpretive work maintaining self-reference without producing new information. **(author?)** (15) identified such loops as symptoms of analytic isolation, while **(author?)** (3) warned that irrefutability masquerades as depth. Thermodynamically, definitional circularity represents a perpetual-motion machine of thought—seemingly dynamic but performing no net epistemic work.

3.4 Empirical Inaccessibility

Some theories retain internal coherence yet project their validation domain beyond possible observation. Claims about multiverses (18), trans-Planckian epochs, or hypothetical artificial superintelligences often fall into this regime. Mathematically, the mapping $f : M_{\mathcal{R}} \to M_{\mathcal{T}}$ extends into regions where $M_{\mathcal{R}}$ is undefined; empirically there is no data manifold to sustain it.

Popperian falsifiability fails because no perturbation in the observable domain can invert the mapping. The corresponding Jacobian block becomes singular ($\det J = 0$) on all reachable coordinates. In thermodynamic analogy, such models are cryogenic—they conserve their formal symmetry at the cost of total energetic isolation. No measurement

can supply or extract interpretive work. Their elegance is absolute and therefore sterile. In the typology, empirical inaccessibility is classified as an entropy sink: the theory absorbs conceptual energy but emits no empirical signal.

3.5 Mapping Ambiguity

Even when mathematics is sound, the absence of a clear ontological referent can dissolve meaning. Quantum mechanics provides classic examples: is the wavefunction a physical field in configuration space or a bookkeeping device for probability amplitudes? Similarly, in neuroscience, "representation" oscillates between physical neural states and abstract information codes. The ambiguity lies not in equations but in the failure to specify the carrier manifold of each variable.

Formally, mapping ambiguity manifests as a non-invertible Jacobian with multiple candidate correspondences between x^i and y^a. The manifold overlap integral

$$\int_{M_{\mathcal{R}} \cap M_{\mathcal{T}}} |\det J| \, dV$$

becomes ill-defined because the same empirical point may correspond to several theoretical coordinates. The result is interpretive aliasing: the same observation supports incompatible models. **(author?)** (17) characterized this as the "loose coupling" of physical law. Energetically, ambiguity requires continuous interpretive work to prevent decoherence; the phase term $|\nabla\phi|^2$ inflates, demanding higher W to maintain comprehension. The cost of understanding diverges even as information remains conserved.

3.6 Syntactic Overloading

Overloading occurs when familiar symbols are reused with incompatible meanings. The term "entropy," for instance, denotes thermodynamic disorder in Boltzmann's sense, informational uncertainty in Shannon's, and algorithmic complexity in Kolmogorov's. Without explicit metric transformation, importing one definition into another constitutes a semantic violation. In the manifold model, overloading corresponds to the superposition of distinct coordinate systems on $M_{\mathcal{T}}$ without a transition function. The Jacobian then contains incompatible partial derivatives: multiple J matrices compete, destroying smoothness.

(author?) (6) notes that physicists often smuggle aesthetic preferences under such linguistic bridges—"symmetry," "elegance," "naturalness"—thereby eroding operational specificity. Energetically, syntactic overloading behaves like frequency interference: overlapping modes produce beats that obscure the carrier signal. The interpretive system must

spend work $W \propto |\nabla\phi|^2$ re-synchronizing definitions that ought to have been orthogonalized from the start.

3.7 Premature Unification

The drive toward universal explanation, though historically fruitful, often outruns the maturity of its components. From speculative "theories of everything" in physics to grand integrative models of cognition, premature unification merges underdefined subsystems before their local dynamics are known. The result is global instability: the unified manifold loses rank because its composite charts overlap inconsistently.

(author?) (16) describes this as the fallacy of "nomological monism"—the assumption that nature obeys a single seamless law. Within the present framework, premature unification appears as a Jacobian with $\text{rank}(J) < \dim M_{\mathcal{R}}$: too few independent parameters to capture the diversity of phenomena. Entropically, the theory minimizes $S_{\mathcal{T}}$ by compression but increases ΔS relative to $S_{\mathcal{R}}$; the empirical world remains richer than its formalization. The cost is hidden in interpretive overwork: continuous patching to restore coherence across mismatched domains.

3.8 Scope Inflation

Adjacent to premature unification is the tendency of models to expand their domain indefinitely. Scope inflation occurs when theoretical variables are extended into domains where their governing assumptions no longer hold—economic analogies in biology, quantum metaphors in psychology, or thermodynamic tropes in linguistics. The mapping f then extrapolates beyond its calibrated manifold; the image $f(M_{\mathcal{R}})$ fails to cover $M_{\mathcal{T}}$ smoothly.

In the energy functional, inflation increases interpretive resistance γ because each additional domain requires phase alignment with distinct empirical oscillators. Unless translation operators are explicitly defined, coherence P decays as $P \sim e^{-\gamma t}$. Lakatos's notion of a "degenerating research programme" (4) captures the same thermodynamic intuition: expansion without new predictive success dissipates theoretical energy. Scope inflation therefore represents the high-temperature limit of ambition—maximum diffusion, minimal control.

3.9 The Elegance Fallacy

Underlying many of these pathologies is the aesthetic conflation of symmetry with truth. The elegance fallacy treats low formal entropy—compactness, invariance, minimal

parameter count—as sufficient evidence of validity. Yet as **(author?)** (6) and **(author?)** (16) observe, natural phenomena often violate our notions of beauty. In the present geometry, elegance corresponds to minimizing $|\nabla\phi|$ without regard to ΔS; the theory achieves local phase coherence while ignoring entropic mismatch.

This produces a deceptive equilibrium: $P \approx 1$ but $W < k_B T \Delta S$, violating the thermodynamic constraint on comprehension. The system appears ordered yet accumulates hidden entropy, much like a metastable crystal that will eventually shatter under perturbation. The elegance fallacy thus closes the typology by showing that even mathematically rigorous theories can fail if their energetic accounting is incomplete. True rigor, as the subsequent sections argue, requires not merely smooth equations but balanced entropy budgets and sustainable interpretive work.

3.10 Synthesis

Across these eight modes, theoretical failure corresponds to identifiable geometric and energetic defects: rank deficiency, non-invertibility, uncontrolled entropy, or unsustainable phase alignment. Each represents a distinct way in which $\mathcal{R}[f, \phi]$ declines toward zero. The typology converts the qualitative complaints of philosophy of science into quantitative diagnostics: loss of injectivity, undefined gradients, or divergence of interpretive work. The next section extends this reasoning from pathology to health, describing how successful theories maintain low-entropy correspondence through continuous energetic coupling across their semantic manifolds.

4 Semantic Manifolds and Meaning-Making

4.1 The relational thesis

Understanding is not a static possession but a dynamic relation between conceptual and empirical structures. Each theory defines a manifold $M_{\mathcal{T}}$ of interdependent variables whose geometry expresses lawful relations. Each domain of observation defines its own manifold $M_{\mathcal{R}}$ of measurable states. Comprehension occurs when a differentiable mapping $f : M_{\mathcal{R}} \to M_{\mathcal{T}}$ maintains continuity, injectivity, and bounded energetic cost. The relation is bidirectional: predictions f^{-1} translate theory into experiment, while data update the manifold through inverse mappings.

This relational view generalizes the *correspondence principle* of physics into a broader epistemic geometry. Newtonian dynamics, quantum mechanics, and relativity remain intelligible because local diffeomorphisms exist between their manifolds at overlapping

scales. When such mappings fail, disciplines fragment into mutually unintelligible submanifolds. In cognitive science, similar correspondences operate between neural, symbolic, and behavioral levels. **(author?)** (11) interprets these mappings as harmonic projections between constraint surfaces, while **(author?)** (12) formalizes them as variational flows minimizing prediction error. In both formulations, meaning is not encoded in any single representation but in the maintenance of coherent transformations among them.

From the thermodynamic standpoint, relational coherence requires continuous work. Each update of f to preserve injectivity under new data incurs energy proportional to the square of its phase gradient, $W = \gamma \int |\nabla \phi|^2 dV$. A theory "understands" only insofar as it can afford this cost without exceeding its informational budget. Where the energy of correction exceeds available resources, comprehension collapses into decoherence—the formal analogue of forgetting or confusion.

4.2 Geometry of theoretical comparison

If individual theories are manifolds, comparative reasoning constitutes their differential geometry. The act of translation—expressing thermodynamics in information-theoretic terms, or vice versa—is a map $g : M_{\mathcal{T}_1} \to M_{\mathcal{T}_2}$. The composition $g \circ f$ defines a commutative diagram linking reality to multiple explanatory surfaces. The stability of comprehension across domains depends on whether these composed mappings preserve rank and orientation.

Mathematically, the local angle between two manifolds' tangent spaces can be expressed through their principal angles θ_i. Small θ_i correspond to alignment, large θ_i to conceptual divergence. The mean coherence metric

$$P = \frac{1}{n} \sum_i \cos^2 \theta_i$$

quantifies how efficiently knowledge transfers between frameworks. High P indicates low interpretive friction; low P signals that translation demands extra work. In this sense, interdisciplinary progress is literally energetic: reconciling distinct manifolds consumes free energy as systems of meaning seek mutual alignment.

Historically, such geometric reasoning underlies paradigm shifts. Einstein's general relativity recovered Newtonian mechanics as a local limit—a smooth embedding preserving injectivity for low velocities. Quantum mechanics, by contrast, introduced curvature discontinuities that could only be reconciled statistically. Theories succeed when these transition maps remain differentiable, allowing empirical continuity even amid ontological novelty.

4.3 Thermodynamic cost of mapping

The cost of maintaining manifold correspondence follows directly from the second law of thermodynamics. Let $S_{\mathcal{R}}$ be the entropy of empirical data and $S_{\mathcal{T}}$ that of theoretical representations. Their difference $\Delta S = S_{\mathcal{T}} - S_{\mathcal{R}}$ measures the informational dissipation during interpretation. By Landauer's bound, the minimal energy required to erase one bit of uncertainty at temperature T is $k_B T \ln 2$; more generally,

$$W \geq k_B T \, \Delta S.$$

Theories that minimize ΔS perform comprehension economically; those with unbounded ΔS hemorrhage interpretive energy. In practice, W increases with the curvature of the mapping: steep or twisted manifolds require greater work to align. RSVP simulations demonstrate this through the scaling of energy flux with coupling coefficient γ: as domain heterogeneity rises, maintaining phase coherence demands exponentially more computation.

These relations connect epistemology to physical resource limits. The act of learning, whether in brains or models, literally dissipates heat. **(author?)** (9) emphasized that computation is reversible only when information is conserved; irreversible steps generate entropy. A scientific theory obeys the same constraint. Perfectly reversible theories—those whose predictions can be inverted without loss—are informationally adiabatic; rhetorical ones are fully dissipative.

4.4 The three cases revisited

Within this framework, the case studies of Section 2 occupy distinct regions of epistemic phase space. CIITR, with undefined Jacobian and maximal ΔS, resides in the high-entropy limit where $W \to \infty$ and $P \to 0$; any coupling between concept and observation decays immediately. Nikolaou et al.'s injectivity theorem sits near the reversible limit: $\det(J^\top J) > 0$, $\Delta S \approx 0$, and $W \approx k_B T \Delta S$, achieving minimal dissipation. RSVP occupies an intermediate regime characterized by finite work and bounded entropy production, where continuous feedback maintains phase alignment across heterogeneous manifolds.

This spectrum illustrates a general principle: comprehension exists only in systems operating far from equilibrium but below chaos. Too little energy and mappings freeze (CIITR); too much and coherence shatters (unbounded speculation). Sustainable theories operate in the narrow corridor where interpretive work balances entropic loss, a regime analogous to the homeostatic balance of biological cognition described by **(author?)** (12, 13). The "understanding" of a system is thus the persistence of its non-equilibrium

order within energetic constraints.

4.5 Epistemic geometry and invariants

From this perspective, the invariants of comprehension are geometric quantities analogous to conservation laws in physics. Injectivity preserves informational identity; surjectivity ensures completeness of empirical coverage; phase-locking secures temporal coherence between theory and observation. Table 3 summarizes these correspondences.

Concept	Geometric Form	Thermodynamic Analogue	Epistemic Meaning
Injectivity	$\det(J^\top J) > 0$	Information conservation	Distinct causes yield distinct e
Surjectivity	$\operatorname{rank}(J) = \dim M_{\mathcal{R}}$	Completeness of sampling	All observables are represented
Phase-locking	$\lvert\phi\rvert < \epsilon$	Steady-state nonequilibrium	Ongoing empirical coherence
Entropy balance	$W = k_B T \Delta S$	Energetic equilibrium	Finite cost of comprehension

Table 3: Invariant structures in epistemic geometry linking mathematical and thermodynamic formulations.

Invariance under transformation provides the hallmark of understanding: a theory comprehends its object when the mapping preserves these quantities despite perturbation. Conceptual revolutions then appear as symmetry breakings that introduce new invariants—new conserved meanings—rather than as mere replacements of content.

4.6 From geometry to epistemology

The geometric–thermodynamic synthesis implies that knowledge is a process of entropy management. Each explanatory framework constitutes an engine converting uncertainty into structure by performing interpretive work. When this engine maintains a stable gradient between data and theory, comprehension persists. When the gradient vanishes—either through exhaustion of novelty or runaway formalism—entropy equilibrates and meaning dissipates.

In philosophical terms, this model reconciles Popper's falsifiability criterion with Bayesian updating and thermodynamic realism. Falsification corresponds to the injection of new entropy from $M_{\mathcal{R}}$ into $M_{\mathcal{T}}$; theory revision performs the compensatory work to restore low-entropy alignment. The more efficiently this cycle runs, the more rigorous the science. Conversely, when feedback loops close—when models absorb anomalies as mere noise—the entropy gap widens until theory ceases to exchange information with reality.

Within RSVP's broader framework, this process is literalized as a set of coupled field equations in which scalar potential Φ encodes conceptual capacity, vector flow $\mathbf{v}$ encodes

causal propagation, and entropy S tracks informational loss. Understanding becomes a conserved current in this field: a flux of low-entropy coherence traversing scales from physics to semantics. The next section translates this insight into practical diagnostics, proposing quantitative and institutional tools for maintaining rigor as a thermodynamic property of inquiry.

5 Practical Diagnostics for Theoretical Rigor

5.1 From typology to toolkit

The typology of failure developed above provides a conceptual grammar; what remains is a set of operational diagnostics that can be implemented in real scientific practice. Each failure mode corresponds to a measurable quantity in the epistemic energy functional $\mathcal{R}[f, \phi]$. Dimensional incoherence is revealed by undefined units in the Jacobian; mapping ambiguity by multi-valued derivatives; elegance fallacy by negative entropy balance. Translating these pathologies into algorithms yields an empirical science of rigor itself.

Historically, methodology has relied on narrative norms—peer review, replication, statistical significance—to maintain epistemic integrity (3; 4). Yet these norms predate the computational age. A modern diagnostic system can formalize them as measurable quantities. Just as thermodynamics replaced intuition about "heat quality" with entropy and temperature, epistemology can replace intuition about "sound reasoning" with injectivity, coherence, and energetic balance. The following subsections outline how this replacement can be operationalized.

5.2 Operational diagnostics

Eight procedural checks correspond to the eight failure modes of Section 3:

(a) **Dimensional audit**: Verify that every quantity possesses defined physical or informational units and that all equations are homogeneous under scaling. Computationally, perform symbolic dimensional analysis to ensure $\text{rank}(J) \neq 0$.
(b) **Primitive-variable isolation**: Identify at least one variable measurable independently of the others. If none exist, the system exhibits definitional circularity.
(c) **Falsifiability horizon**: Quantify the reachable region of the data manifold $M_{\mathcal{R}}$; if predicted observables lie outside it, the theory fails empirical accessibility.
(d) **Mapping table**: Construct an explicit bijection between theoretical parameters and empirical observables. Each row should include the measurement procedure, error bars, and boundary conditions.

(e) **Notation audit**: Cross-check each reused symbol against its original domain definition; flag instances where entropy, information, or energy are invoked without explicit metric transformation.
(f) **Scope limiter**: Document the exact conditions under which mappings remain valid. In simulations, this can be coded as boundary assertions that halt extrapolation beyond calibrated regimes.
(g) **Translation operators**: For interdisciplinary models, specify the transformation functions between domain-specific coordinate systems (e.g., physical informational semantic). If these operators are undefined, mapping ambiguity is inevitable.
(h) **Aesthetic control**: Record all instances where a design decision was justified by "simplicity," "symmetry," or "beauty." Such claims should be accompanied by quantitative performance tests or be treated as priors subject to empirical updating.

Applied systematically, these eight checks convert the typology into a reproducible audit trail. Each yields a numeric or categorical flag indicating potential energy loss in the comprehension process.

5.3 Geometric–thermodynamic diagnostics

At a deeper level, the same logic can be implemented computationally using the metrics defined in Section 0. The core procedure is a three-step evaluation of injectivity, coherence, and entropy balance:

```
function EpistemicAudit(theory_model, data):
    J = jacobian(theory_model.map, data)
    phi = phase_difference(theory_model.predict(data), data)
    W = gamma * integral(norm(grad(phi))**2)
    DeltaS = theory_model.entropy() - data.entropy()
    R = alpha*det(J.T @ J) - beta*norm(grad(phi))**2 - gamma*DeltaS
    return {"detJ": det(J.T@J), "PhasePower": P(phi),
            "EntropyGap": DeltaS, "EnergyFunctional": R}
```

This pseudocode formalizes the rigor test as an executable routine. In practice, it can be implemented within simulation frameworks or model-evaluation pipelines. For example, a cosmological code could compute $\det(J^\top J)$ for its parameter mappings, while a neural network analysis could evaluate P as the cosine similarity between activation trajectories and empirical time series. The resulting diagnostics produce scalar indicators of theoretical health:

$$\det J > 0, \qquad |\phi| < \epsilon, \qquad W \geq k_B T \, \Delta S.$$

When these inequalities hold, the system preserves information, coherence, and thermodynamic feasibility. Violations signal interpretive degeneration: the onset of pseudoscientific drift.

5.4 Minimal success conditions

The results of these tests can be summarized as a compact set of inequalities defining the domain of rigorous theory:

$$\det(J^\top J) > 0, \quad |\nabla \phi|^2 < \Lambda, \quad \Delta S \geq 0, \quad \frac{W}{k_B T \Delta S} \approx 1 \pm \epsilon.$$

Here Λ defines the maximum tolerable phase gradient before comprehension fails. These relations constitute the epistemic equivalent of conservation laws: meaning is neither created nor destroyed but transformed under energetic constraint. The condition $W/(k_B T \Delta S) \approx 1$ expresses the optimal balance between interpretive effort and informational payoff—a theoretical analogue of Carnot efficiency. When this ratio drifts significantly above one, the theory wastes energy in rhetorical elaboration; when it falls below, the mapping underfits reality.

5.5 Institutional application

Beyond individual modeling, these diagnostics can be extended to institutional practice. Research organizations, journals, and funding bodies could employ automated "rigor matrices" to assess theoretical proposals. Each column of such a matrix corresponds to one diagnostic dimension—dimensional grounding, measurability, mapping clarity, entropy balance—and each row to a project or manuscript. Weighted sums yield an overall *rigor index* R_{inst}, allowing comparison across fields.

For example, an internal review system might compute:

$$R_{\text{inst}} = \frac{1}{N} \sum_i \mathcal{R}_i[f_i, \phi_i],$$

averaged over all submissions. Low R_{inst} would indicate systemic drift toward high-entropy speculation, prompting corrective measures such as targeted replication or methodological training. Analogous systems already exist for code quality or energy efficiency; epistemic thermodynamics merely extends the principle to the production of knowledge itself.

Such metrics must, however, be applied reflexively. Over-measurement risks collapsing creative exploration into bureaucratic inertia—another form of entropy. The point is not to police thought but to maintain energy balance in the global research ecosystem, ensuring that interpretive work remains sustainable.

5.6 The epistemic thermodynamics of rigor

Rigor can now be defined quantitatively as low-entropy coupling between mathematics and reality. A rigorous theory minimizes ΔS while maintaining positive work W and finite coherence P. Pseudoscience, by contrast, is a dissipative structure: it consumes interpretive energy without generating predictive power. The distinction mirrors that between engines and friction—both obey the same physics, but only one performs useful work.

In **(author?)** (10)'s sense, scientific communities behave as complex systems evolving toward metastable equilibria. Excessive formalism corresponds to crystallization (rigidity); uncontrolled speculation to turbulence (chaos). The sustainable state lies near criticality, where small perturbations propagate information without runaway divergence. Theories that maintain this critical balance, such as Nikolaou's injectivity theorem or RSVP's field equations, embody the thermodynamic optimum of comprehension.

Philosophically, this redefines rationality as an energetic discipline. To reason well is to manage entropy efficiently—to produce maximal understanding with minimal waste. The measure of a theory is not how beautiful its equations appear, but how economically it converts uncertainty into structure. In this view, the scientific enterprise is a vast non-equilibrium engine: its fuel is curiosity, its exhaust is entropy, and its efficiency is rigor.

5.7 Conclusion

CIITR, Nikolaou, and RSVP illustrate the full energetic spectrum of theoretical behavior: from rhetorical diffusion through mathematical reversibility to dynamic equilibrium. The diagnostics developed here translate that spectrum into measurable criteria applicable across domains. A theory's worth lies in its ability to sustain coherent mapping between manifold and measurement without infinite cost. Where injectivity is preserved, phase coherence maintained, and entropy balanced, meaning endures. Where these conditions fail, mathematics decouples from reality and dissolves into noise.

The conservation law of meaning can thus be stated succinctly:

$$\frac{d}{dt}\mathcal{R}[f,\phi] = 0 \quad \Longleftrightarrow \quad \text{Rigor is sustained.}$$

Scientific understanding persists only when the flux of interpretive energy remains steady. To comprehend is to keep the world and our theories of it phase-locked against the drift of entropy—a task at once physical, cognitive, and moral.

A Derivation of the Phase-Locking Work Equation

The *Work of Understanding* introduced in Section 0.4 can be derived from the continuous dynamics of phase synchronization between theoretical and empirical manifolds. Let $\phi(y,t) = \theta_{\mathcal{T}}(y,t) - \theta_{\mathcal{R}}(y,t)$ denote the instantaneous phase difference between internal theoretical oscillations and empirical data signals. The goal of comprehension is to minimize this phase difference over time.

Assume that ϕ evolves according to a stochastic relaxation equation analogous to a damped Langevin process:

$$\dot{\phi} = -\Gamma \frac{\delta F}{\delta \phi} + \eta(y,t),$$

where Γ is a mobility constant, $F[\phi]$ is a free-energy functional, and η is a Gaussian noise field with correlation

$$\langle \eta(y,t)\eta(y',t')\rangle = 2k_B T\,\Gamma\,\delta(y-y')\delta(t-t').$$

Following **(author?)** (7, 9), the free energy associated with maintaining semantic alignment is

$$F[\phi] = \frac{\gamma}{2}\int |\nabla\phi|^2\, dV,$$

where γ quantifies interpretive resistance: the energetic stiffness of the mapping between manifolds.

The instantaneous power dissipated in reducing phase error is

$$P_{\text{diss}} = \left\langle \dot{\phi}\,\frac{\delta F}{\delta \phi} \right\rangle = \gamma \langle |\nabla\phi|^2 \rangle.$$

Integrating over space and time yields the total work of understanding:

$$W = \int_0^T P_{\text{diss}}\, dt = \gamma \int |\nabla\phi|^2\, dV.$$

This identifies W as the minimal energetic cost required to preserve coherence between

theory and observation under thermal noise. When $\nabla\phi = 0$, the system is in perfect epistemic equilibrium; when gradients steepen, interpretive work scales quadratically. The formalism thus links cognition and computation directly to the statistical mechanics of alignment.

B Example: Applying the Rigor Matrix to a Hypothetical Model

The rigor matrix proposed in Section 5 can be illustrated by evaluating a representative "quantum consciousness" hypothesis, here denoted $\mathcal{T}_{\text{QC}}$. The model claims that neuronal microtubules sustain macroscopic quantum coherence responsible for consciousness. Despite rhetorical appeal, its mapping between theoretical and empirical variables is ill-defined. Table 4 applies the eight diagnostics.

Criterion	Diagnostic Observation	Score (0–3)
Dimensional grounding	"Consciousness energy" has no physical units.	0
Primitive-variable isolation	No measurable primitive apart from subjective report.	0
Falsifiability horizon	Predicted coherence times 10^9 longer than measurable range.	0
Mapping table	No explicit function linking quantum state to cognitive variable.	0
Notation audit	Entropy and decoherence conflated without metric definition.	0
Scope limiter	Extends quantum laws into macroscopic domain without calibration.	0
Translation operators	Absent; physical $\rightarrow$ phenomenological mapping undefined.	0
Aesthetic control	Appeals to "beauty of quantum holism" as justification.	1
Mean Rigor Index	$\bar{R}_{\text{QC}} = 0.13$	

Table 4: Rigor-matrix evaluation of a hypothetical "quantum consciousness" model. Scores 0–3 represent increasing operational rigor.

The resulting $\bar{R}_{\text{QC}} \approx 0.13$ places the theory in the degenerating region of epistemic phase space. In Jacobian terms, $\text{rank}(J) = 0$ and $\det(J^\top J) = 0$, implying total loss of injectivity. The entropy gap $\Delta S = S_{\mathcal{T}} - S_{\mathcal{R}}$ diverges because theoretical complexity increases without corresponding empirical constraints. The model thus functions as an interpretive heat source: it generates rhetorical energy but no information flow.

For contrast, applying the same matrix to Nikolaou et al.'s injectivity theorem yields $\bar{R}_{\text{Nikolaou}} \approx 2.9$; to RSVP, $\bar{R}_{\text{RSVP}} \approx 2.3$. The rigorous theory achieves near-Carnot efficiency ($W \approx k_B T \Delta S$), while the speculative one dissipates unlimited interpretive work ($W \gg k_B T \Delta S$).

This example demonstrates how rhetorical, computational, and physical dimensions of theory can be evaluated quantitatively within a single analytic framework.

C Comparative Entropy Maps of CIITR, Nikolaou, and RSVP Manifolds

To visualize how theoretical coherence varies across domains, we approximate each model as a network of concepts linked by definitional or empirical dependencies. The informational entropy of each network is

$$S = -\sum_i p_i \ln p_i,$$

where p_i is the normalized connectivity degree of node i. Higher S indicates greater conceptual dispersion—more symbols with weak coupling; lower S implies concentrated structure.

Model	$S_{\mathcal{T}}$	$S_{\mathcal{R}}$	$\Delta S = S_{\mathcal{T}} - S_{\mathcal{R}}$
CIITR	7.2	1.1	+6.1
Nikolaou et al.	2.4	2.2	+0.2
RSVP	3.5	3.4	+0.1

Table 5: Approximate entropic balances between theoretical ($S_{\mathcal{T}}$) and empirical ($S_{\mathcal{R}}$) manifolds. Units: nats.

CIITR exhibits maximal theoretical entropy: numerous unconstrained terms yield $S_{\mathcal{T}} \gg S_{\mathcal{R}}$. The interpretive cost of mapping this structure to data diverges, making comprehension thermodynamically unsustainable. Nikolaou's theorem, by contrast, achieves near-equilibrium; theoretical and empirical entropies differ by only 0.2 nats, consistent with reversible computation. RSVP's slightly higher value reflects its cross-domain ambition: it expends small but finite energy to sustain coherence among physics, cognition, and computation.

Plotting these entropies as potentials over conceptual space yields a landscape reminiscent of free-energy surfaces (Fig. C). CIITR lies in a flat, high-entropy plateau; Nikolaou occupies a deep well of low free energy; RSVP traces a valley connecting multiple basins. The geometric picture reinforces the paper's thesis: rigor is a low-entropy channel

through which information flows smoothly between manifolds, while pseudoscience spreads diffusively across unbounded terrain.

D Supplement: Algorithmic Implementation of the Rigor Index

For computational reproducibility, the epistemic energy functional can be implemented as follows:

```
def rigor_index(J, phi, S_theory, S_empirical, alpha=1, beta=1, gamma=1):
    det_term = np.linalg.det(J.T @ J)
    phase_term = np.mean(np.gradient(phi)**2)
    delta_S = S_theory - S_empirical
    R = alpha * det_term - beta * phase_term - gamma * delta_S
    return R
```

Applied across time steps or model versions, this function traces the evolution of theoretical rigor as a dynamic variable. A positive $\dot{\mathcal{R}}$ indicates tightening coherence; a negative derivative signals entropic drift. By analogy with Lyapunov functions in dynamical systems, $\mathcal{R}$ serves as a potential governing the stability of comprehension.

E Supplement: Interpreting $\mathcal{R}[f, \phi]$ as a Cognitive Free Energy

Finally, note that the epistemic energy functional

$$\mathcal{R}[f, \phi] = \alpha \det(J^\top J) - \beta |\nabla \phi|^2 - \gamma \Delta S$$

can be rewritten as a free-energy-like quantity:

$$\mathcal{R} = \text{Information Coherence} - \text{Work Cost} - \text{Entropy Loss}.$$

Maximizing $\mathcal{R}$ corresponds to minimizing surprise and energetic expenditure simultaneously—the same variational objective found in the Freecrypto Energy Principle (12). The correspondence table below summarizes this structural equivalence.

This equivalence closes the theoretical circle. Whether in physics, cognition, or scientific reasoning, the condition for sustainable understanding is the same: maintain injective

[width=0.8]entropy$_l$*andscape*$_p$*laceholder.pdf*

Figure 1: Schematic entropy landscape of theoretical manifolds. Vertical axis: interpretive free energy; horizontal axes: empirical and conceptual dimensions. Rigor corresponds to steep gradients with bounded minima.

mapping, bounded work, and minimal entropy production. Rigor, comprehension, and life itself are special cases of energy-efficient mapping in an entropic universe.

F Appendix F: $\mathcal{R}[f, \phi]$ as a Lyapunov Functional

F.1 Setup and assumptions

Recall the epistemic energy functional (Sec. 0.8)

$$\mathcal{R}[f, \phi] \;=\; \alpha \, \det(J^\top J) \;-\; \beta \int_\Omega |\nabla \phi|^2 \, dV \;-\; \gamma \, \Delta S, \qquad \alpha, \beta, \gamma > 0,$$

with $J = \partial f / \partial y$ the Jacobian of the map $f : M_{\mathcal{R}} \to M_{\mathcal{T}}$, phase field ϕ the theory–data phase difference, and $\Delta S = S_{\mathcal{T}} - S_{\mathcal{R}}$. We consider the joint state $z := (\theta, \phi)$, where θ are parameters of f (weights, PDE coefficients, etc.). Let $\Omega \subset \mathbb{R}^d$ be the spatial domain with either periodic or Neumann boundary conditions so that boundary terms vanish.

Throughout this appendix we assume:

(A1) **Regularity.** $f(\cdot; \theta)$ is $\mathcal{C}^2$ in both arguments and $\det(J^\top J)$ is well-defined and $\mathcal{C}^1$ on the region of interest; $\phi \in H^1(\Omega)$.
(A2) **Boundedness.** ΔS is finite, and the admissible parameter set Θ is closed and bounded, or else regularized so that trajectories remain precompact.
(A3) **Positivity.** Gains α, β, γ and the metric/operator gains introduced below are positive definite.

F.2 Passive (dissipative) interpretive dynamics

We first model *passive* interpretive drift: in the absence of active correction, the mapping and phase relax along the *negative gradient* of $\mathcal{R}$ (i.e., toward lower rigor). Consider the gradient flows

$$\dot{\theta} = -K_\theta \, \nabla_\theta \mathcal{R}(\theta, \phi), \qquad K_\theta \succ 0, \tag{1}$$

$$\partial_t \phi = -\Gamma \, \frac{\delta \mathcal{R}}{\delta \phi} \;=\; -\Gamma \left(-2\beta \, \Delta \phi \right) \;=\; 2\beta \Gamma \, \Delta \phi, \qquad \Gamma \succ 0, \tag{2}$$

where ∇_θ is the Euclidean gradient and $\delta/\delta\phi$ the L^2 variational derivative. Equation (2) is a diffusion (heat) equation for ϕ with diffusivity $2\beta\Gamma$.

Theorem F.1 (Lyapunov decay of $\mathcal{R}$ under passive drift). Under (A1)–(A3) and dynamics (1)–(2), the functional $\mathcal{R}$ is nonincreasing along trajectories:

$$\frac{d}{dt}\,\mathcal{R}(\theta(t),\phi(t)) \;\le\; 0,$$

with equality if and only if $\nabla_\theta\mathcal{R} = 0$ and $\nabla\phi = 0$ a.e. (modulo measure-zero degeneracy of J). Hence $\mathcal{R}$ is a Lyapunov functional; the ω-limit set is contained in the set of critical points of $\mathcal{R}$.

Proof. By the chain rule and calculus of variations,

$$\dot{\mathcal{R}} = \big(\nabla_\theta\mathcal{R}\big)^\top\dot\theta \;+\; \int_\Omega \frac{\delta\mathcal{R}}{\delta\phi}\,\partial_t\phi\,dV = -\big(\nabla_\theta\mathcal{R}\big)^\top K_\theta\big(\nabla_\theta\mathcal{R}\big) \;+\; \int_\Omega\big(-2\beta\,\Delta\phi\big)\,(2\beta\Gamma\,\Delta\phi)\,dV.$$

Both terms are nonpositive: the first is $-\|K_\theta^{1/2}\nabla_\theta\mathcal{R}\|^2 \le 0$; the second equals $-4\beta^2\Gamma\int_\Omega|\Delta\phi|^2\,dV \le 0$. Boundary terms vanish under periodic/Neumann conditions. Equality requires $\nabla_\theta\mathcal{R} = 0$ and $\Delta\phi = 0$, which together with Neumann/periodic conditions implies $\nabla\phi = 0$. □

Remark. Passive drift therefore *decreases* $\mathcal{R}$ (rigor); the system is attracted to low-rigor equilibria (smooth but potentially degenerate mappings). This formalizes the intuitive claim that, without work, coherence decays.

F.3 Stochastic perturbations and expected decay

Let the dynamics be perturbed by mean-zero noise:

$$\begin{aligned}\dot\theta &= -K_\theta\,\nabla_\theta\mathcal{R}(\theta,\phi) + \Sigma_\theta\,\xi_\theta(t),\\ \partial_t\phi &= 2\beta\Gamma\,\Delta\phi + \sigma_\phi\,\xi_\phi(y,t),\end{aligned}$$

with ξ_θ white in time and ξ_ϕ space-time white (interpreted in the mild sense). Under standard Ito calculus and boundedness (A2), one obtains

$$\mathbb{E}\,\dot{\mathcal{R}} \;\le\; -\,\mathbb{E}\Big[\|K_\theta^{1/2}\nabla_\theta\mathcal{R}\|^2\Big] \;-\; 4\beta^2\Gamma\,\mathbb{E}\int_\Omega|\Delta\phi|^2dV \;+\; \tfrac{1}{2}\,\mathrm{Tr}\Big(\Sigma_\theta^\top H_{\theta\theta}\Sigma_\theta\Big) \;+\; \beta\Gamma\,\sigma_\phi^2\,C,$$

where $H_{\theta\theta}$ is the θ-Hessian of $\mathcal{R}$ and C depends on the Green operator of Δ. If the noise terms are sufficiently small compared to the dissipation, the expectation still decays: $\mathbb{E}\,\dot{\mathcal{R}} < 0$.

F.4 Active (controlled) interpretive dynamics

To *stabilize* high rigor, invert the drift by injecting control that performs epistemic work:

$$\dot{\theta} = +K_\theta \, \nabla_\theta \mathcal{R}(\theta, \phi), \qquad K_\theta \succ 0, \tag{3}$$

$$\partial_t \phi = +\Gamma \, \frac{\delta \mathcal{R}}{\delta \phi} \; = \; -\, 2\beta\Gamma \, \Delta\phi. \tag{4}$$

Then

$$\dot{\mathcal{R}} = \|K_\theta^{1/2} \nabla_\theta \mathcal{R}\|^2 \; + \; 4\beta^2 \Gamma \int_\Omega |\Delta\phi|^2 dV \; \geq \; 0,$$

so $\mathcal{R}$ is *increasing* and trajectories ascend toward critical points of $\mathcal{R}$. Interpreting $V := -\mathcal{R}$ as a Lyapunov function yields the standard gradient-descent picture: V decreases monotonically under (3)–(4). Thus, in the controlled regime, V is Lyapunov and $\mathcal{R}$ is an antiderivative of the supplied epistemic work.

Local stability of high-rigor equilibria. Let $z^\star = (\theta^\star, \phi^\star)$ satisfy $\nabla_\theta \mathcal{R}(z^\star) = 0$ and $\Delta\phi^\star = 0$, and suppose the *projected* Hessian of $-\mathcal{R}$ at $z^\star$ is positive definite. Then, under (3)–(4), $z^\star$ is (locally) asymptotically stable by Lyapunov's direct method (or LaSalle's invariance principle for the PDE part). Intuitively, these are the "high-rigor" maxima of $\mathcal{R}$.

F.5 Discrete-time (algorithmic) version

For iterative simulators or learning updates with step size $\eta > 0$,

$$\theta_{k+1} = \theta_k \mp \eta \, K_\theta \, \nabla_\theta \mathcal{R}(\theta_k, \phi_k), \qquad \phi_{k+1} = \phi_k \mp \eta \, \Gamma \, \frac{\delta \mathcal{R}}{\delta \phi}(\theta_k, \phi_k),$$

where the minus sign gives passive decay, the plus sign active control. A standard descent/ascent lemma yields, for sufficiently small η,

$$\mathcal{R}(z_{k+1}) - \mathcal{R}(z_k) \; \approx \; \mp\eta \left\| G^{1/2} \nabla \mathcal{R}(z_k) \right\|^2 + \mathcal{O}(\eta^2),$$

with block-diagonal metric $G = \mathrm{diag}(K_\theta, \Gamma)$. Hence $\mathcal{R}$ is monotonically nonincreasing (passive) or nondecreasing (active) up to $\mathcal{O}(\eta^2)$ terms.

F.6 Interpretation and design guidelines

- **Two regimes, two Lyapunov choices.** For *passive drift*, $\mathcal{R}$ itself is Lyapunov: $\dot{\mathcal{R}} \leq 0$. For *active control*, $V := -\mathcal{R}$ is Lyapunov: $\dot{V} \leq 0$. This mirrors thermodynamics:

without work, order decays; with work, order grows.

- **Gains as *interpretive viscosity.*** K_θ and Γ trade off convergence rate and robustness. Large gains speed alignment (higher work input) but risk overshoot in noisy settings; small gains conserve energy but slow recovery of rigor.
- **Constraints and projections.** If Θ or admissible ϕ are constrained (e.g., PDE stability, physical units), project the gradient step onto the feasible set. Projected gradient maintains monotonicity of the appropriate Lyapunov functional.
- **Noise budgets.** Stochastic bounds above show a critical noise level where passive decay dominates. Active control must at least offset this with work so that $\mathbb{E}\,\dot{\mathcal{R}} \geq 0$.

F.7 Connection to the Free Energy Principle

If we identify $V := -\mathcal{R}$ with a cognitive free energy (Appendix E), then the *active* dynamics (3)–(4) implement gradient descent on V (Friston's variational flows), guaranteeing V decreases monotonically while $\mathcal{R}$ increases. Conversely, the *passive* dynamics (1)–(2) correspond to free evolution without control, in which V increases and $\mathcal{R}$ decays.

F.8 Summary

- Under dissipative (passive) dynamics, $\dot{\mathcal{R}} \leq 0$: rigor decays monotonically; $\mathcal{R}$ is Lyapunov.
- Under controlled (active) dynamics, $\dot{\mathcal{R}} \geq 0$ and $V = -\mathcal{R}$ is Lyapunov: rigor is stabilized near high-$\mathcal{R}$ equilibria.
- These results hold in both continuous time and discrete algorithms (for small steps), with stochastic perturbations handled in expectation.

G Appendix G: Worked RSVP Simulation Check ($\det(J^\top J)$, $|\nabla\phi|^2$, ΔS, P)

G.1 G.1 Data products and notation

Assume a d-dimensional periodic or Neumann lattice $\Lambda \subset \mathbb{Z}^d$ with spacing h and time step Δt. Your simulator outputs per step t_k:

$$\Phi_k(x), \quad \mathbf{v}_k(x) \in \mathbb{R}^d, \quad S_k(x) \qquad (x \in \Lambda),$$

plus any exogenous forcing u_k (e.g., boundary drive, source terms). Let $\widehat{\cdot}$ denote the discrete Fourier transform (DFT), ∇_h the centered finite difference, and Δ_h the discrete Laplacian.

We also assume an *external reference* stream $r_k(x)$ against which phase is aligned (e.g., empirical timeseries, a target field, or a filtered channel of the simulation regarded as "data"). When no external data are available, use a bandpassed proxy of a measured observable (e.g., divergence of $\mathbf{v}$) to define a consistent reference.

G.2 G.2 Phase field and work of understanding

Define analytic signals by Hilbert transform (1D time) or narrowband DFT (spacetime window W). For each lattice site,

$$\theta_{\mathcal{T},k}(x) := \arg\Big(\Phi_k(x) + i\,\mathcal{H}_t[\Phi_\cdot(x)]_k\Big), \qquad \theta_{\mathcal{R},k}(x) := \arg\Big(r_k(x) + i\,\mathcal{H}_t[r_\cdot(x)]_k\Big).$$

Phase difference:

$$\phi_k(x) := \mathrm{wrap}\,\Big(\theta_{\mathcal{T},k}(x) - \theta_{\mathcal{R},k}(x)\Big) \in (-\pi, \pi].$$

Discrete gradient (componentwise):

$$(\nabla_h \phi_k)_j(x) := \frac{\phi_k(x + e_j) - \phi_k(x - e_j)}{2h}, \quad j = 1, \ldots, d.$$

Work of understanding over the domain (Sec. 0.4):

$$W_k \;=\; \gamma \sum_{x \in \Lambda} \|\nabla_h \phi_k(x)\|_2^2\, h^d.$$

Practical notes. (i) Use phase unwrapping per axis before differencing. (ii) Apply a Tukey/Hann taper in time for DFT windows to suppress leakage. (iii) Choose band(s) based on the dominant coherence peak in Sec. G.5.

G.3 G.3 Coherence power P via principal angles

Form m-length time windows (rows) of the reference $R \in \mathbb{R}^{n \times m}$ and target $T \in \mathbb{R}^{n \times m}$ at a chosen site, channel, or spatial average (n windows). Zero-mean and unit-variance normalize each column. Compute SVD of cross-covariance:

$$C := \frac{1}{n-1} R^\top T \;=\; U \Sigma V^\top, \qquad \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_q).$$

Principal angles $\theta_i := \arccos(\sigma_i)$ and the mean coherence power:

$$P_k \;=\; \frac{1}{q}\sum_{i=1}^{q}\cos^2\theta_i \;=\; \frac{1}{q}\sum_{i=1}^{q}\sigma_i^2 \;\in\; [0,1].$$

Guidelines. Use overlapping windows (50–75%) and the same bandpass as in G.2. For field-wide P_k, average σ_i^2 across a representative spatial set or compute via spatial principal subspaces first (reduced q).

G.4 G.4 Entropy balance ΔS

Two robust estimators are useful; pick one and keep it consistent:

(a) Spectral (Shannon) entropy. Let X_k be a scalar observable (e.g., Φ_k or $\nabla\cdot\mathbf{v}_k$) and $\widehat{X}_k(\omega)$ its windowed periodogram. Normalize $p_k(\omega) := \widehat{X}_k(\omega)/\sum_\omega \widehat{X}_k(\omega)$. Then

$$S_{\mathcal{T},k} := -\sum_{\omega} p_k^{(\mathcal{T})}(\omega)\,\ln p_k^{(\mathcal{T})}(\omega), \quad S_{\mathcal{R},k} := -\sum_{\omega} p_k^{(\mathcal{R})}(\omega)\,\ln p_k^{(\mathcal{R})}(\omega), \quad \Delta S_k := S_{\mathcal{T},k} - S_{\mathcal{R},k}.$$

(b) State-space histogram entropy. Form a joint sample $Z_k(x) := (\Phi_k(x), \|\mathbf{v}_k(x)\|_2, S_k(x))$. Bin into a fixed 3D grid (with Freedman–Diaconis or fixed bins), estimate $p_k(z)$, and compute

$$S_{\mathcal{T},k} := -\sum_{z} p_k^{(\mathcal{T})}(z)\,\ln p_k^{(\mathcal{T})}(z),$$

likewise for $S_{\mathcal{R},k}$ from the reference $Z_k^{(\mathcal{R})}$. **Tip.** Add a small ε pseudocount per bin.

G.5 G.5 Jacobian term $\det(J^\top J)$ (discrete proxy)

Two practical proxies work well in simulators; pick one that matches your workflow.

(i) Observation–sensitivity Jacobian. Choose p summary observables (e.g., spatial means of Φ, energy flux $\mathbf{v}\cdot\nabla\Phi$, spectral peaks, P_k) and q controllable inputs (forcing amplitudes, boundary potentials, coupling constants). For each input y_a, apply a small perturbation δy_a at t_k and re-run for one step to get δx_i. Fill

$$J_{ia} := \left.\frac{\delta x_i}{\delta y_a}\right|_{t_k}, \qquad i = 1,\dots,p,\ a = 1,\dots,q.$$

Compute stabilized log-determinant:

$$\lambda_1, \ldots, \lambda_q := \mathrm{eigvals}\Big(J^\top J + \varepsilon I\Big), \quad \log\det(J^\top J + \varepsilon I) = \sum_{a=1}^{q} \ln \lambda_a.$$

Report $\det(J^\top J)$ or its log to avoid under/overflow; fix $\varepsilon \sim 10^{-8} - 10^{-6}$ times the median diagonal.

(ii) Parameter–gradient Fisher proxy. If you already compute parameter gradients $\nabla_\theta x_i$, form a Fisher-style matrix $F = \sum_i (\nabla_\theta x_i)(\nabla_\theta x_i)^\top$. Then use

$$\log\det(F + \varepsilon I) \quad \text{as a surrogate for} \quad \log\det(J^\top J + \varepsilon I).$$

This avoids reruns and exploits autodiff.

G.6 G.6 Epistemic energy functional and checks

Given $\alpha, \beta, \gamma > 0$,

$$\mathcal{R}_k \;=\; \alpha\, \log\det(J_k^\top J_k + \varepsilon I) \;-\; \beta\, \underbrace{\left(\sum_x \|\nabla_h \phi_k(x)\|_2^2\, h^d \right)}_{\|\nabla \phi_k\|_2^2} \;-\; \gamma\, \Delta S_k.$$

Minimal feasibility checks per step:

$$\det(J_k^\top J_k) > 0, \qquad P_k \in [0,1], \qquad W_k \geq k_B T\, \Delta S_k.$$

If you operate in the *active* (controlled) regime (Appendix F), expect $\mathcal{R}_{k+1} - \mathcal{R}_k \gtrsim 0$ on average; in the *passive* regime, $\mathcal{R}$ should drift downwards (noise aside).

G.7 G.7 Numerical stability and parameter choices

- **Regularization:** always use εI in log-det; track ε in logs.
- **Windows:** for DFT/Hilbert phases, use windows of $m \in [64, 512]$ samples with 50% overlap; match band to the dominant coherence peak (from P_k).
- **Derivatives:** use 2nd-order centered differences; for rough fields, apply a single Jacobi or bilateral smooth before ∇_h.
- **Units:** record units/scales for each observable to preserve dimensional homogeneity (see Sec. 5.2a).

G.8 G.8 Reference implementation (pseudocode)

```
def step_metrics(sim_state, ref_window, params):
    # --- Phase & work ---
    phi = wrapped_phase(sim_state.Phi, ref_window) # Hilbert/DFT
    grad_phi_sq = sum_over_axes(central_diff(phi)**2)
    W = params.gamma * domain_sum(grad_phi_sq)

    # --- Coherence P ---
    R, T = build_windows(ref_window, sim_state.Phi_or_obs)
    sigma = svd_singular_values(cross_cov(R, T))
    P = mean(sigma**2)

    # --- Entropy S ---
    if params.entropy_mode == "spectral":
        S_T = shannon_entropy(psd(sim_state.obs))
        S_R = shannon_entropy(psd(ref_window.obs))
    else:
        S_T = histogram_entropy(stack([Phi, norm(v), S]))
        S_R = histogram_entropy(stack([Phi_ref, v_ref, S_ref]))
    DeltaS = S_T - S_R

    # --- Jacobian det proxy ---
    if params.jac_mode == "finite_diff":
        J = estimate_observation_jacobian(sim_state, params.perturb_grid)
        logdet = logdet_sym(J.T @ J + eps*I)
    else:
                F = fisher_proxy_from_grads(sim_state.grads) # autodiff
        logdet = logdet_sym(F + eps*I)

    # --- Energy functional ---
    R_index = params.alpha * logdet - params.beta * domain_sum(grad_phi_sq) \
              - params.gamma * DeltaS

    return {"W": W, "P": P, "DeltaS": DeltaS, "logdet": logdet, "R": R_index}
```

G.9 G.9 Suggested log schema (JSONL)

```
{"t": 12.50, "P": 0.84, "W": 1.73e+02, "DeltaS": 0.12,
```

```
"logdet": 23.41, "R": 18.66, "eps": 1.0e-7,
"band": [4.0, 7.0], "window": 256, "jac":"fwd-diff"}
```


G.10 G.10 Acceptance criteria (per regime)

- **Active (controlled) regime:** $\mathbb{E}[\Delta\mathcal{R}_k] \geq 0$; $P_k \to P^\star \in (0.7, 1]$; $W_k \approx k_B T\, \Delta S_k$ within tolerance.
- **Passive (free) regime:** $\mathbb{E}[\Delta\mathcal{R}_k] \leq 0$; P_k decays to baseline; $W_k / k_B T\, \Delta S_k \to 0^+$ or fluctuates sub-Carnot.
- **Sanity checks:** $\log\det$ not dominated by ε (track $\lambda_a \gg \varepsilon$); phase gradients bounded (no unwrap explosions).

References

[1] Hansen, T.-S., 2024. Comprehension as Thermodynamic Persistence (CIITR CITR). Self-published manuscript.

[2] Nikolaou, A., Kambadur, P., Stern, M., Milosavljevic, B., 2025. Language models are injective and hence invertible. arXiv preprint arXiv:2510.15511.

[3] Popper, K., 1959. The Logic of Scientific Discovery. Routledge.

[4] Lakatos, I., 1978. The Methodology of Scientific Research Programmes. Cambridge University Press.

[5] Kuhn, T.S., 1962. The Structure of Scientific Revolutions. University of Chicago Press.

[6] Hossenfelder, S., 2018. Lost in Math: How Beauty Leads Physics Astray. Basic Books.

[7] Jaynes, E.T., 1957. Information theory and statistical mechanics. Physical Review 106, 620–630.

[8] Landauer, R., 1961. Irreversibility and heat generation in the computing process. IBM Journal of Research and Development 5, 183–191.

[9] Bennett, C.H., 1982. The thermodynamics of computation. International Journal of Theoretical Physics 21, 905–940.

[10] Sethna, J.P., 2006. Statistical Mechanics: Entropy, Order Parameters, and Complexity. Oxford University Press.

[11] Smolensky, P., Legendre, G., 2006. The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar. MIT Press.

[12] Friston, K., 2010. The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 11, 127–138.

[13] Clark, A., 2016. Surfing Uncertainty: Prediction, Action, and the Embodied Mind. Oxford University Press.

[14] Bianconi, G., 2025. Entropy and complexity in networked systems. Physical Review D 102, 042013.

[15] Quine, W.V.O., 1951. Two dogmas of empiricism. The Philosophical Review 60, 20–43.

[16] Cartwright, N., 1983. How the Laws of Physics Lie. Oxford University Press.

[17] Feynman, R.P., 1965. The Character of Physical Law. MIT Press.

[18] Tegmark, M., 2014. Our Mathematical Universe: My Quest for the Ultimate Nature of Reality. Knopf.

[19] Barandes, J.A., 2023. A unistochastic reformulation of quantum theory. Foundations of Physics 53, 119.

[20] Li, J., Li, P., 2025. Formalizing Lacan's RSI topology via active inference networks. Journal of Cognitive Modeling 18, 245–272.