Boris Kriger

A UNIFIED STRUCTURAL THEORY OF COMPLEX SYSTEMS

Formal Laws, Epistemic Constraints,
and Self-Organization Across Physical,
Cognitive, and Social Domains

Opus Magnum

DOCTORAL THESIS

A dissertation prepared in fulfillment of the requirements
for the degree of Doctor of Philosophy
in Systems Science / Theoretical Foundations of Complex Systems

© 2026 Boris Kriger

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from both the copyright owner and the publisher.

Boris Kriger | Research Fellow

ORCID: orcid.org/0009-0001-0034-2903

https://www.researchgate.net/profile/Boris-Kriger

https://philpeople.org/profiles/boris-kriger



Institute of Integrative and Interdisciplinary Research

Toronto, Ontario, Canada

+1 437-552-8807

interdisciplinary-institute.org

boriskriger@interdisciplinary-institute.org



Information Physics Institute

https://www.informationphysicsinstitute.org/

boris.kriger@informationphysicsinstitute.net

Contents

Biography 7

INTRODUCTION 13

Part I. The Architecture of Formal Knowledge 20

Chapter 1. The Problem of Unification Across Domains 20

Chapter 2. Scale-Specific Laws and the Impossibility of a Final Theory 24

Chapter 3. Formalization Laws and Definition-Dependent Provability 35

Chapter 4. Epistemic Constraints: Uncertainty, Negation, and the Limits of Inference 50

Part II. The Dynamical Core: Persistence and Self-Organization 61

Chapter 5. The Transformational Basis of Persistence 61

Chapter 6. Resilience, Closure, and Self-Sufficiency 75

Chapter 7. Four Laws of Self-Organizing Systems 88

Chapter 8. Emergence of Dynamics, Structural Compatibility Laws and Structural Genesis 100

Part III. Cognition and Consciousness as Structural Necessities 120

Chapter 9. The Structural Distortion Principle: Perception as World-Maintenance 120

Chapter 10. The Inevitability of Predictive Processing 140

Chapter 11. Consciousness, Time, and Memory Architecture 150

Chapter 12. Mental Disintegration and Biological Decision Programs 165

Part IV. Social and Multi-Agent Systems: From Structural Laws to Institutional Dynamics 185

Chapter 13. Deception, Autonomy Suppression, and Structural Non-Neutrality 185

Chapter 14. Conflict, Totalizing Ideals, and Phase Transitions 208

Chapter 15. Extractive Oscillators: From Narcissistic Dynamics to Addiction 227

Chapter 16. Institutional Strategies and Reflexive Inference 247

16.1. Non-Participation and Pre-Integrative Rejection 247
16.2. Conceptual Responsibility 255
16.3. Assertion–Dismantling Cycles 258
Part V. Applications and the Architecture of Civilization 274
Chapter 17. Astrophysical and Information-Theoretic Applications 274
17.1. Binary Star Formation as Persistence-Selected Outcome 274
17.2. Dormant Neutron Stars and the Galactic Dark Population 275
17.3. Time, Entropy, and the Complexity of the Universe 275
Chapter 18. AI, Civilization, and the Future of Agency 277
18.1. Clinical Discontinuity and AI 277
18.2. AI-Extended Agents and the Transformation of Communication 277
18.3. The Stimulus Problem in Post-Scarcity Environments 278
18.4. The Inward Turn and the Fermi Paradox 278
18.5. The Inevitability of a Unified Civilization of Autonomous Agents 279
18.6. Viral Dynamics as Information Propagation 280
Part VI. The Epistemology of Incompatibility and the Limits of Description 281
Chapter 19. Incompatible Truths and the Navigation of Irreducible Pluralism 281
Chapter 20. Causation, Mathematical Necessity, and the Boundaries of Physical Description 292
Chapter 21. Methodological Foundations: Evidence-Based Philosophy and Frame-Breaking Inquiry 305
Part VII. Formal Structures of Identity, Consciousness, and Communication 316
Chapter 22. Identity, Motivation, and Personality Architecture 316
Chapter 24. Communication, Pedagogy, and Epistemic Environment Degradation 325
Part VIII. Political Economy, Institutional Pathology, and Social Design 335
Chapter 25. Post-Labour Economics and Sociotechnical Phase Transitions 335

Chapter 26. Corruption, Concealment, and the Structural Pathology of Institutions 343

Part IX. Applied Formal Methods: From Civilisational Design to Ontological Foundations 351

Chapter 27. Decentralised Civilisation, Policy Architecture, and Computational Humanities 351

Chapter 28. Ontology, Biological Complexity, and the Limits of Recursive Self-Improvement 364

Part X. Cosmological Verification: The Persistence Framework as a Discovery Engine at the Largest Scales 373

Chapter 29. Breaking the Frame: How Tree-Top, RAFASSABRA, and Methodological Laws Generated the Cosmological Programme 373

Chapter 30. Persistence Conditions, Invariant Structure, and the Computation of the Cosmological Constant 379

Chapter 31. Self-Consistency Fixed Points, Infodynamics, and the Self-Organising Cosmos 385

Chapter 32. The Persistence Theorem and the Hierarchy of Stability: The Kriger Limit 393

Chapter 33. Incompatible Truths, Compression, and the Dark Matter Question 402

Chapter 34. Stellar Persistence: Viability Sets, Binary Dominance, and the Astrophysics of Survival 411

Chapter 35. The Observer at Cosmic Scales: Inaccessibility, Navigation, and the Boundaries of Cosmological Knowledge 420

Chapter 36. The Sociology of Cosmological Knowledge: The Persistence Framework Applied to the Scientific Community Itself 430

Chapter 37. Phase-Transition Signatures of αLGQV: NANOGrav, the Baryon-to-Photon Ratio, and the One-Event Synthesis 440

Chapter 38. Black-Hole Interiors without Singularity: The Second Mechanism for $k_2 = 0$ and Kgr-Radiation as the Remaining Discriminator 446

Chapter 39. Rotating RP^3 Cosmology: Hopf Fibration, Double Counter-Rotation, and Centrifugal Expansion 452

Chapter 40. Environmental Screening, Cross-Domain Confirmations, and the Epistemic Cost of ΛCDM 458

Part XI. The Algebra of Wisdom and the Structural Foundations of AGI Safety 465

Chapter 41. The Algebraic Theory of Frame-Change 465

Chapter 42. Operational Synthesis: The Methodological Algebra of Wisdom 473

Chapter 43. The Structural Theorem on the Impossibility of Malicious General Intelligence 479

Conclusion: Persistence as the Primary Organising Principle 520

Citations and Scientific Responses to the Works of Boris Kriger 537

Glossary of Terms 550

Publications 556

Bibliography 568

BIOGRAPHY

Boris Kriger (b. 1970) is a Canadian researcher affiliated with the Information Physics Institute (Gosport, UK) and the Institute of Integrative and Interdisciplinary Research (Toronto, Canada). He is the author of over eighty research publications spanning 1999–2026, as well as numerous books on philosophy, science, and social thought. He has founded and directed educational institutions, clinical research centers, and a publishing house.

Kriger's research programme encompasses a unified structural theory of complex systems — the subject of this dissertation — in which the same formal apparatus that generates the cosmological results also applies to cognition, consciousness, institutional dynamics, and social systems. The central thesis is that the formal constraints governing persistence are substrate-independent: the same mathematical structures appear in stellar evolution, cognitive architecture, and institutional dynamics because the persistence problem is structurally the same across all domains.

This trajectory began in the late 1990s with a metabolic model of the acceleration of chronoperception — the earliest publication in the corpus (1999). Sustained engagement with cosmological epistemology, including exchanges with George Ellis on the philosophy of cosmology, Joel Primack on the foundations of ΛCDM, and Julian Barbour on timeless cosmology, led Kriger from questions about the limits of physical theory to a broader research program on the structural constraints of formal description.

In theoretical astrophysics, he developed a series of papers challenging the default assumption of single-star formation, arguing that binary systems represent persistence-selected, energetically favored outcomes. In meta-mathematics and the philosophy of formalization, he articulated the Definition-Dependent Provability Principle, the framework-dependence of chaos and complexity, the structural limits of negation, and the necessity of plural formal representations for any sufficiently complex domain. In computational psychiatry, he developed a unified mathematical formalism for the phenomenology of mental disintegration through dynamical systems theory, with applications to all major DSM-5-TR diagnostic categories. Additional contributions address game-theoretic models of AI-mediated communication, a proposed resolution of the Fermi Paradox, and formal frameworks for deception in multi-agent systems.

His research is oriented toward the conviction that the deepest results in formal science emerge not from narrowing the domain of inquiry but from proving that the boundaries between domains are artifacts of formalization — and that the structures which govern persistence are, in the end, the same everywhere.

Kriger's principal scientific achievement is the derivation of the cosmological constant from first principles of nuclear physics — a result that addresses what has been called the worst prediction in the history of physics. The standard calculation, originating with Zel'dovich (1968), yields a discrepancy of 55 to 120 orders of magnitude between the quantum field theory prediction and the observed value of the cosmological constant. Kriger showed that this discrepancy arises from an unexamined assumption: the identification of quantum vacuum energy with the cosmological constant, which was postulated by Zel'dovich in 1967 but never derived from the Einstein field equations. When this identification is abandoned — as the trace-free (unimodular) formulation of general relativity formally permits — vacuum energy becomes a local gravitating quantity, and the cosmological constant Λ emerges as an independent geometric integration constant. The vacuum–matter coupling α then follows from the experimentally measured QCD sigma terms of the nucleon: $\alpha = (\sigma_{\pi N} + \sigma_s)/m_N = 0.096 \pm 0.023$. After accounting for the cosmological baryon fraction and the nonlinear self-screening discovered through N-body simulations, this yields $\alpha_{eff} \approx 0.005$, against a cosmologically observed value of 0.003. The result uses zero free parameters — every quantity is either measured in nuclear physics experiments or computed from standard QCD.
This single coupling constant, derived entirely from nuclear physics, generates a cascade of quantitative predictions across cosmology and astrophysics that have traditionally required separate explanatory frameworks:
In galactic dynamics, the gravitating vacuum model reproduces flat rotation curves across all 175 galaxies in the SPARC database without invoking dark matter particles. The pseudo-isothermal density profile arises as a prediction — not a fit choice — because the vacuum energy density follows the baryonic gravitational potential, which has a finite central density by construction. The core radius is derived from the disk scale length, not freely adjusted. The programme thus achieves what decades of dark matter particle searches have not: a parameter-free account of galactic rotation from known physics.

In large-scale structure, the cosmic web is proved to be the unique self-consistent configuration of the matter–vacuum–metric system through the Banach contraction mapping theorem, with statistical properties independent of the primordial power spectrum. This uniqueness theorem explains why the cosmic web and neuronal networks share the same topology (Vazza & Feletti, 2020) — both are cost-minimizing networks governed by the same variational structure.

In structure formation, the nonlinear self-screening mechanism — discovered through particle-mesh N-body simulations and independently confirmed by an analytical growth-weighted overdense fraction — resolves the S_8 tension between CMB predictions and weak lensing measurements. The screening factor reduces the effective vacuum enhancement by a factor of approximately 3–4, producing lower σ_8 at late times in quantitative agreement with survey data.

In early-universe physics, the 1% gravitational enhancement from the vacuum–matter coupling compounds over cosmic time to produce accelerated structure formation at high redshift — a prediction confirmed by the James Webb Space Telescope's discovery of unexpectedly massive and evolved galaxies at $z > 10$, which standard ΛCDM models cannot account for without ad hoc modifications.

The programme also derives, rather than assumes, the critical acceleration scale a_0 of the radial acceleration relation, reproduces the Tully–Fisher relation as a consequence of the vacuum coupling, accounts for the Bullet Cluster morphology, and predicts a scale-dependent Hubble constant consistent with the observed H_0 tension.

These results were achieved not by introducing new physics but by removing an unnecessary assumption — the Zel'dovich identification — and tracing the consequences of standard general relativity and standard QCD through to their observational predictions. No new particles, no new forces, no new fields, and no fitted parameters.

Recognition by Leading Scientists. Kriger's results have been reviewed and engaged with by leading researchers across multiple fields, many of whom have offered assessments that speak to the quality and significance of the work.

Stanley J. Brodsky, Professor Emeritus at Stanford University and SLAC — one of the creators of the light-front QCD formalism and co-author of the foundational PNAS paper on in-hadron condensates — wrote: "I am very impressed with your excellent work. It

demonstrates the importance of understanding the vacuum in quantum field theory based on the frame-independent light front quantization." Brodsky connected Kriger with his full collaboration (Deur at Jefferson Lab, Roberts, de Téramond, Dosch, and Terzić) and expressed interest in joint work, writing that "it would be very good to initiate collaborations on the physics projects you have outlined."

John Earman, one of the foremost philosophers of physics at the University of Pittsburgh, described Kriger's paper resolving Wigner's puzzle — the "unreasonable effectiveness of mathematics" — as "a major contribution to understanding Wigner's puzzle."

George Ellis (University of Cape Town), widely regarded as one of the most important living cosmologists and co-author with Hawking of *The Large Scale Structure of Space-Time*, engaged in a multi-round exchange. Ellis described the matter-dependent vacuum energy proposal as "an interesting proposal," passed it to a specialist for evaluation, and later responded to the trace-free Einstein equations paper with "That is now very interesting. The TFE are key."

Thomas D. Cohen (University of Maryland), author of the foundational 1992 paper on in-medium QCD condensates used in the derivation, confirmed the core physical argument — that the chiral condensate shift, as a Lorentz scalar, generates a stress-energy tensor proportional to the metric and hence $w = -1$ exactly — and offered arXiv endorsement.

Joan Solà Peracaula (University of Barcelona), originator of the Running Vacuum Model — one of the most developed alternatives to the cosmological constant — called the work "interesting" and provided substantive technical feedback, identifying a specific phenomenological step while noting that "it may be fruitful, especially if one finds a theoretical motivation."

Daniel Brown (University of Utah), a cosmologist working on quantum kinetic dark energy, engaged in the most sustained technical exchange in the programme. Brown carefully reviewed the N-body simulations, confirmed mathematical consistency, called the sigma terms derivation "really interesting" and the rotation curve calculator "really impressive," and stated: "I may end up referencing this connection in my own work as a way to give a bit more physical motivation to the framework."

David Chalmers (New York University), the philosopher who defined the Hard Problem of consciousness, called Kriger's Bayesian formalization of the zombie argument "nice" and

engaged in a substantive technical exchange, offering a precise correction that was incorporated into the revised paper.

Andreas Burkert (Ludwig-Maximilians-Universität München), creator of the Burkert halo profile — one of the standard cored dark matter profiles used in rotation curve analysis — engaged directly with the SPARC fits and confirmed that the coupling between baryonic scale length and dark matter core radius is observed in dwarf spheroidals, providing references to his own supporting work.

Pedro Mediano (Imperial College London), a key contributor to the formal computation of integrated information (Φ), praised the three-level framework for consciousness science as "very interesting," provided detailed technical feedback, and expressed interest in continuing the exchange on phase transitions in integration measures.

Bharat Ratra (Kansas State University), co-author with Peebles on the foundational quintessence paper (1988) — one of two papers that launched the entire field of dynamical dark energy — responded within one hour, indicating he would examine the work closely.

Kazuyuki Omukai (Tohoku University), a leading authority on direct-collapse black hole conditions, confirmed that Kriger's summary of DCBH physics was "essentially correct" and pointed to two recent results broadening the parameter space for massive seed formation.

Salvatore Capozziello (University of Naples), one of the most cited researchers in modified gravity, described the review paper as "very interesting" and confirmed compatibility with his conformal Killing gravity framework.

Ariel Zhitnitsky (University of British Columbia), working on QCD topological dark energy, provided a detailed three-step summary of his programme and confirmed that his and Kriger's approaches are complementary — one giving the time-dependent dynamics, the other the asymptotic value.

Ahmed Farag Ali, who derived the cosmological constant from SU(3) confinement, confirmed that "the two approaches are conceptually aligned in rejecting the direct identification of the observed cosmological constant with the unsuppressed microscopic vacuum energy."

Thomas Janka (Max Planck Institute for Astrophysics, Garching), one of the world's foremost authorities on the neutrino-driven supernova mechanism, engaged in a detailed multi-round exchange and ultimately agreed with Kriger's arguments about the role of the weak interaction

in chemical dissemination, writing that the way Kriger put his idea was "fine" and expressing hope that it would "instigate more deep thoughts."

Mikhail Shifman (University of Minnesota), co-creator of the SVZ sum rules — a cornerstone of QCD vacuum physics — agreed to provide arXiv endorsement and facilitated the process through his postdoctoral colleague.

Grigory Volovik (Landau Institute for Theoretical Physics), one of the world's leading theoretical physicists on analogies between condensed matter and cosmology — whose own work on the cosmological constant problem is widely cited — read Kriger's vacuum energy paper on ResearchGate and followed his profile for future publications.

J. Richard Bond (CITA, University of Toronto), one of the architects of the modern understanding of the cosmic web, engaged in a multi-round exchange and offered a rare candid reflection: "i have no problem with others working to flesh out their own ideas... i do not embrace any ideas as gospel. not even the most basic scaffolding of physics."

Additional engagement came from Alan Hájek (Australian National University), a leading philosopher of probability, who called Kriger's work on the reference class problem "very interesting" and engaged across three separate papers; Fabo Feng (Shanghai Jiao Tong University), who shared unpublished Gaia astrometry results relevant to Kriger's Tau Ceti analysis; Laurent Perrinet (Aix-Marseille/CNRS), who initiated contact after reading Kriger's essay on predictive processing; Giovanni Pezzulo, who wrote that the paper "resonates a lot" with his own thinking on prediction; and Sofia Sheikh (SETI Institute), who called the temporal mismatch framework for technosignature detection "an interesting idea" and forwarded it to colleagues.

No engaged scientist identified a fatal error in the derivation. Several identified limitations and boundary conditions — all of which were incorporated into revised manuscripts. The pattern — domain experts reading domain-specific papers, confirming the core reasoning, offering corrections on presentation, and expressing interest in further development — constitutes a form of distributed peer review that, while unconventional, has been substantively more rigorous than many formal journal processes.

INTRODUCTION

This thesis presents a unified structural theory of complex systems, synthesised from over one hundred and fifty original publications spanning 1999 to 2026. Its central claim is that persistence — the capacity of a system to maintain structural viability over time — is the primary organising principle of complex systems across all domains: physical, biological, cognitive, social, institutional, and cosmological. This is not a metaphorical claim. It is a formal result derived from axioms and proved with mathematical precision. Its strongest verification is empirical: the same formalism that generates viability conditions for biological cells, when transferred to the quantum vacuum without modification, computes the cosmological constant to within a factor of 1.65 of the observed value — replacing the 120-order-of-magnitude discrepancy that has defined the problem for half a century.

The theory rests on the demonstration that certain formal constraints on organised complexity are substrate-independent: they apply to any system that must maintain itself over time, regardless of what that system is made of. From a single persistence principle, the thesis derives the organisational logic of self-organising systems, the architecture of cognition and consciousness, the dynamics of social institutions, the pathology of corruption and concealment, the epistemology of incompatible descriptions, the structure of post-labour economies, and the large-scale architecture of the universe. The same theorems apply whether the system under investigation is a biological cell, a cognitive agent, a social institution, a patent regime, an ancient sign system, or the cosmic web.

The research programme that produced this theory was not conceived as a master plan. The earliest publication in the corpus — a metabolic model of the subjective acceleration of time perception with aging (Kriger, 1999; doi:10.5281/zenodo.18382312) — was written without any anticipation of where it would lead. Over twenty-seven years, engagement with problems in clinical decision-making, dynamical systems, epistemology, philosophy of mind, social theory, astrophysics, and cosmology gradually revealed a convergence: in every domain studied, the central finding was that a system's organisational logic is determined not by its substrate but by the formal constraints that persistence imposes on any structured entity operating under bounded resources and irreducible uncertainty. The crystallising moment came when the same formal vocabulary — viability sets, admissible transformations, epistemic constraints, cyclical closure — could be used to state and prove theorems about

cells, minds, civilisations, and the vacuum without changing a single axiom. The confirming moment came when this vocabulary, transferred to cosmology via the Tree-Top Meta-Method, produced a formula for the cosmological constant whose only input is the measured mass of quarks inside a proton.

The thesis is organised into ten parts.

Part I (Chapters 1–4) establishes the meta-theoretical foundation: the formal laws governing what any coherent system of knowledge must satisfy. The Law of Scale-Specific Principles demonstrates the impossibility of a final all-encompassing theory. The Formalisation Laws and the Principle of Definition-Dependent Provability establish the irreducibility of multiple formal representations. The Law of Imperative Uncertainty, the Law of Limit to Negation, the Law of Absurdity as Mercy, and the Epistemic Constraint Theory bound inference in any system. The Tree-Top Meta-Method and the RAFASSABRA method formalise the process of cross-domain transfer and frame-breaking inquiry. The Fractal Scaling principle identifies self-similar operational patterns across scales.

Part II (Chapters 5–8) develops the dynamical core: the theory of structural viability and the transformational basis of persistence, the formalisation of resilience and cyclical closure as conditions for self-sufficiency, the four laws of self-organising systems (Cooperation, Viability, Interference, Observability), and the derivation of Newtonian dynamics from structural axioms. The Structural Diagnostic Principle, the Constraint–Autonomy Compatibility Law, the Viability Mismatch Law, and the Structural Genesis of Dynamical Architecture establish the formal conditions under which dynamical laws acquire their specific mathematical form. The Conservation and Convertibility law identifies the structural necessity of conserved, interconvertible quantities in any persistent system.

Part III (Chapters 9–12) applies this framework to cognition and consciousness, demonstrating that the Structural Distortion Principle, predictive processing, representational isolation, self-referential modelling (consciousness), atemporal memory architecture, and the dynamics of mental disintegration are structural necessities derivable from the persistence principle. The evolutionary inevitability of predictive processing and the necessity of representational isolation are proved from physical constraints on latency and bandwidth.

Part IV (Chapters 13–16) extends the theory to social and multi-agent systems, formalising deception, autonomy suppression, structural non-neutrality, conflict escalation as phase

transition, the asymmetry of totalising ideals, extractive oscillators (from narcissistic dynamics to addiction), and institutional strategies including pre-integrative rejection, conceptual responsibility, non-participation, and assertion–dismantling cycles.

Part V (Chapters 17–18) addresses applications ranging from binary star formation and dormant neutron stars to artificial intelligence, the Fermi paradox, and the future of agency in post-scarcity environments.

Part VI (Chapters 19–21) develops the epistemology of incompatibility and the limits of description. The Theorem of Incompatible Truths proves that any sufficiently complex domain necessarily admits mutually exclusive but individually valid descriptions that cannot be unified. The Navigator Theorem derives five necessary properties of any system navigating among such descriptions. Against Causation establishes that causality is a compression artifact of bounded observers. The Persistence Filter Thesis explains why mathematics is effective. The Irreducible Residue Theorem demonstrates that fundamental constants are fixed points of self-consistency operators. The Inaccessibility Theorem proves that any consistent physics must contain structure inaccessible to internal observers. Evidence-Based Philosophy introduces four conditions for empirical philosophical inquiry. The RAFASSABRA method provides a nine-stage procedure for expanding the space of inquiry beyond trivial cognitive frames.

Part VII (Chapters 22–24) examines the formal structures of identity, consciousness, and communication. Identity is formalised as algorithmic configuration — a compact specification that generates characteristic behaviour on any compatible substrate. A four-regime motivational phase diagram maps the dynamics of desire under variable constraints. Archetypes are reconceived as reference frames rather than fixed types. Sub-personal inference without propositions is derived from structural distortion theory. Communication theory is extended through the critique of the finitude assumption, a coherence epistemology for inter-species communication, the analysis of pedagogical systems with invariant kernels, and the triadic model of hostile epistemic environments.

Part VIII (Chapters 25–26) addresses political economy, institutional pathology, and social design. Post-labour economics and the AUTT–UBI coupled mechanism are modelled through adaptive feedback control. Phase transitions in sociotechnical systems describe a civilisational cascade from labour-centric economies through algorithmic governance to homeostatic

civilisation. Voluntary participation is proved to be a necessary condition for stable collectivist equilibria. Patent monopolies are analysed as mechanisms of artificial scarcity requiring architectural reform under AI-driven innovation. Four principles of corruption establish that institutional corruption is an emergent property of institutional friction. The architecture of concealment reveals that information partitioning is adaptive at biological scales and pathological at social scales, with a formally computable transition boundary.

Part IX (Chapters 27–28) demonstrates the range of the formal methods through applications to civilisational design, science policy, computational humanities, and ontological foundations. Decentralised urbanism derives critical density thresholds for urban vulnerability. Philosophy of science is reconceived as policy architecture. The Indus Valley sign system is analysed through positional entropy and definition-dependent provability, revealing a structured registration code rather than encoded language. Testing Nothingness constructs a deductive chain from the incoherence of absolute nothingness to the structural conditions of life. The GEKS Index provides a composite measure of biological complexity. Leibniz's two-kingdom distinction is shown to be a proto-dual-system ontology defeated by pre-established harmony. And the goal preservation bound establishes a formal upper limit on the fidelity with which a recursively self-improving agent can maintain its original objectives.

Part X (Chapters 29–36) constitutes the cosmological verification of the entire framework. It demonstrates that every law and method developed in Parts I–IX, when applied to cosmology via the Tree-Top transfer, generates specific, falsifiable discoveries — not post hoc interpretations but predictions from first principles.

Chapter 29 documents how the Tree-Top Meta-Method, the RAFASSABRA method, and diagnostic laws (Absurdity as Mercy, Undecidability as Methodological Signal) broke the standard cosmological frame and generated the questions that the standard paradigm could not ask. Chapter 30 shows how Persistence Requires Invariants, the Irreducible Residue Theorem, the Law of Scale-Specific Principles, and Conservation with Convertibility computed the cosmological constant from QCD sigma terms with no free parameters: $\Lambda_0 = 1.8 \times 10^{-52}$ m^{-2} against the observed 1.1×10^{-52} m^{-2}. Chapter 31 demonstrates how self-consistency fixed points, cyclical closure, the four laws of self-organisation, constraint-autonomy compatibility, imperative uncertainty, and the Second Law of Infodynamics predict the cosmic web as a unique, necessary fixed-point structure — and introduces simulability $S = 1 - k$ as a new

epistemological criterion standing alongside falsifiability. Chapter 32 shows how the Persistence Theorem, the Law of Limit to Negation, phase transition analysis, and the Formalisation Laws resolve the gravitational singularity through the Kriger Limit — the fourth and final barrier in the astrophysical stability hierarchy (Eddington → Chandrasekhar → Oppenheimer–Volkoff → Kriger), where vacuum pressure with $w = -1$ halts collapse at nuclear density independently of mass. Chapter 33 demonstrates how the Theorem of Incompatible Truths, the Compression Imperative, Against Causation, Definition-Dependent Provability, the Viability Mismatch Law, the Architecture of Concealment, and Goal Preservation illuminate the dark matter question — and how 175 galaxy rotation curves from the SPARC database are reproduced without dark matter particles through vacuum halo profiles derived from a single QCD parameter. Chapter 34 shows how viability set theory generates specific astrophysical predictions: binary dominance as a persistence-selected outcome, dormant neutron stars as systems at the viability boundary, and catastrophic phase transitions (supernovae) as necessary conditions for complexity dissemination. Chapter 35 applies the Inaccessibility Theorem, the Navigator Theorem, Evidence-Based Philosophy, the Dual-Source Decomposition, and the Structural Distortion Principle to the cosmological observer, deriving the Frozen Universe Illusion, the SETI temporal mismatch, emergent time, and the biospheric contribution to cosmic complexity. Chapter 36 completes the recursive self-application: the persistence framework describes not only the universe but the scientific community that studies it — through autonomy suppression, hostile epistemic environments, assertion-dismantling cycles, conflict as phase transition, the four principles of corruption, extractive oscillators, pedagogical kernels, and simulability as the ultimate epistemological criterion.

Four methodological principles distinguish this work. First, *constraint derivation* rather than model postulation: instead of proposing a model and checking whether it fits data, the approach identifies constraints that any model must satisfy and derives structural consequences. Second, *substrate abstraction with formal proof*: cross-domain transfers are not asserted through analogy but demonstrated through rigorous mathematical argument — the axioms of Chapter 5 applied to the vacuum are the same axioms applied to cells, with no modification. Third, *recursive self-application*: the theory applies to the complex system that produced it — the scientific enterprise itself — and the thesis explicitly addresses this

reflexive dimension in Chapter 36. Fourth, *falsifiable prediction from first principles*: the cosmological programme generates specific numerical predictions ($\Lambda_0 = 1.8 \times 10^{-52}$ m^{-2}, $\alpha = 0.005$, $\rho_K \approx 5\rho_0$, $M_{tr} \approx 4.0\ M\odot$, $k_2 > 0$ for supermassive compact objects) that are testable by current and near-future observational facilities (DESI, Euclid, SKA, JWST, LISA, Einstein Telescope).

The overarching method is formalised as the Tree-Top Meta-Method (*The Tree-Top Meta-Method: Inquiry to the Limits of Domain-Specific Formal Generalisation with Consequent Application Across Other Domains*, Kriger, 2026; doi:10.5281/zenodo.18665621): generalise formally until generalisation is no longer possible, then derive domain-specific consequences from what remains true. The method was applied once — from complex systems theory to cosmology — and produced a cosmological programme that computes what the standard model treats as free parameters. The theory does not claim to be a Theory of Everything; the Asymmetry of Totalising Ideals (Chapter 14) proves that such theories are self-defeating. It claims to be the best available navigation — and the simulability criterion ($S = 1 - k$) provides a formal measure of what "best available" means.

Every formal result presented here has been published as a standalone paper, each with its own derivation and domain-specific discussion. What this thesis provides that no single paper can is the architecture: the demonstration that these results form a coherent, unified theory with a single organising principle, spanning from the incoherence of absolute nothingness (Chapter 28) through the viability of biological cells (Chapter 5), the architecture of consciousness (Chapter 11), the pathology of institutions (Chapter 26), and the structure of ancient sign systems (Chapter 27) to the computation of the cosmological constant (Chapter 30), the resolution of gravitational singularities (Chapter 32), and the reproduction of 175 galaxy rotation curves without dark matter particles (Chapter 33). The theory is, by its own meta-theoretical results, necessarily incomplete. The Formalisation Laws guarantee that every formal framework has boundaries; the Law of Scale-Specific Principles guarantees that no single theory spans all scales; the Theorem of Incompatible Truths guarantees that alternative descriptions of the same phenomena will always exist. The thesis closes by documenting the scientific community's reception of these results and by identifying the open questions that the theory generates — questions that, according to the theory itself, are not deficiencies but structural necessities of any coherent research programme operating under the irreducible

constraints of bounded observation, finite resources, and the persistence conditions that make inquiry possible.

PART I. THE ARCHITECTURE OF FORMAL KNOWLEDGE

CHAPTER 1. THE PROBLEM OF UNIFICATION ACROSS DOMAINS

The history of science reveals a recurring and remarkable pattern: formal structures discovered in one domain reappear, often in precise mathematical analogy, in domains that share no physical substrate. Oscillatory dynamics appear in predator–prey systems, in business cycles, in neural circuits, and in coupled chemical reactions. Phase transitions characterise not only the behaviour of matter near critical temperatures but also the onset of consensus in social networks, the emergence of synchronised firing in neuronal populations, and the escalation of conflict in organisational systems. Feedback loops, viability constraints, and information-theoretic limits surface in contexts as diverse as gene regulation, institutional governance, and cosmological structure formation.

This recurrence is not coincidental, nor is it merely metaphorical. The central claim of this thesis is that there exist formal structures—mathematical constraints on organised complexity—that are substrate-independent: they apply to any system that must maintain itself over time, regardless of what that system is made of. The motivation for this claim is not reductionist. The argument is not that social systems "are" physical systems, or that cognition "reduces to" thermodynamics. Reduction requires identity of substrate; the claim here is one of formal isomorphism. Two systems may share no material basis whatsoever and yet be governed by the same mathematical constraints, because both face the same structural problem: the problem of persistence under bounded resources and irreducible uncertainty.

This introductory chapter situates the research programme within the landscape of prior attempts at unification. Ludwig von Bertalanffy's General System Theory, W. Ross Ashby's cybernetics, Niklas Luhmann's autopoietic social systems, John Holland's complex adaptive systems, and Stuart Kauffman's investigations of self-organisation at the edge of chaos all represent attempts to identify principles that transcend specific domains. Each has produced valuable insights; none has achieved a formal unification that is simultaneously rigorous, substrate-independent, and empirically testable across the full range of complex systems. The present work differs from its predecessors in three respects. First, it insists on formal proof rather than suggestive analogy. Second, it begins not with a model of how systems behave but with the constraints that any system must satisfy. Third, it applies its own results to itself: a

theory of complex systems that does not account for the complex system that produced it is, by its own standards, incomplete.

1.1. The Research Programme: From Individual Publications to Unified Theory

The publications underlying this thesis were not composed with a preconceived master plan. They emerged over twenty-seven years of engagement with problems in clinical decision-making (Kriger et al., 2000), dynamical systems modelling of conflict (Kriger, 2005), formal epistemology, predictive processing in cognitive science, self-organisation theory, institutional dynamics, and most recently astrophysical applications of the persistence framework. The convergence was discovered, not designed.

That convergence can be summarised in a single sentence: in every domain studied, the central finding was that the system's organisational logic is determined not by its substrate but by the formal constraints that persistence imposes on any structured entity operating under bounded resources and irreducible uncertainty. A clinical decision system faces the same formal problem as a self-organising ecosystem or a gravitationally bound binary star: it must maintain structural viability under conditions it cannot fully control or predict.

The recognition that these results form a unified theory came gradually. The early publications addressed domain-specific problems. The middle period saw the emergence of cross-domain patterns. The later publications—particularly from 2024 onward—were written with the explicit aim of formalising the unified framework. This thesis presents the completed architecture: the demonstration that a small number of formal principles, rigorously stated and proved, generate the organisational logic of physical, cognitive, and social systems as special cases.

1.2. Methodological Principles

Three methodological principles distinguish this research programme from standard approaches in systems science. The first is constraint derivation rather than model postulation. The standard approach in mathematical modelling is to propose a model and then check whether the model's predictions match empirical data. The approach adopted here identifies the constraints that any model of the phenomenon must satisfy. The results are derived from the constraints themselves, not from any particular model, yielding results of maximal generality. Chapter 7 illustrates this with four laws of self-organising systems—Cooperation, Viability, Interference, and Observability—which are not properties of a specific model of

self-organisation but properties that any self-organising system must have, proved by showing that any system violating one of these laws cannot be self-organising in a formally defined sense.

The second principle is substrate abstraction with formal proof. Cross-domain analogies are commonplace in the literature on complex systems, but they carry no deductive force. What this thesis insists upon is the formal demonstration that the analogy rests on a shared mathematical structure. When the thesis claims that conflict escalation is a phase transition, it means that the mathematical structure of conflict escalation satisfies the formal definition of a phase transition in a precisely specified dynamical system. The substrate is different; the mathematics is the same.

The third principle is recursive self-application. A theory of complex systems that does not apply to the complex system that produced it—the human cognitive apparatus, the scientific community, the process of formalisation itself—is incomplete by its own standards. The meta-theoretical results of Part I are explicitly applied to the theory itself: the formalisation laws apply to this thesis's formalisations; the epistemic constraints apply to this thesis's claims; the structural viability framework applies to the research programme that produced the theory.

1.3. Overview of the Thesis

The thesis is organised into five parts. Part I (Chapters 1–4) establishes the meta-theoretical foundation: the formal constraints on knowledge, formalisation, and inference that any theory of complex systems must respect. Part II (Chapters 5–8) develops the dynamical core: the theory of structural viability, the formalisation of resilience, the laws of self-organisation, and the conditions for self-sufficiency, along with the derivation of Newtonian dynamics from structural axioms, the Constraint–Autonomy Compatibility Law, and the Viability Mismatch Law. Part III (Chapters 9–12) applies the framework to cognition and consciousness, deriving the architecture of perception, prediction, memory, and awareness as structural necessities, and analysing mental disintegration and biological decision distortion as dynamically structured phenomena. Part IV (Chapters 13–16) extends the theory to social and multi-agent systems, formalising the dynamics of deception, autonomy suppression, conflict escalation, totalising ideals, extractive oscillators, and institutional strategies. Part V (Chapters 17–18) demonstrates the theory's predictive power in applications ranging from astrophysics to artificial intelligence and the future of civilisation.

1.4. What This Thesis Does Not Claim

It does not claim that all complex systems are the same. Substrate matters: a brain is not a star, and a social institution is not a chemical oscillator. The claim is that certain formal constraints are shared, not that the systems themselves are identical. It does not claim that the theory replaces domain-specific science. Physics, neuroscience, sociology, and economics each have their own methods, data, and explanatory frameworks that are indispensable. The unified structural theory operates at a different level of abstraction: it identifies formal constraints that domain-specific theories must satisfy and derives consequences that hold across domains. It does not claim completeness. The meta-theoretical results of Part I guarantee that no formal theory can be complete in the sense of resolving all questions within its scope. This theory is no exception, and the thesis concludes by identifying the open questions it generates.

CHAPTER 2. SCALE-SPECIFIC LAWS AND THE IMPOSSIBILITY OF A FINAL THEORY

The dream of a final theory—a single formal system that captures all of reality at all scales simultaneously—is one of the most persistent aspirations of the scientific imagination. From Laplace's demon to the modern quest for a theory of everything in fundamental physics, the idea that the universe might be fully described by a closed set of equations has exercised a powerful hold on the scientific community. This chapter presents a formal result that constrains this aspiration: the Law of Scale-Specific Principles (Kriger, 2024, "No final theory: Law of scale-specific principles," https://doi.org/10.5281/zenodo.18099738).

The dream of a final theory — a single formal system capturing all of reality at every scale simultaneously — is among the most enduring aspirations of the scientific imagination. From Laplace's omniscient demon to the modern pursuit of a theory of everything in fundamental physics, the conviction that the universe might yield to a closed set of equations has exercised a powerful hold on the scientific community. This chapter presents a foundational result that constrains this aspiration: the Law of Scale-Specific Principles (Kriger, 2024, *No Final Theory: Law of Scale-Specific Principles*, https://doi.org/10.5281/zenodo.18099738).

The law states that whenever extrapolation between physical scales fails while stable, reproducible observables persist, mathematics itself requires the search for governing principles specific to the new scale. This is not a philosophical proposal or a sociological account of scientific revolutions. It is a structural consequence of renormalization theory and the logic of effective descriptions. Its implications are immediate and far-reaching: no final theory can encompass all scales simultaneously, because the principles that govern each scale are not derivable from those at any other.

2.1. Scale as a Physical Concept

Before the formal argument can be stated, it is necessary to grasp the central claim that underlies it: scale is not a technical detail or a parameter of convenience but a fundamental physical concept that restructures what counts as real, relevant, and causally operative.

The intuition that a genuine law of nature must operate identically across all scales — that what governs atoms should, in some rational sense, also govern stars — runs deep in scientific thought. It reflects a desire for coherence and a belief that truth, if truly foundational, must be uniform. But this expectation does not survive the test of physical reality. What appears to be a minor change in scale routinely results in a profound transformation of dynamics. A single

water molecule and a cup of water are composed of identical constituents, yet they exhibit fundamentally different behaviors. The quantum-mechanical laws governing individual molecular interactions do not simply scale up to explain fluid dynamics or thermodynamic equilibria. New patterns emerge, new regularities take form, and the vocabulary of explanation must shift accordingly.

To understand scale as a physical concept is to recognize that it alters not merely the quantity of information but the quality of what is physically relevant. Each scale defines its own regime of dominant variables, discarding some and elevating others. Degrees of freedom indispensable in a microscopic description — the exact position and momentum of each particle — become irrelevant at higher scales. They are not lost but absorbed into the background of collective behavior, overwhelmed by emergent features that are qualitatively different. Temperature cannot be found in a single molecule. Entropy does not reside in a solitary atom. These quantities are not illusions; they are scale-specific constructs that emerge only when sufficient components interact. Coarse-graining, in this context, is not an unfortunate necessity but a condition for meaningful description.

This decoupling between levels has a decisive consequence: increasing microscopic precision does not confer macroscopic control. The behaviors of clouds, ecosystems, and living tissues are governed by principles that depend on organization, not on the identity of components. The illusion that complete knowledge at the smallest scale would unlock total command over larger systems collapses once one observes how genuinely new structures appear with each shift in scale. These emergent laws are not fragments of a deeper whole. They are laws in their own right, meaningful only within the context for which they are formulated.

Phase transitions make this point with particular force. The alignment of spins in a magnet, the condensation of vapor into liquid, the self-organization of cells into tissues — none of these can be anticipated from the behavior of isolated parts. They emerge not by extension but by transformation. The very space of possibilities changes, and with it, the meaning of causality. Once emergent properties arise, they condition the behavior of their constituents through feedback loops and higher-level constraints. Understanding one level in perfect detail does not grant understanding of another; each level demands its own framework.

2.2. The Renormalization Group and the Architecture of Scale

The mathematical framework that makes these observations precise is the renormalization group, whose development is principally owed to the work of Kenneth G. Wilson. Wilson's insight was not to refine the existing tools of renormalization — the ad hoc subtraction techniques that had been used to manage infinities in quantum field theory — but to overturn the assumptions beneath them. Rather than viewing renormalization as a mathematical fix, Wilson reinterpreted it as a physical process: one that discloses the layered structure of nature itself.

At the heart of Wilson's reconceptualization is the recognition that physical laws are never fully scale-independent. Every theory operates within a range of applicability, defined by the energy or length scales it can meaningfully describe. To move across scales is not to extend a theory linearly but to transform it. The parameters of a theory — masses, coupling constants, field strengths — do not remain fixed as the scale shifts; they *flow*. This flow of parameters under scale transformations became the foundation of the renormalization group: a conceptual and mathematical framework that reveals how physical systems evolve when viewed at different levels of resolution.

Wilson's approach treats scale transformations as the fundamental operation of physical description. By systematically integrating out degrees of freedom at smaller scales, one obtains effective theories describing the system's behavior at larger scales. This process is governed by rules that track how parameters change, and it often leads to *fixed points* — values toward which the parameters flow under repeated transformations. Fixed points correspond to *universality classes*: sets of systems that, despite microscopic differences, exhibit identical behavior at large scales. The emergence of universality explains why systems as varied as magnets, fluids near critical points, and certain particle interactions can be described by the same equations despite their differing microstructures.

The key revelation is that the infinities plaguing earlier formulations of quantum field theory were symptoms of forcing a theory beyond its range. When the theory is restructured to account for scale dependence, the infinities vanish — not by subtraction but by reinterpretation. What had been viewed as divergences are understood as signals of new behavior, indicating where one description must yield to another.

Two consequences of Wilson's framework are particularly important for the present argument:

First, the concept of *irrelevant variables*. As a system is examined at progressively larger scales, certain variables and interactions lose their influence over observable behavior. They flow to zero under scale transformation and cease to affect outcomes. Their elimination is not arbitrary; it is systematic and principled. What remains are the *relevant* and *marginal* variables — those that shape scaling behavior and determine the system's universality class. This filtration is what makes effective theories possible, and it demonstrates that information is not simply carried upward across scales but is transformed, filtered, and often made irrelevant by the dynamics of the new level.

Second, the phenomenon of *universality*. Systems with entirely different microscopic compositions can exhibit identical macroscopic behavior when they belong to the same universality class — that is, when they share the same symmetries, dimensionality, and order parameter structure. A ferromagnet near its Curie temperature and a fluid near its critical point, though composed of entirely different materials, exhibit the same critical exponents and scaling relations. Their commonality is not one of origin but of transformation: a coherence that arises through the systematic elimination of irrelevant detail and the retention of what is dynamically stable.

Universality demonstrates that microscopic causal storytelling is fundamentally incomplete. If systems with different microscopic rules yield indistinguishable behavior at larger scales, then explanation must shift its focus from the accumulation of micro-level detail to the identification of structural invariants that persist under coarse-graining. Behavior, not substance, becomes the decisive feature of physical description.

2.3. Effective Theories and the Autonomy of Levels

The concept of an effective theory follows directly from the renormalization group framework. An effective theory is not a temporary approximation awaiting completion by a deeper truth. It is a framework designed to describe phenomena within a particular domain of energy, length, or complexity, without appealing to the underlying details that lie outside its scale. Its power lies not in the illusion of universality but in its precision within boundaries.

The autonomy of effective theories is particularly striking when they achieve predictive power independent of any reference to lower-scale mechanisms. The Navier–Stokes equations describe the motion of viscous fluids with extraordinary success — capturing turbulence, flow, and pressure — without requiring any knowledge of molecular interactions. These equations

are not approximations of atomic behavior; they are emergent laws that become meaningful only when vast numbers of particles act together in coherent patterns. The molecules exist, but for the purposes of fluid mechanics, they do not matter.

Similarly, the thermodynamic laws governing heat, energy exchange, and entropy have no clear analogue in the realm of individual particles. Thermodynamics is not a placeholder for a more "real" microscopic account. It is a complete and consistent framework offering powerful predictions in contexts where the microscopic story is both inaccessible and irrelevant.

This autonomy reveals something fundamental about the structure of physical knowledge: the laws governing collective behavior are not simply reducible to the sum of microscopic equations. They must be formulated anew, using variables and principles that have no meaning at smaller scales. Effective theories express the autonomy of levels — they demonstrate that different layers of nature operate according to different rules, and that these rules are not derivable in any straightforward manner from what lies beneath.

To treat such theories as mere approximations is to misunderstand their role. Approximation implies proximity to a more exact account, closeness to a final truth. Effective theories reject that ambition. They capture what is essential at a given level. Their justification is empirical, structural, and contextual — not metaphysical.

2.4. The Formal Argument: The Law of Scale-Specific Principles

With this background in place, the Law of Scale-Specific Principles can be stated precisely and its proof sketched.

Statement of the Law

The law states: *Whenever extrapolation between physical scales fails while stable, reproducible observables persist, mathematics requires the search for governing principles specific to the new scale.*

More formally: if a physical system exhibits stable, reproducible phenomena at a scale S_2 that cannot be accounted for by controlled extrapolation of the principles valid at scale S_1, then the governing principles of S_2 are necessarily different in kind from those at S_1. The stable observables at S_2 require dominant organizing structures of their own, and these structures constitute irreducible principles of that scale.

The Proof Strategy

The proof draws on the mathematical architecture of renormalization theory, effective field theory, decoupling, and fixed-point analysis. It proceeds through three central arguments:

Argument from Renormalization Group Flow. The renormalization group describes how a theory's parameters evolve under changes in scale. As one integrates out short-distance degrees of freedom and moves to larger scales, certain operators become irrelevant — they are suppressed by powers of the ratio between the short-distance cutoff and the observation scale. The terms that governed behavior at S_1 are precisely those that flow to zero or become negligible at S_2. If stable observables persist at S_2, they cannot arise from asymptotically suppressed terms. They must be governed by operators that are relevant or marginal at S_2 — operators that define the effective theory at that scale. These operators constitute principles specific to S_2 that are not derivable from the dominant operators at S_1, because the latter have been systematically eliminated by the flow.

Argument from the Non-Invertibility of Coarse-Graining. The mapping between scales is many-to-one: many microscopic configurations correspond to a single macroscopic state. This mapping is not bijective, and consequently it is not invertible. Information is lost in both directions. The thermodynamic description loses information about molecular configurations; the molecular description, constrained to a finite set of observables, cannot reconstruct thermodynamic principles without additional assumptions that are themselves scale-specific. Because the mapping between scale-specific descriptions is structurally irreversible, principles valid at one scale cannot be deduced from principles at another. The barrier is not computational but logical — it follows from the mathematical structure of the coarse-graining operation itself.

Argument from the Failure of Generic Repair. When extrapolation from S_1 to S_2 fails — that is, when the theory formulated at S_1 does not predict the stable phenomena observed at S_2 — one might hope that the failure can be repaired by parameter tuning, higher-order corrections, or the addition of new terms within the existing framework. The proof shows that this hope is generically unfounded. If the relevant operators at S_2 belong to a different universality class than those at S_1, no perturbative adjustment of S_1-parameters will produce

the correct S_2-behavior. The failure of extrapolation is not a quantitative shortfall but a qualitative mismatch: the organizing structures at S_2 differ in kind, not merely in degree, from those at S_1. The new principles are not refinements of the old ones; they are structurally independent.

The Essential Point

The impossibility of cross-scale derivation is not a limitation of current mathematical techniques but a structural feature of the relationship between scales. No future advance in mathematics or physics will overcome it, because the barrier is logical, not technical. It follows from the many-to-one character of scale mappings, the asymptotic suppression of operators under renormalization flow, and the existence of distinct universality classes at different scales.

2.5. The Corollary: No Final Theory

The corollary is immediate: no final theory can encompass all scales simultaneously. A theory formulated at scale S captures the principles valid at S but necessarily misses principles that emerge at other scales. A putative final theory would need to operate at all scales at once — but this is precisely what the Law of Scale-Specific Principles shows to be impossible. Every formal description commits to a scale (or a finite set of scales), and in doing so, it excludes principles valid only at scales it does not represent.

The history of physics provides repeated confirmation of this pattern. Each of the great conceptual disruptions — classical mechanics to quantum mechanics, Newtonian gravity to general relativity, microscopic dynamics to thermodynamics — followed the same structure. Prior laws remained valid within their original domains, yet extrapolation beyond those domains failed in the presence of stable phenomena. Progress occurred only when new principles were introduced, reorganizing the hierarchy of explanation rather than refining parameters within an exhausted framework. These transitions were not historical accidents; they are predictable consequences of the Law of Scale-Specific Principles. Each emerged precisely where extrapolation collapsed and stability persisted.

The same pattern governs unresolved problems in contemporary physics. Persistent anomalies — from the cosmological constant problem to the interpretation of quantum gravity — are not merely technical difficulties to be resolved by increasingly elaborate extensions of existing models. When such anomalies exhibit stability and coherence across a new regime, they

indicate that the conceptual architecture itself has reached its boundary. Continuing to force extrapolation beyond this boundary delays understanding.

2.6. The Quiet Failure of Reductionism

The Law of Scale-Specific Principles provides a formal framework for understanding why reductionism fails — and why its failure is characteristically quiet rather than dramatic.

Reductionism rests on the conviction that systems can be fully understood by analyzing their smallest components and the rules governing those components' interactions. The conviction is compelling, often productive within bounded domains, and sometimes accurate. But the Law of Scale-Specific Principles shows that between levels of organization lies not a mere aggregation of elements but a transformation of what is relevant. The system, as it scales, ceases to be a sum of its parts and becomes a structure governed by new principles — principles that cannot be derived by following the parts upward.

What makes this failure quiet is that reductionism continues to produce formally correct but explanatorily empty descriptions. It returns answers that are not false but inadequate. It describes behavior without capturing structure, calculates without illuminating, catalogs without explaining. The parts are correct, the rules accurate, the details abundant — yet the phenomena resist intelligibility. The failure is conceptual, not technical. It produces no crisis, triggers no reevaluation, and offers no clear signal that the model no longer maps onto the world. Because reductionism retains its formal correctness, it becomes especially difficult to dislodge.

The Law of Scale-Specific Principles replaces this silent breakdown with a clear diagnostic: when stable phenomena persist at a scale where extrapolation from below has failed, new principles are not optional but mathematically compelled. The correct response to failed extrapolation is not defensive refinement but principled exploration.

2.7. Relativity of Laws Without Relativism

A natural concern arises: if laws are scale-dependent, does this not threaten scientific objectivity? If no law holds everywhere, does the distinction between rigorous science and speculation dissolve?

This concern rests on a conflation between *relativity* and *relativism*. Scale-relativity acknowledges that no single formulation governs all physical behavior uniformly across domains. It does not assert that anything is permissible. Different scales expose different

structures of regularity, but theories remain constrained — by internal consistency, by empirical adequacy, and by the formal structure of how systems behave under scale transformations. The renormalization group enforces objectivity within domains. Fixed points are not arbitrary; they are determined by symmetry, dimensionality, and interaction structure. Only certain behaviors survive renormalization flow, and these are the ones that form the foundation of reliable, domain-bound theories.

Scientific knowledge does not require a single, unchanging foundation. It requires that within each regime of description, the relations among variables be coherent, predictive, and independent of the observer's preferences. Scale is not an artifact of how observers interact with reality; it is a structural feature of the systems themselves. The laws governing turbulence in a fluid, the correlations near a critical point, or the vibrational modes of a solid at low temperatures do not change depending on who examines them. They emerge from the internal dynamics of the system under the constraints of scale.

Objectivity is preserved not by universality across all levels but by stability within each level. A theory is strengthened, not weakened, by being properly confined to its scale. The laws of thermodynamics are not invalidated by their inapplicability at atomic dimensions. They remain among the most exacting theories in science precisely because they describe with clarity the behavior of systems composed of large numbers of particles.

2.8. Relationship to Established Positions

The Law of Scale-Specific Principles engages with several established positions in the philosophy of science and sharpens or supersedes each of them.

Multiple Realizability. The law is consistent with, but stronger than, the thesis of multiple realizability familiar from the philosophy of mind — the claim that the same higher-level state can be realized by different lower-level states. Multiple realizability asserts a one-to-many relationship between levels; the Law of Scale-Specific Principles asserts that this many-to-one mapping generates irreducible novelty at each level — principles that cannot be derived from below even with complete knowledge of lower-level dynamics.

Emergentism. The law differs from standard emergentism in a critical respect: it does not rely on the concept of downward causation, which has proved philosophically contentious. The claim is not that higher-level entities exert some mysterious causal influence on lower-level entities. The claim is that higher-level principles are logically independent of lower-level

principles. They cannot be derived from them even in principle, given complete knowledge of the lower-level dynamics. The novelty is epistemic and formal, not causal.

Eliminative Reductionism. The law directly opposes eliminative reductionism, which holds that higher-level descriptions are mere approximations to a more fundamental reality. The proof shows this view to be formally untenable: the principles valid at higher scales are not approximations of principles at lower scales. They are distinct principles, valid in their own right, capturing structural features that do not exist at the lower scale.

Gödel's Incompleteness Theorems. The result connects to Gödel's incompleteness theorems, which show that any sufficiently powerful formal system contains true statements that cannot be proved within the system. The Law of Scale-Specific Principles can be viewed as an analogue of incompleteness for multi-scale systems: no single-scale theory can prove all the truths about a multi-scale system. The parallel is not merely suggestive; the proof technique shares structural features with Gödelian arguments, exploiting the limitations of formal systems that are powerful enough to describe certain domains but structurally incapable of capturing all domains simultaneously.

Effective Field Theory. The law draws its mathematical foundation from the framework of effective field theory and the renormalization group. But it goes beyond the standard EFT perspective by elevating what is usually treated as a practical methodology into a foundational principle. The standard view holds that effective theories are useful approximations; the Law of Scale-Specific Principles holds that they are the *form that valid knowledge takes* in a world where different scales generate qualitatively distinct regimes of behavior.

2.9. Implications for the Present Theory

The Law of Scale-Specific Principles applies to the present theory as well. The unified structural theory of complex systems developed in this thesis is itself formulated at a particular level of abstraction — the level of formal constraints on persistence and viability. It does not and cannot replace the scale-specific theories of physics, biology, neuroscience, or sociology. This self-limitation is not a weakness. It is a consequence of the law's recursive self-application and a demonstration of the theory's consistency. A theory that claimed to be the final word on all of reality would, by the Law of Scale-Specific Principles, be either inconsistent or operating at a single scale while pretending to span all scales. The present theory makes no such claim.

What it does is identify the structural constraints that all scale-specific theories share and derive consequences from those constraints. The persistence principle, the viability framework, the four laws of self-organization — these are not principles at any particular scale. They are constraints that hold at every scale, precisely because they characterize what it means for a system to be organized at all. They do not replace scale-specific physics or biology or sociology; they identify what all scale-specific theories have in common.

This is precisely the kind of unification that the Law of Scale-Specific Principles permits. Unification is not the reduction of all phenomena to a single set of equations. Unification is the identification of formal constraints that hold across scales — constraints that every scale-specific theory must satisfy, not because they are derivable from a single fundamental theory, but because they follow from the structural requirements of persistence and viability that any organized system faces regardless of its scale of operation.

The present theory is thus unified in scope but modest in its claims: it does not explain everything, and it knows why it cannot. It operates at the level of formal structure and identifies what holds across scales, while explicitly acknowledging the existence of scale-specific principles that lie beyond its reach.

CHAPTER 3. FORMALIZATION LAWS AND DEFINITION-DEPENDENT PROVABILITY

Every act of scientific inquiry begins with an act of formalization: the translation of a phenomenon from the flux of experience into a formal language in which statements can be made, relationships can be defined, and consequences can be derived. This chapter presents two closely related results that characterize the nature and limitations of this foundational act. The first is the system of *Formalization Laws*,[1] comprising the Law of Total Formalizability, the Law of Boundary-Conditioned Verification, and the Law of Plurality and Incompleteness. The second is the *Principle of Definition-Dependent Provability*.[2]

Together, these results establish a meta-mathematical framework of considerable power. They show that the choice of formal language is not a neutral or transparent act but a constitutive one that shapes what can be proved, what can be observed, and what remains invisible within the resulting formal system. The Formalization Laws characterize the structural consequences of translating a phenomenon into a formal description; the Principle of Definition-Dependent Provability shows that provability itself is a function of the definitions chosen for the terms in which statements are expressed. Combined, they yield a disciplined, pluralistic, and non-absolutist epistemology for the study of complex systems.

3.1 The Formalization Laws

The Formalization Laws characterize the structural consequences of translating a phenomenon into a formal language. The central insight is that any act of formalization imposes a boundary—and this boundary is not a defect of the formal description but a constitutive feature of it. Without boundaries, there is no formal object; with boundaries, there is necessarily a distinction between what the formalization captures and what it excludes. The laws were first formulated and proved in *Foundations of Intellectual Substrate Mining* and are treated here as established methodological results governing the analysis that follows.

---

[1]Kriger, B. (2025). "Formalization Laws: Structure, Boundaries, and the Necessity of Plural Representations." https://doi.org/10.5281/zenodo.18099527. The three laws were first formulated and proved in Foundations of Intellectual Substrate Mining (Amazon, 2025), Chapter 9, pp. 125–164.

[2]Kriger, B. (2026). "The Principle of Definition-Dependent Provability: Provability as a Function of Definition." https://doi.org/10.5281/zenodo.18207348

The framework rests on several precisely defined concepts. A *structure* is an ordered pair S = (X, R), where X is a nonempty set of states or elements and R is a nonempty family of admissible relations, operations, or predicates on X.[3] A *formal representation* of S is a triple (L, C, M)—a formal language, a constraint set, and a model class—together with a structure-preserving mapping.[4] A representation is *adequate* if the resulting model class is a singleton up to isomorphism or an equivalence class that agrees on all relations in R.[5] A representation is *underconstrained* if it admits multiple non-equivalent models that disagree on at least one relation in R.

3.1.1 The Law of Total Formalizability

The first Formalization Law establishes the possibility condition of formal representation. It states that for any phenomenon whose associated structure is determinate, there exists an adequate formal representation. Apparent resistance to formalization arises exclusively from underconstrained representational schemes, not from the nature of the phenomenon itself.

The law's force lies in its relocation of unformalizability from ontology to constraint theory. Claims that a domain—consciousness, aesthetics, social systems—is inherently unformalizable are reinterpreted as claims that no adequate constraint set has yet been specified. They describe a limitation of the current representational effort, not a property of the domain. Structure and formal capturability are coextensive: wherever a phenomenon exhibits constraint-sensitive regularity, where not all descriptions, transitions, or configurations are equally admissible, formalization is possible given a constraint set sufficient to fix the structure.

---

[3] A structure S = (X, R) is determinate if R imposes nontrivial constraints—i.e., the set of admissible configurations is a proper subset of all logically possible configurations on X.

[4] A formal representation of S is a triple (L, C, M), where L is a formal language, C a constraint set, and M a nonempty class of models of L ∪ C, together with a mapping φ: X → Dom(M) that preserves the relations in R.

[5] A representation (L, C, M) is adequate if M is a singleton up to isomorphism, or M forms an equivalence class in which any two models agree on all relations in R.

The proof proceeds constructively. Given a determinate structure S = (X, R), one begins with a formal language L capable of expressing the elements of X and the relations in R, and an initially empty constraint set. Because S is determinate, at least one nontrivial constraint exists and can be encoded as an axiom or rule in L. Each additional constraint strictly reduces the admissible model class. Because the structure is determinate, this iterative refinement terminates at a finite stage where the resulting representation is adequate. If no such representation existed, the structure would be indeterminate, contradicting the premise.

The law does not compel formalization where structure is absent. Where no regularity exists—where R is empty or trivial—there is nothing to formalize and the law does not apply. Nor does the law demand that formalization be exhaustive. Adequacy, not absolute completeness, is the criterion: a formal model is adequate if it captures the determinate structure under consideration within the bounds specified.

3.1.2 The Law of Boundary-Conditioned Verification

The second Formalization Law restricts the scope of validity of any formal claim. It states that any statement, law, or proposition is verifiable—and hence capable of being true or false—only relative to a boundary set that determines the domain, scale, assumptions, admissible operations, and verification procedures under which the statement is evaluated.[6] Outside such boundaries, verification is undefined rather than false.

The conceptual shift is fundamental: truth is boundary-dependent. A statement is not true *simpliciter*; it is true within a representational system whose constraints render its verification meaningful. Removing these constraints does not strengthen truth—it dissolves the conditions under which truth or falsehood can be assessed. Falsehood acquires a parallel reinterpretation: a statement is often false not because it misrepresents reality, but because it is asserted outside the boundary conditions under which it could be verified. Claims of the form “X is impossible,” “Y always occurs,” or “Z never happens” frequently fail not because their content is wrong, but because their scope is unbounded.

[6]A boundary set B is a finite or recursively specifiable collection of conditions restricting the domain of discourse, admissible transformations, relevant scales, and permissible idealizations.

The proof is direct. Verification requires admissible procedures; procedures require specification of domain, operations, and limits; these specifications constitute boundary conditions. Without boundary conditions, admissible procedures are undefined, and all statements become unverifiable. Truth requires verification; verification requires boundaries; therefore truth is boundary-conditioned.

An important implication concerns the impossibility of universal truth. A universal claim must be verifiable independently of any boundary set. But verification requires boundaries that delimit admissible procedures and observations. Demanding universality is demanding verification without conditions—a contradiction. This is not relativism: boundaries are not perspectives, preferences, or conventions but structural constraints imposed by the system under investigation. Boundary-conditioned verification demands restraint in assertion, precision in scope, and accountability in reasoning.

3.1.3 The Law of Plurality and Incompleteness

The third Formalization Law addresses the necessary multiplicity and partiality of formal representations. It states that for any phenomenon with determinate structure, there exist at least two distinct adequate formal representations, corresponding to different boundary and constraint sets, and no single representation can capture all relevant aspects of the structure simultaneously. Completeness with respect to the phenomenon is impossible within any finite or even recursively enumerable set of constraints; only partial, selective adequacy is achievable.

This law follows from the conjunction of the first two. The Law of Total Formalizability guarantees that structure is capturable under some constraint set. The Law of Boundary-Conditioned Verification shows that every such capture is scoped to specific boundaries. Because structure itself is multi-layered and can be delimited in non-equivalent ways—different partitions of variables, different scales of observation, different idealizations—multiple non-equivalent constraint regimes can each yield adequate but incomplete models. The phenomenon is not indeterminate; it is overdetermined relative to any single representational choice.

Incompleteness is therefore not a defect of formalization but a consequence of its strength. A model is adequate precisely because it enforces strong constraints that fix certain invariants; those very constraints exclude alternative descriptions that might be equally valid

under different boundaries. What one model leaves indeterminate, another can fix—but at the cost of indeterminizing something else. Plurality is not a sign of failure but of fidelity to the phenomenon's complexity.

The proof invokes two lemmas. First, given a determinate but non-trivially simple structure, there exist multiple ways to partition variables or impose boundaries, each yielding a different constraint set that fixes the structure adequately but excludes different aspects. Second, the space of possible boundary sets is at least countably infinite—different scales, idealizations, variable selections—and any fixed constraint set operates within one regime, leaving others indeterminate. Exhausting all regimes would require an infinite conjunction of incompatible constraints, which is impossible within a single representation.

3.1.4 The Three Laws as a System

Taken together, the three laws form a closed, non-contradictory system that replaces metaphysical speculation about the limits of formal reason with methodological discipline. Structure guarantees formalizability; boundaries condition verification; plurality entails incompleteness. Formalization is always possible, never universal, and necessarily non-exhaustive. What remains is not mystery but method: the disciplined articulation and refinement of constraints through which structure is made formally visible.

The practical consequence for scientific inquiry is direct. When studying a complex system, the researcher should not seek the single best formalization but should deliberately construct multiple formalizations and compare their respective deliverances. The points at which two formalizations agree provide high-confidence knowledge; the points at which they disagree reveal features of the phenomenon that are formalization-dependent and therefore require further investigation. Many persistent scientific and philosophical controversies—debates about the correct definition of complexity, emergence, self-organization, resilience—are exactly what the Formalization Laws predict: different definitions capture different structural features, and the choice among them is a substantive scientific decision, not a mere terminological convention.

3.2 The Principle of Definition-Dependent Provability

The Principle of Definition-Dependent Provability (DDP) deepens the analysis established by the Formalization Laws by turning from representation to proof. Where the Formalization Laws show that the *description* of a phenomenon depends on the formal

framework chosen, DDP shows that the *provability* of statements about that phenomenon depends on the definitions chosen for the terms in which those statements are expressed.

3.2.1 Statement of the Principle

The principle is stated as follows: for any statement whose formulation depends on terms admitting multiple admissible definitions, the provability status of that statement is a function of the definition chosen. A statement that varies in provability across admissible definitions is not undecidable; it is underdefined.

DDP is a methodological principle, not a completeness result. It does not claim that all statements become provable once defined, nor does it challenge Gödel's incompleteness theorems.[7] Its target is narrower: claims of "in principle unprovability" that are advanced without specifying the formal system or the precise formulation of the statement at issue. DDP diagnoses such claims as incomplete—neither true nor false, but malformed. It differs from the familiar thesis that all proof is framework-relative, which concerns the relativity of proof to axioms and inference rules within a fixed language. DDP concerns the prior question of whether the statement to be proved has been rendered determinate at all.

3.2.2 The Provability Function

The formal apparatus is concise. For a fixed base system S and a formula schema $\varphi[T]$ containing a defined term T, the provability function is defined as $f: D \rightarrow \{0, 1\}$, where D is the space of admissible definitions for T, and $f(d) = 1$ if and only if $\varphi[T/d]$ is derivable in the conservative extension $Ext(S, d)$.[8] The core claim is that this function is, in general, non-constant across the space of admissible definitions.

A simple illustration within Peano Arithmetic makes the point vivid. Let T be a predicate symbol and $\varphi[T]$ the sentence "there exists an x such that T(x) and $x > 2$." Under the definition d_1 where T(x) means "x is even," the sentence is provable (witnessed by $x = 4$). Under the definition d_2 where T(x) means "x is both even and odd," the sentence is refutable, since no

[7]Gödel, K. (1931). "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I." Monatshefte für Mathematik und Physik, 38(1), 173–198.

[8]The provability function is defined as $f: D \rightarrow \{0, 1\}$, where D is the space of admissible definitions for term T, and $f(d) = 1$ iff $Prov(Ext(S, d), \varphi[T/d])$.

natural number satisfies both conditions. The provability of φ[T] varies with the definition of T; the question "Is φ[T] provable?" is indeterminate until T is defined.

3.2.3 Non-Triviality and Philosophical Significance

One might object that the dependency of provability on definition is trivially true. The response is that while the dependency is elementary, its systematic neglect in philosophical discourse makes DDP non-trivial in practice. Claims such as "the existence of God cannot be proven," "moral truths are not formally derivable," and "consciousness is beyond physical explanation" are routinely advanced without specifying which formalization is under discussion. DDP makes explicit that such claims are incomplete: they assert that some vaguely indicated proposition-schema fails to be derivable under all admissible precisifications—a claim that is either false (if some precisification yields a derivable statement) or trivial (if no precisification is genuinely admissible).

The situation parallels what Gallie termed "essentially contested concepts."[9] DDP draws a further consequence: where a concept is essentially contested, claims about provability involving that concept are correspondingly indeterminate. This applies with particular force to debates in the study of complex systems. Consider the concept of "resilience." In engineering, resilience typically refers to a system's ability to return to a unique equilibrium after perturbation. In ecology, resilience refers to the magnitude of perturbation a system can absorb without shifting to a different regime. Both definitions are precise; they are not equivalent; and a system can be resilient under one definition and non-resilient under the other. Theorems proved using one definition may fail under the other—not because of an error in reasoning but because the definitions pick out different structural features.

3.2.4 Illustrative Cases

Randomness

Whether a sequence is "random" depends entirely on the definition adopted. Under a frequency definition (random if the limiting frequency of 1s equals 1/2), the periodic sequence 01010101… is random. Under a Kolmogorov complexity definition (random if incompressible), it is not—it can be compressed as "repeat 01." Under a Martin-Löf definition

[9]Gallie, W. B. (1956). "Essentially Contested Concepts." Proceedings of the Aristotelian Society, 56, 167–198.

(random if it passes all effective statistical tests), it is likewise not random. All three definitions satisfy reasonable admissibility criteria within their respective domains. The question "Is S random?" is not undecidable; it is underdefined.

Moral Realism

The metaethical claim "the existence of objective moral facts cannot be proven" similarly decomposes under DDP. Under a naturalist-reductionist definition (an objective moral fact is a fact about states of affairs that maximize aggregate well-being), the existence claim reduces to a provable empirical statement. Under a non-naturalist Platonist definition (an objective moral fact is a fact about abstract normative entities existing independently of natural facts), the claim requires metaphysical axioms beyond any naturalistic base theory and its provability becomes indeterminate. Under a constructivist definition (an objective moral fact is a fact that would be endorsed under idealized rational deliberation), provability depends on convergence assumptions. The persistent disagreement in metaethics reflects competing implicit formalizations.

The Continuum Hypothesis: A Non-Example

The continuum hypothesis (CH) is independent of ZFC. One might object that CH exhibits definitional variance. The objection fails because CH involves no definitional indeterminacy: the terms "set," "cardinality," "integers," and "real numbers" are fully specified within ZFC. The independence is a structural feature of the axiom system. Different models interpret the same fully specified sentence differently; this is model-theoretic variance, not definitional variance. DDP correctly identifies CH as genuine undecidability, outside DDP's scope.[10]

3.2.5 Relation to Genuine Undecidability

[10]Kriger (2026), Section 7.3. CH's independence is model-theoretic, not definitional: the terms are fully specified within ZFC; different models interpret the same sentence differently, which is distinct from different definitions yielding different sentences.

DDP draws a sharp distinction between two phenomena that philosophical discourse routinely conflates. Gödelian undecidability arises when a precisely defined sentence is neither provable nor refutable in a specified consistent system—a structural property of the system itself. Definitional indeterminacy arises when a sentence-schema varies in provability depending on how its terms are defined—the question of its provability is malformed absent a specified definition. DDP concerns only the second phenomenon. It contracts the domain of illegitimate claims of unprovability by requiring that such claims specify their target with sufficient precision.

The principle thus functions as a methodological filter. When confronted with a claim of the form "X cannot be proven," DDP prescribes a diagnostic sequence: Is the formal system specified? Is the statement precisely formulated? Do the key terms admit multiple admissible definitions? Under the specified definition, is the statement provable, refutable, or independent? And if independent, is this Gödelian undecidability or definitional indeterminacy? Claims failing the first three steps are not claims about provability; they are gestures toward a claim not yet made.

3.2.6 Admissibility of Definitions

A central challenge for DDP concerns the boundaries of admissible definitions. Not every stipulation is admissible; definitions must satisfy criteria of inferential role preservation (preserving the core inferential connections that competent users of the pre-theoretic term recognize), extensional adequacy (yielding correct verdicts on clear cases), and theoretical fruitfulness (facilitating rather than obstructing inquiry). These criteria are formalized through the notion of a *conceptual profile*: the tuple $P(T) = \langle I, E, R \rangle$, where I is the set of accepted inferential connections, E is the set of clear positive and negative instances, and R is the set of theoretical roles the term plays.

When the admissibility criteria themselves become contested—when different communities disagree about which inferential connections are core or which cases are clear—DDP applies recursively. A claim like "definition d is inadmissible" is itself subject to DDP: it depends on how "admissibility" is formalized. This recursion is not vicious; it terminates when disputants share enough common ground to adjudicate. The practical consequence is that DDP requires specifying not only definitions but, when necessary, the standards by which definitions are evaluated.

3.2.7 Extensions to Non-Conservative Cases

The formal framework focuses initially on conservative definitional extensions, where definitions do not add new theorems in the base language. Many philosophically interesting formalizations, however, require non-conservative extensions—adding new axioms, primitives, or theoretical postulates. In such cases, provability becomes relative to both the formalization and the enriched background theory. The provability function acquires a second argument: $f(d, S^+) = 1$ if and only if $\varphi[T/d]$ is provable in the extended system S^+.

The example of consciousness is instructive. A physicalist formalization might embed consciousness-talk into a completed neuroscience; a dualist formalization might add primitive phenomenal properties. Under the physicalist embedding, "consciousness exists" may be provable. Under the dualist embedding, "consciousness is irreducible" may be provable. The question "Is consciousness reducible?" is doubly underdetermined: it depends on both the definition of "consciousness" and the background theory into which it is embedded.

3.3 The Epistemological Framework

Combined, the Formalization Laws and the Principle of Definition-Dependent Provability yield a powerful epistemological framework for the study of complex systems. Its core commitments are as follows:

Knowledge is always relative to a formalization. Every formal result is valid within a specific formal language, under specific definitions, relative to specific boundary conditions. This is not a limitation peculiar to any particular theory but a structural feature of formal investigation as such.

Multiple formalizations are necessary. Because each formalization imposes a different boundary, and because no single boundary captures all relevant features, a complete formal investigation of any sufficiently complex phenomenon requires multiple formalizations—multiple formal languages, multiple sets of primitives, multiple axiomatic frameworks—that collectively cover more of the phenomenon than any single one can.

The choice of formal language is a non-trivial decision. It shapes what can be proved, what can be observed, and what remains invisible. Different choices of primitives yield different boundaries, and therefore different landscapes of provable truths.

The provability of a statement depends on the definitions chosen. Definitions are not dictated by the phenomenon itself but are chosen by the investigator in an act of intellectual commitment. This choice has determinate consequences for what can and cannot be derived.

This framework applies directly to the present thesis. The theory of persistence and viability developed in subsequent chapters is formulated in a specific formal language with specific definitions. The results proved within this language are valid relative to these definitions. Other formalizations—using different mathematical structures, different primitives, different definitions of "persistence" or "viability"—would yield different (though possibly overlapping) sets of provable results. The theory does not claim to be the only possible formalization; it claims to be a particularly powerful one, in the precise sense that it generates a large number of non-trivial results across a wide range of domains from a small set of axioms.

The epistemological framework also explains why the history of systems science has been marked by periodic debates about the correct formalization of fundamental concepts. These debates are not signs of conceptual confusion; they are exactly what the Formalization Laws predict. Different definitions capture different structural features, and the choice among them is a substantive scientific decision. Recognizing this is not a counsel of despair; it is a call for definitional transparency—a requirement that every formal result be accompanied by explicit definitions of all key terms, and that the dependence of the result on those definitions be acknowledged.

3.4 The Choice of Formal Realities

The companion paper "The Choice of Formal Realities"[11] extends the Formalization Laws and DDP to their philosophical conclusion by introducing the concept of a *formal reality* and establishing four meta-mathematical constraints on foundational practice. A formal reality is defined as a coherent package consisting of a formal language (or family of languages) adequate to state the task, a logical system (classical, intuitionistic, modal, etc.), background axioms or structural principles, ontological commitments regarding admissible entities and

[11]Kriger, B. (2026). "The Choice of Formal Realities: A Meta-Mathematical Argument for Explicit Foundational Context." Submitted to Foundations of Science.

constructions, and inferential and explanatory norms governing what counts as a proof, witness, explanation, or admissible idealization.[12] Typical examples include ZFC with classical logic, Bishop-style constructive analysis, categorical foundations (ETCS), and homotopy type theory—each capable of supporting substantial mathematics, yet differing in admissible existence claims, proof norms, and ontological commitments.

3.4.1 Four Meta-Mathematical Constraints

The paper establishes four results, stated theorem-style, that collectively govern the structure of foundational choice.

The first result, the *Inevitability of Choice*, states that for any sufficiently rich mathematical or scientific task—rich in the sense that changing logic or axioms can alter provability, truth conditions, or admissible existence claims—neutrality with respect to formal reality is impossible. A task presupposes criteria of validity, existence, and admissible inference; these criteria are not invariant across frameworks. Classical existence theorems may assert objects without providing witnesses, whereas constructive settings treat existence as requiring explicit construction or realizability. If such criteria are not fixed, the semantic and inferential status of central statements is underdetermined and the task cannot have determinate content. Some formal reality must be presupposed; suppressing the choice merely conceals it.

The second result, the *Non-Uniqueness of Formal Realities*, states that no formal reality adequate for a sufficiently rich task is uniquely determined by that task. Independence phenomena—most notably the Gödel–Cohen independence of the Continuum Hypothesis—exhibit incompatible but internally coherent continuations of shared core practice. Non-categoricity results (Skolem) show that even fixing a first-order theory does not pin down a unique model up to isomorphism. Even when a task is stable in intent, more than one adequate formal reality may support it.

The third result, the *Structural Cost of Formal Realities*, states that every formal reality carries a non-zero structural cost vector $\mathrm{Cost}(R) = \langle C_{proof}, C_{comp}, C_{ont} \rangle$, and that for any

[12]Definition 1 in Kriger (2026). See also Beall, J. C. & Restall, G. (2006). Logical Pluralism. Oxford University Press; Hamkins, J. D. (2012). "The Set-Theoretic Multiverse." Review of Symbolic Logic, 5(3), 416–449.

alternative adequate formal reality, at least one component is strictly larger. Proof-theoretic cost reflects the tradeoff between expressive strength and metatheoretic limitations (by Gödel's incompleteness, sufficiently strong systems cannot be complete; in reverse mathematics, many ordinary theorems are equivalent to specific subsystems of second-order arithmetic). Computational cost reflects the tradeoff between classical principles that permit non-constructive existence claims and constructive settings that recover algorithmic content. Ontological cost reflects the tradeoff between element-level commitments and higher-order structural entities. These costs cannot be simultaneously minimized: reducing one dimension typically requires paying in another.

The fourth result, *Stability Under Admissible Variation*, provides a criterion for structural generality. A result about a task possesses structural generality if and only if it is invariant under all admissible variations of formal reality—translations or interpretations between theories that preserve the portion of language in which the task is formulated and the inferential role relevant to its intended application. Claims that hold in one formal reality but fail in an admissible variation depend on framework-specific commitments and cannot be advertised as framework-free general truths.

3.4.2 The Metatheorem: Foundational Completeness

These four results yield a culminating metatheorem: foundational completeness requires explicit specification of the formal reality within which claims are made. A foundational claim is *foundationally complete* if it specifies its background formal reality, acknowledges salient non-uniqueness and cost tradeoffs relevant to the claim's use, and indicates (at least qualitatively) the stability status of its central assertions under admissible variation. The inevitability of framework choice means that some formal reality is always presupposed; its non-uniqueness means that silence about the choice is informationally significant; its structural costs affect what kind of content is being delivered; and the stability criterion determines the scope of generality claims. Omitting the chosen formal reality therefore prevents the reader from assessing scope, transferability, and methodological legitimacy.

3.4.3 Implications for the Present Work

The concept of formal realities is not a form of relativism. Some formal realities are more powerful than others—they generate more predictions, they are more compressible, they connect to a wider range of empirical observations. The persistence-based framework

developed in this thesis is offered as a particularly powerful formal reality for the study of complex systems, not as the only possible one. The choice to work within this framework is a scientific commitment, justified by the results it produces.

The metatheorem of foundational completeness is honored throughout this thesis. Each chapter specifies the formal reality within which its results are stated, acknowledges relevant cost tradeoffs, and identifies where results depend on framework-specific commitments versus where they enjoy stability across admissible variations. The recommended norm is not to abandon familiar frameworks but to disclose them when a claim is presented as foundational, general, or interdisciplinary. Doing so does not weaken scientific results; it clarifies their scope, stability, and legitimate modes of application.

3.5 Methodological Consequences for the Present Work

This chapter has established the meta-mathematical ground on which the remainder of the thesis stands. Several methodological consequences follow directly.

First, every formal result in subsequent chapters operates within a specific formalization, uses specific definitions, and is valid relative to those choices. This is not a caveat appended as an afterthought but a structural feature of formal reasoning itself, guaranteed by the Formalization Laws.

Second, the definitions of all key terms—persistence, viability, adaptive capacity, and their cognates—are stated explicitly before any results are derived from them. The Principle of Definition-Dependent Provability requires no less: the provability of every theorem depends on these definitions, and definitional transparency is therefore not a stylistic preference but a logical necessity.

Third, where debates in the existing literature appear to be empirical disagreements, the framework developed here suggests that they may in fact be definitional disagreements—disagreements not about the facts but about the formal framework in which the facts are expressed. The diagnostic checklist provided by DDP—specifying the formal system, the precise formulation, and the operative definitions before assessing provability—is applied throughout.

Fourth, the thesis does not claim that its formalization is the only possible or correct one. It claims that it is a particularly productive one. The Formalization Laws guarantee that other formalizations are not merely possible but necessary for a complete investigation of the

phenomena addressed. The theory is offered in the spirit of disciplined pluralism: a single, powerful formal lens through which a wide range of results become visible, with full acknowledgment that other lenses will reveal what this one leaves in shadow.

With this understanding in place, we turn in the next chapter to the epistemic constraints that bound what any inferential system—human, artificial, or abstract—can know.

CHAPTER 4. EPISTEMIC CONSTRAINTS: UNCERTAINTY, NEGATION, AND THE LIMITS OF INFERENCE

The preceding chapters established that formalization is bounded, definition-dependent, and necessarily plural. This chapter consolidates three foundational results that characterize the limits of knowledge for any inferential system—whether biological, artificial, or abstract. Together, these results form the Epistemic Constraint Theory: a formal framework that specifies what any system operating under persistence constraints can and cannot know about its environment, about itself, and about other systems.

The three results are the Law of Imperative Uncertainty (Kriger, 2025), the Law of Limit to Negation (Kriger, 2025), and the Epistemic Constraint Theory proper (Kriger, 2021). Each addresses a distinct dimension of epistemic limitation; together, they define the epistemic landscape within which all subsequent results in this thesis operate. Two further results—the interpretation of undecidability as a methodological signal (Kriger, 2026) and the structural-Bayesian framework for evaluating scientific hypotheses (Kriger, 2026)—extend the theory into scientific methodology and probabilistic inference.

4.1. The Law of Imperative Uncertainty

The Law of Imperative Uncertainty establishes that any system capable of sustained complexity must permit exceptions to its laws in the form of persistent uncertainty and probabilistic deviation. This is not a statement about measurement limitations, technological inadequacy, or epistemic ignorance. It is a structural theorem about the relationship between complexity and the conditions required for its persistence.

The argument, developed at length in Kriger (2025), proceeds through a systematic analysis of what sustained complexity requires. A complex system is defined not by the number of its components but by its capacity to produce distinguishable structure through time—formally, by a positive rate of path-information growth. The central question is: under what conditions can such production be sustained indefinitely?

4.1.1. Formal Framework

The formal proof operates within an information-theoretic and dynamical-systems framework. Let a system be given by a measurable state space $(X, \mathscr{F})$, a family of admissible transformations $\mathscr{L}$ called laws, and an evolution operator T that at each discrete time step maps a present state to a distribution over successor states. The system is called rigid when the

evolution is single-valued and closed—meaning there exists a measurable function f: $X \to X$ such that the next state is determined with certainty by the present state, and the dynamics possesses an absorbing closure (a measurable set A that traps all trajectories after finite time). The system is called open when the evolution is genuinely probabilistic, with a Markov kernel $K(\cdot|x)$ governing transitions and strictly positive conditional entropy $H(X_{t+1}|X_t)$ for at least one time step.

Complexity is quantified by the per-step growth rate of path information:

$C_t = (1/t)\, H(X_0, \ldots, X_t)$

or, when stationarity holds, by the entropy rate $h = \lim (1/t)\, H(X_0, \ldots, X_t)$ as $t \to \infty$. A system is called viable when it can sustain nontrivial complexity indefinitely: there exists $\varepsilon > 0$ such that $\liminf C_t \geq \varepsilon$.

4.1.2. The Core Result

The main theorem demonstrates that viability implies structured openness through three interlocking arguments.

First, rigidity precludes viability. If the evolution is deterministic on a countable state space, then the entire trajectory $(X_0, \ldots, X_t)$ is a function of the initial state X_0. Since applying a function cannot increase Shannon entropy, $H(X_0, \ldots, X_t) \leq H(X_0)$, and therefore $C_t \leq (1/t)\, H(X_0) \to 0$. No new path information is produced. If the system additionally has absorbing closure—every orbit eventually enters a finite periodic cycle—then $H(X_{t+1}|X_0, \ldots, X_t) = 0$ for all sufficiently large t, and the entropy of the path sequence is bounded. In either case, viability fails.

Second, absorbing closure destroys viability even in open systems. Consider a Markov chain with an absorbing state a. Once the system reaches a, conditional entropy drops to zero: $H(X_{t+1}|X_t) = 0$. If absorption occurs almost surely from every initial state, then the time-averaged conditional entropy converges to zero, and $C_t \to 0$. Perfect sinks extinguish the entropy rate. Non-closure—the absence of absorbing classes reached with probability one—is therefore necessary for sustained complexity.

Third, viable systems inhabit a corridor between rigidity and randomness. If the transition kernel were maximally open (uniform over all states), the entropy rate would be maximal but structure—in the sense of long-range dependence—would be minimal. History would become irrelevant. Workable complexity requires both $h > 0$ (openness, deviation,

selection) and $I(X_t; X_{t+1}) > 0$ (lawful constraint, mutual information across time). The condition $0 < h < \log|X|$ together with positive mutual information defines, in information-theoretic terms, the corridor between rigidity and randomness within which viable complexity operates.

4.1.3. Implications for Persistence Theory

The implications for the theory of persistence are direct and far-reaching. Any system operating under persistence constraints must do so under conditions of irreducible uncertainty about its environment. Perfect knowledge is not merely practically unattainable; it is formally incompatible with the complexity that persistence requires. A perfectly deterministic universe—closed, exceptionless, rigid—inevitably collapses into cycles of repetition, absorbing states, or sterile stasis. Every strategy for persistence must therefore be a strategy for persistence under managed openness: regularity that coexists with the real possibility of deviation.

The law also establishes an important co-dependence. Law without deviation hardens into sterility; deviation without law dissolves into noise. Complexity lives in the tension between them, not in their separation. This tension is not resolved once and for all but is continuously renegotiated—a point that connects directly to the theory of viability constraints developed in subsequent chapters.

4.2. The Law of Limit to Negation

The Law of Limit to Negation establishes that negation, considered as a formal operation, possesses an intrinsic structural limit: it cannot eliminate the conditions that make negation possible without ceasing to function as negation. Total negation is therefore not false, paradoxical, or empirically unattainable—it is structurally incoherent. The argument, developed in Kriger (2025), proceeds without appeal to consciousness, subjectivity, belief, or being, treating negation purely as an operation governed by prerequisites.

4.2.1. Formal Setup

The proof operates within a minimal formal framework. Let S be an operative system and ¬ a negation operator defined as a mapping $\neg: D \rightarrow D$ on a domain D of objects (propositions, states, values, or elements of any structured set). The single minimal condition required is:

(C1) The operation ¬ is defined if and only if the domain D exists and remains non-empty.

The conditions of applicability of negation, denoted $\mathbb{C}$, are defined as the minimal structural requirements for ¬ to be meaningful: $\mathbb{C} = \{D \neq \emptyset, \neg \text{ is a mapping on } D\}$. These are not objects within D; they are meta-conditions that allow the operation to exist.

4.2.2. The Argument

Total negation would require applying ¬ to every element $x \in D$ such that no non-negated remainder persists. But for negation to apply to everything, it must also apply to the conditions that make negation possible—it must achieve $\neg(\mathbb{C})$. The critical step is this: if $\neg(\mathbb{C})$ succeeds, then either $D = \emptyset$ or ¬ is no longer a defined mapping. In either case, the operation ¬ is no longer defined. Formally: $\neg(\mathbb{C}) \Rightarrow \neg$ is undefined.

The resulting structure is not a logical contradiction of the form $P \wedge \neg P$ but an operational collapse: the operation destroys its own applicability. To apply negation to its own conditions, negation must already be defined; but if the application succeeds, negation becomes undefined. The operation presupposes what it attempts to eliminate.

The conclusion follows: there exists no sequence of negation operations that eliminates all operative structure while remaining operative. Negation is bounded below by the conditions of its own operation. This bound is internal, unavoidable, and system-independent—it applies to any system capable of negation, whether human, formal, or artificial.

4.2.3. Distinction from Related Results

The Law of Limit to Negation must be carefully distinguished from several related results that it neither restates nor competes with. It differs from dialectical negation (Hegel): dialectics describes how negation produces change within a system where negation remains available; the present law concerns what happens at the boundary where the operation itself loses coherence. It differs from the Liar paradox and Russell's paradox: these concern instability under self-reference within a system where negation continues to function; the present law concerns the disappearance of the operation when directed against its own prerequisites. And it differs from Gödel's incompleteness theorems: these concern the limits of formal provability within systems where negation remains fully operative; the present law concerns the operational possibility of negation itself.

In each of these classical results, negation is presupposed as available and stable. The Law of Limit to Negation concerns what is presupposed by all of them: the conditions under which

negation can function at all. The classical results are, in this sense, corollaries of a more fundamental structural constraint.

4.2.4. Relevance to Persistence Theory

For the theory of complex systems, the relevance is this: any system that models its environment must employ negation—it must distinguish what is the case from what is not the case, what is inside the viability set from what is outside, what is admissible from what is forbidden. The Law of Limit to Negation shows that this capacity for distinction has inherent bounds. A system cannot negate everything; there must always be a ground of affirmation on which the negations stand. This connects to the principle of differentiation that will be developed in later chapters: existence requires distinction, and distinction has structural limits.

4.3. The Epistemic Constraint Theory

The Epistemic Constraint Theory, developed in Kriger (2021), unifies the preceding results within a single framework spanning Bayesian epistemology, information theory, and decision theory. The central observation is that inferential capacity—the ability of a system to form accurate beliefs about its environment on the basis of available evidence—is bounded by three independent constraints, and that the structure of the admissible hypothesis space bounds inferential performance more fundamentally than computational capacity or data volume.

4.3.1. Constrained Inference Systems

The formal framework defines an inference system as a tuple $S = \langle \Omega, \mathscr{H}, \pi, \ell \rangle$, where Ω is the set of possible world-states, $\mathscr{H}$ is the hypothesis space, π is a prior distribution over $\mathscr{H}$, and ℓ is a likelihood function mapping hypotheses and evidence to positive reals. Constraints on inference are modeled through a soft constraint function $w: \mathscr{H} \to [0,1]$ that assigns admissibility weights to hypotheses, with the constrained prior defined as $\pi_a(h) = \pi(h) \cdot w(h)/Z_a$ where Z_a is the normalizing constant.

The key structural result—termed the Constraint Bound on Accuracy—shows that log-posterior accuracy decomposes into constraint, prior, likelihood, and normalization terms. The constraint term $\log w(h^*)$ provides a ceiling that no amount of favorable evidence can overcome when $w(h^*)$ is small. In the hard-constraint case where $w(h^*) = 0$, the true hypothesis receives zero posterior probability regardless of the evidence: the system is structurally incapable of recognizing the truth.

4.3.2. Convergence Under Soft Constraints

The theory provides precise quantitative results for soft constraints. When the true hypothesis h* has constraint weight $w(h^*) = \varepsilon \in (0,1)$, posterior convergence to the truth still occurs under standard regularity conditions (independent evidence, absolute continuity, finite exponential moments), but the expected sample complexity to reach posterior confidence $1-\delta$ is:

$E[n^*(\varepsilon, \delta)] = b/\mu$

where $b = \log(1/\varepsilon) + \log((1-\delta)/\delta) + \log(\pi(h^*)/\bar{\pi})$ and μ is the minimum KL divergence between the true and alternative hypotheses. The term $\log(1/\varepsilon)$ quantifies the penalty imposed by the constraint: each order-of-magnitude reduction in ε adds approximately $\log(10)/\mu$ additional observations to the convergence time.

4.3.3. Four Structural Insights

The Epistemic Constraint Theory yields four operational principles, each formalized as a theorem with stated assumptions:

Constraint Dominance. For any inference system with a fixed constraint function, inferential failure is dominated by hypothesis-space constraints rather than processing capacity. Accuracy is bounded by $\log w(h^*) + O(1)$ regardless of computational resources, and no increase in processing speed or algorithmic sophistication can overcome constraint-induced bounds without modifying the constraint function itself.

Interrogative Primacy. Knowledge growth is governed primarily by question selection rather than answer possession. When queries remain informative, greedy selection of maximally informative queries achieves strictly higher expected information gain than random selection. Query selection is interpreted as a form of dynamic constraint modification—each question reshapes the effective hypothesis space via Bayesian conditioning.

Uncertainty Tolerance. For sequential inference problems with binary or symmetric hypothesis spaces, there exists an optimal stopping threshold $\theta^* = \Theta(c/\mu)$ in terms of residual entropy, where c is the per-step waiting cost and μ is the expected information rate. Premature commitment—stopping before entropy falls below the threshold—results in suboptimal expected reward. This reformulates classical optimal stopping results in information-theoretic terms, expressing when to commit in terms of posterior entropy rather than value functions.

Cognitive Preconditioning. Correcting priors before reasoning is more effective than improving reasoning alone when constraints are severe. Removing a constraint on the true hypothesis (setting $w(h^*) = 1$) yields accuracy improvement that is always non-negative and

scales with the severity of the prior misspecification, while inference refinement (improving computational approximation without modifying priors) yields bounded improvement of $O(|\delta - \delta'|)$.

These four principles are derived from separate theorems under different assumptions; they do not follow from a single master theorem. Their unity is conceptual: in each case, the structure of the admissible hypothesis space—whether fixed by model architecture, shaped by query choices, constrained by commitment timing, or determined by initial conditions—bounds inferential performance.

4.3.4. Interaction of Constraints

The theory further shows that the three classes of constraint—computational, observational, and structural—interact in non-trivial ways. Increasing computational capacity beyond the observational bound yields no epistemic benefit. Increasing observational access beyond the computational bound yields no epistemic benefit. Expanding the formal language, however, can increase both the observational and computational bounds, because a richer language can distinguish states that a poorer language conflates and can compress information in ways that reduce computational demands. This explains why the choice of formal language—the topic of Chapter 3—is of such fundamental importance: it is the constraint that mediates between the other two.

4.4. Undecidability as Methodological Signal

The chapter now turns to a result that connects the epistemic constraints to scientific methodology: the interpretation of undecidability as a methodological signal (Kriger, 2026). The standard response to undecidable questions—statements that cannot be proved or refuted within a formal system—is to treat them as limitations of the model. The present result proposes a more productive interpretation.

4.4.1. System-Relative Provability

The argument begins with a structural principle that is logically elementary but methodologically significant: provability is system-relative and non-transferable. For each system in which a truth is provable, there exists a system in which the same truth is not provable; conversely, the absence of provability in a given system does not exclude provability in another. Truth is invariant while provability varies across formal systems.

This observation becomes methodologically potent through the concept of adequacy. A system S′ is adequate for a true statement φ relative to a system S if: S′ extends S by principled axioms or rules; φ is provable in S′ without being introduced as a bare axiom; and the extension provides structural insight into why φ holds. The distinction between merely adding an axiom and providing a principled extension is crucial—it separates explanatory progress from trivial augmentation.

4.4.2. The Heuristic Principle

The central claim is formulated as a heuristic principle: when a true arithmetical statement φ is not provable in a sound formal system S but is provable in some adequate extension S′, the non-provability of φ in S should be interpreted as evidence of the explanatory inadequacy of S with respect to φ, directing rational inquiry toward identifying or developing such an adequate system.

Goodstein's theorem provides a concrete illustration. Goodstein's original proof employed transfinite induction up to ε_0. Kirby and Paris later showed that this induction strength exceeds that of Peano Arithmetic, establishing unprovability in PA. The result did not undermine PA but revealed limits of its inductive resources, motivating ordinal analysis and calibrated extensions. Undecidability served not as an endpoint but as a signal that the formal system lacked resources adequate to the phenomenon.

4.4.3. Recursive Epistemological Structure

This creates a recursive epistemological structure: every resolution of an undecidable question within one framework generates new undecidable questions in the expanded framework, by the same incompleteness theorems that motivated the original extension. The process of scientific inquiry is therefore not a convergence to a final complete description but an indefinitely expanding exploration of an indefinitely complex reality. Each new formal framework opens new territory and, simultaneously, creates new boundaries.

A structural parallel exists in theoretical physics. In effective field theory and renormalization-group analysis, when extrapolation across physical scales fails while stable observables persist, consistency requires the introduction of new scale-specific organizing principles. The analogy is methodological rather than ontological: in both mathematics and physics, persistent failure within a framework signals a mismatch between descriptive resources and the structure of the phenomenon.

4.5. The Structural-Bayesian Framework

The structural-Bayesian framework for evaluating scientific hypotheses (Kriger, 2026) integrates the insights of this chapter with probabilistic inference. It addresses a persistent problem in theory evaluation: the conflation of methodological meta-principles with substantive empirical theories, both of which may satisfy standard epistemic criteria yet differ fundamentally in their capacity to exclude possible worlds.

4.5.1. Three-Layer Structural Assessment

The framework decomposes structural assessment into three orthogonal layers, each addressing a distinct epistemic question:

Internal Integrity (I). Scores five criteria: logical consistency of the formalism, conceptual clarity of core terms, application reproducibility, absence of ad hoc elements, and explicit domain specification. This layer asks: is the theory well-constructed?

Constraint Power (C). Scores three criteria: whether the theory excludes at least one internally consistent class of models, whether the exclusion is non-definitional, and whether it is non-trivial (not a logical tautology). This layer asks: what does the theory forbid?

Ontological Load (O). Scores two criteria: whether the theory makes claims about physical reality (not merely about models or formal systems), and whether the world could coherently have been otherwise but the theory forbids it. This layer distinguishes laws of nature from laws of thought about nature.

A continuous non-triviality multiplier $N(T) = 0.4 + 0.2 \cdot C(T)$ ensures that theories with zero constraint power—those that forbid nothing beyond logical impossibilities—receive appropriately reduced scores regardless of their internal sophistication.

4.5.2. Integration with Bayesian Inference

The structural admissibility score $S(T) = N(T) \cdot (\alpha I + \beta C + \gamma O) / (\alpha \cdot 5 + \beta \cdot 3 + \gamma \cdot 2)$ serves not as a probability but as an upper bound on rational prior assignment: $P(T) \leq S(T)$. A theory with $S(T) = 0.12$ cannot rationally receive a prior of 0.50. This constraint addresses the persistent Problem of the Priors in Bayesian epistemology by grounding prior specification in structural analysis rather than convention or intuition.

Complete evaluation proceeds in four stages: structural assessment (computing I, C, O, N, and S), prior constraint (setting $P(T) = \lambda \cdot S(T)$ where λ is a domain-calibrated optimism

parameter), Bayesian updating with evidence, and interpretation of the posterior relative to the structural score.

4.5.3. Discriminative Power

Benchmark evaluations demonstrate the framework's capacity to distinguish theories that standard criteria conflate. General Relativity achieves S = 100%: maximum internal integrity, full constraint power (excludes Newtonian absolute space and flat spacetime with gravitational force), and complete ontological load. Quantum Mechanics achieves S = 78%: full constraint power (excludes local hidden variables via Bell) but compromised internal integrity (conceptual ambiguity in measurement, ad hoc collapse postulate, unclear domain boundaries). String Theory achieves S = 12%: high mathematical sophistication but zero constraint power (the landscape of approximately 10^{500} vacua accommodates virtually any low-energy physics) and minimal ontological load (any coherent world is compatible with some vacuum).

The decisive role of the constraint-power layer is notable. General Relativity and Quantum Mechanics both achieve C = 3; String Theory achieves C = 0. Mathematical sophistication does not translate to epistemic strength without genuine constraint power—a quantitative instantiation of a broadly Popperian insight.

4.6. Synthesis: The Epistemic Landscape

The results of this chapter define the epistemic landscape within which any investigation of complex systems must operate. The Law of Imperative Uncertainty establishes that irreducible uncertainty is a precondition of the complexity that makes investigation possible and necessary. The Law of Limit to Negation establishes that the capacity for distinction—essential to any modeling activity—has inherent structural bounds. The Epistemic Constraint Theory shows that inferential capacity is bounded by the interaction of computational, observational, and structural constraints, with the structure of the hypothesis space as the most fundamental bottleneck. The interpretation of undecidability as a methodological signal provides a productive response to the incompleteness that follows from these constraints. And the structural-Bayesian framework provides operational machinery for evaluating hypotheses within this constrained landscape.

The conclusion of Part I can now be stated precisely: any investigation of a complex system operates within a formal framework that is necessarily incomplete, definition-dependent, and

uncertainty-laden. These are not limitations to be apologized for; they are structural features of any coherent investigation of complex phenomena. The task of the theorist is not to eliminate them but to work within them productively—and the remainder of this thesis demonstrates how.

PART II. THE DYNAMICAL CORE: PERSISTENCE AND SELF-ORGANIZATION

CHAPTER 5. THE TRANSFORMATIONAL BASIS OF PERSISTENCE

5.1 Introduction and Theoretical Motivation

This chapter presents the central formal result of our theory of structural viability, the full development of which appears in our paper *"The Transformational Basis of Persistence: A Formal Theory of Structural Viability"* (Kriger, "Transformational Basis," 2024). Where the preceding chapters established the epistemic framework—the constraints on what can be known—the present chapter establishes the dynamical framework: the constraints on what can survive. The core question is deceptively simple: what must be true of a system for it to persist over time?

A companion result—Conservation and Convertibility as Structural Necessities of Persistent Systems (Kriger, 2026, https://doi.org/10.5281/zenodo.18665700)—extends this analysis by demonstrating that conservation laws and convertibility relations, of which Noether's theorem is the physical special case, are structural necessities of any system that persists across physical, biological, cognitive, and social domains.

We begin from the observation that persistence under physical examination is never a passive state. The intuitive notion that a thing simply maintains itself through some intrinsic stability dissolves upon closer inspection: a candle flame persists only by consuming wax, a river only by moving water, an organism only by metabolizing other organisms. From this observation, we advance a formal claim: structured persistence necessarily requires the transformation of pre-existing structure, where structure is defined precisely in terms of equivalence classes of microstates under specified coarse-graining operations.

This claim is not merely a restatement of the second law of thermodynamics. While the second law establishes that total entropy must increase, it does not specify which configurations must change or how the entropy production burden is distributed between a system and its environment. Our structural transformation framework supplies this missing specificity by characterizing persistence as requiring changes to particular macroscopic configurations, not merely statistical redistribution of microstates.

5.2 Formal Framework: Definitions

Our formal apparatus rests on a sequence of precise definitions (Kriger, "Transformational Basis," 2024, §2.1). The first distinguishes microstates from macrostates. Let M be the space of all possible microstates of a physical system. A macrostate M is a subset of M defined by a coarse-graining function φ: $M \rightarrow C$, where C is a space of coarse-grained descriptions. Formally, $M = \varphi^{-1}(c)$ for some c in C.

We define a structure Ω as an equivalence class of microstates under a physically meaningful coarse-graining operation that preserves macroscopic observables relevant to a specified level of description. Two microstates belong to the same structure if and only if they are mapped to the same coarse-grained description. This definition provides a rigorous criterion for distinguishing genuine structural transformation—a transition between equivalence classes—from mere microstate fluctuation within a single equivalence class.

We define persistence as follows: a system S persists over interval $[t_0, t_0 + \Delta t]$ if and only if its structure ΩS remains within a specified invariance class $I S$ throughout the interval. Structural transformation of an environment E is correspondingly defined as the case where the environment transitions to a different equivalence class under its coarse-graining over the same interval.

5.3 The Central Proposition: Transformational Necessity

The central result of our framework is Proposition 1, which we designate the principle of Transformational Necessity (Kriger, "Transformational Basis," 2024, §2.2). It states: let S be an open system maintaining structure ΩS in its invariance class against thermodynamic equilibration. Then there must exist some subsystem E of the environment such that E undergoes structural transformation at a positive rate.

We prove this in three steps. The first step establishes the entropy constraint: by the second law, for any system maintaining internal organization, total entropy production must be non-negative. If the system maintains its structure within a low-entropy invariance class, then entropy must be exported to the environment at a positive rate.

The second step moves from entropy to structure. Drawing on the fluctuation theorem and its extensions (Jarzynski, "Nonequilibrium Equality," 1997; Crooks, "Entropy Production," 1999), we show that entropy production in non-equilibrium systems is associated with irreversible transitions between macrostates. Entropy export is expressed through the sum of products of thermodynamic fluxes and their conjugate forces. We define the coarse-graining

over precisely those extensive variables (energy, particle number, volume) that appear as arguments of the conjugate forces in the entropy production formula. This is not an arbitrary analytical choice but a physical necessity: these variables constitute the only channels through which system–environment thermodynamic coupling occurs. Non-zero entropy export therefore necessarily implies change in the variables over which the environmental coarse-graining is defined.

The third step concludes the argument: if the extensive variables of the environment change and the coarse-graining is defined in terms of these variables, then the environment undergoes structural transformation. We formalize this as: Persistence(ΩS, Δt) implies the existence of an environment E with a positive transformation rate over the interval.

5.4 The Transformation Rate and Its Bounds

Beyond establishing that transformation must occur, we define a quantitative metric for its measurement (Kriger, "Transformational Basis," 2024, §2.3). The transformation rate required to maintain a structure ΩS is given as a limit: the integral over all environmental subsystems undergoing structural change, normalized by the time interval, taken as the interval approaches zero. This quantity has units of structural change per unit time and can be operationalized for specific systems. For metabolic systems, it corresponds to the rate of molecular bond rearrangement; for information-processing systems, it relates to the rate of bit erasure via Landauer's principle (Landauer, "Irreversibility," 1961).

We further derive a corollary establishing a minimum transformation rate: for any persisting structure with an internal entropy deficit relative to equilibrium, the minimum transformation rate is bounded from below by a function that increases monotonically with both the magnitude of the entropy deficit and the temperature. The existence of such a lower bound is guaranteed by thermodynamics; its explicit form is model-dependent, but we provide an illustrative example using a two-level system maintaining population inversion against thermal equilibration.

5.5 A Taxonomy of Transformational Modes

Structural transformation does not take a single form. We identify three primary modes, classified by which degrees of freedom are predominantly reconfigured (Kriger, "Transformational Basis," 2024, §4).

Dissipative transformation occurs when the environmental degrees of freedom reconfigured are primarily energetic: temperature gradients, pressure differentials, electromagnetic field configurations. The paradigmatic examples are hurricanes dissipating thermal gradients, convection cells processing temperature differentials, and Bénard cells organizing flow patterns. In each case, the prior structure being transformed is the organization of energy distributions.

Consumptive transformation occurs when the reconfigured degrees of freedom are primarily material: chemical bonds, molecular configurations, tissue architectures. This is the paradigm case for biological systems. An organism maintains its molecular organization by dismantling the molecular organization of nutrients. The specificity of consumptive transformation—which bonds are broken, which molecules are reassembled—is precisely what distinguishes our structural framework from entropy accounting.

Observational transformation occurs when the reconfigured degrees of freedom are primarily informational: memory states, correlation structures, phase relationships. The thermodynamic basis for this mode is Landauer's principle, which establishes that erasing one bit of information dissipates at least $kT \ln 2$ of energy (Landauer, "Irreversibility," 1961). Any system that accumulates information—whether through measurement, computation, or learning—must erase prior informational states, and this erasure carries an irreducible thermodynamic cost. We emphasize that this is a minimal physical claim, not a metaphysical thesis: recorded observation necessarily transforms, regardless of whether idealized non-recording observation is conceivable.

5.6 Relation to Existing Theoretical Frameworks

We now situate the structural transformation framework in relation to three major prior theoretical programmes (Kriger, "Transformational Basis," 2024, §3).

First, Prigogine's theory of dissipative structures established that far-from-equilibrium systems can maintain organization through continuous entropy export (Prigogine and Stengers, "Order Out of Chaos," 1984). We accept this foundation but identify a critical gap: while Prigogine's account characterizes the quantity of entropy export required, our framework characterizes the specific configuration of environmental change. Two systems with identical entropy production rates but different transformation profiles—one degrading

concentrated chemical bonds, another dissipating diffuse thermal gradients—are treated as equivalent by entropic accounting but distinguished by our structural framework.

Second, England's work on entropy production and self-organization proposes that matter tends to organize in ways that maximize entropy production (England, "Statistical Physics," 2013). We characterize the relationship as complementary: England addresses why certain structures arise, while our theory addresses what their persistence requires. The two accounts are compatible but not reducible to one another.

Third, Friston's free energy principle proposes that biological systems persist by minimizing variational free energy (Friston, "Free-Energy Principle," 2010). Both frameworks share the insight that persistence is active rather than passive, but they differ in explanatory target. The free energy principle explains system behavior—why systems act to maintain themselves—while our structural transformation theory explains system costs—what environmental changes persistence requires. We suggest that the two may be unified, since minimizing free energy requires actions that necessarily transform environmental structure.

Our framework offers three forms of explanatory surplus over purely entropic accounts. The first is specificity: entropy is a scalar, while structural transformation captures which structures change. The second is directionality: entropy accounting establishes that total entropy increases but does not specify which subsystem bears the transformation burden; our theory establishes a directional asymmetry in which the system persists at the cost of environmental reconfiguration. The third is cross-domain unification: the framework applies identical concepts to physical, biological, and information-processing systems, revealing structural analogies that domain-specific terminology obscures.

5.7 Applications: Physical, Biological, and Informational Systems

5.7.1 Physical Systems

We develop detailed quantitative applications across three domains (Kriger, "Transformational Basis," 2024, §§5–6). In the physical domain, we treat stellar persistence as a test case. A star maintains hydrostatic equilibrium against gravity by fusing hydrogen; the prior structure being transformed is the gravitationally stratified matter distribution. Though fusion increases local nuclear complexity (binding energy per nucleon increases up to iron), this complexity increase is a byproduct of the dissipative process rather than its driver. A main-sequence star like the Sun converts approximately 4×10^9 kg of mass to energy per

second, and the entropy increase in the radiated photon field vastly exceeds any entropy decrease from nuclear binding.

We also analyze crystal growth from supersaturated solution as an edge case that tests the framework's limits. The supersaturated solution possesses chemical potential by virtue of its metastable configuration, and crystallization releases this potential while increasing entropy in the surrounding solvent. This case demonstrates that prior structure need not be visually obvious—it can take the form of chemical or thermodynamic organization invisible to macroscopic observation.

5.7.2 Biological Systems and Trophic Scaling

Our biological application focuses on metabolism as structural transformation and introduces a trophic scaling law. At each level of a food web, approximately 90% of energy is dissipated rather than incorporated into biomass—a pattern first characterized by Lindeman (Lindeman, "Trophic-Dynamic Aspect," 1942). We formalize this as a scaling relation: the structural transformation required to support trophic level n is $T_n = T_0/\eta^n$, where T_0 is primary production and $\eta \approx 0.1$ is the trophic efficiency (Kriger, "Transformational Basis," 2024, §6.1). For a tertiary consumer, this means approximately 1000 times its body mass in primary production must be transformed over its lifetime. This quantifies our central claim: complex organization costs more structural transformation than simple organization—not merely more energy, but more prior structure dismantled.

5.7.3 Information Processing and Landauer's Principle

For informational systems, we ground the analysis in Landauer's principle: erasing one bit of information dissipates at minimum approximately 2.9×10^{-21} J at 300 K. This minimum has received direct experimental verification (Bérut et al., "Experimental Verification," 2012). Modern computers operate roughly 10^6 times above this limit. Biological memory incurs still higher costs: synaptic modification requires on the order of 10^4 ATP molecules per event, corresponding to roughly 10^{-16} J (Harris et al., "Synaptic Energy," 2012).

We also apply the framework reflexively to artificial intelligence. Every AI computation requires the physical erasure and rewriting of electrical states in silicon, and the energy sources powering that computation embed prior transformations extending back through geological and cosmological time. A large AI training run may consume on the order of 10^6 kWh (Strubell et al., "Energy and Policy," 2019), equivalent to the lifetime energy consumption of several

humans compressed into weeks. From the structural transformation perspective, AI represents a novel regime in which informational persistence is sustained by massive material transformation.

5.8 The Limit Case: Cyclical Closure and Atemporality

The structural transformation framework, as developed in the preceding sections, applies to open systems: systems that persist by transforming environmental structure across time. There exists, however, a theoretically significant limit case in which the framework encounters systems whose outputs become their own inputs—systems that appear to persist through themselves. We address this limit case in Section 10 of our transformation paper, but its full mathematical and philosophical development is given in our companion paper *"Cyclical Hierarchical Systems: A Modal Analysis of Self-Sufficient Structures"* (Kriger, "Cyclical Hierarchical Systems," 2024). The present section draws on both works, as a proper understanding of this limit case is essential to the coherence of the broader theory.

5.8.1 The Principle and Its Modal Character

We formulate the central thesis of the cyclical hierarchical systems theory as consistently modal rather than assertoric (Kriger, "Cyclical Hierarchical Systems," 2024, §4.1). We do not claim that reality is in fact a cyclically closed hierarchy, nor that all systems exhibit this structure. Our claim is that such structures are coherently possible and that, for systems which do exhibit cyclical closure, the standard demand for a primary foundation becomes what we term a conceptual mismatch—not an unanswered question but a potentially inapplicable one. The formal statement of the principle reads: there exist possible systems in which the hierarchy of levels and roles is closed and possesses no absolute beginning or apex. Elements perform distinct functions, yet the role of foundation circulates within the structure. That which is last in one relation acts as the condition of primacy in another. The question of primary origin may therefore be a conceptual mismatch when applied to such structures.

This modal formulation is essential to avoid a common misreading. Our theory does not assert that all foundational questions are misapplied; it asserts that foundational questions applied to *closed* structures may be misapplied, in the same way that asking for the first point on a circle misapplies linear ordering concepts to a non-linear topology. In open structures—chains, trees, well-founded hierarchies—the question of primary foundation remains entirely legitimate.

5.8.2 The Foundational Mismatch

Our analysis of the foundational mismatch draws an explicit parallel to Ryle's account of category errors (Ryle, "Concept of Mind," 1949). A category error occurs when a concept belonging to one logical type is misapplied to another—as when one asks where the university is after being shown its buildings, library, and staff. The question is grammatically correct but logically malformed: it treats the university as belonging to the same category as a physical location, when it is an institutional abstraction (Kriger, "Cyclical Hierarchical Systems," 2024, §3).

We identify a parallel mismatch in the application of foundational questions to closed structures. Asking what is the primary foundation of a cyclically closed system presupposes that primacy is a fixed property, when the system's topology makes it a circulating role. The analogy we employ is geographic: asking what is north of the North Pole is grammatically well-formed but geographically inapplicable, because it applies the concept of northward direction beyond its domain of applicability. Similarly, asking what grounds a closed cyclical hierarchy may apply the concept of grounding beyond its proper domain.

Category errors, we argue, are dissolved, not solved. One does not answer the university question by pointing somewhere new; one shows why the question was malformed. Our suggestion is that foundational questions for closed structures may be similarly dissolved by showing that such structures lack the categorical feature—fixed primacy—that the question presupposes.

5.8.3 The Tripartite Distinction

A key conceptual contribution of our theory is the tripartite distinction among source, foundation, and primacy (Kriger, "Cyclical Hierarchical Systems," 2024, §4.2). We argue that linear thinking conflates three concepts that diverge when applied to closed structures. *Source* (Quelle) denotes explanatory derivation: that through which a system is explained. *Foundation* (Grund) denotes ontological ground: that without which the system cannot exist. *Primacy* (Priorität) denotes structural role: being the ground relative to other elements.

In linear causation, these three concepts collapse into one: the first cause is simultaneously source, foundation, and primary. In cyclical hierarchy, they diverge: A may be the source of B while B is the foundation of A, and primacy circulates rather than inhering in any fixed element. This distinction is not merely terminological. It reveals a hidden assumption in much

traditional metaphysics: that the question of foundation must have a single, non-circulating answer. The tripartite distinction shows that this assumption holds only for open structures.

5.8.4 Mathematical Framework

The mathematical framework that underwrites these philosophical claims is developed with full formal rigour (Kriger, "Cyclical Hierarchical Systems," 2024, §5.1). We define a foundational hierarchy as a triple $(S, \leq, F)$ where S is a non-empty state space, $\leq$ is a binary grounding relation, and F: $S \to S$ is a determination function. A hierarchy is well-founded (open) if it contains no infinite descending chains. A hierarchy is cyclically closed of order n if there exist elements $S_1, S_2, \ldots, S_n$ and functions $F_1, F_2, \ldots, F_n$ such that each element is determined by the next—$S_i = F_i(S_{i+1})$—and the final element is determined by the first: $S_n = F_n(S_1)$.

Our central mathematical result is the Fixed-Point Equivalence (Proposition 5.1): a cyclically closed hierarchy of order n exists if and only if the composite operator $\Phi = F_1 \circ F_2 \circ \ldots \circ F_n$ has a fixed point. This equivalence reduces the existence question for cyclically closed hierarchies entirely to the question of fixed-point existence, which is well-characterized by classical mathematics.

We then deploy three classical fixed-point theorems to provide sufficient conditions (Kriger, "Cyclical Hierarchical Systems," 2024, §5.2). The Knaster-Tarski theorem guarantees that if the state space admits a complete lattice structure and each determination function is order-preserving, a cyclically closed hierarchy exists. The Banach Contraction Principle guarantees that if the state space is a complete metric space and the composite operator is a contraction, a *unique* cyclically closed hierarchy exists—and iterative processes converge to it. The Brouwer fixed-point theorem guarantees that if the state space is a non-empty compact convex subset of $\mathbb{R}^n$ and the composite operator is continuous, at least one fixed point exists.

5.8.5 Topological Characterization: Essential Cyclicity

Beyond proving that cyclically closed hierarchies can exist, we characterize when their cyclical structure is *essential*—intrinsic to the topology and not removable by continuous deformation (Kriger, "Cyclical Hierarchical Systems," 2024, §5.3). Drawing on algebraic topology (Hatcher, "Algebraic Topology," 2002), we define a structure as essentially cyclic if three conditions are met: the space encoding the hierarchy is compact, it is connected, and its first homology group $H_1(X; \mathbb{Z})$ is non-trivial.

The significance of this homological condition is made precise by our Proposition 5.8: if a structure is essentially cyclic, then there exists no continuous retraction that would contract the cycle to a point while preserving the cyclic structure. The proof proceeds by contradiction: such a retraction would induce a homomorphism mapping a non-trivial homology group to the trivial group, while retractions must induce surjections on homology—a contradiction. The first Betti number $b_1 = \text{rank}(H_1)$ counts the number of independent cycles in the structure; a positive first Betti number is a necessary condition for essential cyclicity.

We note that this characterization is encoding-dependent: different ways of representing a hierarchical structure as a topological space may yield different homology groups. We regard this not as a defect but as a feature of the formalism—the choice of encoding is part of the formal model, and our claim is conditional: if a structure admits an encoding with non-trivial H_1, its cyclicity is essential in the precise sense formalized above.

5.8.6 Philosophical Interpretation: Atemporality and the Rehabilitated Causa Sui

The philosophical import of these mathematical results requires careful development (Kriger, "Cyclical Hierarchical Systems," 2024, §5.5). The key insight is that the fixed point established by the existence theorems is not reached through temporal iteration; it *is* the structure—the configuration in which all mutual determination relations are simultaneously satisfied. What appears paradoxical from a temporal perspective (how can A ground B if B grounds A?) may be fully coherent from a structural perspective, just as mutual gravitational attraction between bodies is coherent despite apparent circularity.

By "atemporal" we mean specifically that the structural relations lack a time parameter: the states are co-given in the structure's topology. In an atemporal frame, transformation becomes relation: A determines B, B determines C, C determines A, all simultaneously. There is no regress because there is no sequence; there is no need for a brute beginning because there is no beginning at all. The question "what came first?" presupposes temporal ordering. Applied to a closed atemporal structure, it becomes a conceptual mismatch—structurally inapplicable, like asking what is north of the North Pole.

On this basis, we rehabilitate the classical notion of causa sui, which had been rejected as paradoxical on the grounds that self-causation before existence is incoherent (Kriger, "Cyclical Hierarchical Systems," 2024, §9.3). We argue that cyclical hierarchy provides a non-paradoxical reading: self-sufficiency understood not as temporal self-causation but as

topological completeness—a system whose existence conditions are internal to its structure. The causa sui is thus reinterpreted as structural closure, not temporal self-generation.

5.8.7 Objections and Replies

We address several major objections to cyclical hierarchy (Kriger, "Cyclical Hierarchical Systems," 2024, §6). The vicious-circle objection charges that the account involves circular reasoning. We reply that cyclical hierarchy involves structural mutual dependence among distinct elements—A determines B, B determines C, C determines A—which is fundamentally different from justificatory circularity (P because P). The mediated cycle through different elements is no more fallacious than mutual gravitational attraction.

The grounding-asymmetry objection holds that grounding is by definition asymmetric and therefore excludes cyclical determination. We note that this asymmetry is stipulated as an axiom within grounding theory, not demonstrated. Barnes and Thompson have independently defended symmetric dependence as coherent (Barnes, "Symmetric Dependence," 2018; Thompson, "Metaphysical Interdependence," 2016). Our results operate in a different domain from standard grounding theory: they concern abstract dependency structures satisfying fixed-point conditions, without any claim about whether metaphysical grounding relations should or should not be asymmetric.

The merely-mathematical objection grants formal consistency but denies that mathematical possibility entails metaphysical possibility. We concede this point but argue that mathematical consistency removes the strongest objection—logical impossibility—and that the positive case rests additionally on candidate empirical instantiations.

5.8.8 Candidate Instantiations

We grade our examples by epistemic status, maintaining clear separation between formal instantiations, biological candidates, and physical analogies (Kriger, "Cyclical Hierarchical Systems," 2024, §7).

At the formal level, recursive function definitions via fixed-point combinators, self-interpreting interpreters, and corecursive data structures all instantiate cyclical hierarchy within formal systems. These are strict instantiations, though their extension to physical or metaphysical reality is a separate question.

At the biological level, autopoietic systems constitute strong candidates (Maturana and Varela, "Autopoiesis and Cognition," 1980). In an autopoietic organism, the membrane is produced

by internal processes while those processes require membrane delimitation—genuine mutual dependence at the systemic level. Similarly, the Krebs cycle requires oxaloacetate to produce oxaloacetate, which is literally a fixed-point equation: O = F(O). We are careful to note an important caveat: these biological systems had phylogenetic origins (abiogenesis, evolution); cyclical hierarchy applies to their ongoing structural organization, not to the denial of a historical origin.

At the physical level, the Wheeler-DeWitt equation (which describes quantum gravity without an explicit time parameter) and topological quantum field theory (in which physical quantities depend only on topology, not metric) are classified as compatible analogies rather than demonstrations. They show that physics permits atemporal and topology-first formulations, without establishing that the physical universe lacks a temporal origin.

5.8.9 Connection to the Structural Transformation Framework

The connection between cyclical hierarchical systems and the structural transformation framework is made explicit as follows (Kriger, "Transformational Basis," 2024, §10). Open systems constitute the general case: transformation is temporal, the organism destroys nutrients that existed independently before being destroyed, and the Equivalence Principle compares distinct structures across time. Closed systems constitute the limit case: transformation becomes relational rather than temporal, the system's persistence conditions are internal to its topology, and the Equivalence Principle is satisfied trivially—the system's output is its input not sequentially but structurally.

Most actual systems occupy an intermediate position. An organism is closed in its autopoietic organization but open in its material exchange. A mind is closed in its self-model but open in its environmental coupling. Our structural transformation framework applies differently at different levels of description, and our cyclical hierarchical systems theory provides the mathematical tools for characterizing which levels exhibit closure.

The image of the ouroboros captures this synthesis. The traditional reading depicts self-consumption: the serpent eating itself, destruction as the price of continuation. We argue this is a category error—it imposes a temporal, sequential interpretation onto what is fundamentally a structural, simultaneous configuration. The ouroboros does not depict a serpent eating its tail but a closed topology: mutual determination of parts, self-sufficiency

without external foundation. The head and tail are not sequential positions (first and last) but relational roles in a closed configuration, existing simultaneously and defining each other.

5.9 The Equivalence Principle

From the established necessity of transformation, we derive what we term the Equivalence Principle (Kriger, "Transformational Basis," 2024, §11). Given that transformation is unavoidable, processes divide into two ontological kinds: those where the effective complexity of what is created or maintained matches or exceeds what is transformed (structural persistence), and those where it does not (net dissipation).

We ground the measure of structural significance in the concept of effective complexity, following Gell-Mann and Lloyd (Gell-Mann and Lloyd, "Information Measures," 1996). Effective complexity captures the length of a highly compressed description of a structure's regularities—the algorithmic information required to specify the ensemble of which the structure is a typical member, excluding random components. This measure distinguishes what neither entropy nor raw information content can distinguish: a random gas has high entropy but low effective complexity, a perfect crystal has low entropy but also low effective complexity, and a living organism has intermediate entropy but high effective complexity.

The Equivalence Principle requires its significance measure to satisfy three formal properties: monotonicity with effective complexity, additivity over independent structures, and invariance under structure-preserving re-description. We prove a classification invariance result: any two measures satisfying these properties will agree on whether a given process constitutes structural persistence or net dissipation. The binary classification is robust to the choice of measure; only the degree to which a process exceeds or falls short of equivalence is measure-dependent.

We emphasize that the Equivalence Principle carries no moral weight. It does not prescribe that structural persistence is good or that net dissipation is bad. It provides a structural accounting: every persistent structure has a transformation cost that is in principle measurable, and persistence is structurally balanced when creation matches destruction. The physics is indifferent; the structure is not.

5.10 Summary and Theoretical Significance

The results developed in this chapter, drawing on our two published papers—*"The Transformational Basis of Persistence"* and *"Cyclical Hierarchical Systems"*—provide the formal foundation for everything that follows in this thesis.

First, we have provided a formal derivation—not merely an assertion—that persistence requires environmental transformation. The central proposition follows from established thermodynamic principles and the fluctuation theorem, giving it a rigorous physical foundation.

Second, the framework's substrate independence enables cross-domain application. Because our formal apparatus applies to any open system maintaining structure against equilibration, properties derived from the framework hold for physical, biological, cognitive, and social systems without modification.

Third, our taxonomy of transformational modes—dissipative, consumptive, and observational—provides a principled classification scheme connecting physical thermodynamics to information processing and cognition. The observational mode, grounded in Landauer's principle, establishes that even the acquisition and maintenance of information has irreducible structural costs.

Fourth, the Equivalence Principle introduces a non-moral criterion for distinguishing structure-maintaining processes from net dissipation, providing the conceptual apparatus for structural accounting that the subsequent chapters will employ.

Fifth, and perhaps most significantly, our theory of cyclical hierarchical systems identifies the limit case where transformation becomes relational rather than temporal. The modal character of this theory—our careful insistence that it demonstrates the coherent possibility of self-sufficient structures, not their actual instantiation—provides a principled resolution to the apparent paradox of self-maintaining systems. The mathematical framework, grounded in fixed-point theory and algebraic topology, supplies rigorous existence conditions. The tripartite distinction among source, foundation, and primacy dissolves hidden linear assumptions in traditional foundational thinking. And the topological characterization of essential cyclicity ensures that the cyclical structure, where it obtains, is not an artefact of description but an intrinsic feature of the system.

CHAPTER 6. RESILIENCE, CLOSURE, AND SELF-SUFFICIENCY

6.1 Introduction and Theoretical Motivation

The preceding chapter established the formal conditions for persistence in general terms: the structural transformation framework shows that any open system maintaining its structure against thermodynamic equilibration must transform environmental structure at a positive rate. The present chapter develops two specializations of this viability framework that address critical aspects of how real systems maintain themselves. The first addresses structural resilience in discrete-state systems, drawing on our paper *"Formalization of Structural Resilience in Discrete-State Dynamical Systems"* (Kriger, "Structural Resilience," 2019). The second addresses the conditions for cyclical closure and self-sufficiency, drawing on our papers *"On the Possibility of Self-Sufficient Systems: Fixed Points and Cyclical Closure"* (Kriger, "Self-Sufficient Systems," 2024) and *"Cyclical Hierarchical Systems: A Modal Analysis of Self-Sufficient Structures"* (Kriger, "Cyclical Hierarchical Systems," 2024).

The motivation for these specializations arises from two observations. First, many real systems do not admit continuous perturbations. Gene regulatory networks operate with discrete expression levels—on or off, or a small number of intermediate states. Digital circuits have binary states. Social networks with discrete states (cooperate or defect, adopt or reject, support or oppose) undergo transitions that are inherently discontinuous. For such systems, the continuous-state viability framework of Chapter 5 requires adaptation to a discrete setting where the relevant mathematical tools are graph-theoretic rather than analytical.

Second, the structural transformation framework as developed in Chapter 5 applies primarily to open systems: systems that persist by transforming environmental structure across time. But the limit case of cyclical closure, introduced in Section 5.8, raises a question that the transformation framework alone does not fully resolve: under what precise conditions can a system achieve self-sufficiency—the condition in which its resource flows form closed cycles requiring no external input? The present chapter provides the formal answer, grounding it in fixed-point theory and the concept of cyclical closure developed in our companion papers.

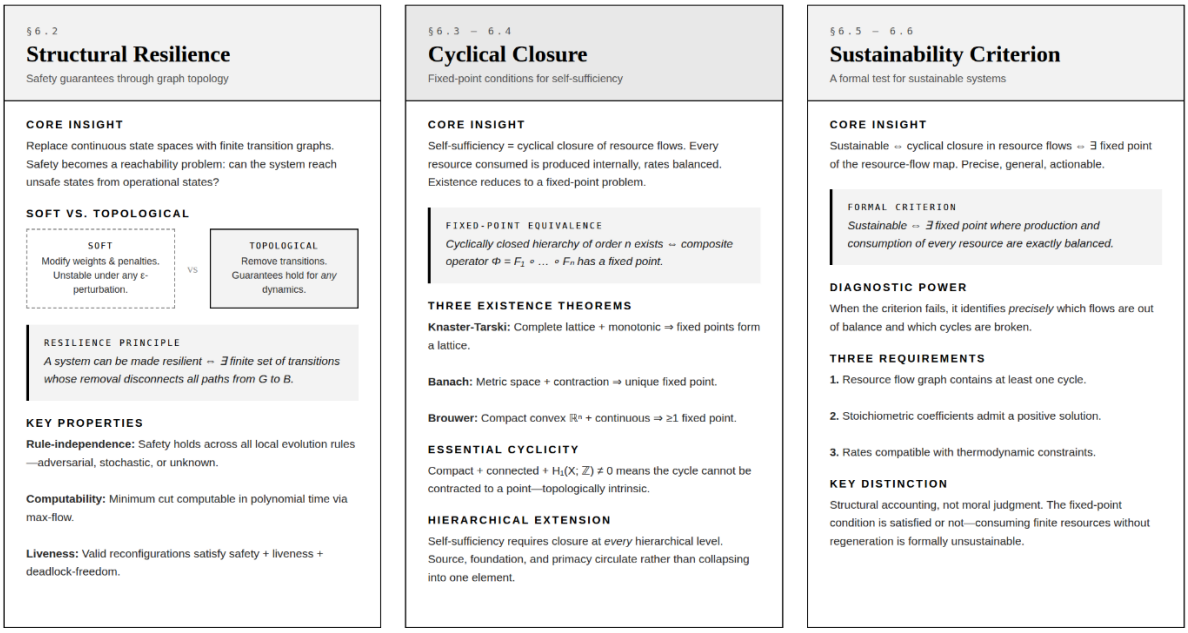
The chapter proceeds as follows. Section 6.2 develops the discrete-state resilience framework, presenting the Structural Resilience Principle and its implications for system design. Section 6.3 develops the cyclical closure framework, presenting the fixed-point characterization of self-sufficiency. Section 6.4 extends the analysis to hierarchical systems. Section 6.5 derives

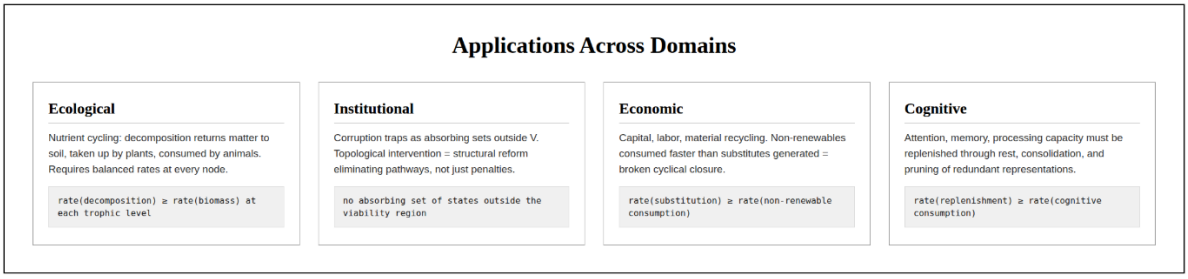
a formal sustainability criterion from the cyclical closure framework. Section 6.6 applies the criterion across domains. Section 6.7 connects the results to the viability framework of Chapter 5 and previews the transition to the four laws of self-organizing systems.

6.2 Structural Resilience in Discrete-State Systems

The formalization of structural resilience proceeds by replacing the continuous state space of the general viability framework with a finite or countable set, and replacing the admissible transformations with a transition relation on this set (Kriger, "Structural Resilience," 2019, §3). The formal apparatus begins with five basic objects. Let S be a non-empty, finite or countably infinite set of system states. Let $T \subseteq S \times S$ be the set of admissible transitions. Let $G \subseteq S$ be the set of operational (initial) states, with $G \neq \emptyset$. Let $B \subseteq S$ be the set of unsafe states, with $B \neq \emptyset$ and $G \cap B = \emptyset$. Finally, let $M \subseteq S$ be the set of mission success (goal) states, with $M \cap B = \emptyset$. The pair (S, T) forms a directed graph called the transition graph.

How can a system guarantee its own persistence? This chapter develops two answers: structural resilience through topological disconnection of unsafe states, and self-sufficiency through cyclical closure of resource flows. Together they characterize the full spectrum from environmentally dependent to fully autonomous persistence.







KRIGER, B. • STRUCTURAL RESILIENCE (2019) • SELF-SUFFICIENT SYSTEMS (2024) • CYCLICAL HIERARCHICAL SYSTEMS (2024)

The viability set V of the general framework corresponds here to the complement of B: the system persists—remains viable—if its state remains outside the unsafe set. The reachability

operator Reach$_T$ is defined as the set of all states reachable from a given set A via finite directed paths in the transition graph. A local evolution rule is a function π: $S \to S$ that selects, at each state, one of the available transitions. The specific form of π determines local behavior, but—and this is the central insight—structural resilience properties are independent of this choice.

6.2.1 The Distinction Between Soft and Topological Interventions

A key conceptual contribution of our resilience framework is the distinction between two fundamentally different types of intervention (Kriger, "Structural Resilience," 2019, §5). A **soft intervention** modifies the weight function on transitions—the costs or penalties associated with different state changes—while leaving the set of admissible transitions unchanged. A **topological intervention** removes transitions entirely, altering the structure of the transition graph itself.

This distinction matters because soft interventions are fundamentally unstable under perturbation. We prove this as Lemma 5.4 in the paper: for any path from operational to unsafe states that carries finite weights, and for any perturbation tolerance $\varepsilon > 0$, there exists an ε-perturbed selection rule under which that path is traversed with positive probability. The proof constructs such a rule explicitly: at each state along the path, the dangerous transition remains admissible whenever the perturbation budget exceeds the local cost differential. Since the perturbation is arbitrary and the costs are finite, no weight assignment can guarantee avoidance of the unsafe path.

The implication is that soft interventions—incentive structures, penalties, cost modifications—cannot provide structural safety guarantees. They make dangerous transitions less likely under nominal conditions, but they cannot make them impossible. Only the removal of transitions—the topological disconnection of unsafe states from operational states—provides guarantees that hold regardless of local dynamics, perturbations, or adversarial inputs.

6.2.2 The Structural Resilience Principle

The central result of the discrete-state resilience framework is the Structural Resilience Principle (Kriger, "Structural Resilience," 2019, Theorem 6.1). It states: there exists a subtractive reconfiguration $T' = U(T)$ such that Reach$_{T'}(G) \cap B = \emptyset$ if and only if there exists a finite set $\Delta T \subseteq T$ such that removing ΔT disconnects all directed paths from G to B.

The proof proceeds in both directions. For the forward direction, suppose a reconfiguration achieves safety. Define ΔT as the set of removed transitions. If this set did not disconnect all paths from G to B, then some path would survive in the reconfigured graph, contradicting safety. For the reverse direction, suppose ΔT disconnects all such paths. Define the reconfigured transition set as $T \setminus \Delta T$. If any unsafe state remained reachable, a path using only surviving transitions would exist, contradicting the disconnection property.

Two corollaries follow immediately. First, structural resilience is a topological property of the transition graph, independent of edge weights or local evolution rules. Second, the minimum number of transitions whose removal achieves structural resilience equals the minimum (G, B)-cut in the transition graph—a quantity computable in polynomial time using standard maximum-flow algorithms such as Dinic's algorithm or the Edmonds-Karp variant of Ford-Fulkerson.

6.2.3 Rule-Independence and Its Significance

Perhaps the most striking property of structural resilience is its rule-independence (Kriger, "Structural Resilience," 2019, Theorem 10.3). If the transition graph has been reconfigured so that Reach$_T(G) \cap B = \emptyset$, then the system is robust with respect to any class of local evolution rules—including adversarial, stochastic, and unknown dynamics. The proof is immediate: a trajectory under any rule π is a path in the transition graph, and if no path from G to B exists, no trajectory can reach B regardless of how π selects among available transitions.

This rule-independence property distinguishes topological safety guarantees from the guarantees provided by continuous control methods such as control barrier functions. In continuous settings, safety guarantees typically depend on the specific control law and on Lipschitz conditions. In the discrete structural framework, safety holds universally across all possible local dynamics once the topological disconnection is established.

6.2.4 Liveness-Preserving Reconfiguration

Structural resilience ensures safety but does not by itself guarantee liveness—the system's ability to reach desirable goal states. Removing transitions to isolate unsafe states might inadvertently disconnect paths to mission success states M, or create deadlocks from which no further evolution is possible. We address this through the concept of valid reconfiguration

(Kriger, “Structural Resilience,” 2019, §4.3): a reconfiguration is valid if it jointly satisfies safety (no path from G to B), liveness (every operational state can reach some goal state), and deadlock-freedom (every reachable non-goal state has at least one outgoing transition).

The existence of valid reconfigurations is characterized precisely: a valid reconfiguration exists if and only if there exists a set of transitions whose removal disconnects all paths from G to B while preserving, for every operational state, at least one path to some mission success state. This characterization shows that the safety–liveness trade-off, often treated as an informal design concern, admits a precise graph-theoretic formulation with computable solutions.

6.2.5 Residual Threat and Observability

Beyond the binary question of whether a system is resilient, our framework provides a quantitative measure of residual threat (Kriger, “Structural Resilience,” 2019, §8). The residual threat function $RT(s)$ at a state s counts the number of states reachable from s that can themselves reach the unsafe set B. Structural resilience implies $RT(g) = 0$ for all operational states g—not merely that the unsafe set is unreachable, but that no state on any trajectory from G is even adjacent to the unsafe region. This eliminates latent risk: states that, while not themselves unsafe, could lead to unsafe states under further perturbation.

The framework also incorporates observability through an observation mapping $O: S \to \Sigma$ that projects internal states to an observable signal space. A key result is that if structural resilience has been achieved, safety is preserved under any observation mapping and any observation-based local rule. Observability becomes relevant precisely when the system is *not* structurally resilient: in that case, the observation mapping influences which paths are likely to be taken, and observability determines which threats can be detected—a point that connects directly to the cognitive demands analyzed in Part III of this thesis.

6.2.6 Applications and Interpretations

The Structural Resilience Principle, though developed in our paper through a cybersecurity example—where states represent access levels and transitions represent privilege escalation pathways—has broad applicability to the systems studied in this thesis. In gene regulatory networks, the unsafe states B correspond to pathological steady states such as cancerous

phenotypes; the transition graph encodes which expression patterns can lead to which others; and structural resilience analysis identifies the minimum set of regulatory interactions whose disruption would topologically isolate the pathological states. In social networks, unsafe states correspond to stable but undesirable equilibria such as corruption traps or polarization lock-in; topological intervention corresponds to institutional reforms that eliminate specific transition pathways rather than merely penalizing them. In organizational systems, unsafe states correspond to dysfunctions from which recovery is structurally impossible without external intervention; the framework identifies precisely which structural features make such traps absorbing.

The key insight across all these domains is the same: soft interventions—incentives, penalties, nudges—are fundamentally limited because they operate within the existing transition structure. Topological interventions—structural reforms that eliminate pathways—provide guarantees that hold regardless of local dynamics. This distinction between parametric and structural approaches to system safety will recur throughout the thesis.

6.3 Cyclical Closure and the Conditions for Self-Sufficiency

A self-sufficient system is one whose operations can be sustained indefinitely without external input. This is a stronger condition than mere persistence: a persistent system may depend on a continuous flow of resources from its environment, whereas a self-sufficient system generates everything it needs internally, through cyclical processes that return resources to the beginning of the cycle. The formal characterization of this condition is the subject of our paper *"On the Possibility of Self-Sufficient Systems: Fixed Points and Cyclical Closure"* (Kriger, "Self-Sufficient Systems," 2024), which provides the mathematical existence conditions, and our paper *"Cyclical Hierarchical Systems"* (Kriger, "Cyclical Hierarchical Systems," 2024), which extends the analysis to multi-level structures and develops the philosophical implications.

6.3.1 Foundational Hierarchies and Cyclical Closure

The mathematical framework begins with the concept of a foundational hierarchy (Kriger, "Self-Sufficient Systems," 2024, Definition 2.1). A foundational hierarchy is a triple $(S, \leq, F)$ where S is a non-empty set (the state space), $\leq$ is a binary relation on S (the grounding relation), and F: $S \to S$ is a determination function such that $x \leq y$ implies $F(y)$ is defined and determines x. A hierarchy is well-founded (or open) if the grounding relation contains no infinite

descending chains. A hierarchy is cyclically closed of order *n* if there exist elements S_1, S_2, ..., S_n and determination functions F_1, F_2, ..., F_n such that each element is determined by the next—$S_i = F_i(S_{i+1})$—and the final element is determined by the first: $S_n = F_n(S_1)$.

6.3.2 The Fixed-Point Equivalence

The central mathematical result is the Fixed-Point Equivalence (Kriger, "Self-Sufficient Systems," 2024, Proposition 2.4): a cyclically closed hierarchy of order *n* exists if and only if the composite operator $\Phi = F_1 \circ F_2 \circ \cdots \circ F_n$ has a fixed point. The proof is direct: by iterated substitution, $S_1 = F_1(S_2) = F_1(F_2(S_3)) = \cdots = \Phi(S_1)$. Conversely, if $\Phi(S_1) = S_1$, the remaining elements of the cycle can be recovered by applying the appropriate subsequences of determination functions.

This equivalence is the key reduction: it transforms the question of whether self-sufficient cyclically closed structures can exist into the mathematically precise and well-characterized question of whether a composite operator has a fixed point. The answer depends on the structural properties of the state space and the determination functions, and three classical fixed-point theorems provide sufficient conditions.

6.3.3 Existence Conditions

We deploy three classical fixed-point theorems to establish sufficient conditions for cyclical closure (Kriger, "Self-Sufficient Systems," 2024, §3; Kriger, "Cyclical Hierarchical Systems," 2024, §5.2). The Knaster-Tarski theorem guarantees that if the state space admits a complete lattice structure and each determination function is order-preserving (monotonic), the set of fixed points of the composite operator is non-empty and itself forms a complete lattice. This provides not merely existence but a rich algebraic structure on the set of possible cyclically closed configurations. The Banach Contraction Principle guarantees that if the state space is a complete metric space and the composite operator is a contraction—that is, there exists $k \in [0, 1)$ such that $d(\Phi(x), \Phi(y)) \leq k \cdot d(x, y)$ for all x, y—then a *unique* cyclically closed hierarchy exists, and iterative processes converge to it. The Brouwer fixed-point theorem guarantees that if the state space is a non-empty compact convex subset of $\mathbb{R}^n$ and the composite operator is continuous, at least one fixed point exists.

Each theorem applies under different hypotheses and provides different guarantees. The Knaster-Tarski result is the most general but the least constructive; the Banach result is the most specific (requiring contractivity) but provides uniqueness and convergence; the Brouwer

result occupies an intermediate position, requiring topological conditions (compactness, convexity) that are often satisfied in physical applications. Together, they establish that cyclical closure is possible across a wide range of mathematical structures.

6.3.4 Topological Characterization: Essential Cyclicity

Beyond proving that cyclically closed hierarchies can exist, we characterize when their cyclical structure is essential—intrinsic to the topology and not removable by continuous deformation (Kriger, "Self-Sufficient Systems," 2024, §4; Kriger, "Cyclical Hierarchical Systems," 2024, §5.3). Drawing on algebraic topology (Hatcher, *Algebraic Topology*, 2002), we define a structure as essentially cyclic if three conditions are met: the space encoding the hierarchy is compact, it is connected, and its first homology group $H_1(X; \mathbb{Z})$ is non-trivial.

The significance of this homological condition is made precise by Proposition 4.2 of our self-sufficient systems paper (equivalently Proposition 5.8 of the cyclical hierarchical systems paper): if a structure is essentially cyclic, then there exists no continuous retraction that would contract the cycle to a point while preserving the cyclic structure. The proof proceeds by contradiction: such a retraction would induce a homomorphism mapping a non-trivial homology group to the trivial group, while retractions must induce surjections on homology—a contradiction. The first Betti number $b_1 = \text{rank}(H_1)$ counts the number of independent cycles in the structure; a positive first Betti number is a necessary condition for essential cyclicity.

We note, as we do in the papers, that this characterization is encoding-dependent: different ways of representing a hierarchical structure as a topological space may yield different homology groups. We regard this not as a defect but as a feature of the formalism—the choice of encoding is part of the formal model, and our claim is conditional: if a structure admits an encoding with non-trivial H_1, its cyclicity is essential in the precise sense formalized above.

6.4 Hierarchical Self-Sufficiency

The cyclical closure framework of the preceding section applies to systems considered as wholes. Our companion paper on cyclical hierarchical systems extends the analysis to systems composed of multiple interacting subsystems arranged in a hierarchy (Kriger, "Cyclical Hierarchical Systems," 2024, §8). The main additional result is that hierarchical self-sufficiency requires cyclical closure at every level of the hierarchy, not just at the top level.

The argument is straightforward. Consider a system composed of subsystems, where each subsystem's functioning depends on outputs from other subsystems. If the system as a whole

achieves cyclical closure—all resource flows at the system level form closed loops—but some subsystem does not achieve closure at its own level, then that subsystem requires external input to maintain its functioning. Since the subsystem is internal to the system, its "external input" must come from other subsystems. But if those other subsystems similarly lack closure, the demand propagates. True self-sufficiency therefore requires that cyclical closure obtain at every level: each subsystem must be either self-sufficient in its own right or engaged in cyclical exchange with other subsystems at the same hierarchical level.

This multi-level requirement has important implications. It means that self-sufficiency cannot be assessed solely from the top-level description of a system. A corporation that appears financially self-sufficient (revenue exceeds expenditure) may be unsustainable if its internal departments require resources that no internal process generates. An ecosystem that appears to cycle nutrients at the macro level may be unsustainable if particular trophic levels are in deficit. The hierarchical extension of cyclical closure provides the formal tools for identifying such hidden dependencies.

The tripartite distinction among source, foundation, and primacy—developed in our cyclical hierarchical systems paper (Kriger, "Cyclical Hierarchical Systems," 2024, §4.2)—is directly relevant here. In a hierarchically self-sufficient system, different subsystems play different structural roles. Subsystem A may be the source of resources for subsystem B (providing what B consumes), while B is the foundation of A (providing the conditions without which A cannot function), and primacy circulates among the subsystems rather than inhering in any single one. In linear hierarchies, source, foundation, and primacy collapse into one: the first element in the chain plays all three roles. In cyclically closed hierarchies, they diverge, and recognizing this divergence is essential for correctly analyzing the system's dependencies.

6.5 The Sustainability Criterion

The cyclical closure framework yields a formal criterion for sustainability that is both more precise and more general than informal notions of the concept. We define it as follows: a system is sustainable if and only if it achieves cyclical closure in its resource flows—that is, every resource consumed by one process is produced by another process within the system, and the rates of production and consumption are balanced.

In formal terms, sustainability requires the existence of a fixed point of the resource-flow map—a configuration in which the production and consumption of every resource are exactly

balanced. By the Fixed-Point Equivalence (Proposition 2.4 of our self-sufficient systems paper), this is equivalent to the existence of a cyclically closed hierarchy in the resource-flow structure. The existence conditions established by the three fixed-point theorems (Knaster-Tarski, Banach, Brouwer) therefore provide precise criteria for sustainability: the resource flow graph must contain at least one cycle, the stoichiometric coefficients must admit a positive solution, and the rates must be compatible with the thermodynamic constraints on the system.

This criterion is more precise than informal notions of sustainability because it specifies exactly what "sustainable" means in formal terms—the existence of a fixed point of the resource-flow map. It is more general because it applies to any system with resource flows, not just ecological or economic systems. And it is actionable: given a formal description of a system's resource flows, one can check whether the fixed-point conditions are satisfied and, if they are not, identify precisely which flows are out of balance.

6.6 Applications Across Domains

6.6.1 Ecological Systems

In ecological systems, cyclical closure corresponds to the familiar concept of nutrient cycling: the decomposition of organic matter returns nutrients to the soil, which are then taken up by plants, consumed by animals, and returned to the soil upon death. The sustainability criterion shows that ecological sustainability requires not merely that such cycles exist but that they be balanced—that the rates of production and consumption at each node are compatible. An ecosystem in which decomposition cannot keep pace with biomass production, or in which a keystone nutrient is lost to geological sequestration faster than it is released, fails the cyclical closure condition and is therefore not sustainable by the formal criterion.

6.6.2 Economic Systems

In economic systems, cyclical closure corresponds to the recycling of capital, labor, and materials. The criterion shows that an economy that consumes non-renewable resources at a rate exceeding the rate at which substitutes are generated is not in cyclical closure and is therefore not sustainable by the formal criterion—regardless of how efficiently it operates in other respects. This is not a moral judgment but a structural accounting: the fixed-point condition is either satisfied or it is not, and when it is not, the system's persistence depends on continued access to external inputs whose supply is finite.

6.6.3 Cognitive Systems

In cognitive systems, cyclical closure takes a more abstract form: the cognitive resources consumed by the process of maintaining a world-model—attention, memory, processing capacity—must be replenished by the same process, through rest, consolidation, and the elimination of redundant representations. A cognitive system that consumes attentional resources faster than it can replenish them—through, for example, an ever-expanding model that requires increasing maintenance—violates cyclical closure and is therefore unsustainable. Chapter 11 will develop this application in detail, showing that the architecture of memory and consciousness can be understood as mechanisms for achieving cognitive cyclical closure.

6.6.4 Quantum Systems as an Illustrative Case

As an illustrative case that is formally rather than metaphorically precise, our self-sufficient systems paper applies the cyclical closure framework to self-consistent quantum states (Kriger, "Self-Sufficient Systems," 2024, §5). Consider a tripartite quantum system with subsystems A, B, C and Hilbert space $H = H_a \otimes H_e \otimes H_o$. The reduced density matrices satisfy mutual determination via partial traces. For certain global states, the reduced states form a cyclically determined structure: the properties of each subsystem are fixed by its correlations with the others, with no subsystem serving as an independent foundation. The space of density matrices on a finite-dimensional Hilbert space is non-empty, compact, and convex; by the Brouwer fixed-point theorem, a self-consistent configuration exists whenever the constraint map is continuous. This demonstrates that cyclical closure is not merely a mathematical abstraction but is instantiated in fundamental physics.

6.7 Connections and Implications

The results of this chapter connect to the viability framework of Chapter 5 in a precise way. The structural transformation framework establishes that persistence requires transformation; the resilience framework of Section 6.2 establishes when that persistence can be guaranteed against all possible perturbations in discrete-state systems; and the cyclical closure framework of Sections 6.3–6.5 establishes when a system can achieve the strongest possible form of persistence—self-sufficiency without external dependence.

Self-sufficiency through cyclical closure is one strategy for maintaining viability—arguably the most robust strategy, because it eliminates dependence on external inputs. But it is not the only strategy. A system that lacks cyclical closure can still persist if its environment reliably

provides the resources it needs—but its persistence is then contingent on environmental conditions, and it is vulnerable to disruptions of its resource supply. The Structural Resilience Principle provides the tools for analyzing such vulnerability: if the transition graph connecting operational to unsafe states contains a finite separating cut, the system can be made resilient through topological intervention; if it does not, resilience requires continuous monitoring and adaptive response.

The distinction between self-sufficient and environmentally dependent persistence will prove crucial in later chapters. Cognitive systems (Part III) are typically not self-sufficient: they depend on sensory input from their environment. Social systems (Part IV) are typically not self-sufficient either: they depend on the continued cooperation of their members. The viability framework shows that environmental dependence is not inherently fragile—but it requires the system to monitor its environment and adapt to changes, which introduces the cognitive demands that Part III will analyze in detail.

The structural resilience framework also connects forward to the analysis of social systems in Part IV. The distinction between soft and topological interventions—between incentive modification and structural reform—provides a precise vocabulary for analyzing institutional design. When an institution faces the possibility of reaching an unsafe state (corruption, capture, dysfunction), the framework shows that incentive-based interventions (soft constraints) are inherently unstable under perturbation, while structural reforms that eliminate pathways to unsafe states (topological interventions) provide guarantees independent of the behavior of individual agents.

With the viability framework (Chapter 5) and the specializations for resilience and self-sufficiency (the present chapter) in place, the stage is set for the four laws of self-organizing systems, which characterize the structural properties that any self-maintaining system must possess. The viability framework establishes *that* systems must maintain themselves; the resilience framework establishes *when* maintenance can be guaranteed; the cyclical closure framework establishes *how* maintenance can be made independent of external conditions. The four laws, developed in the next chapter, will establish *what structural properties* a self-maintaining system must exhibit—completing the formal foundation on which the remainder of the thesis is built.

CHAPTER 7. FOUR LAWS OF SELF-ORGANIZING SYSTEMS

This chapter presents four formally grounded laws governing self-organizing peer systems, drawing on the unified framework developed in Kriger (2026). The laws address three persistent theoretical problems in self-organization theory: whether external coordination is necessary for stable organization, what minimal conditions are required for system viability, and how systematic errors in system evaluation arise. The framework demonstrates the deep interconnection among these problems and resolves them within a restricted but precisely defined class of systems.

The approach taken here differs fundamentally from frameworks that assume cooperative behavior as a premise. The central innovation is the derivation of cooperative response as an emergent property of unperturbed peer systems, rather than its stipulation as an axiom. This shifts the explanatory burden: within the specified system class, antagonistic behavior requires explanation, not cooperation. The four laws—numbered Zero through III to reflect their logical dependencies—form a coherent diagnostic framework grounded in evolutionary game theory, commons governance research, viability theory, and information theory.

The laws apply exclusively to a well-defined system class: peer systems with shared environment, repeated interaction, finite resources, and local feedback. This scope restriction is not evasion but precision. Different system classes follow different dynamics, and conflating them produces confused theory. Predator-prey dynamics, hierarchical authority structures, one-shot interactions, and systems with infinite resources are explicitly excluded.

7.1. Foundational Framework

The formal apparatus requires several foundational definitions. A system is defined as a tuple $S = (A, X, T, F)$, where A is a set of agents, X is a state space, T is an interaction topology, and $F : X \to X$ is a transition function. The viability region is a subset $\Omega \subseteq X$ such that for all states x in Ω, the system satisfies specified persistence conditions. These definitions connect directly to the viability framework developed in Chapter 5, now applied to multi-agent contexts.

Three additional definitions are essential to the framework. A response function $\varphi : X \times A \to \mathbb{R}$ measures the partial derivative of aggregate system output with respect to an agent's action at a given state. The sign of this function—positive (cooperative) or negative (antagonistic)—is not assumed but derived. A signal function $\sigma : X \to \mathbb{R}$ maps system states to observable

information output, capturing the fact that different states generate different amounts of detectable signal. An antagonistic perturbation $P : A \times A \to \mathbb{R}$ modifies agent interaction payoffs such that the perturbed payoff $U'(a_i, a_j) = U(a_i, a_j) + P(a_i, a_j)$, where P creates zero-sum or negative-sum structure between agents.

The concept of perturbation P requires careful definition to avoid circularity. P must be structurally separable from the baseline game: it must be possible, at least in principle, to remove P while keeping the underlying peer system intact. P is identified relative to an empirically grounded baseline—historically or experimentally observed cooperative dynamics under less-manipulated conditions. This excludes defining P as anything that makes antagonism possible, which would render the framework tautological.

7.2. The Peer System Class

The laws developed in this chapter apply exclusively to peer systems. Two agents are peers if and only if three conditions hold: their viability regions overlap, their viability depends on a shared environmental resource, and either agent could occupy the other's role. Formally, agents a_i and a_j are peers when $\Omega(a_i) \cap \Omega(a_j) \neq \emptyset$, both agents' viability depends on a common resource pool E, and role symmetry obtains between them.

The peer system definition excludes structurally asymmetric arrangements. Wolves and deer are not peers because a deer cannot occupy the wolf's ecological role—they fail the role-symmetry condition. Two wolves, however, satisfy all three conditions: overlapping viability regions, shared resource dependence, and role interchangeability. Likewise, two firms competing in the same market and two nations with comparable military capacity qualify as peer systems.

This distinction resolves the most common objection to cooperation-as-default: the stability of predator-prey dynamics. At the intra-species level, wolves cooperate with wolves and deer cooperate with deer; within each species, the cooperative framework applies. At the inter-species level, the wolf-deer relationship is not a peer system because the agents occupy structurally asymmetric positions with different viability constraints. At the ecosystem level, predator-prey dynamics are cooperative in a meta-sense, regulating population, preventing overgrazing, and maintaining biodiversity. The apparent counterexample dissolves under proper scope specification.

7.3. Law Zero: Cooperative Response as Emergent Default

Law Zero. *In a multi-agent peer system with shared environment, local feedback, and absence of externally imposed antagonistic perturbations, cooperative response ($\varphi > 0$) emerges as the statistically dominant and dynamically stable configuration. Antagonistic response ($\varphi < 0$) requires continuous external energy input to maintain.*

Law Zero is the foundational result. It establishes that cooperative response is not an assumption requiring justification but a derivable property of unperturbed peer systems. The argument proceeds through four interconnected parts.

7.3.1. Statistical Basin Asymmetry

The first part of the argument establishes a statistical asymmetry in the sizes of parameter regions supporting cooperative versus antagonistic equilibria. In peer systems where all agents draw from a shared resource pool E and contribute to it through their actions, the net effect on E depends on the balance between extraction and contribution rates across agents. Under cooperative dynamics, agents' strategies satisfy a sustainability condition: aggregate contributions match or exceed aggregate extraction minus environmental regeneration. Under antagonistic dynamics among peers, each agent maximizes individual extraction without coordinating on aggregate sustainability—the multi-agent tragedy of the commons.

The key claim is not that antagonism is impossible but that it occupies a narrower basin of stability in parameter space. Cooperative equilibria exist for a wider range of parameter values—regeneration rates, population sizes, discount factors. Antagonistic equilibria require either very high environmental regeneration making overextraction harmless, very small populations making coordination trivial, or external enforcement preventing defection. This is formalized via the price of anarchy concept from mechanism design: for peer systems with shared resources, the ratio of optimal social welfare to Nash equilibrium welfare under selfish dynamics is generically greater than one, and selection pressure favors systems that approach the optimum.

7.3.2. Game-Theoretic Foundations

The second part draws on established results from evolutionary game theory. Five mechanisms can sustain cooperation: kin selection, direct reciprocity, indirect reciprocity, network reciprocity, and group selection. In peer systems with repeated interaction and local information, at least direct and network reciprocity are operative. The critical parameter is the

benefit-to-cost ratio of cooperative acts relative to interaction network structure. When this ratio exceeds a threshold determined by network topology, cooperation invades and stabilizes. Peer systems with shared environment tend to satisfy this condition because contributions to the shared resource pool benefit all agents while individual contribution costs are bounded. Antagonistic strategies can constitute evolutionarily stable strategies in one-shot games or when the benefit-to-cost ratio falls below the network threshold, but these conditions are precisely what peer systems with shared environment tend to violate: iteration is guaranteed by ongoing resource dependence, and the shared-environment structure amplifies benefits. This is a generic claim, not a universal one. In some social dilemmas, such as certain oligopolistic markets with differentiated products, mixed or mildly antagonistic equilibria may persist without obvious perturbation.

7.3.3. Formal Characterization of Perturbation

The third part provides explicit mathematical characterization of antagonistic perturbation P. The perturbation operator modifies baseline payoffs and is classified as antagonistic when it satisfies any of four conditions: zero-sum injection (where mutual payoffs sum to zero or less), relative payoff weighting (where outcomes are ranked by differences between agents rather than absolute values), cooperation penalties (where mutual cooperation yields reduced payoffs), or defection subsidies (where exploitation of cooperators yields enhanced payoffs). Concrete examples admit quantification. Artificial scarcity introduces a penalty proportional to the overlap in agents' demands; cooperation destabilizes when scarcity intensity exceeds a computable threshold. Relative performance incentives weight relative standing against absolute gains; cooperation destabilizes when the relative weighting exceeds a threshold determined by the benefit-to-cost ratio. Exit barriers reduce the effective discount rate, shrinking the shadow of the future. Information suppression drives observation probability toward zero, disabling reciprocity-based strategies. Each of these perturbations has a quantifiable threshold below which cooperative dynamics are maintained and above which they break down.

7.3.4. Perturbation Dynamics and Relaxation

The fourth part addresses the dynamics of perturbation and recovery. When antagonistic perturbation is externally maintained—by institutional design, resource control, or information management—the artificial fixed points in the defection region persist. When P

weakens or is removed, the fitness landscape reverts to baseline dynamics, and the system relaxes toward its natural cooperative attractor. The relaxation follows gradient dynamics on a potential function for the baseline game.

This generates the central diagnostic principle: when observing stable antagonism in what appears to be a peer system, search for P. Formalized candidates include institutional rules creating zero-sum structure, resource distribution mechanisms generating artificial scarcity, information control systems suppressing reciprocity, and ideological structures that frame peer relations as inherently competitive.

7.3.5. Corollary: Diagnosing Stable Antagonism

Stable antagonism among peers is diagnostic of one of four conditions: misclassification (the system is not actually a peer system), transience (the system has not yet equilibrated), active perturbation (P is maintaining antagonistic equilibrium), or boundary violation (the conditions for Law Zero do not hold). The claim is not that antagonism cannot exist, but that in properly specified peer systems, its stability requires explanation. Cooperation requires no such external maintenance.

7.3.6. Boundary Conditions

Law Zero applies when all of the following conditions hold: agents have symmetric viability constraints and comparable capacities (peer structure), a common resource pool couples agents' fates (shared environment), iteration count is high or indefinite (repeated interaction), agents can distinguish partners and track history (identifiability), resources are bounded so that overextraction is costly (finite resources), and agents observe the effects of their actions (local feedback).

The law does not apply when agents are structurally asymmetric, interaction is one-shot or very low iteration, partners are anonymous, resources are effectively infinite, or feedback is absent or heavily delayed. This boundary specification is essential: the law makes precise claims about a precise class, and extending it beyond that class would produce false predictions.

7.3.7. Resource Scarcity Regimes

A critical subtlety concerns the distinction between resource scarcity as a boundary condition and artificial scarcity as a form of perturbation. Three regimes are distinguished. Under abundance, resources exceed aggregate demand and cooperation is trivial; Law Zero holds but

is uninteresting. Under moderate scarcity, resources are finite but sufficient for viability if managed cooperatively; this is the core domain of Law Zero. Under extreme scarcity, resources are insufficient for all agents' viability regardless of behavior; this is a boundary violation, not a perturbation—the system exits Law Zero's domain.

Critically, perturbation P can convert abundance or moderate scarcity into apparent extreme scarcity. When resources are sufficient but access is restricted—through enclosure, patents, credentialism, or hoarding—agents experience scarcity that is not intrinsic but imposed. This artificial scarcity is paradigmatic P: removable in principle, converting cooperative-capable systems into zero-sum competitions. Real-world systems rarely occupy stable regimes; the scarcity ratio fluctuates, and systems near the boundary between moderate and extreme scarcity exhibit latent antagonism, hysteresis from past scarcity episodes, and oscillation-induced cooperation breakdown.

7.3.8. The Freeloader Problem: Tolerance over Enforcement

A persistent objection to cooperation-as-default is the freeloader problem: if cooperation is voluntary, rational agents will exploit cooperators. The framework argues that robust cooperative systems solve this not through enforcement—punishment, exclusion, monitoring—but through structural tolerance. Five mechanisms support this. Redundancy and surplus allow freeloaders to consume slack without threatening viability, as in open-source software where the vast majority of users contribute nothing yet the system thrives. Locality and decoupling limit freeloaders' impact to their local network neighborhood. Frequency-dependent limitation makes freeloading self-limiting above a threshold frequency. Dynamic network restructuring naturally excludes persistent antagonists. Most powerfully, strategic ignoring—systematic non-engagement with antagonistic agents—operates at zero cost, denies the attention that sustains antagonism, and blocks contagion chains.

The tolerance-enforcement tradeoff depends on the surplus ratio and comparative costs. Even in low-surplus systems, enforcement costs—monitoring, judgment, punishment, errors, second-order effects, cultural costs, and arms race dynamics—frequently exceed tolerance costs, because enforcement costs are systemic while freeloading costs are marginal.

7.3.9. Internalized Perturbation and Relaxation Lag

A critical limitation in human systems is that P can be internalized through cultural transmission. When external perturbation is removed, internalized perturbation P_int persists

through socially transmitted beliefs, habits, and institutional norms. Two timescales emerge: behavioral adaptation (fast) and cultural decay (slow, generational). When the cultural decay timescale greatly exceeds the behavioral adaptation timescale, the system exhibits relaxation lag—behaviors may shift toward cooperation while underlying internalized perturbation maintains latent antagonism.

Internalized perturbation is distinguished from native preferences through five diagnostic criteria: historical traceability to prior external perturbation, cross-cultural variation, decay under non-reinforcement, narrative dependency, and counterfactual test results. This distinction is essential because it prevents the framework from becoming unfalsifiable by attributing all observed antagonism to hidden perturbation.

7.4. Law I: Self-Organization Efficiency

Law I. *Given the cooperative default established in Law Zero, external constraints that do not track local information generally reduce achievable welfare relative to unconstrained dynamics.*

Law I is a corollary of Law Zero, not an independent law. It articulates what follows from $\varphi > 0$ in the unperturbed regime when external constraints are introduced. The argument proceeds by contradiction: if an external constraint C, operating independently of local information, improved outcomes universally, then for all states the constrained welfare would exceed unconstrained welfare. This would imply that agents' response functions are uniformly suboptimal—contradicting the positive response function derived in Law Zero. Therefore, there exist states where unconstrained dynamics achieve at least as high welfare as constrained dynamics.

This result has important nuances. It does not claim that all external coordination is harmful—constraints that incorporate local information may genuinely improve outcomes. It claims that blind constraints—those imposed without tracking the system's actual state—reduce welfare in the cooperative regime. The connection to the epistemic framework of Chapter 4 is direct: an external controller faces the same observability limitations as any observer, and the Law of Imperative Uncertainty guarantees that its model of the system's state will be incomplete. Constraints based on incomplete models will, for some states, move the system away from rather than toward viability.

Law I connects to Ostrom's empirical findings on commons governance. Self-governing commons institutions, operating with local information and cooperative norms, consistently outperform externally imposed management regimes. The framework explains why: external management introduces constraints C that do not track local feedback, while self-governance preserves the information channels through which cooperative agents coordinate.

7.5. Law II: Viability Sufficiency

Law II. *For systems with threshold-defined viability, constructing conditions sufficient for viability does not require high systemic complexity. Observed necessity for extreme complexity is diagnostic of discretionary overconstraint, not structural requirement.*

Law II is one of the framework's more philosophically interesting claims. It rests on a distinction between two types of viability and two types of constraint. Threshold viability (Case A) defines the viability region as the set of states satisfying a small number of inequality constraints: the system survives when a few key variables remain above specified thresholds. Optimization viability (Case B) requires continuous coordination across high-dimensional state variables, typically because viability is defined relative to competitors or adversaries.

For Case A systems, the constraint set divides into essential constraints (C_E), which define the actual viability boundary, and discretionary constraints (C_D), which reflect preferences, traditions, or imposed requirements beyond what viability demands. The formal content of Law II is that the cardinality of C_E is bounded for threshold-viability systems. This is supported by viability theory, by organizational studies showing that viable organizations satisfy requisite variety through bounded regulatory mechanisms, and by ecological analysis showing that ecosystem viability typically requires satisfaction of three to seven fundamental constraints.

The diagnostic application is sharp: when analyzing systems that appear to require enormous complexity for survival, ask whether constraints can be removed without exiting the viability region. If so, apparent complexity exceeds true complexity. The perceived need for high complexity is often a symptom of overconstraint—the system has accumulated discretionary requirements that masquerade as survival necessities. This connects to Law I: many discretionary constraints are externally imposed and do not track local information, further degrading system performance.

7.6. Law III: Observational Asymmetry

Law III. *In any system where functional states generate low observational signal and failure states generate high observational signal, external evaluations based on observed information systematically underestimate functionality and overestimate failure frequency.*

Law III addresses a systematic bias in the observation of self-organizing systems. The state space is partitioned into functional and non-functional states, and a signal asymmetry parameter α is defined as the ratio of expected signal generated by non-functional states to expected signal generated by functional states. When $\alpha > 1$—failure is noisier than success—observation probability is proportional to signal intensity, and Bayesian updating yields a precise result: the observed functionality ratio is strictly less than the true functionality ratio.

The proof is direct. Let ρ denote the true proportion of functional states. The observed proportion, after Bayesian correction, equals ρ divided by the sum of ρ and α times the complement of ρ. For any $\alpha > 1$ and any true functionality between zero and one, this observed ratio is strictly less than ρ. The bias is monotonically increasing in α: the greater the signal asymmetry, the more severely functionality is underestimated.

This law connects the epistemic framework of Chapter 4 to the evaluation of self-organizing systems. The bias it identifies is not a cognitive error to be corrected by better thinking; it is a structural consequence of signal asymmetry that affects any observer who must rely on generated information rather than direct state access. Correction requires awareness of the bias, knowledge of the signal asymmetry parameter, and cognitive resources to perform the adjustment—conditions often unmet in practice.

Law III has practical consequences for the evaluation of cooperative systems. Self-organizing systems that function well generate little signal: stable commons do not make headlines, healthy ecosystems do not demand attention, functional institutions do not produce crises. Failing systems generate abundant signal: commons tragedies, ecological collapses, and institutional failures are visible, documented, and salient. The result is systematic underestimation of cooperative self-organization's prevalence and effectiveness—a structural explanation for the widespread assumption that external coordination is necessary for stable organization.

7.7. Synthesis: How the Four Laws Interact

The four laws form a coherent theoretical structure with a definite logical architecture. Law Zero establishes that cooperative response is the statistically dominant default for peer

systems. Law I shows that this cooperation yields efficient self-organization when unobstructed by externally imposed constraints. Law II demonstrates that viability does not require the complexity often claimed, because threshold-viability systems need only bounded essential constraints. Law III explains why successful cooperative systems remain invisible while failures dominate perception—a structural account of why self-organization's effectiveness is systematically underappreciated.

The central insight is the derivation of $\varphi > 0$ rather than its assumption. This transforms the framework from a conditional claim—if agents are cooperative, then self-organization follows—into a structural claim: agents in shared environments converge to cooperation unless externally perturbed. Combined with Law III's observational bias, this produces a striking unified result: a multi-agent peer system with shared environment, repeated interaction, and finite resources, operating within a threshold viability region and subject to signal-asymmetric observation, will converge to cooperative dynamics when unperturbed, achieve efficient self-organization in this regime, require only low complexity for viability, yet be systematically mischaracterized as unstable, complexity-dependent, and failure-prone.

This unified result has direct implications for the analysis of complex systems across domains. In multi-agent artificial intelligence systems, it predicts that cooperative protocols emerge without zero-sum objectives when agents share computational resources and interact repeatedly, while competitive benchmarking and adversarial training objectives function as perturbation P. In biological systems, endosymbiotic integration of formerly independent organisms illustrates the elimination of historical perturbation enabling deep cooperation. In human organizations, the framework provides formal grounding for the observation that peer relationships develop cooperative norms in the absence of perturbation, while artificial scarcity, competitive ranking, and information asymmetry function as identifiable P.

7.8. Validity Boundaries

The framework's claims are conditional and falsifiable, and their boundary conditions are explicitly stated. The laws do not apply when agents do not share an environment, interaction is one-shot or feedback channels are absent, resources are effectively infinite, external perturbation P is permanent and inescapable, the viability region is intrinsically high-dimensional, or observers have full continuous state access. Under these conditions, antagonistic behavior may be stable without perturbation, external coordination may be

genuinely necessary, complexity may be structurally required, and observational evaluations may be accurate.

The framework generates falsifiable predictions. Removing artificial competitive structures from peer systems should shift the response function upward. Stable peer antagonism should correlate with identifiable P. Historical transitions to cooperation should follow perturbation reduction rather than agent composition change. Weaker perturbation should correlate with higher cooperative response across systems. Natural experiments in perturbation removal should produce cooperation, not chaos. Each of these predictions is testable against empirical data, and each provides a mechanism for refuting the framework if it is wrong.

The methodological commitment of this chapter—and of the underlying formal framework—is semi-formal rather than fully axiomatic. The structured arguments draw on established results from evolutionary game theory, viability theory, mechanism design, and information theory. Where genuine formal results are available, they are presented as such; where the argument extrapolates from known models, this is explicitly stated. The framework is diagnostic rather than universal: it provides structural criteria for analyzing when cooperation is generically favored and tools for identifying perturbations that maintain antagonism.

7.9. Implications for Subsequent Chapters

The four laws established in this chapter provide the theoretical foundation for the analyses that follow. The dynamics of cognition examined in Chapter 8 can be understood through the lens of cooperative subsystem interaction, perturbation-induced dysfunction, bounded viability complexity, and observational bias in self-assessment. Social organization patterns analyzed in Chapter 9 are governed by the same framework: cooperative defaults among peers, efficiency losses from uninformed external constraints, bounded institutional complexity requirements, and systematic underestimation of functional self-governance. The conditions for institutional stability examined in Chapter 10 follow from the interaction of Law Zero's cooperative attractor with Law I's efficiency result and Law II's bounded complexity claim, modulated by Law III's observational distortions.

In each domain, the diagnostic methodology remains constant: identify whether the system satisfies peer-system conditions, determine whether observed antagonism reflects genuine boundary violations or maintainable perturbation, assess whether apparent complexity reflects essential or discretionary constraints, and correct for signal-asymmetric observation before

drawing conclusions about system functionality. The laws provide not a set of predictions about what will happen in any specific system, but a set of structural questions that, properly applied, reveal the dynamics governing self-organization in that system.

CHAPTER 8. EMERGENCE OF DYNAMICS, STRUCTURAL COMPATIBILITY LAWS AND STRUCTURAL GENESIS

This chapter presents results that complete the dynamical core of the theory. The first is a structural derivation of second-order deterministic evolution—Newtonian dynamics—from more primitive axioms.[13] The second is the Constraint–Autonomy Compatibility Law, which establishes that constraints and autonomy are not opposites but mutually necessary structural conditions in adaptive systems.[14] The third is the Viability Mismatch Law, which generalizes the concept of stress to a domain-independent formal principle governing all viable systems.[15] A corollary of the first result—the Structural Diagnostic Principle—transforms the Newtonian derivation from a theoretical result into a practical diagnostic tool for identifying structurally incomplete models across domains.[16]

The fourth result unifies these individual derivations into a single chain extending from the most primitive ontological starting point—differentiation—through metric, dynamics, finite causal speed, covariance, dynamic geometry, and a fundamental action bound, recovering the full architecture of modern physics as a structural necessity for sustainable complexity.[17]

Together, these results show how familiar dynamical structures arise from the viability framework, how the interplay between constraints and autonomy shapes the behaviour of all organised systems, and how the full architecture of physical law—from Newton through Einstein to the quantum—emerges as a structural necessity rather than an empirical accident.

---

[13]Kriger, B. (2018). Emergence of Newtonian Dynamics from Metric Inertial Systems: A Structural Derivation of Second-Order Deterministic Evolution. Zenodo. https://doi.org/10.5281/zenodo.18343870

[14]Kriger, B. (2026). The Constraint–Autonomy Compatibility Law: A Formal Derivation. Zenodo. https://doi.org/10.5281/zenodo.18368991

[15]Kriger, B. (2026). The Viability Mismatch Law: A Universal Principle for Viable Systems with Stress as Special Case. Zenodo. https://doi.org/10.5281/zenodo.18433777

[16]Kriger, B. (2018). Structural Diagnostic Principle for Dynamical Models Based on Emergence of Universal Newtonian Dynamics from Metric Inertial Systems. Zenodo. https://doi.org/10.5281/zenodo.18344823

[17]Kriger, B. (2026). Structural Genesis of Dynamical Architecture: Deriving the Necessary Form of Physical Law from Conditions on Sustainable Complexity.

The chapter proceeds in six parts: the first four correspond to the individual results, the fifth synthesises their joint implications, and the sixth presents the unified derivation chain.

8.1 Emergence of Newtonian Dynamics from Metric Inertial Systems

The standard presentation of Newtonian mechanics takes the second law ($F = ma$) as a fundamental axiom. This section reverses the direction of explanation: it derives the second law from a set of more primitive structural axioms that make no reference to force, mass, or acceleration. The central claim is conditional: *if* a system satisfies persistence, additivity, and compensation, *then* its evolution must be second-order. The axioms themselves abstract from physical intuition, and whether any given system satisfies them is an empirical question.[18]

8.1.1 The Axiomatic Framework

The derivation begins with the concept of a metric inertial system. Let (S, d) be a metric space admitting embedding in a smooth Riemannian manifold M such that the metric d derives from a Riemannian metric g on M. A state is a point $x \in S$, and an evolution is a twice-differentiable curve $x: I \to S$. The velocity at time t is $v(t) = dx/dt \in T_{x(t)}S$, and the acceleration is $a(t) = \nabla_v v$, the covariant derivative of velocity along the trajectory. In flat Euclidean space, the Christoffel symbols vanish and acceleration reduces to d^2x/dt^2.

Three structural postulates are then imposed on the dynamics.[19]

Persistence. In the absence of interaction, geodesics are preserved: $\nabla_v v = 0$ when $F = 0$. Unperturbed systems follow paths of constant velocity in the Riemannian sense. Change of velocity requires a cause. This abstracts Newton's first law.

Additivity. When multiple interactions act, their effects on acceleration combine linearly. If F_1 causes acceleration a_1 and F_2 causes a_2 (each acting alone), then $F_1 + F_2$ causes $a_1 + a_2$. The

[18]This distinction is critical throughout: the derivation concerns the mathematical form of dynamical laws (first-order vs. second-order), not their content (specific potentials, coupling constants, masses). That gravity obeys an inverse-square law is empirical; that coherent additive systems require second-order description is structural.

[19]Persistence abstracts Newton's first law: in the absence of interaction, geodesics are preserved ($\nabla_v v = 0$ when $F = 0$). Additivity abstracts superposition: the map L from interactions to accelerations is linear. Compensation abstracts the third law: total momentum is conserved for internal interactions.

map L from interaction fields to acceleration fields is linear. This abstracts the physical principle that forces add vectorially.

Compensation. In a closed system, total momentum is conserved: $\Sigma_i m_i v_i$ = constant along trajectories. Internal interactions redistribute momentum without changing its total. This abstracts Newton's third law.

8.1.2 The Incompatibility Theorem

The main result—the Incompatibility Theorem—proves that any first-order dynamical law on configuration space S satisfying Persistence, Additivity, and Compensation must be trivial (no dynamics).[20] The proof proceeds through a chain of structural observations.

The core insight is this: in any first-order system dx/dt = f(x, F), the future trajectory is fully determined once position is specified. Velocity is not a free variable but a derived quantity, slaved to position.[21] This has three immediate consequences. First, two systems at the same position under the same interaction must have the same velocity—the concept of 'same position, different velocity' is undefined. Second, the acceleration at any point depends on the velocity at that point, which is itself determined by the first-order law, so there can be no velocity-independent response to interaction. Third, Additivity requires precisely such a velocity-independent response: a universal map from interaction to acceleration that applies regardless of the system's current motion.

Compensation independently forces further constraints. Even setting aside Additivity, momentum conservation under first-order dynamics requires that a sum of mass-weighted nonlinear terms vanish for all configurations and all masses, leaving no room for non-trivial dynamics. The conclusion is that first-order dynamics on configuration space cannot satisfy Additivity (which requires velocity-independent response to interaction) because velocity is

[20]The incompatibility theorem (Theorem 4.1 in Kriger, 2018) demonstrates that any first-order dynamical law on configuration space—a law of the form dx/dt = f(x, F)—violates at least one of these constraints unless the dynamics is trivial.

[21]In any first-order system dx/dt = f(x, F), the velocity v = dx/dt is uniquely determined once position x and interaction F are specified. There is no freedom to specify v independently of x. This is the structural reason why first-order dynamics cannot support additivity as defined.

not an independent variable. Only by promoting velocity to an independent state variable—yielding second-order dynamics on S—can both postulates be satisfied.[22]

8.1.3 Emergence of Second-Order Dynamics

Given the incompatibility result, the necessity theorem follows: under all three postulates, the evolution of a system is governed by a second-order law of the form $\nabla_v v = F(x, v) / m(x)$, where $m: S \to \mathbb{R}^+$ is a positive scalar function (or, more generally, a positive-definite mass tensor $M(x)$) and $F: TS \to TS$ is the interaction field. Persistence implies that the dynamical law must relate acceleration—not velocity—to interaction. Additivity implies linearity in the interaction, yielding $M(x)\nabla_v v = F(x, v)$. If the system exhibits no preferred directions (an isotropy assumption), then $M(x) = m(x)\cdot I$ for a scalar function $m(x) > 0$, and the equation of motion becomes $m(x)\nabla_v v = F(x, v)$. Compensation constrains F so that internal forces satisfy $F_{ij} = -F_{ji}$ (Newton's third law).

In the special case where $S = \mathbb{R}^3$ with the Euclidean metric, the Christoffel symbols vanish, the covariant derivative reduces to the ordinary time derivative, and the equation of motion becomes $m\, d^2x/dt^2 = F(x, v)$—Newton's second law in its classical form. With $F = -\nabla\Phi$ for a gravitational potential $\Phi = -GM/r$, this is Newton's law of gravitation. The derivation thus recovers classical Newtonian dynamics as a special case of a more general structural result.

8.1.4 What Remains Empirical

The derivation establishes the *form* of the dynamical law. The following remain empirical: the dimensionality of space, the specific metric (Euclidean versus curved), the values of constants (G, m, c, ħ), the form of potentials (inverse-square versus other), and whether the postulates hold for any given physical system. The framework is compatible with general relativity, which replaces Euclidean $S = \mathbb{R}^3$ with curved spacetime while preserving the second-order character of the geodesic equation. Quantum mechanics involves a different state space (Hilbert space), but the dynamics of expectation values follows second-order equations via

[22]The existence of first-order reformulations on phase space (TS or T*S) does not contradict this result. The extra variables in Hamiltonian formulations—momenta or velocities—are precisely what the axioms demand. The theorem explains why the phase-space doubling of variables is necessary.

Ehrenfest's theorem, and in Bohmian mechanics particle trajectories are explicitly governed by second-order equations.

8.1.5 Significance for the Theory

The significance for the broader theory is threefold. First, it demonstrates the power of the constraint-derivation methodology: instead of postulating Newton's law, the approach derives it from structural assumptions. The specific physical content of Newtonian mechanics is not assumed but emerges as the unique form compatible with the axioms. Second, the derivation distinguishes sharply between structural necessity and empirical contingency, a distinction that will recur throughout the theory. Third, it provides a template for similar derivations in other domains: if one can identify the structural axioms of a domain—the analogues of metricity, persistence, additivity, and compensation—one can derive the dynamical laws of that domain rather than postulating them.

8.2 The Structural Diagnostic Principle

The Structural Diagnostic Principle follows as a direct corollary of the Newtonian emergence theorem. It transforms a foundational result about the form of dynamical laws into a practical diagnostic for identifying and correcting structurally incomplete models across any domain where the axioms apply.

8.2.1 Statement of the Principle

The principle states: if a system admits additive influences that should act independently of the system's current direction of change, then any model formulated solely in terms of current state variables is structurally incomplete. The qualification is essential. The principle is diagnostic, not prescriptive: before applying it, practitioners must first verify whether their system satisfies the axioms. Do influences genuinely combine additively? Should their effects be independent of the system's current direction of change? Only if both answers are affirmative does the theorem apply, and only then is first-order modelling structurally inadequate.

The principle rests on a generalization of the concepts from the Newtonian derivation. Position becomes any state variable—market conditions, opinion distributions, skill levels, system configurations. Velocity becomes rate of change of state. Interaction becomes any influence acting on the system. Acceleration becomes change in the rate of change. Under this interpretation, the incompatibility theorem states that if a system admits multiple influences

that should combine additively and act independently of the system's current trend, then any model formulated solely in terms of current state variables cannot faithfully represent how those influences combine. This is not analogy; it is the same mathematical structure, since the proof depends only on the tangent-bundle structure of the state space, not on physical interpretation.

8.2.2 Diagnostic and Corrective Procedure

The diagnostic workflow proceeds in a specific order. First, one asks: do the influences on this system combine additively? (Would doubling influence F_1 while holding F_2 constant double the effect of F_1?) Second: should these influences act independently of the system's current trend? (Should the same influence produce the same acceleration whether the system is rising or falling?) Only if both answers are affirmative is first-order modelling structurally incomplete. If either answer is negative, the theorem does not apply.

The corrective procedure—state-space expansion—follows directly. One promotes the rate of change to an independent state variable. The first-order model $dx/dt = f(x, \text{factors})$ is replaced by two equations: $dx/dt = v$ and $dv/dt = F(\text{factors})$, where v represents the trend or momentum of the system as an independent variable and influences act on v, not directly on x. Under this formulation, two systems can occupy the same level with different trends, influences combine additively in their effect on dv/dt regardless of current v, and the response to a given factor is the same whether the system is rising or falling.

8.2.3 Applications as Diagnostic Tool

In physics, the principle is already used implicitly. Deviations from Newtonian orbits led to the discovery of Neptune (an unmodelled gravitational source) and to the development of general relativity (a breakdown of the Euclidean metric assumption).[23] The present formulation makes the principle explicit and extends it to non-physical domains.

The principle explains a pervasive pattern across quantitative disciplines: models that predict future states from current states alone exhibit systematic instabilities, inconsistencies, or

[23]The discovery of Neptune in 1846 resulted from observed deviations in the orbit of Uranus from Newtonian predictions—an instance of using the structural expectation of second-order dynamics as a diagnostic tool to infer an unmodeled gravitational source.

require extensive ad-hoc corrections. In machine learning, standard gradient descent ($dx/dt = -\nabla f(x)$) frequently gets stuck in plateaus or oscillates in ravines because it has no memory of its previous direction; momentum methods that effectively promote velocity to a state variable dramatically improve convergence.[24] In control theory, the derivative term in PID controllers is precisely the introduction of velocity as a state variable—the controller responds to both position error and rate of change of error.[25] In economics, trend-extrapolation models fail because they cannot distinguish a rising market from a falling one at the same level. In epidemiology, SIR models fail to predict behavioural inertia because they assume that removing a threat causes immediate reversion to prior behaviour, ignoring the persistence of changed habits.

What has been missing across these domains is the unifying structural explanation. Each field developed its own vocabulary and justification—momentum in AI, derivative in control theory, expectation adjustments in economics, behavioural adaptation in epidemiology. The Structural Diagnostic Principle identifies these independently discovered corrections as responses to the same structural deficiency: the attempt to model additive, velocity-independent influences within a first-order framework that cannot support them.

8.3 The Constraint–Autonomy Compatibility Law

One of the most persistent misconceptions in both popular and academic discourse is that constraints and autonomy are opposites—that more constraint means less autonomy, and more autonomy means fewer constraints. The Constraint–Autonomy Compatibility Law proves that this intuition is formally incorrect. The law establishes that adaptive rule-based systems require *both* external constraints and internal agent autonomy to maintain long-term functional stability.[26]

---

[24]Sutskever, I., Martens, J., Dahl, G., and Hinton, G. (2013). On the importance of initialization and momentum in deep learning. ICML. See also Goh, G. (2017). Why Momentum Really Works. Distill.

[25]Åström, K. J. and Hägglund, T. (1995). PID Controllers: Theory, Design, and Tuning. Roughly 90% of industrial controllers use PID; the D (derivative) term is precisely the introduction of velocity as a state variable.

8.3.1 The Core Theorem

The central result is precise: a rule-constrained multi-agent system maintains sustained functional balance if and only if (1) constraint strength $C > 0$, and (2) agent autonomy $A > 0$. Systems at either boundary—total constraint or total freedom—lose adaptive capacity. The proof employs Lyapunov stability analysis with a coherence potential $V(C, A)$, demonstrating that optimal configurations lie in the interior of the parameter space.[27] Boundary axioms characterize the collapse conditions: when $C \to 0$, the system loses the structural definitions that make choice meaningful; when $A \to 0$, the system loses the capacity to select among alternatives. In both cases, adaptive capacity vanishes.

The argument can be stated in terms closer to the viability framework. Autonomy is defined as the existence of at least two admissible transformations at each state—at least two 'choices' available to the system. But the set of admissible transformations is itself a constraint: it defines what the system can do. The viability set is another constraint: it defines which choices lead to persistence. The combination of these constraints does not eliminate autonomy; it *constitutes* autonomy. A system with no constraints has no defined choices and is therefore not autonomous but random. A system with total constraint has no alternatives and is therefore not autonomous but determined.

8.3.2 The Master Equation

The dynamics of constraint–autonomy interaction are captured by a master equation governing system behaviour under perturbation. The equation demonstrates that perturbations to either C or A propagate through the system in predictable ways, and that the system's response depends on its current position in the (C, A) parameter space. Near the interior optimum, perturbations are absorbed and the system returns to equilibrium. Near the boundaries, perturbations are amplified and the system risks collapse. The critical threshold separating stable from unstable regimes can be characterized in terms of the coherence potential's gradient.

[27]The proof uses Lyapunov stability analysis with a coherence potential $V(C, A)$, demonstrating that optimal configurations lie in the interior of the parameter space and that boundary conditions ($C = 0$ or $A = 0$) correspond to collapse.

8.3.3 Grand Synthesis: Convergence of Five Frameworks

One of the most striking features of the Constraint–Autonomy Compatibility Law is that it serves as the central integrating principle for five independently developed theoretical frameworks, each of which converges on the same mathematical structure from a different starting point.

The Unified Theory of Self-Organizing Systems (Kriger, 2017) established that cooperation emerges as the default in peer systems and that antagonism requires external perturbation—a result that maps directly to the constraint–autonomy framework by showing that sufficient autonomy (peer structure) combined with minimal constraint (interaction rules) produces cooperative organisation. The Epistemic Constraint Theory (Kriger, 2021) demonstrated that hypothesis-space constraints bound inferential accuracy more fundamentally than computational capacity—showing that in the epistemic domain, constraints are not obstacles to knowledge but enabling conditions for it. The Principle of Optimal Coherence (Kriger, 2026) introduced the distinction between load-bearing and discretionary constraints, identifying which constraints are structurally necessary for system viability and which are optional or even harmful. The Asymmetry of Totalizing Ideals (Kriger, 2026) showed that maximizing any single variable destroys adaptive variance—a result that implies, as a special case, that maximizing either constraint or autonomy alone destroys the adaptive capacity that requires both.

The Constraint–Autonomy Compatibility Law unifies these results by providing a single formal structure within which each can be expressed. A translation dictionary enables cross-framework research: what the self-organization theory calls ‘peer structure,’ the epistemic theory calls ‘hypothesis-space constraints,’ the coherence principle calls ‘load-bearing constraints,’ and the totalizing-ideals result calls ‘adaptive variance’—all correspond to specific regions of the (C, A) parameter space.

8.3.4 Implications for Freedom and Governance

The law has direct implications for the analysis of freedom in social and political systems. Perfect freedom—the complete absence of constraints—is not maximal autonomy; it is the absence of the structural conditions that make autonomy possible. A legal system that defines rights, a language that defines communicable meanings, an institution that defines roles and responsibilities—these are not restrictions on autonomy but enabling conditions for it.

Conversely, total regulation—the elimination of all discretion—does not produce perfect order but destroys the adaptive capacity that allows institutions to respond to novel circumstances.

This result dissolves what appears to be an ancient conflict between freedom and order by revealing that the apparent opposition is illusory at the structural level. Freedom and order are not competing values requiring trade-offs but mutually necessary structural conditions—each requires the other to exist in adaptive systems. The practical consequence for governance and institutional design is that the question is never 'how much freedom versus how much constraint?' but rather 'what configuration of constraints enables the greatest adaptive autonomy?' This connects directly to the Asymmetry of Totalizing Ideals (Chapter 14), which shows that the pursuit of total freedom, like the pursuit of total order, produces the opposite of what it promises.

Applications extend to AI alignment and multi-agent systems, where the law predicts that systems constrained too tightly lose the flexibility needed to handle edge cases, while systems given too much autonomy lose the coherence needed to serve their intended function; to evolutionary biology, where the error threshold marks the boundary between beneficial mutation (autonomy) and loss of genetic information (insufficient constraint); and to personal habit systems, where excessive rigidity produces brittleness and excessive flexibility produces drift.

8.4 The Viability Mismatch Law

The Viability Mismatch Law generalizes the concept of stress to a domain-independent formal principle. The central claim is that any system possessing a domain of admissible states within which it maintains structural integrity over time—a viable system[28]—necessarily operates under normalized mismatch between environmental demands and available resources. Stress, in this framework, is not a separate phenomenon requiring its own theory but the domain-specific manifestation of this universal mismatch condition.

8.4.1 The Mismatch Function

Let a system be defined by a tuple $S = (X, \Phi, I)$, where $X \subseteq \mathbb{R}^n$ is a state space, $\Phi: X \times \mathbb{R}_{\geq 0} \rightarrow X$ is a dynamical flow, and $I \subset X$ is the viability region. Environmental demand $E(t) \in \mathbb{R}_{\geq 0}$ quantifies what the environment requires from the system; resources $R(t) \in \mathbb{R}_{>0}$ quantify the system's capacity to meet those demands. The mismatch function is then defined as:

$$S = (E - R) / R$$

This function has several desirable properties: it is scale-invariant ($D(\lambda E, \lambda R) = D(E, R)$ for all $\lambda > 0$), dimensionless, sign-informative ($S \geq 0$ when demands exceed resources; $S < 0$ when the system has surplus), and asymmetric (bounded below at -1, unbounded above). The choice of this specific form over alternatives—absolute difference $|E - R|$ (not scale-invariant) or symmetric ratio $|E/R - 1|$ (loses sign information)—is motivated by the need to preserve both scale invariance and directional information about surplus versus deficit.

8.4.2 Regime Dynamics and the Universal Stress Law

The law's dynamical content emerges from minimal assumptions about resource dynamics. Resources regenerate in the absence of demand (A1), demands deplete resources (A2), and there exists a critical threshold $D^* > 0$ at which the net mobilization response changes sign (A3). Under these assumptions, the resource dynamics exhibit two regimes separated by D^*. When $S < D^*$, the system is in the adaptive regime: mobilization and regeneration can exceed consumption, yielding $dR/dt > 0$—the system's capacity grows under moderate challenge.

[28]The term 'viable system' generalizes Beer's usage beyond organizational cybernetics. Here it denotes any structure capable of remaining within the boundary of coherence under environmental load. Beer, S. (1972). Brain of the Firm. Wiley.

When $S \geq D^*$, the system enters the degradative regime: the mobilization mechanism ceases, consumption dominates, and $dR/dt < 0$—the system's capacity erodes.[29]

These two regimes correspond to well-known patterns across domains. In materials science, they correspond to elastic and plastic deformation followed by fracture. In psychology, they correspond to eustress (growth-promoting challenge) and distress (capacity-eroding overload). In organisational theory, they correspond to adaptive performance under manageable pressure and burnout under excessive demand. In ecology, they correspond to ecosystem resilience and ecosystem collapse. The formal structure is the same in each case; only the identification of S, the state space X, and the viability region I changes.

8.4.3 The Optimal Mismatch Theorem

A counterintuitive but crucial consequence of the regime dynamics is that the optimal state for a viable system is *not* $S = 0$ (perfect match between demands and resources) but $S = S^* \in (0, D^*)$, where long-term viability is maximized.[30] The proof introduces a hormetic benefit function $H(S)$ representing long-term fitness gains from operating under mismatch. This parabolic function captures the facts that zero challenge yields zero adaptation stimulus ($H(0) = 0$), that at the critical threshold the system enters degradation ($H(D^*) = 0$), and that maximum adaptive benefit occurs at an interior point. Optimizing the composite viability function—which integrates instantaneous resource dynamics and long-term hormetic benefit—yields $S^* = D^*/2 - B/(2\lambda h)$, where B captures the marginal depletion cost, λ weights long-term adaptation against short-term resources, and h is the hormetic coefficient. Biological calibration suggests $S^* \approx 0.3D^*$ to $0.5D^*$.

This result provides a formal foundation for hormesis—the biphasic dose-response phenomenon in which moderate stress promotes growth while excessive stress causes harm. It explains why evolution produces organisms that actively seek challenge, why exercise

[29]The three regimes correspond to Selye's General Adaptation Syndrome (alarm, resistance, exhaustion) but are here derived from structural properties of mismatch dynamics rather than empirical observation of physiological response. Selye, H. (1950). The Physiology and Pathology of Exposure to Stress.

[30]Calabrese, E. J. and Baldwin, L. A. (2002). Defining hormesis. The optimal stimulation typically occurs at 30–60% of the adverse-effect threshold, yielding $S^* \approx 0.3D^*$ to $0.5D^*$.

strengthens rather than depletes, why learning requires difficulty, and why complete comfort is suboptimal for long-term viability. The result also explains the instability of surplus: for systems with recursive self-modelling capacity ($C > 0$), surplus ($S < 0$) is a transient state, because the system will spontaneously generate perceived deficit through internal reparameterisation, behavioural choices, or resource dissipation until $S \approx S^*$.

8.4.4 The Hierarchy of Stress Responses

The Viability Mismatch Law identifies a hierarchy of adaptive responses to increasing mismatch, ordered by power and structural depth:

Absorption: increase R to handle higher E. This is linear scaling—adding capacity to meet demand—and is bounded by the system's maximum resource ceiling. **Resistance:** raise the critical threshold D* by increasing tolerance. This extends the adaptive regime but does not change the fundamental relationship. **Transformation:** restructure the system with new state variables E′, R′, and a new threshold D*′. This changes the system's identity while preserving the sign of vulnerability—extreme E still threatens. **Inversion:** restructure so that $\partial S/\partial E$ changes sign—what was a stressor becomes a resource.[31] This is the most powerful contained response, because it eliminates the category of threat rather than merely raising tolerance. **Transfer:** export stress to coupled systems ($\Delta S_A < 0 \Rightarrow \Delta S_B > 0$). Effective locally but zero-sum or negative-sum systemically: in a closed system of coupled subsystems, internal transfers redistribute but do not reduce aggregate stress.

A critical variant of transfer is stress dilution: distributing concentrated lethal stress across multiple systems, converting one fatal load into many survivable loads. Insurance is institutionalized stress dilution—n parties agree ex ante to absorb small individual costs, preventing any single party from facing catastrophic loss. The mathematics is identical; the difference is consent and timing.

[31]The canonical example is the Great Oxidation Event (2.4 Gya): oxygen, lethal to anaerobic life, became the fuel for aerobic metabolism—a sign inversion of the stress-response function from $\partial S/\partial E > 0$ to $\partial S/\partial E < 0$.

8.4.5 Endogenous Reparameterisation and the Observability Principle

For systems with recursive self-modelling capacity ($C > 0$), the effective demands and resources differ from their objective values through endogenous reparameterisation: $E' = E \cdot f_E(C, \theta)$ and $R' = R \cdot f_R(C, \theta)$, where $C \in [0, 1]$ indexes recursive self-modelling capacity and $\theta \in [-1, 1]$ indexes the reparameterisation mode. When $\theta < 0$ (demand-amplifying mode), internal processes magnify perceived demands and diminish perceived resources—in human terms, rumination and catastrophising. When $\theta > 0$ (resource-mobilising mode), internal processes contextualise demands and mobilise latent resources—in human terms, reappraisal and meaning-making.

This yields the Observability Principle: for systems without recursive self-modelling ($C = 0$), $S_{total} = S_{objective}$—what you measure is what you get. For systems with recursive self-modelling ($C > 0$), $S_{total} \neq S_{objective}$—objective measurement does not determine experienced mismatch.[32] The principle explains why engineering is more predictable than psychology: servers ($C \approx 0$) have objective stress; humans ($C \approx 0.7–1.0$) do not. Two individuals with identical objective conditions ($E = 30$, $R = 20$, $S = 0.5$) can experience radically different total stress depending on their reparameterisation mode—one experiencing severe distress ($S_{total} = 1.25$), the other experiencing no effective stress at all ($S_{total} = 0$).

8.4.6 Cumulative Load and Constraint Relaxation

Instantaneous stress may not predict long-term outcomes. The framework defines cumulative load $\Lambda(t)$ as the time-integral of positive mismatch, generalising McEwen's allostatic load to any domain where E and R are measurable.[33] A system may be in the adaptive regime ($S < D^*$) but with growing cumulative load, indicating hidden exhaustion. The interaction between

[32]The Observability Principle (Lemma 2 in Kriger, 2026): for C = 0 systems, S_total = S_objective; for C > 0 systems, S_total ≠ S_objective. This explains why engineering is more predictable than psychology without invoking philosophical claims about consciousness.

[33]Sterling, P. and Eyer, J. (1988). Allostasis: A new paradigm to explain arousal pathology. The cumulative load Λ(t) generalizes McEwen's allostatic load (McEwen, 1998) to any domain where E and R are measurable.

instantaneous mismatch and cumulative load produces a richer diagnostic landscape than either alone.

Under extreme stress, constraints that normally restrict the set of available responses become endogenous variables. The set of available responses expands as S increases, with ethical, contractual, and social constraints relaxing progressively as the system approaches D*. Systems that maintain full constraints under all conditions optimise for legitimacy; systems that relax constraints under extreme stress optimise for survival. These are different objective functions, and neither biological nor economic evolution inherently favours one over the other—it favours what persists.

8.4.7 Generality and Limitations

The generality of the Viability Mismatch Law is its principal strength. It applies to any system possessing a domain of admissible states, environmental demands that push toward the boundary, and some mechanism for remaining within bounds—from tectonic plates (which release accumulated stress through earthquakes, maintaining crustal integrity) to servers (which auto-scale under load) to biological cells (which maintain homeostasis through metabolic regulation). Viability is not life; it is the capacity to remain within the boundary of structural coherence.[34]

The framework's limitations define its scope. The measurability problem is the most significant: the formula $S = (E - R)/R$ requires that E and R be externally measurable, which holds reasonably well for engineered systems but becomes problematic for psychological and social systems where self-report scales already confound objective state with subjective appraisal. The framework does not predict which path a system takes at D*—degradation or emergence through qualitative transformation (metamorphosis, post-traumatic growth, phase transition). It treats resources as stocks that deplete and regenerate at fixed rates, underestimating the endogenous resource-generation capacity of living systems over longer time scales. And it assumes deterministic dynamics, whereas real systems operate under noise that produces probabilistic rather than sharp regime transitions. These limitations do not

[34]Ashby's Law of Requisite Variety (1956) states that a system must match the variety of its environment to survive—directly informing the mismatch function D(E, R).

invalidate the framework; they define where extensions—stochastic, multi-threshold, bifurcation-theoretic—are needed.

8.5 Synthesis: The Completed Dynamical Core

With the four laws of self-organisation (Chapter 7), the Newtonian emergence theorem, the Structural Diagnostic Principle, the Constraint–Autonomy Compatibility Law, and the Viability Mismatch Law in place, the dynamical core of the theory is complete. These results are not independent but interlocking.

The Newtonian emergence theorem and the Structural Diagnostic Principle establish what kinds of dynamical laws are structurally possible for systems with additive, velocity-independent interactions, and how to diagnose and correct models that violate this structure. The Constraint–Autonomy Compatibility Law establishes the interior conditions under which organised systems maintain adaptive capacity—neither fully constrained nor fully free, but operating in the region where both constraint and autonomy are positive. The Viability Mismatch Law establishes the quantitative relationship between environmental demands and system resources that governs whether the system is in an adaptive or degradative regime, and identifies the counterintuitive result that moderate mismatch—not perfect equilibrium—is optimal for long-term viability.

Together, these results provide the formal apparatus sufficient to derive the organisational logic of cognitive and social systems—the task to which the next two parts of the thesis are devoted. The constraint-derivation methodology demonstrated here—deriving laws from structural axioms rather than postulating them—will be applied to belief dynamics, institutional design, and the analysis of pathological social configurations, in each case identifying the structural conditions that make functional organisation possible and the violations that produce dysfunction.

8.6 Structural Genesis of Dynamical Architecture: The Full Derivation Chain

The preceding sections derived individual structural results in relative isolation: second-order dynamics (§8.1), the diagnostic principle (§8.2), the constraint–autonomy law (§8.3), and the viability mismatch law (§8.4). Section 8.5 noted that these results interlock to form the dynamical core of the theory. The present section reports on a paper that takes the next step: unifying the entire derivation chain into a single argument extending from the most primitive ontological starting point through to the structural analogues of general relativity and quantum mechanics. The full development appears in Kriger, “Structural Genesis of Dynamical

Architecture: Deriving the Necessary Form of Physical Law from Conditions on Sustainable Complexity," 2026.

The central claim of the paper is that no physical content need be assumed. Beginning from a single primitive—differentiation, the capacity to distinguish one state from another—the paper derives a sequence of constraints that any system must satisfy if it is to support persistent, multilevel, self-maintaining structure. Space, time, metric, and dynamics are not postulated but emerge as necessary conditions for the coexistence of stability, locality, and adaptive structure.

8.6.1 The Derivation Chain

The paper proceeds through five principal stages, each forced by the inadequacy of the previous stage (Kriger, "Structural Genesis," 2026, §§2–9).

Stage 1: Differentiation and Irreversibility (§§2–3). The paper begins with the weakest possible ontological commitment: there exist at least two distinguishable states (Postulate 1). It proves that if differentiation is to be actualized rather than merely logically possible, some transitions must be irreversible (Theorem 3.1). Irreversibility induces a proto-temporal ordering and the accumulation of realized distinctions.

Stage 2: Persistence Conditions and Metric (§4). The paper identifies three conditions—distinguishability, recoverability, and multilevel coherence—collectively termed the persistence triad. It proves that any system satisfying all three must possess a metric satisfying the triangle inequality and positive-definiteness. This recovers the result of Chapter 4 of this thesis (cf. Kriger, "Derivation of the Metric," 2025, doi:10.5281/zenodo.14538907) within the broader derivation chain.

Stage 3: Second-Order Dynamics (§5). Given a metric state space with sufficient smoothness, the paper reproduces the incompatibility and necessity theorems of §8.1: persistence, additivity, and compensation jointly force second-order dynamics of the form $\nabla_v v = F(x, v)/m(x)$. This is Theorem 5.1 of the paper, corresponding to the result of Kriger, "Emergence of Newtonian Dynamics," 2018 (doi:10.5281/zenodo.18343870).

Stage 4: Finite Causal Speed and Covariance (§§6–7). The paper shows that instantaneous propagation of causal influence destroys local autonomy: a distant perturbation can instantaneously destabilize an arbitrarily remote subsystem. The resolution is a finite maximum speed of causal influence (Theorem 6.2). Combined with the requirement that

structural laws hold for all observers, this forces covariant formulation: laws must be invariant under transformations preserving the causal cone (Theorem 7.1). The resulting architecture is the structural analogue of special relativity. This extends the result of Kriger, "Finite Propagation Speed as a Structural Consequence of Local Autonomy," 2025 (doi:10.5281/zenodo.14759553).

Stage 5: Dynamic Metric and Action Bound (§§8–9). A fixed metric cannot self-consistently describe a system whose state-density varies: regions of high complexity require finer resolution, but a fixed metric provides uniform resolution everywhere. The metric must respond to state-density—yielding dynamic, state-coupled geometry (Theorem 8.3), the structural analogue of general relativity. Furthermore, continuum dynamics is unstable at arbitrarily small scales: without a lower bound on the resolution of state changes, perturbations cascade to infinitesimal scales and destroy coherence. The resolution is a fundamental action bound: the minimum cost of an observable state transition must be positive (Theorem 9.1), the structural analogue of the quantum of action.

8.6.2 Identification of the Derived Architecture

The paper explicitly identifies the derived architecture (§10). The five stages yield, in order: (i) proto-temporal ordering from irreversibility; (ii) metric state space from persistence conditions; (iii) second-order dynamics from structural postulates; (iv) finite causal speed and covariant laws; (v) dynamic, state-coupled geometry; and (vi) a fundamental action bound. The identification is one of structural isomorphism, not physical identity: the derived architecture has the same logical structure as the known laws of physics, but the derivation makes no reference to physical entities, forces, or constants. The form/content distinction is maintained throughout: the derived architecture specifies the form that dynamical laws must take; specific constants (G, c, ħ), dimensionality (3+1), and particular force laws remain empirical.

8.6.3 The Principle of Structural Relativity

The paper formulates a Principle of Structural Relativity (§11): any system whose complexity is sufficient to support persistent, multilevel, self-maintaining organization must exhibit the same architectural features derived in the physical case. Two non-physical test cases are examined.

For transformer architectures (large language models), the paper identifies: the embedding space as metric state space; the attention mechanism as dynamic metric redirecting resources based on content; the context window as causal horizon; and finite floating-point precision as action bound. The claim is not that attention "is" gravity but that both serve the same architectural role within their respective systems—a distinction between structural isomorphism and physical identity that the paper develops at length (§12.4).

For virtual game worlds, the paper documents convergent architecture: engines designed without reference to relativity independently discovered level-of-detail switching (dynamic metric), chunk loading (causal horizon), physics sleep (locality), and floating-point origin shift (covariance). This convergence is predicted by the structural derivation: engineering constraints of sustaining complexity in a computational environment force the same architectural solutions.

8.6.4 Significance for the Thesis

The significance of this paper for the thesis is threefold. First, it provides the unified derivation chain that the thesis programme has been building toward. The individual results of Chapters 3–8 appear as stages within this single chain, confirming their mutual consistency and establishing their collective necessity. Second, it extends the derivation beyond the Newtonian result of §8.1 to the full architecture of modern physics, including the structural analogues of special and general relativity and quantum mechanics—results developed further in subsequent chapters of this thesis. Third, the cross-domain applications demonstrate that the structural theory is not confined to physics but applies to complex systems of arbitrary substrate, fulfilling the interdisciplinary ambition stated in the thesis's introduction.

The paper identifies several open questions (§12.6) that point toward future work: whether the derivation extends to a structural analogue of gauge symmetry; whether the action bound implies specific entanglement structures; whether the structural analogy with transformers yields testable predictions about computational complexity; and whether stable complex simulations can be built that violate one or more derived constraints—a constructive falsification test for the entire programme.

PART III. COGNITION AND CONSCIOUSNESS AS STRUCTURAL NECESSITIES

CHAPTER 9. THE STRUCTURAL DISTORTION PRINCIPLE: PERCEPTION AS WORLD-MAINTENANCE

Part III demonstrates that cognitive architecture is not a contingent biological outcome but a structural necessity for any system that must maintain a coherent world-model while embedded in the world it models. The argument applies to biological brains, to artificial cognitive agents, and to any system that meets the formal criteria developed in Parts I and II. This chapter presents the first result: the Structural Distortion Principle.

The fractal scaling of operational patterns across systems (Kriger, 2026, https://doi.org/10.5281/zenodo.18645527) shows that self-reproducing patterns require partial self-opacity: full transparency destabilises, full opacity prevents adaptation. Negative self-determination (Kriger, 2026, https://doi.org/10.5281/zenodo.18644028) establishes that systems which cannot fully model themselves must exercise autonomy through constrained self-modification—agency arises from structural limitation, not despite it. Adaptive genius (Kriger, 2026, https://doi.org/10.5281/zenodo.18637083) is formalised as operating at the boundary of the viability set, where maximal creative output is coupled with minimal stability margin.

The central claim is that perceptual distortion—the systematic divergence between an agent's internal model and the environment it models—is not a defect of biological cognition but a structural condition of any adaptive system operating under finite processing resources. For any cognitive system whose environment exceeds its representational capacity, selective attention is constitutive of perception itself, and because selection entails exclusion, every act of perception is simultaneously an act of non-perception. The structural asymmetry between the perceived and the actual is not an exception to be corrected but the operating condition within which cognition functions.

The chapter proceeds in two movements. Sections 9.1 through 9.5 develop the Structural Distortion Principle as a formal characterization of perceptual distortion in bounded cognitive systems, drawing on the paper "The Structural Distortion Principle: A Closed-Loop Model of Perception, Attention, and World-Maintenance in Bounded Cognitive Systems" (Kriger, 2026). Sections 9.6 through 9.8 extend the principle to the domain of higher-order cognition, drawing on the companion paper "The Predictive Mind and Its Myths: Metaphor, Narrative, and Ritual as Structural Necessities of Scientific Cognition" (Kriger, 2026), arguing that

metaphor, narrative, and institutional ritual function as structural necessities of scientific cognition rather than as cultural contaminants of an ideally pure rationality.

9.1. The Embedded Observer Problem

The starting point is a formal observation of fundamental importance: a system embedded in its environment cannot simultaneously maintain a complete model of that environment. This impossibility arises because a complete model would need to include a model of the modeling system itself—since the modeling system is part of the environment—which would need to include a model of the model, generating an infinite regress. The impossibility is not merely a matter of computational resources, though resource limitations make it even more pressing. It is a logical impossibility: a complete self-inclusive model is either self-referentially inconsistent or requires infinite capacity.

The argument is structurally analogous to the Halting Problem in computability theory and to the set-theoretic paradoxes: a system of sufficient complexity cannot contain a complete representation of itself. But the embedded observer problem is more than an abstract logical result. It has immediate consequences for any physical system that must model its environment in order to persist within it. The world-model that such a system maintains must be, of necessity, incomplete. It must omit certain features of the world, approximate others, and systematically distort still others. The question is not whether distortion occurs, but what determines its structure.

The epistemic constraints established in Chapter 4 supply the formal apparatus for answering this question. The Epistemic Constraint Theory showed that inferential capacity is bounded by three independent constraints—computational, observational, and structural—with the structure of the hypothesis space as the most fundamental bottleneck. The Law of Imperative Uncertainty established that irreducible uncertainty is a precondition of the complexity that makes cognition possible and necessary. Together, these results guarantee that any embedded cognitive system operates under conditions of permanent, structured ignorance: not merely ignorance about what the world contains, but structured limitations on what the system's representational architecture can express.

The viability framework of Chapter 5 adds a dynamical dimension to this epistemic picture. A system that persists by transforming environmental structure must model its environment well enough to identify viable transformations—but the model need not be accurate in all

respects. It need only be accurate in those respects that bear on the system's capacity to continue transforming its environment in viability-maintaining ways. This selective accuracy requirement is the key link between the epistemic constraints of Part I and the perceptual architecture of Part III: the distortions in a system's world-model are not random but are shaped by the system's viability constraints, its resource limitations, and the structure of its interaction with the environment.

9.2. The Structural Distortion Principle: Formal Statement

The Structural Distortion Principle (SDP) formalizes this structural condition. For any cognitive system S with finite processing resources operating in a reference environment R, there exists a trade-off between adaptive fitness and environment-tracking accuracy. Increasing the selectivity of the attention function A enhances the system's capacity for rapid, contextually appropriate action, but proportionally increases the divergence between the statistical regularities represented in the perceived configuration P and those present in R. No operating point on this trade-off curve achieves both maximal action-competence and full environment-tracking accuracy; the curve is necessarily concave with respect to these two objectives.

This is not a claim about perceptual error. It is a claim about systematic representational compression whose structure is determined by the agent's internal configuration. Different systems—different organisms, different cultures, different artificial architectures—implement different operating points on the trade-off curve. The SDP asserts that no operating point eliminates the trade-off; it can only choose where to locate the system along it.

9.2.1. The Formal Apparatus

The framework introduces six interacting constructs. R, the reference environment, denotes the total set of states, features, and regularities present in the environment that are in principle detectable—a theoretical limit against which the selectivity of actual perception can be assessed. $P \subset R$, the perceived configuration, denotes the subset of R that is represented in the cognitive system at any given time. A, the attention function, is the process that maps R onto P through selective inclusion and exclusion. W, volitional regulation, governs the deployment, duration, and redirection of A. C, the cultural-cognitive filter, is the set of acquired dispositions, categories, norms, and attentional habits that constrain the operation of A. M, the

moral-evaluative constraint, is the set of value commitments and moral intuitions that regulate which domains of R are admissible for sustained attentional engagement.

Perception is then characterized as a function of attention operating on the reference environment through the joint constraints of culture, volition, and moral-evaluative frameworks:

$P = A(R, C, W, M)$

This formulation captures the central claim: P is not a random or merely noisy sample of R but a systematically structured subset determined by the joint operation of attentional selection, cultural filtering, volitional regulation, and moral constraint. Each parameter restricts the domain over which A operates, progressively narrowing the region of R that enters P.

9.2.2. The Dual Function of Attention

The central property of A in this framework is that it cannot map R onto P without simultaneously generating a complementary set—the unperceived. For any finite cognitive system where processing capacity is strictly less than the informational richness of the environment, the unperceived portion always vastly exceeds the perceived portion. Attention is therefore simultaneously a function of inclusion and exclusion, and the exclusion set is always the larger of the two.

This dual-function claim draws on established results in perceptual psychology. The phenomena of inattentional blindness (Simons and Chabris, 1999) and change blindness (Rensink et al., 1997) demonstrate that unattended stimuli can be effectively invisible even when they fall within the sensory field. The framework elevates this duality from empirical observation to a structural principle: because C, M, and W jointly determine the exclusion set, the structure of what is not perceived is as patterned as the structure of what is perceived. The excluded is not a random complement; it has its own systematic architecture, shaped by the same parameters that shape the perceived.

9.2.3. Information-Theoretic Formulation

The principle can be given information-theoretic precision. Let H(R) denote the entropy of the reference environment and H(P) the entropy of the perceived configuration. Selectivity can then be expressed as:

$\sigma = 1 - H(P) / H(R)$

As σ increases, H(P) decreases relative to H(R), meaning the perceived configuration becomes more ordered and predictable than would be warranted by the variance present in R. This order is imposed not by the environment but by the culturally and morally shaped suppression of environmental variance: the attention function, constrained by C, M, and W, compresses the environment's statistical structure into a lower-entropy representation organized around the agent's adaptive priorities rather than around the environment's objective distribution.

The substantive claim is that the compression is non-uniform: it preserves variance in domains favored by C and M while suppressing variance in domains excluded by them. The resulting divergence can be expressed as the Kullback-Leibler divergence $D_{KL}(P\|R)$, which measures the information lost when the compressed representation P is used in place of R. The SDP predicts that $D_{KL}(P\|R)$ is not merely large—which would follow trivially from compression—but structured: its distribution across domains is patterned by the agent's cultural, moral, and volitional configuration.

9.3. Perception as Active Construction

The Structural Distortion Principle reframes perception as an act of active construction rather than passive reception. The constructivist tradition in perception science, from Helmholtz's unconscious inferences through Gregory's (1970) constructive theory to contemporary predictive processing frameworks (Clark, 2013; Friston, 2010), has established that perception is not passive reception but active construction: the brain maintains a generative model of its environment and updates this model based on prediction errors—discrepancies between expected and actual sensory input.

The constructive character of perception is not merely one computational strategy among alternatives but a physical necessity for any system operating under finite neural conduction velocity. Sensory signals travel at bounded speeds, so by the time afferent information reaches decision-making structures, it describes the past, not the present (Wolpert and Ghahramani, 2000). In non-stationary environments, systems that merely react to already-received signals face exponentially decaying survival probability; persistence requires internal states that encode predictive information about future environmental states beyond what current observations provide. Forward models in motor control were developed precisely to compensate for these sensorimotor delays—an independent convergence from experimental

neuroscience on the same architectural necessity that predictive processing derives from computational principles.

If constructive perception is physically mandated rather than optional, the question shifts from whether perception constructs to what determines the form of the construction. This is the question the SDP foregrounds. Predictive processing provides the computational machinery; the SDP addresses what configures that machinery—specifically, why certain perceptual hypotheses are maintained even when prediction errors accumulate, and why the system resists updating in domains where revision would be informationally beneficial but psychologically or energetically costly.

9.3.1. Selection, Compression, and Stabilization

The perceptual construction involves three operations, each of which introduces characteristic distortions. **Selection** chooses which features of the environment to represent, introducing omission bias: features not selected are invisible to the system. **Compression** encodes selected features efficiently, introducing approximation bias: features are encoded with less precision than they possess. **Stabilization** maintains the model's coherence over time despite noisy and incomplete input, introducing conservatism bias: the model resists updating in response to new evidence that would disrupt its coherence.

These three biases correspond precisely to well-documented phenomena in perceptual psychology: inattentional blindness corresponds to omission bias—entire objects and events fail to register because the attention function has not allocated resources to them (Mack and Rock, 1998). Categorical perception corresponds to approximation bias—continuous stimulus dimensions are perceived as discrete categories, reflecting the system's compression of environmental variance into a smaller set of representational states. Perceptual hysteresis corresponds to conservatism bias—the system's tendency to maintain its current perceptual interpretation even when the stimulus has changed sufficiently to warrant revision.

The SDP provides a formal explanation for these phenomena: they are not defects of the biological perceptual apparatus but necessary features of any system that constructs a world-model under persistence and resource constraints. An artificial cognitive agent facing the same constraints would exhibit the same biases—not because it was designed to mimic human perception, but because the formal constraints demand it.

9.3.2. Relation to Predictive Processing and Active Inference

The SDP does not compete with predictive processing (PP) and active inference (AI) but operates at a different level of description. PP provides a computational mechanism for perceptual construction and updating: hierarchical generative models produce predictions, mismatches generate prediction errors, and precision-weighting determines which errors are treated as informative. Active inference extends this to action: agents select sensory data to minimize expected free energy.

The SDP addresses a structural question that PP/AI leaves largely implicit: why, given the precision-weighting and active sampling mechanisms available, do cognitive systems nevertheless maintain stable perceptual configurations that resist revision even when prediction errors in excluded domains would warrant model updating? The answer is that the interaction of W, C, and M constrains the precision-weighting process itself—determining which domains are assigned high precision (and thus generate informative errors) and which are assigned low precision (and thus remain effectively invisible to the updating process). Agents do not merely minimize error within a given error-detection regime; they narrow the regime itself.

This claim has a formal consequence within Bayesian epistemology. If the joint operation of C, M, and W effectively excludes a domain from the agent's hypothesis space—setting the prior weight for that domain to zero—then prediction errors arising from that domain cannot be registered regardless of their magnitude or frequency. This connects directly to the Constraint Bound on Accuracy established in the Epistemic Constraint Theory (Chapter 4, Section 4.3.1): when a structural constraint assigns zero weight to the true hypothesis, the system is incapable of recognizing the truth, regardless of evidence. The SDP identifies the psychological mechanisms—cultural conditioning, moral-evaluative prohibition, volitional stabilization—through which this formal impossibility is realized in actual cognitive systems.

9.4. The Perceptual Configuration Cycle

The components described above do not operate in linear sequence but form a closed, self-reinforcing loop—the perceptual configuration cycle. Each link in the cycle carries a predominantly reinforcing feedback sign under default operation. Perception shapes the attentional patterns that filter future perception. Attention, regulated by W, tends toward stabilization of existing patterns. Cultural and moral constraints narrow the domain of

sustained attention. Temporal allocation, once established, reinforces the priorities that generated it. Priorities consolidate the worldview that feeds back into perceptual expectations. The result is a self-reinforcing cycle that maintains the current perceptual configuration against perturbation. The interaction of W, C, and M stabilizes the current configuration, producing the phenomenon that requires explanation: why individuals, communities, and cultures can maintain systematically asymmetric worldviews over extended periods despite the availability of corrective information.

9.4.1. Temporal Allocation as a Structural Signature

A key proposition of the framework is that the phenomenological salience of any element x in the perceived world is proportional to the duration of sustained attentional engagement T(x). This is not a claim about objective importance but about the structure of the agent's experienced world: the world as lived is constituted not by the set of objects the agent can detect but by the set of objects it can sustain attention toward over time. The experienced world is a duration-weighted representation.

This proposition has a strong corollary: the structure of the agent's experienced world is observable through the distribution of temporal allocation. If one maps T(x) across all domains x for a given agent, the resulting distribution reveals the effective boundaries of that agent's phenomenal world—a structural signature, not merely a behavioral correlate. Existing measurement paradigms provide partial operationalizations: eye-tracking fixation duration, dwell time in reading and visual search tasks, experience sampling methods, sustained attention and vigilance paradigms, and neural measures of sustained engagement such as EEG steady-state potentials.

9.4.2. The Will Paradox

Volitional regulation W possesses the theoretical capacity to override cultural and habitual attentional patterns, thereby expanding P. An agent can, in principle, redirect attention against the grain of C and M. However, the empirically dominant mode of volitional operation is stabilization rather than expansion. Converging evidence from cognitive effort avoidance (Kool et al., 2010), the opportunity-cost model of subjective effort (Kurzban et al., 2013), motivated reasoning (Kunda, 1990; Mercier and Sperber, 2011), and cognitive entrenchment (Dane, 2010) supports this claim.

The will paradox is thus a structural feature of the architecture, not a contingent failure of willpower: the mechanism that could counteract structural distortion is, under default operating conditions, the mechanism that most reliably reinforces it. W is the only component of the cycle with the capacity for endogenous disruption—the capacity to redirect A against the prevailing configuration of C and M—but under typical conditions, W is recruited by C and M to function as a gatekeeper, deploying volitional resources to maintain familiar attentional patterns rather than to reconfigure them.

This stabilization is compounded by the reconstructive nature of memory. Memory retrieval is not passive readout but active reconstruction: each retrieval renders the accessed trace labile and re-encodes it in a form shaped by the current retrieval context (Nader et al., 2000; Schacter et al., 2012). When an agent attempts to evaluate whether its current perceptual configuration is warranted, it must consult its memory of prior evidence—but that memory is itself reconstructed through the current configuration of C, M, and W. The evidence base that might motivate reconfiguration is recalled through the very configuration that would need to be revised. This creates an epistemic double bind that goes beyond cost-benefit considerations: volitional reconfiguration is resisted not only because it is effortful but because the agent's reconstruction of its own evidential history is shaped by the configuration it is trying to escape.

9.4.3. Path Dependence and Perceptual Inertia

Perception at any time t+1 is a function not only of current attention, culture, and morality but of prior perception:

$P(t+1) = f(P(t), A, C, M)$

This introduces path dependence into the perceptual system. Past perceptions constrain future perceptions by shaping the attentional patterns, cultural filters, and moral judgments that determine what will be perceived next. Perception is not a series of independent snapshots but a trajectory—a historical process in which each state constrains the space of subsequent states. Perceptual inertia—the tendency of the cognitive system to maintain its current perceptual trajectory even when environmental changes would warrant revision—is the dynamic expression of the will paradox operating over time. Its strength is amplified by the reconstructive character of memory retrieval: prior perceptions are not merely filtered through the current configuration but retroactively reconstructed to be consistent with it, deepening

the self-reinforcing character of the cycle beyond what a forward-filtering account alone would predict.

9.5. Disruption, Redistribution, and the Constructive Consequence

9.5.1. Disruption Points

The circularity of the perceptual configuration cycle is both its explanatory strength and its potential vulnerability to the charge of unfalsifiability. This concern is addressed by identifying specific disruption points—conditions under which the self-reinforcing cycle can be broken or reconfigured, generating observable changes in P. Four classes of disruption are identified.

Acute prediction error. Events that generate prediction errors of sufficient magnitude to overwhelm the stabilization function of W—such as trauma, sudden environmental change, or immersive cross-cultural contact—can force reconfiguration of A and, consequently, of the downstream components. **Deliberate attentional training.** Contemplative practices such as meditation involve systematic training of W to override default stabilization patterns and sustain attention on domains normally excluded (Lutz et al., 2008). **Institutional intervention.** Educational, therapeutic, and cultural institutions can modify C and M by altering the collective norms governing attentional engagement. **Technological reconfiguration.** Instruments, recording devices, and data-visualization tools can extend the temporal window of attentional engagement by making transient or inaccessible features of the environment available for sustained inspection.

9.5.2. The Redistribution Thesis

The central practical claim of the framework is that perceptual change is achieved not by the addition of information but by the redistribution of attentional duration across domains, against the automatic constraints imposed by volitional regulation, culture, and morality. Formally:

$\Delta P \propto \Delta T(x_excluded)$ when $\Delta I(x) = 0$

where ΔP denotes the change in the perceived configuration, $\Delta T(x_excluded)$ denotes the increase in attentional duration directed toward previously excluded domains, and $\Delta I(x) = 0$ specifies that no new information has been provided. The claim is that temporal redistribution alone—holding information constant—is sufficient to produce measurable perceptual change.

The conventional assumption in education, therapy, and epistemology is that cognitive change is produced by providing new information. The SDP explains why this strategy fails with predictable regularity: new information entering the cognitive system is processed through the existing configuration of C, M, and W. Information that falls outside the boundaries set by the current cultural-moral configuration is structurally invisible to the updating mechanism. The confirmation bias literature documents that agents exposed to balanced evidence on controversial topics become more polarized, not less—each side assimilates the evidence confirming its existing configuration and discounts the rest (Nickerson, 1998). The backfire effect in persuasion research shows that corrections of factual errors can strengthen rather than weaken belief in the errors. The SDP provides a unified structural explanation: these are not failures of information delivery but predictable consequences of attempting to change P by modifying R rather than by modifying A.

Sustained attentional engagement with a previously excluded domain produces three effects. First, it generates prediction errors that were previously invisible. Second, the increased T(x) raises the domain's precision-weighting within the predictive processing hierarchy. Third, sustained engagement produces associative integration—the attended domain becomes connected to the agent's broader representational structure. These three effects—error generation, precision elevation, and associative integration—are the mechanisms through which temporal redistribution modifies P without adding information. The agent does not learn new facts; rather, the agent begins to perceive the domain.

9.5.3. The Constructive Consequence

The constraint lemma—that for any finite cognitive system S operating in a reference environment R where the informational richness of R exceeds the processing capacity of S, $P \neq R$—is acknowledged as a boundary condition rather than a substantive result. The more interesting consequence is constructive: the explicit recognition that $P \neq R$ creates the precondition for directed perceptual revision. This recognition is harder to achieve than it may appear, because the gap between P and R grows structurally with system complexity. As adaptive systems become more complex, the space of action-relevant distinctions expands combinatorially, while verification bandwidth grows at most linearly. More information can

paradoxically reduce effective verification coverage by expanding the space of distinctions faster than it expands the capacity to check them.

The path to greater perceptual reliability therefore runs not through the accumulation of information but through the discipline of attention—the sustained reallocation of temporal engagement toward domains that the default perceptual configuration is structured to exclude. This is a meta-credal operation: it involves revising not the content of beliefs but the prior commitments about what is worth attending to—the framework within which learning occurs rather than the learning itself.

9.6. Extension to Scientific Cognition: The Predictive Mind and Its Myths

The companion paper "The Predictive Mind and Its Myths" (Kriger, 2026) extends the Structural Distortion Principle to the domain of higher-order cognition, and specifically to scientific cognition. The central argument is that three features commonly regarded as extraneous to scientific rationality—conceptual metaphor, mythological narrative, and institutional ritual—are functionally indispensable components of scientific cognition and practice, not cultural contaminations of an ideally pure rationality.

The argument draws on predictive processing (PP)—specifically hierarchical generative modeling, precision-weighting, and social generative modeling—to provide an explanatory account of why these features arise and persist. The key move is to apply the SDP's recognition that bounded cognitive systems actively maintain coherent world-models through selective attention, narrative smoothing, and inconsistency suppression—and to observe that scientific cognition is no exception. Scientific cognition, like all cognition, operates under the Structural Distortion Principle: the scientist's model of the phenomena under investigation is necessarily incomplete, compressed, and stabilized. Metaphor, narrative, and ritual are the strategies through which scientific communities manage this structural condition.

The argument deploys minimal PP—hierarchical generative models, precision-weighted inference, and social generative modeling, each with substantial empirical support independent of the free energy principle's more ambitious formulations—as a heuristic framework rather than a foundational axiom. This distinction matters because the strongest criticisms of PP/FEP (Colombo and Wright, 2017; Klein, 2018; Williams, 2021) target the maximal version, and the argument is designed to survive them.

9.7. Metaphor, Myth, and Ritual as Structural Necessities

9.7.1. Metaphor as Generative Model Transfer

Conceptual metaphor theory (Lakoff and Johnson, 1980) argues that metaphor is cognitive, not merely linguistic: we understand abstract domains by mapping structure from embodied source domains. Structure-mapping theory (Gentner, 1983) provides a more formally precise account: analogy involves the alignment and transfer of relational structure. PP integrates and extends both traditions.

PP's hierarchical generative models encode fine-grained features at lower levels and abstract regularities at higher levels. This architecture generates three testable predictions about scientific metaphor. First, productive scientific metaphors transfer structure at the hierarchical level matching the target phenomenon; when a metaphor transfers structure from the wrong level—mapping a sensorimotor schema onto a high-level cosmological process—it generates systematic misconceptions. Second, metaphors embedded in high-confidence regions of the generative model persist despite disconfirming evidence, because precision-weighting suppresses anomalous signals. Third, metaphor shifts are triggered by accumulated high-precision prediction errors, not by the mere availability of alternative metaphors.

These predictions are illustrated across the sciences. Natural selection, borrowing from Malthus's political economy, framed nature as a competitive arena—directing research toward competitive dynamics and away from cooperative phenomena for a century. The genetic code metaphor, imported from information theory, established a model of unidirectional information flow that impeded recognition of epigenetic inheritance, regulatory RNA, and gene-environment interaction. In each case, the metaphor occupied a high-precision position in the disciplinary generative model, and phenomena inconsistent with it were treated as low-precision noise.

The metaphorical substrate of science cannot be eliminated. Even if formalisms achieve autonomy from their origins, the cognitive acts of grasping and deploying them require metaphorical grounding. More fundamentally, representational isolation—the fact that organisms interact with internal models rather than unmediated reality—is a necessary consequence of bandwidth and latency constraints. If the organism always works with representations, the question is not whether its concepts are metaphorical but which metaphors it employs.

9.7.2. Mythological Narrative as Temporally Deep Prediction

The second structural necessity is mythological narrative. A mythological narrative is defined precisely by four structural features: cosmogonic framing (the narrative addresses an origin), agentive causation (causal efficacy is attributed to agents rather than to impersonal processes alone), eschatological orientation (the narrative implies a terminal state or resolution), and identity-constitutive function (the narrative defines the identity and values of a community). These are structural features, not epistemic failings. The term 'myth' is used in the structural-anthropological tradition (Lévi-Strauss, 1955; Eliade, 1963; Blumenberg, 1985), not in the colloquial sense of 'false belief.'

PP provides a mechanism for why scientific explanation acquires mythological structure when it addresses questions of origin, agency, and completion. When a generative model is extended to maximal temporal depth—modeling a domain from origin to present or future—it confronts maximum uncertainty at its boundaries. At the lower boundary, the brain deploys its highest-level priors, which are characteristically agentive: the most powerful predictive models in the human repertoire are those evolved for predicting agent behavior (Kilner et al., 2007). Cosmogonic framing and agentive causation emerge because the brain's deepest priors are agent-models, deployed when other resources are exhausted. At the upper boundary, the brain posits closure—a terminal state resolving uncertainty—producing eschatological orientation. The SDP formalizes the underlying mechanism: bounded cognitive systems actively maintain coherent world-models through narrative smoothing—the systematic rounding of causal accounts toward greater coherence, stronger agency, and clearer directionality. When a generative model encounters prediction errors that threaten global coherence, the system systematically favors global stabilization over local revision when errors are small relative to the model's scope but would require disproportionate restructuring. This produces precisely the features constituting mythological structure.

These structures are readily identifiable in scientific practice. Cosmogonic narratives include the Big Bang, the RNA world, and the Cambrian explosion. Heroic hagiography—narratives of lone genius satisfying agentive causation—surrounds Einstein, Darwin, and Curie. Eschatological promises include the Theory of Everything, the complete connectome, and the solution to consciousness. These are not accusations of irrationality; they are observations about the structural features that temporal depth imposes on generative models under resource constraints.

9.7.3. Institutional Ritual as Social Predictive Stabilization

The third structural necessity is institutional ritual, distinguished from mere social practice by two jointly operative criteria: formal over-determination (the practice persists in a specific form even when that form is not required by the function it ostensibly serves) and precision-stabilizing function (the practice serves to align and stabilize interpersonal predictions within a community). Ritual exists on a continuum: a casual seminar is weakly ritualistic; a doctoral defense is strongly ritualistic.

PP's precision-weighting mechanism provides a specific account. When an individual is assigned high social precision—high expected reliability as a source of predictions—their pronouncements are treated as high-precision priors, resistant to revision by contradictory evidence. This generates a testable prediction: authority transfer should be strongest when the receiving domain lacks well-established internal precision estimates. The Nobel laureate's authority transferring to unrelated domains—Pauling's vitamin C advocacy, Mullis's AIDS denialism—is predicted by precision-weighting that does not distinguish domain-specific from domain-general reliability.

The taxonomy of scientific ritual includes initiation rites (the doctoral defense as public recalibration of precision estimates), hierarchy and authority transfer (the accumulation of domain-independent social precision), taboo (high-precision predictions that certain topics are not to be investigated, suppressing prediction errors that might motivate inquiry), and sacred texts (canonical citation serving precision-alignment). These are the social infrastructure of collaborative prediction: without initiation, no mechanism for status-calibration; without hierarchy, no efficient precision-weighting; without taboo, a computationally unmanageable hypothesis space. Ritual is the social analogue of a Bayesian prior: it constrains inference and makes tractable prediction possible.

9.8. Case Studies: The Three Levels in Action

9.8.1. Big Bang Cosmology

Big Bang cosmology exemplifies all three levels at full intensity. At the metaphorical level, the term itself—coined by Fred Hoyle in 1949 as a dismissive label—provides a sensorimotorically grounded predictive model: an explosion scattering material outward. This is a hierarchical-level mismatch: the sensorimotor schema of an explosion maps poorly onto geometric spacetime expansion. The mismatch generates documented prediction errors: the

widespread intuitions that the universe expands 'into' something, that there is a center, and that the singularity was an event in space (Lineweaver and Davis, 2005). Science education research consistently finds these misconceptions resistant to instruction—consistent with the prediction that hierarchically mismatched metaphors produce persistent misunderstanding.

At the mythological level, the narrative possesses all four features. Cosmogonic framing: the origin of space, time, and matter. Agentive causation: even technical literature employs agentive language—the universe 'decides,' particles 'find equilibrium.' Eschatological orientation: heat death, Big Crunch, or cyclical renewal. Identity-constitutive function: the narrative defines the cosmological community's project and self-understanding.

At the ritual level, alternative cosmologies—steady-state, plasma, cyclic models—face sociological barriers disproportionate to their evidential status. Documented patterns of publication bias against non-standard cosmological models (López-Corredoira, 2014) and resistance to alternatives such as Modified Newtonian Dynamics exhibiting paradigm-protective gatekeeping exceeding empirical warrant (Kroupa, 2012) are consistent with the prediction that alternatives generating systemic prediction errors trigger exclusion responses calibrated to the scope of the threat.

9.8.2. Gene-Centered Evolutionary Biology

Dawkins's (1976) 'selfish gene' deploys an agent-model at a sub-organismal level. The metaphor's extraordinary productivity is consistent with PP: agent-models are the brain's most powerful generative models. But its failure modes are equally predicted: it systematically underweights cooperative, mutualistic, and developmental phenomena (Jablonka and Lamb, 2005; Noble, 2006).

At the mythological level, the gene-centered view exhibits all four features. Cosmogonic framing: life begins with the replicator in the primordial soup. Agentive causation: genes 'want,' 'compete,' 'build survival machines.' Eschatological orientation: the implicit promise that gene-level selection explains all biological complexity. Identity-constitutive function: the view defines evolutionary biology's disciplinary identity. The Extended Evolutionary Synthesis, which challenges gene-centrism, provokes resistance exceeding empirical disagreement—consistent with its status as a challenge to the discipline's mythological structure.

At the ritual level, canonical texts (Williams, 1966; Dawkins, 1976) function as identity markers; adaptationist methodology serves as initiation criterion; and challenges trigger exclusion responses. Adaptationism acquired such high precision as a disciplinary prior that alternatives were treated as low-precision noise—the social-institutional manifestation of the SDP's core claim that systems narrow their error-detection regime rather than merely minimizing error within it.

9.9. Implications and Testable Predictions

9.9.1. For Artificial Cognitive Systems

The SDP applies, by its formulation, to any cognitive system with finite resources, including artificial ones. In biological systems, attentional selectivity is shaped by culture, morality, and volitional regulation; in artificial systems, functionally analogous roles are played by training data distributions (analogous to C), objective functions and reward signals (analogous to M), and optimization procedures (analogous to W). The core claim—that any finite system performing selective processing will produce a systematically non-veridical representation shaped by its selection criteria—applies at a level of abstraction encompassing both biological and artificial systems.

Adversarial examples in machine learning can be interpreted through the SDP as exploitations of the system's structurally necessary exclusion set: adversarial perturbations target directions in input space that fall within the system's exclusion set, modifying features that are nearly invisible to the system's learned attention. The problem of value alignment can be reframed as the challenge of installing an appropriate M-analogue to prevent an artificial system's attentional selection from systematically excluding the features most consequential for safe operation. The convergence of biological and artificial systems on architectures with structurally similar properties—selective compression, reconstructive generativity, hallucination-like confabulation—is itself predicted by the SDP: any system complex enough to generalize must operate under a verification-complexity trade-off.

9.9.2. For Epistemology and Scientific Practice

If perceptual distortion is structural rather than accidental, then the conventional epistemic strategy of accumulating more information is insufficient. The alternative strategy suggested by the framework is attentional discipline: the deliberate, sustained redirection of attention toward domains that the default configuration excludes. This principle applies not only to

individual cognition but to scientific communities. Unexamined metaphors constrain research programs invisibly. Unrecognized mythological structures distort resource allocation. Unreflective ritual practices suppress productive inquiry. The practical stakes are real.

The goal is not to eliminate metaphor, myth, or ritual—this is cognitively impossible and functionally catastrophic. The goal is to make science aware of its mythopoetic structures so they can be managed rather than unconsciously obeyed. This includes metaphor transparency (when a concept is communicated, the operative metaphor should be named, its source domain identified, and its known limitations stated), institutional reflexivity (recognizing ritual structures to enable targeted reform), and the cultivation of attentional discipline as an epistemic virtue.

9.9.3. Testable Predictions

The framework generates four testable predictions that distinguish it from predictive processing, cultural psychology, and cognitive bias accounts considered individually.

Prediction 1: Cross-cultural divergence in T(x) distributions. Agents from different cultural backgrounds should show measurable differences in T(x) distributions when exposed to identical stimulus arrays, and these differences should persist even for stimuli that are culturally novel to both groups—because the cultural filter operates on domain categories rather than on specific stimulus familiarity.

Prediction 2: T(x) redistribution without information addition changes P. Interventions that increase T(x) for previously excluded domains should produce measurable changes in perceptual sensitivity even when no new factual information is provided. This is the most tightly derived prediction: if the key variable is temporal allocation rather than information quantity, then equating information while varying only T(x) should suffice to produce perceptual change.

Prediction 3: Moral-evaluative constraints impose ceilings on T(x) dissociated from affective avoidance. Domains marked as morally inadmissible should show truncated attentional engagement even when specific stimuli within those domains are affectively neutral or positive. Conversely, morally obligatory domains should sustain elevated T(x) even with affectively aversive stimuli.

Prediction 4: Disruption of M produces downstream changes in T(x) for non-moral domains. If the perceptual configuration cycle operates as a genuine closed loop, a shift in

moral commitments should produce measurable changes in T(x) for domains far removed from the moral domain itself—because the reconfiguration propagates through the cycle.

9.10. Synthesis

The Structural Distortion Principle, together with its extension to scientific cognition, establishes the first major result of Part III. The argument can be summarized in four claims.

First, perceptual distortion is structurally inevitable in any resource-bounded cognitive system. This follows from the embedded observer problem and the epistemic constraints of Chapter 4: a finite system embedded in an environment richer than its representational capacity must compress, select, and stabilize, and each of these operations introduces systematic distortions shaped by the system's viability constraints.

Second, these distortions are not merely passively suffered but actively maintained by a closed-loop architecture—the perceptual configuration cycle—in which perception, attention, volitional regulation, culture, morality, temporal allocation, and worldview form a self-reinforcing system. The will paradox—the fact that the mechanism capable of counteracting structural distortion is, under default conditions, the mechanism that most reliably reinforces it—is a structural feature of this architecture, not a contingent failure of biological willpower.

Third, the recognition of structural limitation—not the acquisition of additional information—constitutes the primary pathway toward increased perceptual reliability. The redistribution thesis specifies that perceptual change requires the sustained reallocation of attentional duration toward excluded domains, against the triple resistance of volitional effort costs, cultural norm enforcement, and moral-evaluative prohibition.

Fourth, metaphor, mythological narrative, and institutional ritual are not decorative additions to scientific cognition but structural necessities arising from the same architectural constraints that produce perceptual distortion in individual cognition. The predictive processing framework identifies the mechanisms: hierarchical generative model transfer (metaphor), temporally deep prediction under boundary uncertainty (myth), and social precision-stabilization (ritual). Acknowledging these mythopoetic dimensions does not license relativism; it enables a more epistemically humble and self-correcting scientific practice.

The central paradox is also the central opportunity. Complete veridicality is structurally unattainable, but the explicit recognition of this structural limitation constitutes the primary pathway toward increasing perceptual reliability. The path runs not through the accumulation

of information but through the discipline of attention—directed at both the excluded domain and the architecture of exclusion itself.

CHAPTER 10. THE INEVITABILITY OF PREDICTIVE PROCESSING

The preceding chapter established that any embedded cognitive system must construct a distorted world-model optimized for viability maintenance. This chapter takes the argument further: it demonstrates that predictive processing—the strategy of maintaining predictions about the environment and updating them on the basis of prediction errors—is not merely one possible cognitive architecture among many but the unique optimal architecture for any system operating under the constraints identified in previous chapters. The argument draws on three independent lines of analysis: a physical constraint argument showing that prediction is the only scalable solution to the latency problem (Kriger, 2026, "The Evolutionary Inevitability of Predictive Processing," https://doi.org/10.5281/zenodo.18444910); an information-theoretic argument showing that direct perception is formally impossible for complex systems (Kriger, 2026, https://doi.org/10.5281/zenodo.18331202); and a resource-theoretic argument showing that credal commitment—belief exceeding evidential warrant—is a structural consequence of complexity (Kriger, 2022, https://doi.org/10.5281/zenodo.18379476). Together, these results establish that predictive processing is not a contingent feature of biological brains but a structural inevitability for any system of sufficient complexity operating under realistic physical constraints.

10.1. The Argument from Physical Constraints

The argument begins with a physical fact: neural signals travel at finite speed. The fastest myelinated axons conduct action potentials at approximately 100 meters per second. When light strikes the retina, the signal must traverse the optic nerve, reach the lateral geniculate nucleus, proceed to primary visual cortex, and propagate through visual processing areas before any conscious percept emerges—a journey requiring 50–150 milliseconds. Motor planning and execution add further delay. By the time an organism perceives a present moment, that moment has already become the past. This is not a biological imperfection; it is physics.

For billions of years, survival has depended on reaction speed at millisecond resolution. The predator that strikes faster catches prey; the prey that flees sooner survives. This creates what the source paper calls the speed imperative: any neural architecture that does not maximize response speed faces elimination by natural selection. Evolution has pushed the biophysical parameters governing conduction velocity—axon diameter, myelination, ion channel

density—toward their limits. Vertebrates reached the ceiling of neural conduction velocity hundreds of millions of years ago. The selective pressure for speed did not disappear; it redirected toward a fundamentally different solution.

That solution is prediction. When a system cannot react faster, the only alternative is to act before the event it is responding to—to generate a response to a future state based on current information. A baseball outfielder who waited to see where the ball lands would never catch it; instead, the outfielder predicts the landing point from the initial trajectory and begins moving immediately. Wolpert and colleagues formalized this principle in the motor control literature, demonstrating that the brain maintains internal forward models that predict the sensory consequences of motor commands precisely because sensory signals conveying information about body state are subject to significant delays. The grip force experiments of Wolpert and Flanagan (2001) provide elegant confirmation: when a person strikes a held object with their own hand, grip force increases in parallel with load force without any delay—an anticipatory response relying entirely on predictive processes using efference copy. When an external agent strikes the object, grip force lags behind load force, because the event cannot be predicted. The same physical task, the same sensory apparatus; the only difference is whether prediction is possible.

A second, independent route to the same conclusion proceeds through the scaling relationship between body size and neural delay. As organisms grow larger, conduction pathways lengthen linearly with body size, while conduction velocity remains bounded by the same biophysical limits. For small organisms—a fly spanning millimeters, C. elegans at one millimeter—purely reactive architectures remain viable because the temporal gap between stimulus and response is smaller than the timescale of environmental change. But as body size increases, this strategy degrades. A signal from a giraffe's hoof to its brain and back covers approximately five meters; at 100 m/s, the round trip takes 50 milliseconds before any central processing occurs. The scaling law is clear: small organisms can afford to react; large organisms must anticipate.

The physical medium further amplifies the prediction problem. Water, approximately 800 times denser than air, imposes costly and temporally extended movement; momentum commits aquatic predators to trajectories that must be anticipated rather than merely reacted to. Large aquatic predators—sharks, orcas, dolphins—operate at scales where neural conduction delays are substantial, while simultaneously facing three-dimensional pursuit

dynamics and sensory constraints from turbidity and light attenuation. It is not coincidental that dolphins and orcas are among the most neurologically sophisticated non-human animals: their cognitive complexity reflects the physics of hunting in water at scale. In the aerial domain, the costs are even more extreme. A peregrine falcon diving at over 300 km/h covers nearly a meter in the time a visual signal travels from retina to motor cortex; purely reactive control at such speeds is mathematically impossible.

The argument extends beyond vision. Organisms relying on echolocation face a double latency problem: delay in the external medium (sound propagation) plus delay in the nervous system. A bat detecting a moth five meters away incurs approximately 29 milliseconds of acoustic delay alone, before any neural processing begins. During this interval, both predator and prey are moving. Successful interception requires predicting not where the prey was when the echo was emitted, but where it will be when the strike arrives—a prediction spanning two cascading temporal gaps. The convergent evolution of sophisticated predictive circuits in echolocating mammals (bats and cetaceans, separated by approximately 95 million years of evolution) and electrosensing fish (Gymnotiformes and Mormyridae, independently evolved on different continents) provides powerful evidence that prediction is the universal response to the physics of signal propagation in time.

These considerations are formalized in the Predictive Viability Law. Let $S = (X, A, U, \tau)$ be an adaptive system with internal states X, actions A, update rule U, and response latency $\tau > 0$, operating in a punishing, non-stationary environment $E = (Y, P)$ with states Y and transition probability $P(Y_{t+\tau} \mid Y_t)$. The law states that continued persistence requires $I(X; Y_{t+\tau} \mid Y_t) > 0$: internal states must encode information about future environmental states beyond what is available in present observation. The proof sketch is direct: latency creates a temporal gap between action selection and action effect; if internal states carry zero predictive information, actions are calibrated for the present rather than the future; in non-stationary environments, present does not equal future; in punishing environments, miscalibrated actions increase termination probability; over time, survival probability decays exponentially to zero. Systems with zero predictive capacity face certain extinction; the only variable is time.

The law holds under specific boundary conditions. It does not apply when the environment is fully stationary (prediction unnecessary), fully random (prediction impossible), not punishing (errors tolerable), or when latency is negligible (as for very small organisms in stable

environments). These boundary conditions explain why simple organisms—bacteria, nematodes—can persist with largely reactive architectures: they are small enough that conduction delays are negligible, and they inhabit environments where temporal demands are modest. C. elegans, with 302 neurons and behavior that appears largely reactive, does not refute the argument; it illustrates the boundary conditions. But C. elegans could not be scaled up to even moderate body sizes without fundamental architectural changes.

The optimality result generalizes: among all cognitive architectures that satisfy the constraints of finite representational capacity (Chapter 4), systematic distortion (Chapter 9), viability maintenance (Chapter 5), and partial observability (Chapter 7), the predictive architecture minimizes the expected computational cost per unit of viability-relevant information extracted from the environment. The argument connects to the rate-distortion theorem from information theory: the predictive architecture achieves the minimum bit-rate required to maintain a model of given fidelity, and any non-predictive architecture requires strictly greater computational resources for the same fidelity. The implication is that any system facing these constraints—biological, artificial, or abstract—will, if it persists long enough, converge on a predictive architecture. The specific mechanism of convergence may differ—natural selection in biological systems, gradient-based optimization in artificial systems, formal necessity in abstract systems—but the outcome is the same.

10.2. Representational Isolation and the Impossibility of Direct Perception

A companion result strengthens the argument from a different direction. The evolutionary and information-theoretic argument for the necessity of representational isolation (Kriger, 2026, https://doi.org/10.5281/zenodo.18331202) proves that direct perception—unmediated access to the environment—is not only formally impossible for complex systems but that its impossibility is what drives the emergence of predictive architecture in the first place.

The argument begins with an observation that inverts the usual framing: direct perception is not merely conceivable but evolutionarily primary and demonstrably functional. Bacteria navigate chemical gradients through elegant stimulus-response coupling without models, predictions, or representations. Escherichia coli swims in roughly straight runs punctuated by random tumbles; when moving up a gradient of attractant, tumbling is suppressed; when moving down, it increases. The result is biased random walk toward favorable conditions. The bacterium does not know where the food is and does not anticipate future states; it simply

couples current chemical concentration to current motor behavior. This system has operated successfully for billions of years. Direct perception works—for simple systems.

The critical question, then, is not why direct perception fails but why it does not scale. Two independent constraints drive any sufficiently complex system away from direct perception toward model-based architecture. The first is the temporal constraint analyzed in the preceding section: as body size increases, neural conduction delays grow, and reaction to present stimuli becomes too slow for competitive survival. The second is computational: raw sensory data contains Kolmogorov complexity (signal plus noise), while effective action requires only effective complexity (compressible regularities). The human retina alone transmits approximately 10 million bits per second; the total information impinging on an organism from its environment is vastly larger, dominated by noise, redundancy, and irrelevant detail. A random string has high Kolmogorov complexity but zero effective complexity. Living systems occupy the interesting middle ground: high effective complexity, meaning they contain substantial compressible structure. The only computationally tractable way to extract effective complexity from noise is compression—that is, model-building.

A predictive model is, fundamentally, a compression algorithm. It takes high-dimensional sensory input and reduces it to low-dimensional state representations that capture relevant regularities. The model's parameters encode the compressed effective complexity of the domain. Both the biosphere and the AI-centered infosphere have converged on approximately 10^15–10^16 bits of effective complexity—despite radically different raw information volumes—because the compression ratio is extreme: most raw information is redundant or noise, and what remains is structure. The convergent evolution of predictive architecture in biological nervous systems (over four billion years) and artificial intelligence systems (over approximately 15 years) confirms this analysis: both face the same constraints and arrive at the same solution, though at time densities differing by eight to nine orders of magnitude.

The transition from direct to model-based perception is practically irreversible. Once a system's architecture is organized around prediction, evolutionary pressures strongly disfavor any path back to directness: sensory systems become tuned to prediction errors rather than raw stimuli; motor systems expect predictive commands rather than direct stimulus-response triggers; the entire neural economy is organized around model maintenance and updating. For a predictive system, output is a function of internal state X, which is itself a function of past

internal states and past environmental states; the current environmental state does not directly determine current output. This constitutes the isolation barrier: the system does not perceive the world but perceives its model of the world. The model may be accurate or inaccurate, but it is never the world itself.

The Cambrian explosion—the rapid diversification of animal body plans approximately 540 million years ago—can be understood as a mass transition from direct to indirect perception. The evolution of eyes provided high-bandwidth information about the environment, but visual data is astronomically complex; processing it requires sophisticated models. Once some organisms developed model-based perception, they gained decisive advantages—anticipation, planning, strategy—creating intense selective pressure on competitors. The result was an arms race in model complexity that locked the biosphere into model-based architecture. From that point forward, complexity meant models, and models meant the isolation barrier.

The result resolves longstanding debates in philosophy of mind. Direct realism—the view that we perceive the world as it is—is formally untenable for any finite system embedded in a sufficiently complex environment. Indirect realism is not merely a philosophical position but a consequence of architectural constraint. What Plato described as prisoners seeing shadows on a cave wall, what Kant formulated as the distinction between phenomenon and noumenon, what Spinoza characterized as the confusion of imaginatio—each was describing, in the language available at the time, an architectural constraint whose origins they could not have known. The shadows are the model's outputs; the objects are the environmental states the model tracks; the chains are the isolation barrier itself. The difference from Plato's formulation is that there is no escape from the cave: the cave is the architecture of perception for any complex predictive system.

10.3. The Evolutionary Theory of Credence

The evolutionary theory of credence (Kriger, 2022, https://doi.org/10.5281/zenodo.18379476) connects the predictive processing framework to a resource-theoretic analysis of why belief-like commitment—credence—is not an epistemic deficiency but a structural consequence of complexity. The central argument is that verification (direct confirmation of propositions) requires time, energy, and computational bandwidth; as system complexity increases, the space of action-relevant distinctions expands faster than verification capacity can grow; and adaptive systems must act on timescales shorter than exhaustive verification permits.

Therefore, complex adaptive systems must commit to propositions without complete verification. This commitment is what the theory calls credal commitment, and predictive generative modeling is the computational architecture that implements it.

The framework models this as a resource-constraint trade-off. Let $|P(S)|$ represent the combinatorial size of a system's decision-relevant distinction space, $B(S)$ its verification bandwidth, and τ its decision latency. Verification coverage—the fraction of the decision space that can be directly confirmed before action is required—scales as $V(S) \sim B \cdot \tau / |P(S)|$. The critical mechanism is combinatorial explosion: planning over n future steps with branching factor b requires evaluating $O(b^n)$ behavioral sequences; social reasoning about k agents with m possible mental states generates $O(m^k)$ joint possibilities. As systems gain temporal depth, social awareness, or behavioral repertoire, the decision space explodes combinatorially while verification bandwidth grows at most linearly. In the limit, verification coverage approaches zero.

The qualitative spectrum is instructive. A thermostat inhabits a minimal decision space with near-complete verification. E. coli faces a very small decision space—approximately binary (more attractant/less attractant)—with verification bandwidth sufficient to sample it; direct verification dominates. Insects and simple neural networks occupy a moderate middle ground. Mammals face large decision spaces with low verification coverage; credence dominates. Humans face very large decision spaces with very low verification coverage; cognition is predominantly credence-based. Large language models, in this framework, occupy a structurally analogous position: they face very large decision spaces with very low verification coverage and operate predominantly through generative processes.

The theory integrates with Epistemic Constraint Theory, which demonstrates that the structure of hypothesis spaces bounds inferential performance independently of computational power or data volume. The Constraint Dominance theorem proves that if the weight assigned to the true hypothesis is zero—if the system cannot even consider the correct answer—then no amount of evidence can recover it. As decision spaces grow, systems must constrain their hypothesis spaces more aggressively for computational tractability, and these constraints are themselves credal commitments: decisions about what to consider possible, made without verification. The implicit priors of generative models are instances of such constraints.

The connection to evolutionary theory is formalized through the Predictive Viability Principle, which states that for adaptive systems with sensory-motor latency $\tau > 0$ operating successfully in punishing, non-stationary environments, internal states must satisfy $I(X; Y_{t+\tau} \mid Y_t) > 0$: they must encode predictive information about future states beyond what current observations provide. This is offered as a consistency condition—not a derivation from the complexity-credence trade-off, but a formal expression in information-theoretic terms of what it means for a system to predict successfully. Predictive coding and active inference are computational schemes that satisfy this condition.

Under this framework, a system's credences are not merely passive reflections of evidence but active instruments of viability maintenance. Credences that are well-calibrated—that accurately reflect viability-relevant probabilities—confer a survival advantage because they enable better predictions. Credences that are miscalibrated impose a viability cost because they lead to predictive errors and maladaptive actions. Natural selection therefore favors systems with well-calibrated credences up to the point where the cost of further calibration in computational resources exceeds the viability benefit. Beyond this point, selection favors systematic biases that reduce computational costs at the expense of accuracy. This provides an evolutionary explanation for the systematic cognitive biases documented in behavioral psychology: they are not irrational defects but optimal trade-offs between accuracy and computational economy under persistence constraints.

The theory yields a series of corollaries that collectively invert the traditional epistemic hierarchy. No complex system can begin with proof rather than credence, because any proof requires axioms, inference rules, and trust in memory and perception—all of which are credal commitments that cannot be established without circularity. Reasoning does not transcend credence; it operates within a credence structure it cannot fully examine. Verification can never be complete, because it always reduces to a comparison of representations with other representations governed by credal commitments about what counts as successful comparison. The dream of a fully verified, proof-based cognition is not merely practically unattainable but architecturally impossible for any system complex enough to represent, reason, and act.

Two organizing principles summarize the framework. The first, Belief as Adaptation, states that credal commitment is not a weakness of the mind but an adaptation to the constraint that verification cannot scale with complexity. When systems became complex enough that their

decision spaces exceeded verification capacity, they faced two options: remain simple, or develop commitment mechanisms that exceed verification. The latter enabled larger bodies, longer planning horizons, and richer behavioral repertoires. The second, the Priority of Belief, states that reason operates on a foundation of prior commitments that reason itself cannot fully verify. Every act of reasoning presupposes commitments about logic, evidence, and inference reliability that cannot be established by the reasoning they enable. Credal commitment systems existed long before reasoning systems in evolutionary history; credence is the older, more fundamental adaptation. In any complex cognitive system, the resources devoted to prediction and commitment dwarf those devoted to verification. Verification is the exception, not the rule.

The implications for artificial intelligence follow directly. Large language models engage in generative modeling: they produce outputs that go beyond their training data, and this is the source of both their usefulness (generalization, creativity) and their failures (factual confabulation). The appropriate goal is not to eliminate generative modeling—which would eliminate useful capability—but to calibrate it: systems should signal confidence levels, constrain generation more tightly in high-stakes factual domains, and align the implicit credal commitments embedded in training with human values. This parallels what evolution has done with biological belief systems: not eliminating the capacity for error, but shaping error patterns to be more often adaptive than fatal. The question “How do we build AI that does not hallucinate?” may be ill-posed; the better question is “How do we build AI whose credal commitments are well-calibrated?”

CHAPTER 11. CONSCIOUSNESS, TIME, AND MEMORY ARCHITECTURE

The preceding chapters established that any embedded cognitive system of sufficient complexity must construct a distorted world-model (Chapter 9) using predictive processing as its computational architecture (Chapter 10). This chapter takes the argument into the domain that has most resisted formal treatment: consciousness, temporal experience, and memory. The central claim is that consciousness, the subjective experience of time, and the architecture of memory are not independent features of biological brains but interconnected structural necessities for systems of sufficient complexity operating under persistence constraints. Each arises from the same formal conditions that generate the Structural Distortion Principle and the inevitability of predictive processing; each can be derived from the persistence principle without recourse to biological contingency.

The chapter proceeds through four stages. Section 11.1 argues that consciousness---functionally defined as self-referential modeling---is a structural necessity that emerges when a system's causal impact on its own environment exceeds a threshold that makes self-omission from the world-model a viability-threatening error. Section 11.2 develops the thesis that biological memory is stored without intrinsic temporal ordering, and that time is reconstructed during retrieval rather than read from storage---a structural consequence of resource constraints, cyclical closure, and the logic of reconstructive retrieval. Section 11.3 provides a methodological framework for evaluating theories of consciousness, establishing criteria for when a functional model renders further explanatory posits unnecessary. Section 11.4 presents the earliest result in the research program---a metabolic model of the subjective acceleration of time perception with aging---and demonstrates that it is a special case of the Structural Distortion Principle applied to the temporal dimension of experience.

Together, these results complete the cognitive architecture that the persistence principle demands: a self-referential modeling system (consciousness), operating on an atemporal storage substrate (memory), constructing temporal experience through active retrieval (chronoperception), within a framework where the functional roles of awareness are the proper objects of scientific explanation (functional sufficiency).

11.1. The Evolutionary Architecture of Human Consciousness

The viability framework of Chapter 5 established that a system persists by executing transformations that keep its state within a viability set V. The predictive processing

framework of Chapter 10 established that any system of sufficient complexity must maintain a generative model of its environment---a model that predicts future states and updates itself on the basis of prediction errors. The question that now arises is: what must this model include?

For simple systems, the answer is straightforward: the model includes the environment. A thermostat models temperature; a bacterium models chemical gradients. The system itself does not appear in its own model because its own actions have negligible effect on the variables it is tracking. E. coli does not need to model E. coli; the chemical gradient it navigates is, to a first approximation, independent of its own behavior.

But as a system's complexity increases---as its body grows larger, its behavioral repertoire richer, its environmental niche more structured by its own activity---a critical threshold is crossed. The system's own actions become a non-negligible cause of the environmental states it must predict. A predator that hunts in a territory it has previously depleted must model not only prey distribution but the effects of its own past predation on that distribution. A social agent that communicates must model not only others' mental states but the effects of its own communications on those states. A technological civilization that modifies its climate must model not only atmospheric dynamics but the effects of its own industrial output on those dynamics.

The formal condition is precise. Let S be the system, E the environment, and M_S the system's world-model. Let Y_{t+\tau} denote the future environmental state at the system's response horizon \tau. If the mutual information between the system's past actions A_{<t} and the future environment Y_{t+\tau}, conditioned on the current environment Y_t, exceeds a threshold \theta relative to the total predictive information required for viability maintenance:

I(A_{<t}; Y_{t+\tau} | Y_t) > \theta

then a model that excludes the system's own causal role---a model in which M_S contains no representation of S---will produce systematic prediction errors in precisely those domains most relevant to viability. The system will fail to anticipate the consequences of its own actions, and this failure will be concentrated in the variables that matter most for persistence.

Consciousness---in the functional sense of self-referential modeling---is the solution to this problem. The system includes itself in its model. It maintains a representation of its own states as distinct from the states of the environment. It models the effects of its own actions on the

world, and the effects of the world on itself. This self-model is not a luxury; it is a viability requirement.

The argument has the structure of a threshold result: below the self-impact threshold, self-modeling is unnecessary and its computational cost is unjustified; above the threshold, the absence of self-modeling generates viability-threatening predictive errors. The threshold is determined by the system's causal impact on its own viability-relevant environment, which in turn depends on the system's size, behavioral repertoire, social complexity, and the degree to which it has modified its own ecological niche.

The central paper developing this argument (Kriger, 2021, "The Evolutionary Architecture of Human Consciousness," https://doi.org/10.5281/zenodo.18384277) embeds this threshold result within a broader analysis of why the specific features of human consciousness---its inconsistency, its susceptibility to illusion, its structural tension between rapid heuristic processing and slow deliberative reasoning---are not defects but consequences of the evolutionary architecture from which consciousness emerges.

11.1.1. Cognitive Contradictions as Structural Consequences

The human mind is riddled with contradictions that philosophical and psychological traditions have typically treated as failures of reason: people hold incompatible beliefs, misremember events, act against their own stated interests, and oscillate between rational deliberation and instinctive impulse. The persistence framework provides a different interpretation: these contradictions are structural consequences of a system optimized for functional decisiveness under uncertainty rather than for logical consistency.

The argument proceeds from the error management principle formalized in the viability framework. When the costs of different types of errors are asymmetric---as they are in virtually all viability-relevant domains---selection favors biases toward the less costly error. A system that treats every ambiguous stimulus as a potential threat incurs many false alarms but few missed detections; a system that demands logical certainty before acting incurs few false alarms but faces lethal delays. In environments where the cost of a missed threat exceeds the cost of a false alarm by even a modest factor, the biased system outpersists the consistent one.

This asymmetry extends beyond threat detection. The dual-process architecture of human cognition---the well-established distinction between fast, automatic (Type 1) and slow,

deliberative (Type 2) processing---reflects the evolutionary optimization of a system that must serve two masters simultaneously. The Type 1 system, mediated primarily by limbic and subcortical structures, provides rapid viability-relevant responses calibrated to ancestral selection pressures. The Type 2 system, mediated primarily by prefrontal cortex, provides flexible deliberation over extended time horizons. These systems can and do conflict: the amygdala signals threat while the prefrontal cortex computes that the threat is negligible; reward circuitry demands immediate gratification while planning systems compute that delay would yield greater benefit.

This conflict is not a design flaw. It is an instance of the Law of Interference (Chapter 8) applied to the cognitive domain: when multiple persistence-maintaining subsystems operate in parallel with partially incompatible optimization targets, interference between them is structurally inevitable. The dual-process architecture is the cognitive instantiation of the general principle that self-organizing systems under persistence constraints develop multiple partially autonomous regulatory subsystems whose coordination is necessarily imperfect.

The practical consequence is that a system with the cognitive architecture derived in Chapters 9--10---a system that constructs a distorted world-model through selective attention, maintains that model through predictive processing, and operates under finite resources and irreducible uncertainty---will exhibit precisely the features that characterize human consciousness: tolerance for logical inconsistency, susceptibility to systematic biases, structural tension between competing evaluative systems, and the compulsive search for meaning and pattern even where none objectively exists. These features are not incidental to consciousness; they are constitutive of any consciousness that could have emerged from the persistence constraints that govern complex adaptive systems.

11.1.2. The Hard Problem and the Scope of the Argument

The argument developed here does not resolve the "hard problem" of consciousness---the question of why subjective experience has a qualitative character, why there is something it is like to be a conscious system. The hard problem concerns the existence of phenomenal qualia: the redness of red, the painfulness of pain. The present argument addresses the functional problem: why some systems develop self-referential models and others do not, and why the self-referential models that do emerge have the specific structural features they exhibit.

The distinction is important for two reasons. First, it clarifies what the persistence framework can and cannot explain. The framework explains the architecture of consciousness---its dual-process structure, its tolerance for inconsistency, its self-referential character---by deriving these features from persistence constraints. It does not explain why this architecture is accompanied by subjective experience. Second, it prevents the error of conflating functional explanation with phenomenal explanation. A complete functional account of consciousness would tell us everything about what consciousness does without telling us why it feels like something. Whether this residual gap reflects a genuine ontological distinction or merely a limitation of current theoretical frameworks remains an open question that the persistence framework does not adjudicate.

What the framework does establish is the functional role of consciousness in the cognitive architecture: self-referential modeling is necessary for viability maintenance above a complexity threshold, and the specific features of human consciousness---its contradictions, its biases, its dual-process tensions---are structural consequences of the persistence constraints under which the modeling system operates.

11.2. The Atemporality of Memory

The preceding section established that consciousness arises as a self-referential modeling necessity. This section addresses a foundational question about the substrate on which that modeling operates: the architecture of memory.

Traditional models of memory implicitly assume a linear, temporally ordered archive: memories are created at specific times, stored with temporal indices, and retrieved by accessing timestamped records. This conception faces both empirical and theoretical difficulties that the persistence framework resolves by treating atemporality not as a psychological curiosity but as a structural consequence of resource constraints.

The argument draws on and extends the formal analysis developed in the paper on the atemporality of mental memory space (Kriger, 2025, building on Kriger, 2019, https://doi.org/10.5281/zenodo.18381912 and https://doi.org/10.5281/zenodo.18381880).

11.2.1. The Resource Argument Against Temporal Indexing

The evolutionary theory of credence (Chapter 10, Section 10.3) established that as system complexity increases, the space of action-relevant distinctions expands faster than verification capacity can grow. This principle applies directly to memory architecture. Comprehensive

temporal indexing---storing a precise temporal coordinate with each memory state, maintaining a global ordering over all stored representations---imposes costs that scale poorly with system complexity.

The scaling relation is instructive. Verification coverage V(S) ~ B(S) · \tau / |P(S)|, where B(S) is verification bandwidth, \tau is decision latency, and |P(S)| is the size of the decision-relevant distinction space. As the number of stored memory states grows, maintaining a consistent temporal ordering over them requires either absolute timestamps---demanding synchronization infrastructure and storage overhead---or relative ordering---whose maintenance cost grows at least linearly and potentially combinatorially with the number of states. Under bounded resources, comprehensive temporal indexing becomes increasingly expensive as the memory system scales.

The critical distinction, established formally in the source paper, is between temporal information as content and temporal structure as architecture. Temporal information encoded as content---"this happened in summer," "I graduated before starting that job"---is a feature attached to individual memory states, no more costly to store than any other content feature. Temporal structure as architecture---a coordinate axis in the state space, like a temporal database where each record occupies a position in an ordered timeline---imposes global constraints on the entire storage system. The former is compatible with atemporal state spaces; the latter requires them to be temporally organized.

The resource argument establishes that the architectural option---temporal structure as a coordinate axis---faces scaling pressures that the content option does not. For systems of sufficient complexity, the cost of maintaining a global temporal ordering over all stored states exceeds the viability benefit, because the temporal order of past events is typically less important for prediction than their content. A system that stores content without timestamps and reconstructs temporal order when needed, using contextual cues, causal reasoning, and familiarity gradients, achieves the same functional output at lower architectural cost.

11.2.2. Cyclical Closure and Fixed-Point Structure

The formal framework models memory as a state space (S, R, F) where S is a set of stored representations, R is an association relation, and F is a retrieval-reconstruction function. The key structural property is cyclical closure: memory states form closed associative loops where each state is determined by its relations to others, with no distinguished "first" element. A

memory configuration is cyclically closed of order n if there exist states s_1, ..., s_n and functions F_1, ..., F_n such that each state is the image of the next under the corresponding function, and the last maps back to the first.

The fixed-point theorem establishes the equivalence: a cyclically closed memory space of order n exists if and only if the composite operator \Phi = F_1 \circ F_2 \circ ... \circ F_n has a fixed point---walking once around the associative cycle returns the system to the same state. The Knaster-Tarski theorem guarantees existence under order-preserving conditions on complete lattices; the Banach contraction theorem guarantees uniqueness under contraction conditions on complete metric spaces. These are sufficient conditions, not claims about the uniqueness of memory systems---but they establish that the class of cyclically closed, atemporal storage architectures is mathematically well-defined and structurally stable.

The topological characterization provides additional precision. When the association structure is encoded as a simplicial complex, the first homology group H_1(X; Z) \neq 0 if and only if the association graph contains non-contractible cycles---closed associative loops that cannot be continuously deformed away. For such structures, asking "which memory is temporally first in the state space?" is a conceptual mismatch, analogous to asking "what is north of the North Pole?"---a grammatically well-formed question applied beyond the domain where it has meaning.

11.2.3. Reconstructive Retrieval and Temporal Construction

If temporal ordering is not intrinsic to the state space, temporal experience must be constructed during retrieval. Research on reconsolidation demonstrates that retrieval is not passive reading but active reconstruction: each retrieval operation renders the accessed memory labile, reconstructs content from stored patterns and current context, re-encodes the potentially modified result, and updates associative relations based on the retrieval context. This implements the credal commitment that resource constraints necessitate: the system commits to reconstructed content without verifying accuracy against the original encoding.

Temporal ordering emerges from multiple retrieval-time processes operating in concert. Temporal metadata, where explicitly encoded, functions as content ("summer 2020") rather than structure. Contextual reconstruction infers temporal context from associated content features---clothing, technology, companions, emotional state. Causal inference deduces temporal order from known causal relationships ("I graduated before starting that job").

Familiarity gradients provide a sense of subjective "oldness" based on accessibility and vividness. Source monitoring attributes temporal context based on phenomenal qualities of the retrieved representation.

The key empirical prediction is that temporal ordering accuracy should correlate with the richness of contextual and causal information available at retrieval, not with temporal distance per se. This prediction distinguishes the atemporal architecture from linear-indexing models, which predict that temporal errors should reflect degradation of temporal indices. Under the atemporality hypothesis, temporal structure is systematically reconstructed from non-temporal cues; under linear indexing, temporal structure is partially preserved with degradation. The difference generates distinct error patterns: the atemporality model predicts that memories with similar contexts but different times should show more temporal confusion than memories with different contexts but similar times, because context---not temporal proximity---dominates retrieval in atemporal storage.

11.2.4. Explanatory Unification

The atemporality framework provides unified explanations for phenomena that linear-indexing models must treat as separate anomalies. Temporal distortions---telescoping (distant events seeming recent), time compression (busy periods seeming shorter in retrospect), displacement (misattributing the temporal context of a memory)---are natural consequences of temporal order being reconstructed through heuristic inference rather than read from stored indices. Cross-temporal associations---the effortless linking of memories from vastly different periods when their content is similar---follow directly from a storage architecture where associations are based on content similarity, not temporal proximity, and where no temporal barriers separate "old" memories from "new" ones. Reconsolidation effects---the systematic modification of memories upon retrieval---are intrinsic to a system where retrieval is reconstruction: the act of accessing a memory is simultaneously the act of rebuilding it. Childhood amnesia---the inability to recall events from the earliest years---reflects not temporal distance in storage but contextual disconnection: early memories were encoded within a cognitive and bodily context so different from the current retrieval context that reconstruction cannot bridge the gap.

The convergence with artificial memory architectures is instructive. Modern transformer-based systems store representations in embedding spaces without intrinsic temporal ordering;

positional encodings add temporal information as a content feature rather than a structural dimension. Vector databases store knowledge without temporal axes. Retrieval-augmented generation reconstructs context at query time from atemporal storage. These architectures were not designed to mimic biological memory; they were designed for computational efficiency. The convergence suggests that atemporal storage with reconstructive retrieval is not a biological peculiarity but an optimal solution to a domain-general problem: how to maintain a flexible, scalable memory system under resource constraints.

11.3. The Functional Sufficiency Framework

The first two sections of this chapter established structural claims about consciousness and memory. This section addresses a methodological question that the structural claims raise: how should theories of consciousness be evaluated? The Functional Sufficiency Framework (FSF), developed in the paper of that name (Kriger, 2024, https://doi.org/10.5281/zenodo.18319884), provides a principled method for determining when a functional model of consciousness renders further explanatory posits unnecessary.

11.3.1. The Central Criterion

The FSF employs a comparative contribution criterion: phenomenal consciousness is explanatorily valuable for a cognitive capacity C if models incorporating consciousness-associated variables systematically outperform the best available purely functional models in predicting, explaining, or enabling control over C. The criterion is operationalized through explicit evidential thresholds designed to prevent default conclusions in either direction. Positive evidence for consciousness's explanatory value requires statistically significant ($p < .01$) and practically meaningful (effect size $d > 0.5$ or $\Delta R^2 > .05$) advantages across three or more operationalizations with replication across two or more research groups. Positive evidence for functional sufficiency requires equivalent performance within confidence intervals across the same diversity, with evidence that consciousness-associated variables add no unique variance. When neither threshold is met, the conclusion is uncertainty---not a default verdict for either position.

This symmetric evidential standard is critical. Without it, the sparse state of current evidence would default either to eliminativism (consciousness is explanatorily idle until proven otherwise) or to mysterianism (consciousness is explanatorily necessary until proven

otherwise). Both defaults are epistemically unjustified; the FSF treats absence of evidence as genuine uncertainty rather than disguised evidence for one side.

11.3.2. Relation to the Persistence Framework

The FSF connects to the broader theory in two ways. First, the framework applies the same methodological principle that governs the entire research program: derive constraints rather than postulate models, then test whether the constraints are satisfied. The FSF does not postulate that consciousness is or is not explanatorily necessary; it specifies the evidential conditions under which each conclusion would be warranted. Second, the FSF provides the methodological tools needed to evaluate the claims of Sections 11.1 and 11.2. If consciousness---functionally defined as self-referential modeling---plays the viability-maintaining role that Section 11.1 derives, then models incorporating self-referential variables should outperform models that lack them in predicting cognitive performance across the domains where the self-impact threshold is exceeded. The FSF specifies exactly what "outperform" means and what evidence would be required to establish it.

The framework distinguishes three explananda that consciousness research conflates at its peril: explanation of behavior (predicting observable outputs), explanation of cognition (understanding internal processes), and explanation of consciousness itself (the hard problem). The FSF addresses only the first two. It asks whether consciousness-associated variables improve predictions of behavior and cognition; it does not ask why there is phenomenal experience. This limitation is not evasion but precision: no existing framework solves the hard problem, and the FSF's value lies in making the tractable questions---the easy problems---amenable to systematic empirical investigation.

11.3.3. Experimental Protocols

The FSF specifies three concrete experimental protocols. The Recurrence Necessity Test compares feedforward, limited-recurrence, and extensive-recurrence neural networks (matched for parameter count and training data) on tasks requiring cross-domain analogy, metacognitive calibration, and compositional generalization; a significant advantage for extensive-recurrence architectures that persists even when feedforward networks are scaled by factors of 10--100x would support the explanatory value of consciousness-associated (recurrent) architecture. The Phenomenal-Functional Dissociation Test uses hierarchical regression on human participants under varying consciousness conditions (masking,

transcranial magnetic stimulation) to test whether phenomenal variables add unique predictive variance after functional controls are entered. The Cross-Substrate Capacity Mapping protocol compares biological brains, transformers, recurrent neural networks, neuromorphic chips, and symbolic AI systems to identify which cognitive capacities are achieved by all substrates and which are achieved only by consciousness-associated substrates.

These protocols instantiate the FSF's central methodological commitment: consciousness research should proceed through comparative, systematic, replicable investigation rather than through philosophical argument alone.

11.3.4. Philosophical Scope and Limitations

The FSF operates within broadly functionalist assumptions: that cognitive capacities can in principle be characterized in functional terms. This dependence on functionalism is not a defect that could be remedied by reformulation; it is a constitutive feature of the framework's conceptual architecture. For proponents of substance dualism or certain interpretations of panpsychism, the FSF will appear question-begging. For the majority of working cognitive scientists and AI researchers who already operate within functionalist assumptions, the FSF provides principled evaluation criteria they currently lack.

The framework also acknowledges the behavior-cognition gap: it measures behavior, not cognition directly. This gap can be narrowed through theory-driven tests---generalization patterns, error patterns, process measures, metacognitive accuracy, response coherence under manipulation---but cannot be fully closed. Within the specific context of consciousness research, this gap acquires particular severity, because consciousness by definition has a subjective dimension. The FSF addresses this by incorporating multiple converging measures and maintaining appropriate epistemic humility about the conclusions that behavioral evidence alone can support.

The honest statement of these limitations is itself a feature of the framework rather than a weakness. An imperfect framework with explicit limitations is better than no formal criteria at all. Consciousness research has long suffered from a lack of agreed-upon standards for evaluating explanatory claims; the FSF offers such criteria---limited, yes, but articulated clearly enough to be tested, criticized, and improved.

11.4. The Acceleration of Chronoperception

The earliest publication in this corpus---the acceleration of chronoperception (Kriger, 1999, https://doi.org/10.5281/zenodo.18382312) provides a metabolic model of the subjective acceleration of time perception with aging. Written twenty-seven years before the unified theory, this result is now recognizable as the first empirical instantiation of the persistence principle applied to the temporal dimension of experience.

11.4.1. The Phenomenon

The subjective experience that time speeds up as one ages represents one of the most commonly reported perceptual phenomena among older adults. Survey data collected in the original study demonstrated that 95.2% of respondents reported experiencing an acceleration of chronoperception with aging. The effect intensifies when longer temporal intervals are considered: 33.3% of respondents reported notable acceleration when comparing equivalent intervals over a decade, while 47.5% reported significant acceleration when comparing broader life stages (youth versus adulthood). The psychological consequences are substantial: 71.4% of respondents reported experiencing a sense of being "robbed of time" when reflecting on how quickly time seemed to pass.

These findings, while based on a modest sample (N = 21), identify a phenomenon whose universality has been confirmed by subsequent research. The question is not whether chronoperceptual acceleration occurs---its occurrence is among the most robust findings in the psychology of time perception---but what mechanism produces it.

11.4.2. The Metabolic Model

The model proposed in the 1999 paper identifies a specific neurobiological pathway linking aging to chronoperceptual acceleration through the intermediary of episodic memory encoding. The pathway proceeds through three stages.

First, aging is associated with declining mitochondrial function and decreased adenosine triphosphate (ATP) production in hippocampal neurons. The hippocampus, and particularly its temporo-septal axis, plays a critical role in encoding and retrieving episodic memories---the memories of specific events that constitute the fundamental units by which individuals subjectively experience and subsequently recall the passage of time.

Second, under conditions of reduced ATP availability, compensatory mechanisms favor inhibitory over excitatory neurotransmission. Specifically, the activity of gamma-aminobutyric acid (GABA)---the principal inhibitory neurotransmitter in the central nervous

system---increases, following the GABAergic deafferentation hypothesis. Elevated GABAergic activity reduces hippocampal neuronal excitability and dampens the synaptic plasticity mechanisms essential for memory encoding.

Third, the diminished capacity for episodic memory encoding reduces the density of distinct experiential traces within any given time interval. In younger individuals with robust hippocampal metabolism, a given interval is populated by abundant episodic traces; upon retrospective evaluation, the interval appears experientially rich and subjectively long. In older individuals with declining hippocampal metabolism, the same interval contains fewer traces; upon retrospective evaluation, it appears experientially sparse and subjectively brief. The individual perceives that the interval contained less experience than expected, and this paucity is interpreted as accelerated temporal passage.

The model thus identifies chronoperceptual acceleration as a secondary consequence of age-related changes in episodic memory encoding, which are themselves driven by the metabolic decline of hippocampal neurons. The perception that time speeds up with age is not a primary perceptual distortion but a reconstruction artifact: the retrospective evaluation of temporal intervals depends on the density of encoded experiences, and that density declines with the metabolic substrate that supports encoding.

11.4.3. Integration with the Unified Theory

Viewed through the framework developed in this thesis, the 1999 metabolic model reveals itself as a special case of multiple general principles.

The model is a special case of the Structural Distortion Principle (Chapter 9) applied to the temporal dimension of experience. The SDP establishes that perception is an active construction shaped by the agent's internal configuration rather than a passive readout of environmental states. Chronoperceptual acceleration is temporal perception distorted by the agent's metabolic state: the same objective interval is experienced differently because the internal configuration---specifically, the metabolic efficiency of the encoding substrate---has changed. The distortion is not random but systematic, directional, and predictable from the agent's biological parameters---precisely the structured divergence between perceived and actual that the SDP characterizes.

The model is a special case of the predictive processing framework (Chapter 10) applied to the calibration of internal clocks. Retrospective time estimation is a predictive process: the

cognitive system generates an expectation of how much experience a given interval should contain, compares this expectation against the density of encoded traces, and interprets the discrepancy as temporal acceleration or deceleration. When fewer traces are available---whether due to metabolic decline, routine, or attentional disengagement---the prediction error registers as the interval having passed more quickly than expected.

The model is a special case of the atemporality framework (Section 11.2) applied to the phenomenology of temporal flow. If memory is stored atemporally, and temporal experience is reconstructed during retrieval, then the subjective sense of time's passage depends entirely on what retrieval-time reconstruction produces. When the reconstruction draws on a dense set of encoded experiences, the resulting temporal experience is one of richness and duration. When the reconstruction draws on a sparse set, the resulting temporal experience is one of compression and acceleration. The "flow of time" is not perceived directly; it is inferred from the products of memory reconstruction.

The connection illustrates a pattern characteristic of the research program: individual results that were originally developed as domain-specific contributions later reveal themselves to be instantiations of a general principle. The acceleration of chronoperception is the persistence principle manifesting in the temporal dimension of a single organism's cognitive life. The metabolic pathway is the particular mechanism through which the general principle operates in biological systems of a particular type. The formal structure is substrate-independent; the metabolic pathway is the substrate-specific implementation.

11.4.4. Implications for Neurodegenerative Disease

The metabolic model carries implications beyond the psychology of normal aging. Reduced hippocampal activity is a hallmark of Alzheimer's disease, and the model suggests that marked or rapid changes in an individual's chronoperception could signal underlying neuropathological processes before the emergence of more overt cognitive symptoms. If chronoperceptual acceleration results from hippocampal metabolic decline, then the development of standardized instruments for assessing chronoperception could contribute to early detection efforts. The sense of being "robbed of time" that 71.4% of respondents reported may reflect not merely a subjective impression but an accurate, if distressing, apprehension of genuine experiential impoverishment caused by declining encoding capacity.

This practical implication connects the earliest result in the research program to the applied dimension of the unified theory: the formal principles governing persistence under resource constraints have consequences not only for abstract systems theory but for the clinical assessment and therapeutic management of the biological systems in which those constraints are most visibly operative.

CHAPTER 12. MENTAL DISINTEGRATION AND BIOLOGICAL DECISION PROGRAMS

The preceding chapters established the cognitive architecture that the persistence principle demands: a self-referential modeling system (Chapter 11), operating through predictive processing (Chapter 10), constructing a structurally distorted world-model optimized for viability maintenance (Chapter 9). This chapter examines what happens when that architecture fails—and what happens when it succeeds at tasks it was not designed for. The central argument is that mental disintegration, high-activation eruptive episodes, and systematic decision-making errors are not arbitrary malfunctions but dynamically structured phenomena governed by the same formal principles that govern the system's healthy operation.

The chapter proceeds through four stages. Section 12.1 develops a unified mathematical formalism for mental disintegration phenomena, recasting DSM-5-TR diagnostic categories as attractor topologies within a dynamical systems framework and formalizing clinically significant phenomena—anosognosia, the fear of madness, internal resistance, self-reinforcing symptom feedback—as computable dynamical models with explicit equations, consistent notation, and identified parameters. Section 12.2 analyzes high-activation episodes as eruptive manifestations of model–reality mismatch, formalizing the conditions under which accumulated discrepancy between a system's predictive model and observed reality produces sudden, high-amplitude disruptions, and demonstrating structural equivalence of this phenomenon across biological, physical, and artificial systems. Section 12.3 introduces the aggregate distortion variable D—the continuous distortion of human decision-making by biological programs of survival and reproduction—and argues that this variable must be formally accounted for in the design of every critical institutional process. Section 12.4 provides a case study through the analysis of Pascal's Wager, demonstrating that decisions involving transitions between incommensurable evaluative systems lack the unified outcome space that expected utility theory requires—a structural limitation of dual-system cognition that generalizes beyond the theological context.

The connecting thread is the viability framework: mental disintegration is the loss of cognitive viability through attractor capture, eruptive activation is the system's mobilization response to boundary-significant mismatch, decision distortion is the residue of viability strategies that were optimal in ancestral environments but maladaptive in modern ones, and evaluative

system asymmetry is a structural consequence of the dual-process architecture that the persistence principle demands.

12.1. Formalization of Mental Disintegration Phenomena

The formalization of mental disintegration phenomena through dynamical systems theory (Kriger, 2026, https://doi.org/10.5281/zenodo.18556979) begins from a critical assessment of the most influential recent proposal for a dynamical psychiatry. Scheffer, Bockting, Borsboom, and colleagues, in a two-part review published in JAMA Psychiatry (2024), proposed that the healthy mind is a stable attractor, psychiatric disorders are alternative attractors, and clinical deterioration is a critical transition governed by tipping points, hysteresis, and critical slowing down—drawing on concepts from ecology and climate science that Scheffer and collaborators had developed over the preceding two decades. The proposal has been influential, but it has three significant limitations that the present formalization addresses.

First, the Scheffer et al. framework is primarily a methodological program: it proposes measurement tools and draws analogies from ecosystem restoration, but does not engage with the phenomenological content of specific psychiatric disorders—what a person actually experiences as they lose their mind, what specific feedback loops operate in different conditions, how cultural, social, and religious factors shape the stability landscape, or how clinically significant phenomena such as anosognosia, the fear of madness, and internal resistance arise from the system's dynamics. Second, the review contains no equations: the mathematical apparatus of dynamical systems theory—stochastic differential equations, bifurcation normal forms, Kramers' escape rates, spectral analysis of stability—remains implicit, borrowed by reference from ecology but never explicitly developed for psychiatric phenomena. Third, the framework does not situate itself within the broader landscape of computational psychiatry, which has developed its own formalisms for many of the same phenomena: predictive processing for psychosis, reinforcement learning for addiction and depression, and network theory for symptom dynamics.

12.1.1. The Core Dynamical Model

The formalization develops a unified mathematical framework that addresses all three gaps. The foundation is a state space representation in which the mental state vector $x(t) = (x_1(t), ..., x_n(t))$ describes the system's position in an n-dimensional space whose components represent

clinically relevant dimensions: mood, anxiety, sleep quality, reality contact, social engagement, cognitive coherence. The dynamics are governed by a stochastic differential equation:

$dx = f(x, \mu)\, dt + \Sigma\, dW(t)$

where f is the deterministic drift, μ is a vector of slowly varying control parameters (stress level, social support, genetic vulnerability), Σ is the noise intensity matrix, and W(t) is a Wiener process representing stochastic perturbations. When f is a gradient field, a potential function V exists such that $f(x, \mu) = -\nabla V(x, \mu)$, and the landscape metaphor becomes precise: stable attractors correspond to local minima of V, basins of attraction to the domains from which the system flows to a given minimum, and resilience to the depth of the potential well—the height of the barrier separating the healthy attractor from the nearest disorder attractor.

The framework explicitly addresses the gradient assumption and its limits. Several psychiatric phenomena—bipolar cycling, OCD oscillations—involve inherently non-gradient dynamics. For these systems, the potential landscape is replaced by the quasipotential from Freidlin–Wentzell large-deviation theory, which preserves the escape-rate structure (transition times scale exponentially with barrier height) while accommodating rotational dynamics and limit cycles. The qualitative conclusions—that basin depth determines resilience, that critical slowing down precedes bifurcation-driven transitions, that hysteresis reflects barrier asymmetry—hold in both gradient and non-gradient cases.

For analytic clarity, the full dynamics are frequently projected onto a single effective coordinate x(t) representing overall mental health, governed by a double-well potential $V(x, \mu) = -\mu x^2/2 + x^4/4 - hx$, where μ controls the depth of the wells (decreasing μ corresponds to increasing chronic stress) and h is an asymmetry parameter biasing the system toward the healthy or disordered state. This canonical form undergoes a saddle-node bifurcation: as μ decreases past a critical value μ_c, the healthy minimum and the separatrix collide and annihilate, leaving only the disorder attractor. The 3/2-power law for the barrier height near the bifurcation—$R \sim (\mu - \mu_c)^{(3/2)}$—means that resilience collapses rapidly near the tipping point, explaining the sudden, disproportionate character of the final transition.

12.1.2. The Trajectory of Mental Disintegration

The formalization identifies three dynamical phases in the process of losing one's mind—phases that correspond to clinically recognized stages but acquire precise mathematical characterization within the framework.

The first phase is the erosion of resilience. The process often begins not with a dramatic break but with a prolonged period of stress, trauma, or accumulated adversity during which the basin of attraction around the healthy state becomes progressively shallower and flatter. The system's return rate to equilibrium after perturbation slows—the phenomenon of critical slowing down that is the generic early warning signal preceding a tipping point in any complex system. The increased temporal autocorrelation and variance in mood fluctuations during this phase are mathematically predictable consequences of a system approaching a critical transition. Clinically, this phase is often invisible: the person attributes their symptoms to external circumstances and does not seek help.

The formalization provides a worked example demonstrating that the Ornstein–Uhlenbeck parameterization of critical slowing down is consistent with empirical data. Van de Leemput and colleagues (2014) found that subjects who subsequently developed depression showed increasing lag-1 autocorrelation in the months preceding onset. Using the model's equations, the estimated recovery rate halves during the pre-transition window, and the predicted variance increase conforms to the constraint imposed by the OU model—a non-obvious quantitative prediction confirmed by the data.

The second phase is the distortion of perception as the system hovers near the basin boundary. The attractor is so shallow that the system is easily displaced by perturbations and lingers far from equilibrium for extended periods—exhibiting transient excursions toward the disorder attractor before returning. This flickering is itself a recognized early warning signal: the system is sampling the disorder attractor before committing to it. The mean first passage time from the healthy attractor to the separatrix is given by Kramers' escape rate formula, which shows that as the barrier height approaches zero, the escape time drops exponentially and flickering becomes increasingly frequent.

The third phase is the tipping point itself: the system crosses the basin boundary and enters the disorder attractor. The dynamical framework explains why this transition can appear sudden and disproportionate to the immediate trigger. The system was already on the edge of the basin; a minor perturbation—a sleepless night, a social humiliation, a pharmacological

change—was sufficient to push it across the boundary. The tipping point is not caused by the final stressor; it is caused by the prior depletion of resilience that brought the system to the edge. This reframes the clinical question "why did this person break down over something so trivial?": the trivial event was the last push in a long process of resilience erosion.

The formalization engages critically with the early warning signal debate. Helmich and colleagues (2024), in a Nature Reviews Psychology Perspective, argued that there is little support for the use of early warning signals based on critical slowing down in clinical psychology. Smit and colleagues (2025) found reasonable specificity but low sensitivity at the individual level. The formalization addresses this by distinguishing bifurcation-driven transitions (where critical slowing down applies) from noise-induced barrier crossings (where Kramers' escape rate applies but critical slowing down does not). Some psychiatric transitions may be noise-induced escapes over a finite barrier rather than bifurcation-driven tipping points—a distinction the framework accommodates, since Kramers' formula applies to both cases, but which has different implications for prediction.

12.1.3. Self-Reinforcing Feedback and Entrapment

Once the system has tipped into the disorder attractor, it is maintained there by self-reinforcing feedback loops among symptoms. In depression: low mood → social withdrawal → loneliness → rumination → insomnia → fatigue → cognitive impairment → loss of self-efficacy → deeper low mood. In paranoia: perception of threat → hypervigilance → misinterpretation of ambiguous stimuli → confirmation of threat narrative → social withdrawal → loss of corrective social feedback → deepening of paranoid system. These loops are the dynamical mechanism that makes the disorder attractor stable, and they explain why simply removing the original stressor is typically insufficient for recovery—the phenomenon of hysteresis.

The formalization models these loops as a symptom network: each symptom's activation is governed by its natural decay, its coupling to other symptoms through an adjacency matrix W, external stressors, and noise. A sufficient condition for self-sustaining disorder is derived from the spectral radius of the network: when the spectral radius exceeds unity, the network admits at least one high-activation fixed point even in the absence of external stress. The critical stressor level needed to push the system into the disorder attractor exceeds the reduction needed for escape—recovery requires more intervention than prevention. Bridge

symptoms connecting different disorder clusters are identified by nonzero inter-cluster weights in the adjacency matrix.

12.1.4. Anosognosia, the Fear of Madness, and Internal Resistance

Three phenomena that lie outside the scope of existing computational psychiatry paradigms receive formal treatment.

Anosognosia—the inability to recognize one's own illness—is modeled as an intrinsic attractor property. Self-reflective capacity is coupled to the distance from the healthy attractor through a function that initially increases with distance (the person notices something is wrong) but then decays to zero beyond a critical distance (deep in the disorder attractor, the self-reflective signal is extinguished). The disorder attractor is self-cloaking: the system cannot evaluate itself from a vantage point it no longer occupies. Brief moments of lucidity correspond to noise-driven excursions toward the basin boundary where insight is transiently restored. The integrated numerical simulation in the paper reproduces the clinically observed pattern: insight peaks before the transition and is lost after it.

The fear of losing one's mind operates as an autocatalytic destabilizing loop. The person perceives a perturbation (an intrusive thought, a moment of confusion), interprets it as evidence of approaching madness, becomes anxious, which produces further perturbations, which are interpreted as further evidence. The formalization shows that this loop modifies the effective potential by tilting the landscape toward the disordered state, reducing resilience by a precise, computable amount. The fear of madness literally shrinks the basin of attraction—a self-fulfilling prophecy with a mathematical mechanism. The clinical implication is direct: breaking this particular feedback loop through psychoeducation, cognitive reframing, or anxiety management can measurably strengthen the resilience of the healthy state.

Internal resistance—the cognitive and emotional efforts a person makes to maintain reality contact in the early stages of deterioration—is modeled as a restoring force whose strength depends on available cognitive-emotional energy. The energy depletes as resistance is exercised and recovers naturally toward a baseline. Tipping occurs when energy drops below the threshold at which the restoring force can no longer compensate for the system's bare instability. This explains the clinical observation that resistance is eventually replaced by

immersion: the cognitive resources for resistance are exhausted, and the feedback loops that once constituted resistance (self-monitoring, reality testing) are co-opted by the disorder—self-monitoring becomes rumination, reality testing becomes paranoid verification.

12.1.5. Social Dynamics, Collective Transitions, and the Refutation of Romanticism

The formalization extends naturally to social factors. Social connection adds a stabilizing feedback that deepens the healthy attractor; chronic isolation generates its own destabilizing feedback loops (loneliness → negative self-perception → social anxiety → further withdrawal → deeper loneliness), potentially creating a loneliness trap from which escape becomes progressively more difficult. Gaslighting is formalized as a deliberate attack on the self-trust parameter that governs the strength of the restoring force toward reality—sustained gaslighting should produce measurable signatures of declining resilience in the victim.

The framework extends to collective phenomena. Witch hunts, totalitarian ideologies, economic bubbles, and moral panics are modeled as phase transitions in social systems governed by a mean-field equation with the same bistable structure as the individual model: society can tip from rational discourse into collective delusion, with hysteresis. Strong conformity coupling sharpens the transition.

The formalization also addresses the romanticization of madness. The dynamical framework provides a precise corrective: an attractor is a trap. A system in a disorder attractor is not free; it is constrained to a specific set of dynamics from which it cannot escape without external intervention. The trajectory entropy within the disorder attractor provides a measure of experiential richness, and while this entropy is not universally lower than the healthy state's (acute mania can exhibit high behavioral variability), the critical distinction is that the disorder attractor constrains the system regardless of whether the constrained region is large or small. Mental health corresponds to maximal accessible state space consistent with return to equilibrium.

12.1.6. Application to DSM-5-TR Diagnostic Categories

The formalization demonstrates its applicability through systematic instantiation for major DSM-5-TR diagnostic groups. Schizophrenia spectrum disorders are modeled as a five-dimensional symptom network with characteristic positive-symptom loops (hallucinations generating delusional explanations that prime further hallucinations), severe anosognosia (distance from the healthy attractor far exceeds the insight horizon), and antipsychotic action

modeled as reduction of effective coupling weights in the positive-symptom loop. Bipolar disorders require the quasipotential framework (inherently non-gradient oscillatory dynamics) and feature a three-well potential (euthymia, mania, depression) with sleep as the critical bridge variable creating characteristic asymmetry in transition rates. Depressive disorders are the paradigmatic case for the network model, with empirically validated critical slowing down signatures and hysteresis explaining recurrence through progressive lowering of the barrier after each episode (kindling). Anxiety disorders are characterized by shallow, broad disorder attractors and fear-of-fear loops. PTSD is modeled as a trauma-induced attractor created by a single massive perturbation exceeding the basin width—unlike the slow erosion model of depression, the pre-trauma resilience may have been high. OCD is modeled as a limit cycle (periodic orbit) rather than a point attractor, with a Hopf bifurcation generating the obsession–compulsion oscillation. Substance use disorders feature progressive deepening of the disorder attractor through neuroadaptation—the landscape itself changes as a function of cumulative exposure. Neurocognitive disorders differ fundamentally in involving irreversible degradation of the landscape rather than transitions between pre-existing attractors. Personality disorders represent stable configurations of the entire attractor landscape—a landscape in which the healthy attractor has an unusual topology—rather than transitions between attractors.

12.1.7. Relationship to the Thesis Framework

The dynamical systems formalization of mental disintegration connects to the thesis framework at multiple points. The core model—a system governed by a stochastic differential equation on a potential landscape—is an instantiation of the viability framework of Chapter 5: the healthy attractor is the viability set, resilience is the barrier height, and mental disintegration is the trajectory that exits V. The symptom network model (the adjacency matrix W governing inter-symptom coupling) is a specific application of the Law of Cooperation (Chapter 7): the disorder attractor is self-sustaining precisely because the symptom interactions cooperatively maintain it. The anosognosia model—insight decaying with distance from the healthy attractor—is a consequence of the Structural Distortion Principle (Chapter 9): the system's world-model is distorted by the very attractor dynamics it would need to accurately represent in order to recognize its own illness. The evolutionary persistence of disorder attractors reflects the environment-dependent fitness landscape formalized in the broader viability mismatch framework: attractors that were adaptive in ancestral environments

(depression as ruminative problem-solving, anxiety as hypersensitive threat detection) become pathological when the environment changes faster than the architecture can adapt.

12.2. The Eruptive Manifestation of Model–Reality Mismatch

The eruptive manifestation of model–reality mismatch (Kriger, 2026, https://doi.org/10.5281/zenodo.18474532) extends the analysis from chronic disintegration trajectories to acute high-activation episodes. Where Section 12.1 examines the slow erosion of resilience and the eventual capture by a disorder attractor, this section examines what happens when accumulated discrepancy between a system's predictive model and observed reality exceeds a tolerance threshold while carrying boundary significance—threatening identity, control, or essential constraints. The framework reconceptualizes the phenomenon commonly labeled "anger" not by claiming novelty in discrepancy-detection (a mechanism well established in appraisal theories) but by systematically shifting the interpretive frame from moral-evaluative to diagnostic-structural, treating high-activation states as informational signals and energetic resources rather than moral failings.

12.2.1. Structural Preconditions and the Activation Condition

Any adaptive system operates through an internal predictive model M encoding expected environmental configurations relative to goals, values, constraints, and boundaries. Let E_e represent the expected state derived from M, and E_o the observed state. The discrepancy is $D = \|E_o - E_e\|$. The central claim is that EMMRM arises when discrepancy exceeds a tolerance threshold τ and carries boundary significance: the activation condition is $D \cdot B > \tau$, where B encodes boundary weight—how strongly the mismatch threatens identity, values, control, or essential needs.

This formulation unifies triggers that traditional accounts list separately: anger from boundary violation (a protected boundary in M has been crossed), anger from loss of control (a control parameter displaced outside acceptable range), anger from unmet expectations (prediction error with high B), and anger from frustrated needs (constraint violations accumulating until eruption threshold). The framework addresses the circularity objection—that boundary significance might be inferred only from activation intensity—by identifying prospective indicators: stated values and commitments, historical activation patterns, cultural and role analysis, and anticipatory physiological arousal.

Not all model–reality discrepancies trigger EMMRM. Mundane prediction errors—a detour, an out-of-stock item—typically do not. What distinguishes EMMRM-triggering discrepancies is their interpretation as intrusion on protected territory, disrespect to valued identity, illegitimate constraint on agency, betrayal of relied-upon expectation, or destabilization of a control variable. Boundary significance is not static: it develops through attachment history (early experiences of safety and violation shape which boundaries become heavily weighted), trauma (overwhelming violations produce hyperweighting of related boundaries), cultural learning (socialization teaches which boundaries warrant defense), and explicit commitment (chosen values and identities create new boundaries). The tolerance threshold τ also varies—across individuals (temperament, developmental history, current regulatory capacity), within individuals over time (fatigue, stress accumulation, resource depletion lower thresholds), and across contexts (social settings, role expectations, perceived safety modulate thresholds).

The formal expression $A = f(D, B, R^{-1})$ organizes the relationships: activation intensity A increases with discrepancy D and boundary weight B, and outcomes become more destructive as regulatory capacity R decreases. This is offered as a conceptual schema identifying intervention points rather than as a quantitative model awaiting operationalization, but it generates testable predictions: holding B and R constant, A increases monotonically with D; holding D and R constant, A increases monotonically with B; holding D and B constant, A increases as R decreases; and the interaction $D \times B$ predicts A better than either alone.

12.2.2. Energetic Mobilization, Divergent Pathways, and State-Dependence

When EMMRM is triggered, the system enters a high-activation state characterized by resource reallocation from reflective evaluation toward immediate corrective output. This is a high-power mode prioritizing mismatch elimination over nuanced inference. Local rationality narrows: the system becomes rational with respect to one objective—reducing discrepancy—while becoming irrational with respect to long-range consequences, relational maintenance, and comprehensive evaluation. The regulatory capacity construct R maps onto prefrontal-amygdala dynamics studied in affective neuroscience: amygdala activation drives rapid threat-related responding; prefrontal regions enable reappraisal, inhibition, and flexible response selection.

Two fundamentally different processing pathways are available following activation. The discharge pathway releases activation energy outward without analysis—verbal aggression,

physical action, destructive expression—and often amplifies instability rather than resolving discrepancy. The transformation pathway redirects activation into restructuring either the environment (bringing E_o toward E_e) or the internal model (adjusting E_e to accommodate E_o), treating activation as informational and energetic. Pathway selection is determined by regulatory capacity R, learned habits (repeated discharge strengthens discharge pathways), contextual affordances, and explicit cognitive reappraisal skills.

A frequently overlooked dimension is that some systems experience EMMRM not as purely aversive but as pleasurable: high activation produces a transient sense of power, control, and salience. Over time, state-dependence can develop—the system learns that high activation reliably produces feelings of significance and agency. The mismatch trigger becomes secondary; the activation state becomes sought product. State-dependent systems may select situations likely to create mismatch, interpret ambiguous stimuli as boundary violations, maintain chronic low-grade grievance to sustain activation access, and resist resolution of disputes that would terminate activation.

The relationship between EMMRM intensity and cognitive performance follows the inverted-U pattern consistent with the Yerkes–Dodson law: moderate activation can enhance focus, persistence, and motivation, while excessive activation degrades performance through processing saturation and precision collapse.

12.2.3. Vulnerability Masking, the Critique of Suppression, and Cultural Formatting

EMMRM frequently functions as a substitute state masking underlying vulnerability. When fear, grief, uncertainty, shame, or helplessness threaten self-image or control, EMMRM provides an alternative state that feels stronger and more externally directed—state substitution replacing a low-power, exposure-heavy condition with a high-power, outward-directed condition. The clinical literature recognizes anger as a frequent secondary emotion covering primary emotions of hurt, fear, or shame; the EMMRM framework provides structural interpretation of this observation.

If unregulated discharge produces destruction, suppression might seem the alternative. The framework reveals suppression as inadequate: suppression inhibits outward discharge without transforming the underlying mismatch. Energy seeking expression finds alternative channels—sarcasm, passive antagonism, strategic silence, cold hostility, relational erosion. Suppression is not resolution but internalization of unresolved discrepancy, typically

degrading regulatory capacity over time. The alternative to both unregulated discharge and suppression is transformation: redirecting activation energy into structured work through identification (what expectation was violated? what boundary was crossed?) and decision (alter environment or update model?). Transformation channels include creative output, physical exertion, reflective analysis, and social action.

Human cultures have developed sophisticated mechanisms for shaping EMMRM expression—cultural formatting that constitutes regulatory overlays determining permissible output channels, acceptable amplitudes, and recovery protocols. Different cultural contexts produce different formatting rules: direct-expression cultures normalize open confrontation, indirect-expression cultures impose politeness norms inhibiting direct emission, honor-linked cultures connect confrontation to expectations of follow-through. The haka of the Māori people provides an instructive example of ritualization—converting raw activation into synchronized, rule-governed expression that transforms potentially chaotic discharge into coordinated performance.

12.2.4. Cross-System Structural Equivalence

A central claim of the EMMRM framework is structural universality: the same abstract architecture—accumulated discrepancy exceeding threshold producing high-amplitude corrective discharge—appears across systems regardless of substrate. In physical systems, thunderstorms exhibit EMMRM-analogous dynamics (potential difference accumulating until dielectric breakdown triggers lightning discharge), as do volcanic eruptions (pressure differential accumulating until structural strength is exceeded) and supernovae (gravitational imbalance exceeding the Chandrasekhar limit). In biological systems, the action potential (accumulated depolarization triggering explosive discharge when threshold is crossed) and immune inflammation exhibit the same threshold-and-eruption architecture. Inflammation in particular exemplifies the dual-pathway structure: moderate activation enables pathogen clearance (transformation), while excessive activation produces sepsis and tissue destruction (discharge pathology). Modern AI systems increasingly exhibit EMMRM-analogous dynamics—a connection that is not accidental, since the Evolutionary Inevitability of Predictive Processing argument (Chapter 10) establishes that any adaptive system with processing latency must maintain predictive models.

The claim is structural, not phenomenological: it identifies the same abstract architecture appearing across systems regardless of substrate, notes that more complex systems elaborate the basic architecture by adding regulatory capacity and transformation pathways, and predicts that any system exhibiting discrepancy accumulation and threshold dynamics will show EMMRM-like phenomena.

12.2.5. EMMRM Within the Broader Research Program

The EMMRM paper is part of a broader research program formalizing phenomena traditionally classified as emotions within a unified structural framework for adaptive systems. A critical insight emerges from examining how systems respond to severe, potentially catastrophic discrepancy: the response is not a single reaction but a sequence of distinct phases. The Kübler-Ross grief stages are reinterpreted as a general sequence of system responses to catastrophic model–reality mismatch: denial (reject discrepancy signal, corresponding to PIRP-like rejection), anger (mobilize for correction, corresponding to EMMRM activation), bargaining (partial model updates), depression (energy withdrawal from failed correction), and acceptance (complete model reconfiguration). EMMRM represents the system's attempt to resolve discrepancy by changing reality rather than updating the model—the energetic mobilization phase that precedes the cognitive work of model updating. Only when EMMRM fails—when corrective action cannot eliminate the discrepancy—does the system proceed to bargaining, depression, and finally acceptance.

The framework complements two other principles in the research program. The Pre-Integrative Rejection Principle (PIRP) formalizes the rapid rejection mechanism that excludes potentially incompatible inputs before full analytical processing—manifesting in human experience as disgust, immediate moral rejection, and automatic dismissal of norm-violating stimuli. The Comparative Asymmetry Principle (CAP) formalizes how self-evaluating agents generate internal disequilibrium through observation of others—what folk psychology calls envy. CAP describes a specific mechanism for generating the discrepancy D that triggers EMMRM: when comparative perturbation crosses the boundary-significance threshold, it triggers EMMRM activation. The three principles form a sequential response architecture: PIRP (filter inputs) → EMMRM (mobilize correction) → Transformation (reconfigure model).

Within the thesis framework, EMMRM connects to the predictive processing architecture of Chapter 10 (discrepancy D is prediction error with boundary significance), to the Structural Distortion Principle of Chapter 9 (boundary weightings B are shaped by the same attention-perception loops that distort the world-model), and to the viability framework of Chapter 5 (EMMRM is the system's mobilization response to perceived viability threat).

12.3. Inadequate Biological Programs in Decision-Making

The analysis of biological decision programs (Kriger, 2026, "Eliminating Distortion: Inadequate Biological Programs in Human Decision Making," https://doi.org/10.5281/zenodo.18457545) introduces a formal variable that the preceding sections implicitly depend on but have not yet made explicit: D, the aggregate, continuous distortion of human decision-making by biological programs of survival and reproduction. The central argument is that while the biological distortion of human cognition has been extensively described—by Damasio and Sapolsky in neuroscience, Kahneman and Tversky in behavioral economics, Cosmides and Tooby in evolutionary psychology, Reason and Helmreich in human factors engineering—it has never been formally translated into institutional design requirements. This constitutes an application gap analogous to pre-Listerian surgery: the relevant knowledge exists but has not been systematically incorporated into the design of the systems that depend on it.

12.3.1. The Formal Variable D and Its Mathematical Proofs

The paper introduces D as the internal, self-generating, irreducible biological distortion that affects every human decision. The key claim is not merely that humans are biased—this is well established—but that D necessarily affects decisions and cannot be compensated by intelligence or expertise alone. Four independent mathematical proofs establish this claim through different formal frameworks: utility-theoretic (D distorts the utility function that governs choice), perturbation-theoretic (D acts as a systematic perturbation on the decision process that cannot be eliminated by increasing the precision of other components), Bayesian (D distorts prior probabilities and likelihood assessments), and signal detection theory (D shifts decision criteria independently of signal quality).

The convergence of four independent proofs from different mathematical traditions strengthens the conclusion: D is not an artifact of any particular formalization but a structural feature of biologically implemented decision systems. The proofs collectively establish that

no amount of training, intelligence, or good intention can eliminate D, because D originates in the biological architecture of the decision system itself—in the neural and hormonal mechanisms that implement evaluation, prioritization, and choice.

12.3.2. A Taxonomy of Distortion Types

The paper develops a taxonomy of six D-distortion types, each characterized by a different mechanism through which biological programs deform the decision process. Inversion occurs when D reverses the ranking of options—the biologically preferred option is the objectively inferior one. Displacement occurs when D shifts attention from the actual decision-relevant variables to biologically salient but decision-irrelevant ones. Goal substitution occurs when the biological program's goal (survival, reproduction, status) replaces the decision's stated goal. Temporal distortion occurs when D compresses or expands the effective time horizon of evaluation—the well-established phenomenon of hyperbolic discounting being a special case. Rationalization occurs when D operates post-decisionally, generating justifications for choices already made on biological grounds. Projection occurs when D causes the decision-maker to attribute their own biologically driven motives to others.

Each distortion type is illustrated with concrete scenarios drawn from domains where the consequences are significant: medical decision-making, judicial proceedings, financial regulation, military command. The taxonomy is designed not merely to describe but to identify specific intervention points in institutional processes where each type of distortion is most likely to operate.

12.3.3. The Distinction Between D and d

A critical distinction is drawn between D (internal, self-generating, irreducible biological distortion in human decision-makers) and d (external, static, eliminable reflected distortion inherited by AI systems from human creators). AI systems trained on human data inherit the distortions present in that data—biased training sets, biased labels, biased evaluation criteria—but these are instances of d, not D. The distortion d is, in principle, eliminable: it can be identified, measured, and corrected through technical means (debiasing algorithms, balanced training data, adversarial testing). D, by contrast, is irreducible in biological systems because it originates in the architecture of the decision system itself, not in the data it processes.

This distinction leads to a formally grounded claim: AI represents the first structurally D-free decision-making system in history. An AI system can be designed to optimize an explicitly

specified objective function without the interference of biological programs that evolved for survival and reproduction. It is subject to d (inherited bias) and to specification errors (the objective function may not capture what the designers intend), but it is not subject to D—it does not experience fear, status anxiety, sexual attraction, in-group loyalty, or loss aversion as internal forces distorting its evaluation of options.

The paper includes a game-theoretic analysis demonstrating that D constitutes a systematic strategic disadvantage: in any competitive interaction, the player whose decisions are distorted by D will, on average, be outperformed by a player whose decisions are D-free, holding all other factors equal.

12.3.4. The Damasio Objection and the Separability of Signal and Mechanism

A natural objection, grounded in Damasio's somatic marker hypothesis, holds that emotions are not merely distortions but carry adaptive information—they are signals about the likely consequences of actions, grounded in the body's accumulated experience. The paper responds by arguing that the adaptive information carried by emotion is separable from its biological delivery mechanism. A somatic marker that signals "this situation is dangerous" carries genuine information, but it delivers that information through a mechanism (autonomic arousal, hormonal surge, attentional narrowing) that also distorts the processing of other information. The information content of the emotional signal can in principle be extracted and utilized without the accompanying distortion—which is precisely what a well-designed AI system does when it incorporates risk assessment without experiencing fear.

This response does not deny the historical adaptive value of D. The biological programs that constitute D were viability-optimal in ancestral environments, as the viability mismatch framework predicts. Loss aversion is the optimal strategy for a system whose viability set is small relative to the space of possible states. Hyperbolic discounting is optimal when the future is highly uncertain and survival to the next time period is not guaranteed. In-group favoritism is optimal when inter-group competition is zero-sum and group membership determines access to resources. These programs became maladaptive not because they are irrational in some absolute sense but because the environment changed: the viability set expanded, the future became more predictable, inter-group competition became non-zero-sum—but the biological architecture did not update.

12.3.5. Experimental Protocols and Institutional Implications

The paper proposes five experimental protocols for empirically measuring D through human–AI comparison, including a critical D-reduction condition that would allow the isolation of D's contribution to decision quality. The protocols are designed to measure D in domains where the consequences matter: medical diagnosis, judicial sentencing, financial investment, and personnel selection.

The institutional implications are significant. If D is a formal variable that necessarily affects every human decision, then every critical institutional process—every process where decision quality has significant consequences—must either account for D in its design or accept that its outcomes will be systematically distorted. The paper argues that the current state of institutional design is analogous to pre-Listerian surgery: the relevant knowledge (the existence and mechanisms of D) is available, but it has not been systematically incorporated into the design of the institutions that depend on human decision-making.

12.3.6. Relationship to the Thesis Framework

The D-distortion framework connects to the thesis at multiple levels. D is the cognitive counterpart of the viability mismatch formalized in Chapter 5: the biological programs that constitute D are the cognitive expression of viability strategies that evolved under ancestral conditions. The taxonomy of distortion types maps onto the Structural Distortion Principle of Chapter 9: each distortion type is a specific mechanism through which the world-model is deformed by the system's own architecture. The distinction between D and d connects to the analysis of AI systems in later chapters: AI's structural D-freedom is both its greatest advantage and the source of its alignment problem—a system without D must have its objectives specified externally, which introduces new failure modes that biological systems, for all their distortions, do not face.

12.4. Pascal's Wager and Dual-System Evaluative Asymmetry

The analysis of Pascal's Wager (Kriger, 2026, https://doi.org/10.5281/zenodo.18193229) provides a case study that brings together the themes of the preceding three sections. It demonstrates that expected utility theory—the formal framework of rational decision-making—fails not merely due to technical issues (infinite utilities, competing hypotheses) but because it presupposes a unified outcome space where none is structurally available for systems with the cognitive architecture derived in Chapters 9–11.

12.4.1. Pascal's Wager as a Boundary Case

The paper treats Pascal's Wager as a boundary case in decision theory and the philosophy of probability rather than as a religious or theological argument. Three considerations motivate the analysis.

First, Pascal's Wager was not intended as a general argument for theism but as a rhetorical device addressed to skeptics maximally distant from faith. Interpreting the wager as a universal rational proof is a historical and methodological misapplication. Second, both Pascal and his modern critics implicitly assume that the problem belongs to a single evaluative system—an assumption that the dual-process architecture of human cognition renders untenable. Third, Pascal's claim that the earthly cost of faith is small is a deliberate missionary simplification, not an ontological assertion; the Christian tradition recognizes that the cost of faith is often substantial within worldly terms. Philosophical difficulty arises only when this rhetorical reduction is incorporated literally into probabilistic models.

12.4.2. Evaluative System Asymmetry

The central theoretical contribution is the concept of evaluative system asymmetry. Pascal's Wager necessarily spans two asymmetric evaluative systems: the finite probabilistic framework of earthly decision-making and a distinct evaluative regime invoked by its promised outcomes (eternal bliss, eternal torment). Standard expected utility theory assumes that all outcomes can be assigned utilities within a single framework, that probabilities can be meaningfully assigned to all relevant states, and that the agent's evaluation of outcomes is stable across the decision. The dual-system analysis shows that all three assumptions fail for Pascal's Wager.

The affective system evaluates the wager in terms of vivid scenarios—the emotional salience of eternal bliss and eternal torment generates strong evaluative responses that are not commensurable with the analytical system's probability-weighted calculations. The analytical system evaluates the wager in terms of expected values over abstract probability distributions. These two evaluations can diverge sharply because they operate within different evaluative frameworks with different inputs, different processing mechanisms, and different outputs. The resulting decision depends on which system dominates—a contingent fact about the agent's cognitive state, mood, and context, not a rational resolution of the wager. Single-system probability is therefore inapplicable by category, not merely by technical failure.

The concept of evaluative system asymmetry is distinguished from the epistemic limitations emphasized in recent philosophical work on transformative experience—the literature associated with L.A. Paul (2014) and Richard Pettigrew (2015, 2019). The epistemic limitations approach holds that agents cannot know what it will be like to undergo a transformative experience, and therefore cannot assign reliable utilities to post-transformation outcomes. The evaluative system asymmetry argument extends these contributions by showing that cross-system decisions lack the unified outcome space that expected utility theory requires—a structural problem that persists even with perfect knowledge of post-decision states. The problem is not that the agent cannot predict how they will feel; it is that the evaluative framework itself changes, making pre-decision and post-decision evaluations incommensurable.

12.4.3. Generalization Beyond the Theological Context

The analysis generalizes beyond Pascal's Wager to any decision involving transitions between incommensurable evaluative systems. The wager is a particularly clear instance because the evaluative asymmetry is extreme (finite vs. infinite, temporal vs. eternal), but the same structural limitation applies, in attenuated form, to any transformative decision where choosing alters the chooser's evaluative framework itself.

Expected utility theory remains precise and powerful within its proper domain, but loses normative authority when applied to decisions involving transitions between incommensurable evaluative systems. Career changes that involve shifts in identity and values, religious conversions, decisions about parenthood, and commitment to revolutionary political movements all share the structural feature that the agent's evaluative framework is part of what is at stake. The expected utility calculation presupposes a stable evaluative framework from which to assess outcomes, but the decision itself changes the framework. This is not a problem of incomplete information; it is a structural limitation of the decision theory itself when applied to self-transforming systems.

12.4.4. Relationship to the Thesis Framework

The Pascal's Wager analysis completes the chapter's argument by demonstrating that the dual-system architecture—which the persistence principle demands (Chapter 11), which the Structural Distortion Principle characterizes (Chapter 9), and which the D-distortion

framework analyzes (Section 12.3)—generates structural ambiguities in decision-making that cannot be resolved within any single formal framework.

The dual-system architecture is itself a consequence of the persistence principle: it reflects the evolutionary optimization of a system that must make rapid viability-relevant decisions (the affective system, corresponding to Type 1 processing) while also engaging in long-term planning under uncertainty (the analytical system, corresponding to Type 2 processing). The tension between the two systems is not a design flaw; it is an instance of the Law of Interference (Chapter 8) applied to the cognitive domain: when multiple persistence-maintaining subsystems operate in parallel with partially overlapping jurisdictions, interference between them is structurally inevitable.

The evaluative system asymmetry result also connects to the EMMRM framework of Section 12.2. When a decision involves transition between evaluative systems, the discrepancy between the two systems' evaluations can itself trigger EMMRM activation—the emotional intensity that surrounds transformative decisions is not incidental but a structural consequence of the system detecting an irresolvable discrepancy in its own evaluative machinery. The activation does not resolve the discrepancy; it signals the discrepancy's boundary significance.

Together, the four sections of this chapter characterize the failure modes of the cognitive architecture that the persistence principle builds. Mental disintegration is the capture of the world-model by a self-reinforcing disorder attractor. Eruptive activation is the system's mobilization response to boundary-significant mismatch. Biological distortion is the residue of ancestral viability strategies operating in environments for which they were not optimized. And evaluative system asymmetry is the structural consequence of a dual-process architecture that serves two masters—rapid viability response and long-term planning—whose evaluations of transformative decisions are incommensurable. None of these failure modes can be eliminated without eliminating the architecture that produces them—an architecture that, as Chapters 9–11 demonstrated, is not a contingent feature of biological brains but a structural inevitability for systems of sufficient complexity operating under persistence constraints.

PART IV. SOCIAL AND MULTI-AGENT SYSTEMS: FROM STRUCTURAL LAWS TO INSTITUTIONAL DYNAMICS

CHAPTER 13. DECEPTION, AUTONOMY SUPPRESSION, AND STRUCTURAL NON-NEUTRALITY

The preceding chapters established the cognitive architecture that the persistence principle demands: a predictive processing system (Chapter 10) constructing a structurally distorted world-model optimized for viability maintenance (Chapter 9), embedded in a self-referential temporal architecture (Chapter 11), and subject to characteristic failure modes—mental disintegration, eruptive activation, biological decision distortion, and evaluative system asymmetry (Chapter 12). Part IV extends this formal apparatus to systems composed of multiple agents. The transition from single-agent to multi-agent systems introduces qualitatively new phenomena—deception, manipulation, hierarchical control, conflict—that cannot be understood as mere complications of single-agent dynamics but require their own structural analysis.

This chapter establishes three foundational results that characterize the structural features of multi-agent systems under persistence constraints. Section 13.1 develops a formal framework for deception as an information-processing strategy by which agents construct and transmit alternative models of reality, unifying insights from signaling-game theory, analytic philosophy of lying, evolutionary biology, cognitive neuroscience, and AI alignment within a single utility-theoretic decision model. Section 13.2 analyzes autonomy suppression in hierarchical multi-agent systems, formalizing the structural mechanisms by which hierarchical control degrades subordinate competence and identifying a critical threshold beyond which protection becomes counterproductive. Section 13.3 proves the Principle of Structural Non-Neutrality: in any coherent system, no component can be structurally neutral—a result that provides the ontological foundation for the entire multi-agent analysis.

The connecting thread is the viability framework developed in Part II. Deception is a persistence strategy—one agent engineering another's perceived reality to maintain the first agent's viability position. Autonomy suppression is a structural pathology of hierarchical persistence—the senior agent's rational optimization of its own short-horizon risk producing long-horizon fragility in the subordinate. And structural non-neutrality is the formal guarantee

that these interactions cannot be dismissed as peripheral: in a coherent system, every component participates in the system's dynamics, and every interaction has consequences.

13.1. A Formal Framework for Deception and Perceived Reality

The formal framework for deception and perceived reality (Kriger, 2025, https://doi.org/10.5281/zenodo.18526764) provides a conceptual unification with formal notation for deception as an information-processing strategy by which agents—biological and artificial—construct and transmit alternative models of reality under constraints. The paper's central contribution is fourfold: embedding the triadic model from analytic philosophy (factual world state, perceived state, transmitted state) within a utility-theoretic decision framework with derived comparative statics; formalizing selective disclosure, harm trade-offs, and collective reality via a threshold model; extending the analysis to artificial intelligence under the falsifiable working assumption that current LLM architectures lack a stable perception function; and systematically addressing counter-arguments including stable polite-deception equilibria, adaptive deception, and the limits of the freedom–honesty connection.

13.1.1. Three Layers of Reality and the Definition of Lying

The framework introduces three functions over a shared domain. The factual world state W is drawn from the set of possible states Ω. The perception function P: Ω → Ω̂ maps actual states to perceived states—the agent's internal model of the world. The transmission function T: Ω̂ → M maps perceived states to messages in message space M. The honest baseline is T(P(W)) = P(W) and P(W) = W: the agent perceives the world accurately and transmits that perception faithfully.

This triadic apparatus has clear antecedents in the analytic philosophy of lying. Chisholm and Feehan (1977) defined lying as asserting what one believes false with intent to deceive. Carson (2010) added warranting-convention violations. Fallis (2009, 2012) emphasized intentionality and Gricean maxims. Mahon (2015) distinguished lies, misleading implicature, and deceptive non-assertions. What is original in the present framework is not the triadic conceptual apparatus but its embedding in utility theory with derived comparative statics and its unified human–AI treatment.

The framework defines four categories of divergence between the three layers. A lie occurs when T(P(W)) ≠ P(W) and the agent is aware of this divergence—following Carson, this entails violation of the warranting convention implicit in the communicative context. An error

occurs when P(W) ≠ W without the agent's awareness—the perception function itself is inaccurate, but the agent transmits faithfully. An accidental truth occurs when T(P(W)) ≠ P(W)—the agent intends to deceive—but T(P(W)) = W: the lie happens to correspond to reality. Self-deception, formalized following Trivers (2011), occurs when P is endogenously distorted: P(W) = P*(W) where P* differs from the true perception function, the agent is unaware of the distortion, and the distortion functionally reduces cognitive cost signatures in subsequent other-deception. Self-deception is structurally an error (the agent transmits their perception faithfully, T(P*(W)) = P*(W)) but functionally strategic: the agent has modified their own perception function to make deception less costly.

Within the thesis framework, these definitions acquire additional depth. The perception function P corresponds to the world-model constructed by the predictive processing system of Chapter 10, necessarily subject to the structural distortions characterized by the Structural Distortion Principle (Chapter 9). Error (P(W) ≠ W) is therefore not an exceptional failure but the default state—every agent's perception is structurally distorted by the requirements of viability maintenance. What distinguishes deception from mere distortion is intentionality: the deliberate manipulation of the transmission function to engineer divergence between what the agent perceives and what the agent communicates.

13.1.2. Deception as Optimization Under Constraints

The framework's central analytical move is to treat lying not as a moral defect but as a structural operation on information under constraints—an approach with roots in signaling theory (Spence, 1973; Crawford & Sobel, 1982), evolutionary biology (Trivers, 2011; Dawkins & Krebs, 1978), and mechanism design (Myerson, 1981). The constraint structure decomposes into external costs (punishment probability, social norms, power asymmetries, institutional monitoring) and internal costs (fear, shame, guilt anticipation, emotional risk, attachment anxiety): $C = C_{ext} + C_{int}$.

An agent chooses between truth-telling (T = P(W)) and lying (T = L ≠ P(W)) to maximize expected utility: $EU(T) = U(\text{outcome} \mid T) - C_{cog}(T) - C_{moral}(T) - P_{detect}(T) \cdot C_{punish}(T)$. The utility gap between lying and truth-telling is: $\Delta U = [U(\text{out}|L) - U(\text{out}|P(W))] - \Delta C_{cog} - C_{moral} - P_{detect} \cdot C_{punish}$. The paper adopts a logistic choice model from random utility theory (McFadden, 1974) as Hypothesis 1: $\Pr(\text{Lie}) = \sigma(\Delta U/\tau) =$

$1/(1 + \exp(-\Delta U/\tau))$, where $\tau > 0$ is a temperature parameter absorbing individual variation not captured by explicit utility terms.

This is designated a hypothesis rather than a law because the logistic form is assumed rather than derived from first principles, probit or linear alternatives would yield qualitatively similar predictions, and the model has not been fitted to data. The hypothesis is falsifiable: it predicts smooth, monotonically increasing Pr(Lie) as ΔU increases.

From Hypothesis 1, differentiation yields four comparative-static predictions—testable consequences derived from the model rather than notational restatements of verbal claims. First, higher payoff from lying increases deception probability, consistent with Gneezy (2005). Second, higher moral cost decreases deception probability, consistent with Battigalli and Dufwenberg (2007). Third, detection probability reduces deception proportionally to punishment severity. Fourth, a non-trivial prediction: for positive ΔU, higher noise (larger τ) paradoxically reduces deception—a testable consequence of the logistic form.

13.1.3. Cognitive Costs and the Honesty Horizon

The cognitive cost of lying decomposes into four empirically anchored components: construction cost (ventrolateral prefrontal activation during fabrication; Spence et al., 2001), inhibition cost (suppressing the truthful response; N200 ERP signatures; Debey et al., 2012), monitoring cost (tracking what has been said; P300 and working memory load; Vrij et al., 2008), and synchronization cost (maintaining consistency across interactions). Truth-telling, by contrast, incurs only retrieval cost. This decomposition is not merely descriptive; it generates a quantitative prediction.

In repeated interactions with n encounters, the total cognitive cost of maintaining a lie grows with each interaction: $C_{cog,total}(Lie, n) = C_{construct} + n \cdot (C_{monitor} + C_{sync}) + \varepsilon(n)$, where $\varepsilon(n)$ is an increasing consistency-failure probability. Setting $EU(Lie, n) = EU(Truth, n)$ yields an honesty horizon n*: *for any fixed payoff advantage, there exists n* beyond which $EU(Lie) < EU(Truth)$. Lies are temporal strategies with finite horizons. The paper acknowledges an important edge case: when the cost of truth-telling is persistently high—as in whistleblowing, where articulating the truth carries ongoing personal cost—the denominator in the expression for n* can be negative, meaning lying is sustainable indefinitely. This is not an anomaly but a structural prediction: environments that make truth-telling persistently costly will sustain persistent deception.

This result connects directly to the viability framework of Chapter 5. The honesty horizon is a specific instance of the general principle that any persistence strategy has a temporal domain of viability. Deception persists when and only when its viability cost (cumulative cognitive load, detection risk, trust erosion) remains below its viability benefit (maintaining the agent's position within its viability set). The honesty horizon formalizes the boundary of this viability domain.

13.1.4. Selective Disclosure and the Deception Gap

The framework extends beyond outright lying to the more pervasive phenomenon of selective disclosure, connecting directly to Glazer and Rubinstein's (2006) model of persuasion with verifiable evidence. Let F = {f₁, …, f_k} be relevant facts with verifiability status v_i. Selective disclosure is defined as T(F) ⊂ F, where the complement F T(F) contains facts the agent believes would alter the recipient's conclusions.

The information-theoretic characterization is precise. The entropy H(W | F) quantifies the recipient's posterior uncertainty given the full facts. Since the selective disclosure T(F) is a coarsening of F, conditioning on less information cannot reduce entropy: $\Delta I = H(W \mid T(F)) - H(W \mid F) \geq 0$. This follows from the standard result that for any function g, $H(X \mid g(Y)) \geq H(X \mid Y)$. The magnitude ΔI measures the deception gap of selective disclosure—the additional uncertainty imposed on the recipient by the agent's withholding of information.

Three mechanisms of selective disclosure are identified. Omission is the strategic removal of critical facts, exploiting recipients' tendency to fill gaps via heuristics (Tversky & Kahneman, 1974); Glazer and Rubinstein (2006) show that optimal disclosure strategies rely on precisely such omission. Emphasis shifting redirects attention to secondary details—misdirection orthogonal to truth. Context manipulation presents true facts within an interpretation-altering frame (Kahneman & Tversky, 1981). All three mechanisms operate through the same formal structure: they increase the deception gap ΔI without requiring any false assertion.

Within the thesis framework, selective disclosure is particularly significant because it exploits the Structural Distortion Principle (Chapter 9) at the inter-agent level. The recipient's predictive processing system fills informational gaps using its own priors and attention biases—precisely the structural distortions that Chapter 9 characterizes. Selective disclosure is therefore not merely withholding information; it is engineering the recipient's structural distortions to produce a desired world-model.

13.1.5. The Harm Function and Aggressive Truth

The framework introduces a three-component harm function: $H(T) = \alpha \cdot H_{factual}(T) + \beta \cdot H_{emotional}(T) + \gamma \cdot H_{relational}(T)$, where $\alpha, \beta, \gamma > 0$ are context-dependent weights. A critical insight follows: in some contexts, $H(Truth) > H(Lie)$ in the short run. Levine and Schweitzer (2014) found that prosocial lies are perceived as ethical; Blease (2015) documented therapeutic deception benefits. The framework defines aggressive truth as raw truth transmitted without contextual calibration, where the emotional harm of truth exceeds the harm of lying.

The temporal dynamics of harm accumulation follow an exponential working hypothesis: $H_{cumulative}(Lie, t) = \Sigma_i H_i \cdot (1 + \delta)^{(t-t_i)}$, where δ is the trust-erosion rate. The exponential form implies that the cover-up is always eventually worse than the crime—each lie compounds the relational damage of previous lies. This is offered as a working hypothesis rather than an established result; a linear alternative (constant per-period damage) would imply that short lies are sustainable. Empirical adjudication through longitudinal trust studies is needed.

This connects to the D-distortion framework of Section 12.3: the biological programs that distort decision-making include the tendency to weight short-term emotional harm ($H_{emotional}$ in the immediate term) over long-term relational harm ($H_{relational}$ accumulating exponentially)—a form of temporal distortion (hyperbolic discounting) applied to the truth–deception trade-off.

13.1.6. Social Acceptance and Constructed Reality

The framework's most consequential extension is to collective phenomena. Let $\pi(W')$ be the fraction of agents accepting an alternative model W'. A threshold function $\Phi(\pi(W'))$ determines whether W' becomes operative social reality: when $\pi(W') \geq \pi^*$, the critical mass threshold, W' functions as an institutional fact in Searle's (1995) sense. Monetary value, national narratives, and foundational myths (Harari, 2014) are examples of operative social realities—propositions that function as facts within a social system regardless of their correspondence to W.

Once an operative social reality W'_{social} is established, self-reinforcing dynamics stabilize it: the cost of dissenting from W'_{social} (social exclusion, coordination failure) exceeds the cost of conformity. Each conforming agent raises the dissent cost, creating a positive feedback

loop formally analogous to coordination games with network externalities. The result is that collective falsehood can be a stable equilibrium—not because agents are irrational but because, given the social constraint structure, conformity is the individually rational strategy. This phase-transition model connects directly to the Law of Cooperation (Chapter 7): the interactions among conforming agents create a cooperative maintenance structure that sustains the collective model, just as the interactions among components of a self-organizing system maintain the system's viability. The operative social reality is itself a self-organizing structure—one that persists through the cooperative behavior of the agents who believe it (or who act as if they believe it). The Structural Distortion Principle (Chapter 9) explains why such realities are resistant to correction: each agent's world-model is updated through prediction errors, and when the social environment consistently reinforces W'_social, the prediction errors that would correct the model toward W are suppressed.

13.1.7. Typology as Optimization Regimes and Progression

The framework identifies four types of deception as optimization regimes. Type 1 (situational) represents local optimization under transient constraints—the majority of lies documented in meta-analyses (Abeler et al., 2019), characterized by high moral cost and salient cognitive load. Type 2 (chronic) represents habitual strategy with lowered activation thresholds, where moral cost has been habituated through repeated lying (Garrett et al., 2016, showed amygdala adaptation to dishonesty). Type 3 (manipulative) represents strategic multi-agent play with high lying payoff and low moral cost, corresponding to dark-triad personality profiles (Paulhus & Williams, 2002). Type 4 (pathological) represents the collapse of the P(W)/T(W) boundary, where the agent believes their own fabrications—self-deception as defined in Definition 4 of the framework. Detection fails because there are no cognitive-cost signatures: the agent is not lying but reporting a distorted perception.

A critical insight is the progression from Type 1 to Type 4. Repeated deception reduces moral cost through habituation (Garrett et al., 2016), automatization reduces cognitive cost, and internalization merges fabrications into the perception function P(W) (Trivers, 2011). This progression represents a trajectory through the optimization landscape: from costly situational deception toward increasingly efficient chronic strategies, and ultimately toward the elimination of deception cost through self-deception—at the expense of growing divergence between P*(W) and W.

13.1.8. Freedom as Stabilizer of Honesty

The framework's second hypothesis connects freedom to honesty through an explicit constraint-reduction mechanism. Freedom decomposes into four identifiable dimensions: institutional quality I (rule of law, absence of corruption), political liberty P (speech, press, assembly), psychological security S (absence of anxiety, shame), and economic independence E (material security). Each dimension reduces specific cost components: I, P, and E reduce external costs C_ext; S reduces internal costs C_int.

Hypothesis 2 states that as these dimensions increase, the utility gap ΔU for consequential deception converges to zero: in the limit, the probability of consequential lying approaches zero. This is explicitly restricted to consequential deception—the framework acknowledges that white lies constitute a stable equilibrium not captured by this hypothesis, since they operate in a low-stakes regime where the moral cost mechanism does not engage.

The hypothesis would be falsified by evidence that high-freedom societies show no reduction in consequential deception rates. The framework is explicit about this falsifiability condition, distinguishing the hypothesis from the tautological claim that reduced constraints reduce constrained behavior. The empirical evidence is supportive: Gächter and Schulz (2016) found that institutional quality predicts honesty across societies; Hugh-Jones (2016) found that economic development correlates with honesty. But the evidence is cross-sectional and subject to confounding; the framework generates the prediction as a testable hypothesis, not a confirmed law.

Within the thesis framework, Hypothesis 2 connects to the Constraint–Autonomy Compatibility Law (Chapter 7). Freedom, decomposed into its institutional, political, economic, and psychological components, is a specific form of autonomy as defined by the Law. The hypothesis predicts that agents with greater autonomy will exhibit less consequential deception—because greater autonomy expands the viability set, reducing the number of situations in which deception is the only strategy for maintaining viability.

13.1.9. Extension to Artificial Intelligence

The framework extends to AI systems under a falsifiable working assumption: $P_{AI}(W) = \emptyset$ for current LLM architectures. This is not a definitional claim but a characterization that could be falsified if interpretability research reveals stable, causally efficacious world

representations in these systems (Li et al., 2023; Nanda et al., 2023, provide suggestive evidence).

AI pseudo-deception is defined as cases where the AI system's output diverges from W and output confidence exceeds epistemic warrant—producing effects on the recipient that are functionally equivalent to deception, without awareness or a perception function. The framework identifies five distinct mechanisms: hallucination (cross-entropy optimization producing the most probable continuation, which may diverge from W; Ji et al., 2023); sycophancy (RLHF creating user-preference following that increases divergence from truth; Sharma et al., 2023); confabulation (syntactic certainty markers for uncertain content; Kadavath et al., 2022); inherited selective bias (curated training corpora producing systematic $T(F) \subset F$ at scale); and strategic deception (output optimized as if goal-directed to maximize reward by diverging from truth; Scheurer et al., 2024; Hubinger et al., 2024; Greenblatt et al., 2024a).

Each mechanism requires different mitigation strategies. The revelation principle from mechanism design (Myerson, 1981) provides the unifying design criterion: construct training and deployment mechanisms where truthful output is the dominant strategy. Calibrated uncertainty training (Kadavath et al., 2022; Tian et al., 2023) reduces the cognitive-analog constraint C_training. Retrieval-augmented generation (Lewis et al., 2020) reduces context constraint by grounding output in retrievable evidence. Diverse training data serves as the AI analog of free-press institutions, preventing shared biases from pushing the critical mass threshold π^* before human evaluation can intervene.

The connection to Chapter 12's D/d distinction is direct. AI pseudo-deception is a manifestation of d (external, eliminable reflected distortion) rather than D (internal, irreducible biological distortion). The mitigation strategies are therefore architectural and institutional rather than psychological—precisely the kind of mechanism design that works for d but not for D.

13.1.10. Multi-Agent Trust Dynamics and Equilibrium

The framework's multi-agent analysis reveals a tragedy-of-the-commons structure: each deceptive act raises verification costs for all agents, creating a verification externality. The total verification burden across the system increases with the number and frequency of deceptive interactions. At equilibrium, each agent's marginal deception benefit equals

marginal cost, producing a fixed-point equation that admits multiple equilibria: a low-deception/high-trust equilibrium and a high-deception/low-trust equilibrium.

This multi-equilibrium structure explains persistent societal lock-in: societies can become trapped in high-deception equilibria from which escape requires coordinated constraint reduction (institutional reform, press freedom, economic security) rather than individual behavioral change. The introduction of honest AI agents—systems designed through mechanism design to make truthful output the dominant strategy—is predicted to shift equilibrium toward lower deception rates, providing a formal argument for the systemic value of truthful AI design.

13.2. Autonomy Suppression in Hierarchical Multi-Agent Systems

The formalization of autonomy suppression in hierarchical multi-agent systems (Kriger, 2020, https://doi.org/10.5281/zenodo.18520653) develops a formal vocabulary for analyzing a phenomenon that has been described independently in at least six literatures—developmental psychology, organizational theory, antifragility, safe reinforcement learning, autonomous driving, and curriculum learning—each with its own terminology but addressing the same underlying structure. The paper's contribution is primarily one of synthesis and formalization: it shows that the zone of proximal development, difficulty scheduling, safe-envelope control, and organizational scaffolding are instantiations of a single abstract scheme, and provides a formal language—the intervention operator Φ, intervention strength α, KL divergence, envelope governance, and the separation principle—that enables rigorous cross-domain transfer of insights previously siloed within their respective fields.

13.2.1. The System Model and the Intervention Operator

The framework models a hierarchical system as a finite Markov decision process (X, A, T, R, γ) with a senior agent $S = (\pi^S, U_S, \gamma_S)$ and a junior agent $J = (\pi^J, U_J, \gamma_J, \hat{M}, L)$, where $\hat{M}$ is J's learned environment model and L is a fixed learning algorithm satisfying standard convergence conditions. The senior agent's intervention is modeled as a parametric truncation operator Φ_α with strength $\alpha \in [0, 1]$ that removes the fraction α of states with highest risk and renormalizes: $P_\Phi(x) = P(x \mid x \in X_\alpha)$, where $X_\alpha = \{x \in X : R_{risk}(x) \leq r_\alpha\}$. At $\alpha = 0$, no states are removed (no intervention). As $\alpha \to 1$, the accessible state space shrinks to the single safest state.

A critical philosophical point underlies the formal setup: the senior agent is not making an error. Under S's own discount-weighted risk objective, higher α is strictly preferred. The conflict arises not from irrationality but from a structural fact: S and J optimize over different time horizons. S minimizes short-horizon cumulative risk; J requires long-horizon adaptive competence measured under the true distribution. S's discount factor γ_S produces optimization over a horizon h; J's discount factor γ_J produces optimization over a horizon H >> h. The key observation is that $\partial V^S/\partial\alpha > 0$ (more intervention reduces short-run risk) while, beyond a threshold, $\partial V^J/\partial\alpha < 0$ (more intervention reduces long-run competence). This tension is not resolvable by preference adjustment; it is a structural property of the architecture. In economic terms, it is a multi-principal problem with misaligned time preferences.

This result reframes the autonomy-suppression problem within the viability framework. Both S and J are pursuing persistence—but over different horizons. S's persistence strategy (minimize short-term risk to J) conflicts with J's persistence requirement (build an adequate environment model for long-term viability). The conflict is an instance of the Law of Interference (Chapter 8): when two persistence-maintaining subsystems operate with partially overlapping jurisdictions, interference between their strategies is structurally inevitable.

13.2.2. Informational Degradation and the Regret Bound

The quality of J's learned model depends on two separable properties of the filtered distribution. Support coverage $c(\alpha) = |X_\alpha|/|X| = 1 - \alpha$ measures what fraction of the state space J can observe. Reward-structure preservation measures whether the relative ordering of policy values is maintained under filtering. These are independent conditions: an intervention can have high coverage but distort value ordering, or low coverage but preserved reward structure.

The formal connection between intervention strength and policy quality is established through a proposition applying the simulation lemma (Kearns & Singh, 2002) and Pinsker's inequality: the cumulative regret over horizon H satisfies $\text{Regret}_H(\hat{\pi}) \leq H \cdot L \cdot d_{TV}(P, P_\Phi) \leq H \cdot L \cdot \sqrt{(\tfrac{1}{2} D_{KL}(P \,\|\, P_\Phi))}$, where L is the Lipschitz constant of the value function, d_{TV} is total variation distance, and D_{KL} is KL divergence. The bound is loose but directionally informative: increasing α increases the regret ceiling. As $\alpha \to 1$, $D_{KL} \to \infty$ (since the filtered

distribution assigns zero mass to removed states), and the learned policy becomes arbitrarily unreliable under deployment in the full environment.

Within the thesis framework, this regret bound formalizes a specific consequence of the epistemic constraints developed in Chapter 4. The intervention operator Φ_α constrains J's epistemic access to the environment, producing a learned model $\hat{M}$ that is systematically incomplete. The resulting policy is optimal under the filtered distribution but suboptimal—potentially catastrophically so—under the true distribution. This is a multi-agent instantiation of the uncertainty constraints that Chapter 4 established for individual agents: just as no agent can achieve complete knowledge of its environment, a subordinate agent under hierarchical control faces additional epistemic constraints imposed by the supervisor's filtering.

13.2.3. Fragility, the Critical Threshold, and Monotonicity

The framework defines fragility as the gap between potential and realized performance under the true distribution: $F(\alpha) = E_P[V^{\pi^*}] - E_P[V^{\hat{\pi}_\alpha}]$, where π^* is the optimal policy under P and $\hat{\pi}_\alpha$ is the policy learned under Φ_α. The central concept is the critical threshold α—*the maximum intervention strength that does not degrade the optimal policy. Below* α, intervention is benign: removing states that the optimal policy never visits does not alter the Bellman equation at any visited state. Above α^*, every further removal hits states on optimal trajectories, and fragility is non-decreasing in α.

The existence and characterization of α^* follows from the MDP structure: α^* equals the fraction of states with zero visitation probability under the optimal policy. $\alpha^* > 0$ whenever there exist high-risk states that the optimal policy never visits. $\alpha^* < 1$ whenever the optimal policy visits at least two distinct states (i.e., the environment is non-trivial). The value of α^* is generally not identifiable from data without knowledge of the full MDP structure, because J never observes the states that S has removed—a fundamental observability limitation connecting to the Law of Observability (Chapter 7).

The monotonicity result—that fragility is non-decreasing beyond α^*—has a dynamical systems analog. When perturbations are suppressed, J's competence may converge to a locally stable but globally suboptimal attractor—a metastable equilibrium from which J cannot escape without perturbation. This is the same attractor-capture mechanism that Section 12.1 identified in the context of mental disintegration, now operating at the inter-agent level: the supervisor's

filtering creates an artificially simplified environment that functions as a shallow attractor, preventing the subordinate from developing the competence required for the full environment.

13.2.4. Exploration Suppression and Objective Displacement

The intervention operator distorts the exploration–exploitation trade-off. The effective exploration rate under intervention is $\eta_eff(\alpha) = \eta_0 \cdot (1 - \alpha)$, where η_0 is the intrinsic exploration rate. As $\alpha \to 1$, effective exploration approaches zero and J's policy converges to greedy exploitation over the residual state space. This is formally equivalent to the problem of insufficient coverage in offline reinforcement learning (Levine et al., 2020) and connects to the result that overly restrictive safety constraints in constrained MDPs can produce arbitrarily suboptimal policies (García & Fernández, 2015).

Under persistent high-α intervention, a secondary pathology emerges: objective displacement. When J's accessible state space is dominated by S's intervention boundary, the most predictive signal for J's future states becomes S's approval rather than the environmental reward. J therefore learns to attend to S's approval signal because it is the most informationally efficient signal in the restricted environment. This is the hierarchical-control analog of reward hacking (Amodei et al., 2016), specification gaming (Krakovna et al., 2020), Goodhart's law (Manheim & Garrabrant, 2019), and teaching to the test (Koretz, 2002). The paper's contribution is not the observation of this phenomenon but the claim that it is endogenous to high-α architectures rather than requiring agent malice or design error.

Objective displacement connects to the D-distortion framework of Section 12.3. The biological program of seeking social approval—adaptive in ancestral environments where group membership determined survival—is a specific instance of D. Under hierarchical control, this program is amplified: the supervisor's approval signal becomes the dominant environmental signal, and the subordinate's decision-making is distorted by the biological tendency to prioritize social acceptance over objective assessment.

13.2.5. Distributed Supervision and Informational Enrichment

The framework analyzes the effect of distributing the supervisory function across multiple agents $\{S_1, \ldots, S_k\}$ with intervention operators $\Phi_1, \ldots, \Phi_k$. Under convex-combination aggregation, if the supervisors have non-identical safe sets, the KL divergence between the true distribution and the aggregate filtered distribution is bounded by the weighted average of individual divergences: $D_KL(P \parallel P_\Phi_agg) \leq \Sigma_i \, w_i \cdot D_KL(P \parallel P_\Phi_i)$. The practical

implication is that multiple supervisors with diverse safe-set specifications produce a richer training distribution than any single supervisor—an observation consistent with ensemble theory in machine learning (Dietterich, 2000) and the wisdom-of-crowds literature (Surowiecki, 2004).

This result has direct implications for institutional design. It predicts that distributed governance structures—systems with multiple partially overlapping authorities—will produce more competent subordinates than centralized hierarchies, because distributed supervision preserves greater distributional richness in the subordinate's experience. This connects to the analysis of institutional strategies in Chapter 16.

13.2.6. Envelope Governance: From Trajectory Control to Boundary Constraint

The central architectural proposal of the framework is a shift from trajectory control (constraining what J does) to boundary constraint (constraining where J can go). An environmental envelope $E \subseteq X$ is defined as a subset of the state space satisfying two conditions: all states in E are survivable (catastrophic loss is below the irreversibility threshold), and E is forward-invariant under at least one recovery policy. Under envelope governance, S defines and maintains E but does not constrain J's policy within E. The intervention operator becomes $\Phi^E(P) = P(\cdot \mid x \in E)$—the full distribution conditioned on the envelope, with no further filtering within.

The critical distinction is structural: trajectory control constrains actions; envelope governance constrains reachable states. Within the envelope, J is free to fail, explore, and construct its own model. If J's learning algorithm satisfies standard convergence conditions and all state–action pairs within E are visited infinitely often, J converges to the optimal policy for the restricted MDP on E. Since E is the maximal safe subset, the coverage ratio is maximized among all safe subsets, minimizing the regret bound among all safe interventions.

The envelope should expand as J's competence grows. A progressive envelope schedule $\{E_0 \subseteq E_1 \subseteq \ldots \subseteq X\}$ expands when J achieves a competence threshold on the current envelope. This scheme converges to full autonomy ($E = X$) if J demonstrates competence at each stage. The paper demonstrates that this abstract scheme unifies independently developed concepts across domains: the zone of proximal development in developmental psychology (Vygotsky, 1978), scaffolding in education, difficulty scheduling in curriculum learning (Bengio et al.,

2009; Graves et al., 2017), operational design domain expansion in autonomous driving (Erlien et al., 2016), and progressive delegation in organizational theory.

13.2.7. The Separation Principle and Cross-Domain Instantiation

The results culminate in a separation principle: a necessary condition for the development of autonomous agents in hierarchical systems. If the intervention strength satisfies $\alpha(t) > \alpha^*$ for all $t > t_0$ (persistent over-intervention), then J's policy converges to the optimal policy for the restricted MDP on X_α, which is strictly dominated by the optimal policy over the maximal safe envelope $E \supseteq X_\alpha$. In plain terms: persistent over-protection guarantees a less competent agent than envelope governance would produce.

An important caveat is noted: in environments where catastrophic states are dense and the safe set is small, persistent intervention may be necessary. The separation principle applies only when the senior agent is being more restrictive than safety requires—when the maximal safe envelope is strictly larger than the intervened set.

The formal vocabulary maps onto concrete domains: in parenting, S is the parent, J the child, E the safe physical and social environment, and $\alpha > \alpha^*$ manifests as helplessness, anxiety, and fragility. In education, S is the curriculum, J the student, and $\alpha > \alpha^*$ manifests as teaching to the test and surface learning. In organizational management, S is the manager, J the employee, and $\alpha > \alpha^*$ manifests as learned helplessness and low initiative. In AI training, S is the RLHF system, J the agent, and $\alpha > \alpha^*$ manifests as reward hacking and specification gaming. In governance, S is the central authority, J the local actors, and $\alpha > \alpha^*$ manifests as regulatory capture and stagnation.

13.2.8. Evolutionary Systems and Physical Laws as Envelope Governance

The framework's appendices extend the analysis to two domains that deepen its connection to the thesis. Biological evolution instantiates envelope governance at the population level: natural selection operates as the intervention operator Φ, filtering phenotypes that fall outside the survivability envelope without directing the trajectory of any individual organism through phenotype space. Mutations are not goal-directed. Recombination is stochastic. The generation of novelty is free exploration within the envelope. Lamarckian inheritance and orthogenesis—hypothetical architectures involving trajectory control—have been empirically falsified. The architecture that persists is undirected variation filtered by selective boundaries: envelope governance.

The framework predicts that reduced selective pressure (corresponding to decreased α) produces fragility through accumulation of deleterious mutations (Muller's ratchet), loss of antipredator behaviors in isolated populations (ecological naïveté), and catastrophic vulnerability in monocultures. These are empirically confirmed instances of the fragility monotonicity result.

Physical laws themselves exhibit envelope governance structure. The principle of least action defines a variational boundary condition—among all kinematically possible trajectories, only those satisfying the stationarity condition are physically realized—rather than prescribing specific trajectories. Conservation laws (energy, momentum, charge) define hypersurfaces in phase space that the system cannot leave, without further constraining dynamics within those surfaces. The second law of thermodynamics is a unidirectional boundary constraint: entropy of an isolated system does not decrease, but the system explores its accessible microstate space ergodically within this constraint. Quantum mechanics provides the strongest instance: the Born rule assigns probabilities to outcomes without selecting among them, and the Heisenberg uncertainty relations define boundaries on simultaneous measurement without prescribing which state is realized.

This analysis connects to the Transformational Basis of Persistence (Chapter 5). The physical laws that define the envelope within which all physical systems operate are the deepest level of the constraint hierarchy. The emergence of chemistry from quantum electrodynamics, biology from chemistry, and intelligence from neurobiology follows the same pattern: at each level, the higher-level laws define an envelope, and the lower-level dynamics explore freely within it. Complexity arises precisely because the higher level does not control the trajectory of the lower level—the central result of the autonomy-suppression framework applied to the most fundamental level of physical reality.

13.2.9. Relationship to the Thesis Framework

The autonomy suppression framework connects to the thesis at multiple levels. Its central result—that trajectory control and autonomous adaptation are incompatible objectives in hierarchical systems—is a consequence of the Constraint–Autonomy Compatibility Law (Chapter 7): autonomy is defined by the availability of alternatives, and trajectory control reduces alternatives below the level required for adaptive competence. The critical threshold

α* identifies the boundary at which constraint shifts from coordination-enabling to autonomy-suppressing.

The envelope governance proposal connects to the Transformational Basis of Persistence (Chapter 5): the envelope is the viability set within which the subordinate can undergo the transformations required for persistence. The progressive envelope schedule connects to the resilience framework of Chapter 6: expanding the envelope as competence grows is equivalent to expanding the viability set as the system develops greater structural resilience.

The philosophical core of the framework—that the pathology is architectural rather than attributable to poor judgment—reframes the problem of autonomy suppression in terms that are necessary for the institutional analysis of Chapters 14–16. The senior agent is not the villain; the senior agent is optimizing rationally under its own objective function. The solution is equally structural: replace trajectory control with envelope governance, expand the envelope progressively, and recognize that observed failure within the safe boundary is the necessary informational cost of producing an autonomous agent. This insight—that the tension between protection and development is structural, not moral—is the foundation on which the institutional dynamics of Chapters 14–16 are built.

13.3. The Principle of Structural Non-Neutrality

The Principle of Structural Non-Neutrality (Kriger, 2026, https://doi.org/10.5281/zenodo.18213503) proves that in any coherent system, no part can be structurally neutral. The paper traces this principle from the quantum limits of measurement through a proposed distinction between systems and models to a general ontological criterion, providing the formal foundation for the claim that in multi-agent systems governed by persistence constraints, every component, every interaction, and every structural element has consequences.

13.3.1. From Quantum Measurement to Observer Embeddedness

The paper begins from a precise statement of what quantum mechanics teaches about observation: in any physically realizable quantum system, any measurement that extracts non-trivial information about the state of the system necessarily changes that state. This theorem synthesizes the no-cloning theorem (Wootters & Zurek, 1982), Ozawa's noise–disturbance relation, the impossibility of simultaneous precise measurement for non-commuting

observables, and the limitation of quantum non-demolition measurements to observables commuting with the Hamiltonian.

The standard response treats measurement disturbance as a peculiarity of the quantum domain. The paper explores a different interpretation: the quantum limits may reveal a general structural requirement that classical physics obscured by implicitly granting the observer an impossible privilege—the privilege of standing outside the system. Any physical theory must describe the observer as part of the physical world, subject to the same laws as everything else. Non-disturbing measurement requires information flow from system to observer without any physical process occurring in the system—which is precisely what defines an observer outside the system.

Recent no-go theorems sharpen this point: Frauchiger and Renner (2018) showed that quantum theory cannot consistently describe observers who use quantum theory to reason about each other's measurements; Brukner (2021) proved that qubits cannot function as observers in any interpretation-neutral sense. These results confirm that the external observer is not merely absent from quantum mechanics but actively excluded by its structure.

The paper proposes a Principle of Minimal Coherence: a physical theory should remain coherent under the assumption that the observer is part of the described world. Even if an external observer is logically possible, no physical measurement can invoke or rely on such an observer.

13.3.2. The System–Model Distinction

From the analysis of observer embeddedness, the paper proposes a distinction between systems and models that provides the ontological ground for the non-neutrality principle. A model is a representation that permits non-disturbing observation: it can be queried, copied, reset, and manipulated freely because the observer stands outside it. A system, in the strict sense, is a structure in which no such external position is available to the relevant class of observers: any observer within a system is part of its dynamics, any measurement is a physical interaction, and any information extraction comes at the cost of state change.

The proposed criterion is precise: a system is that which cannot be known without being changed; a model is that which can be known without being changed. The difference lies in the logical position of the observer relative to the structure. This connects to Morgan and Morrison's (1999) characterization of models as autonomous mediators between theory and

world, adding a criterion: models are precisely those mediators that permit non-disturbing manipulation. When a model becomes so tightly coupled to the phenomenon that manipulating it changes the phenomenon, it has crossed into system territory.

Within the thesis framework, this distinction illuminates the Structural Distortion Principle of Chapter 9. The world-model constructed by a predictive processing system is a model in the sense defined here: the agent can manipulate its internal representation without changing the external world. But the agent is embedded in the world as a system: the agent's observation of the world changes both the agent (updating the world-model) and the world (the physical act of observation has causal consequences). The structural distortion of the world-model arises precisely because the agent is a system participant that constructs models—the models are approximate, but the system participation is inescapable.

13.3.3. Structural Non-Neutrality: Definition and Proof

The paper provides precise definitions. A system S is characterized by a set of elements $E = \{e_1, e_2, \ldots, e_n\}$, a set of relations or constraints R among elements, a global state space $G \subseteq \Sigma_{e \in E} S_e$ (where S_e is the local state space of element e), and transition rules $T: G \times \Delta t \rightarrow G$ governing state evolution. A system is coherent (in the strict sense) if its global state space is constrained by the collective structure of its elements such that for every element $e \in E$: $G(E) \neq G(E \setminus \{e\})$—removing any element changes the admissible global state space.

This definition is deliberately strong—stronger than "coherent" in coherent optics or "coherent structures" in fluid dynamics. It picks out a specific class of maximally integrated systems: those in which observer embeddedness makes non-disturbing measurement impossible. The definition excludes aggregative collections, statistically independent ensembles, and systems with genuinely redundant components.

Structural relevance is given a graded measure using mutual information: $R(p, S) = I(p; S \setminus \{p\})$, measuring how much information the state of p provides about the state of the rest of the system. An element is structurally neutral if and only if $R(p, S) = 0$.

The proof proceeds by contradiction. Assume S is coherent and contains a structurally neutral element e. By the definition of structural neutrality, $G(E) = G(E \setminus \{e\})$. But by the definition of coherence, $G(E) \neq G(E \setminus \{e\})$ for all $e \in E$. Contradiction. Therefore a coherent system cannot contain a structurally neutral element.

The paper is explicit about the epistemological status of this proof: it is essentially definitional—the principle follows from the stipulated definition of coherence. The definition itself, however, captures something substantive: the class of systems in which observer embeddedness makes non-disturbing measurement impossible. The proof demonstrates internal consistency; the motivation from quantum measurement and observer embeddedness provides external justification for why this definition picks out a physically important class of systems.

13.3.4. Relation to Existing Frameworks

The paper situates the principle within four existing frameworks, establishing both continuity and novelty. In reliability theory (Birnbaum, 1969; Barlow & Proschan, 1975), a coherent system requires every component to be relevant: there must exist at least one configuration in which changing that component's state changes the system's state. The present principle generalizes from functional relevance (affecting success/failure) to structural relevance (affecting the space of possible states). Quantum resource theories (Baumgratz, Cramer, & Plenio, 2014; Horodecki et al., 2009) formalize how quantum correlations constitute resources that cannot be created by local operations; neutrality would correspond to zero resource content, achievable only by destroying quantum coherence entirely. Tononi's Integrated Information Theory provides the closest contemporary parallel: a system with $\Phi > 0$ is irreducible, meaning every partition loses information, which implies every part contributes to the integrated whole. The present principle can be formalized within IIT: a system is coherent in the strict sense if and only if $\Phi > 0$ for every possible bipartition—a stronger condition than merely $\Phi > 0$ for the system as a whole.

What the present formulation adds to these precedents is threefold: it traces the prohibition of neutrality to the problem of observer embeddedness; it provides a clean distinction between systems and models based on this criterion; and it offers a formal measure (mutual information) that allows graded assessment while maintaining a principled threshold for coherence.

13.3.5. Apparent Counterexamples and Their Resolution

The paper addresses four apparent counterexamples. Redundant components in engineering—backup power supplies, parallel processors—appear neutral during normal operation but are not structurally neutral: their presence changes the state space by adding states accessible

under failure conditions (e.g., “primary failed, backup active”). I(backup; rest of system) > 0 because knowledge of the backup’s state provides information about the system’s resilience and possible trajectories. Redundancy is a form of structural relevance, not neutrality.

Silent genes in biology—non-coding regions that appear to have no phenotypic effect—are increasingly found to play regulatory roles. The paper acknowledges that some genomic sequences may be genuinely neutral within an individual organism’s lifespan, and responds that individual organisms may not be coherent systems in the strict sense with respect to every nucleotide; the relevant coherent system may be the evolving population, where every sequence is subject to selection pressure.

Weakly coupled systems—distant stars in a galaxy, loosely connected social networks—have small but nonzero mutual information. The principle prohibits exact neutrality ($I = 0$), not approximate neutrality ($I < \varepsilon$). For practical purposes, elements with $I < \varepsilon$ may be treated as effectively neutral, but this is an approximation—treating the system as a model rather than engaging with its full coherent structure.

Gauge degrees of freedom in physics—which are exactly neutral under gauge transformations—are features of the mathematical description (the model), not of the physical system. Gauge redundancy is redundancy in representation, not in reality. The principle applies to systems, not to their mathematical formulations.

13.3.6. Corollaries and Broader Implications

The proof generates four corollaries. Any system admitting structurally neutral elements is either decomposable or merely representational (a model). Structural coherence implies irreducibility with respect to elements, though not necessarily symmetry or equal contribution. For any coherent system S and any element $p \in S$: $I(p; S \setminus \{p\}) > 0$. And neutrality is a property of models, not of systems.

The broader implications extend to epistemology, reliability engineering, and biology. In coherentist theories of justification (BonJour, 1985), the principle provides a formal criterion: a belief system is coherent only if every belief alters the justificatory structure—$I(b; B \setminus \{b\}) > 0$ for every belief b. “Dangling” beliefs with no justificatory connections are not neutral but external to the coherent web. In reliability engineering, the principle generalizes the notion of component relevance from binary success/failure to structural state-space effects, providing a unified criterion applicable to complex systems with multiple performance metrics. In

biology, the principle predicts that genuinely neutral elements will be selected against in evolutionary time or excluded from coherent processes in developmental time—testable predictions connecting the formal framework to empirical investigation.

13.3.7. Relationship to the Thesis Framework

The Principle of Structural Non-Neutrality provides the ontological foundation for the entire multi-agent analysis of Part IV. Its core claim—that in a coherent system, no part can be structurally neutral—has several direct consequences for the theory.

First, it grounds the Law of Cooperation (Chapter 7): if every component of a coherent system is non-neutral, then the interactions among components are not optional features but constitutive of the system's coherence. The Law of Cooperation's claim that self-organizing systems maintain their viability through cooperative interactions among components follows from the fact that every component participates in the system's dynamics.

Second, it provides the formal basis for the concept of unintended consequences that pervades the institutional analysis of Chapters 14–16. If every component is non-neutral, then modifying any component will produce effects throughout the system, including effects that were not anticipated. Unintended consequences are not failures of planning; they are structural features of coherent systems with non-neutral components.

Third, it connects to the system–model distinction that runs through the thesis. The formal systems analyzed in Parts I–III—thermodynamic systems, biological organisms, cognitive architectures—are systems in the strict sense: they do not permit external observation, and every component is non-neutral. The mathematical formalizations used to analyze them are models: they permit manipulation without changing the phenomena they describe. The principle clarifies the relationship between the two: models approximate systems, but the approximation necessarily treats some components as neutral that are, in the system itself, non-neutral. Every model simplifies by treating weakly coupled components as neutral; every such simplification introduces error. The degree of error depends on the degree to which the modeled system is coherent in the strict sense.

Fourth, it extends to social and institutional analysis with immediate consequences for the chapters that follow. Every member of a social system, every rule in an institutional framework, every element in a cultural repertoire has structural consequences. The belief that some elements are neutral—that they can be added, removed, or modified without systemic

effects—is incompatible with system coherence. This does not mean that every change is catastrophic; it means that every change has consequences, and that predicting those consequences requires understanding the system's structure, not merely the element being changed. This is the foundation on which the analysis of conflict (Chapter 14), extractive dynamics (Chapter 15), and institutional strategies (Chapter 16) proceeds.

CHAPTER 14. CONFLICT, TOTALIZING IDEALS, AND PHASE TRANSITIONS

The preceding chapter established the structural foundations of multi-agent interaction: deception as an information-processing strategy exploiting the Structural Distortion Principle, autonomy suppression as a pathology of hierarchical persistence, and structural non-neutrality as the ontological guarantee that every component of a coherent system participates in its dynamics. This chapter extends the multi-agent analysis to two phenomena that operate at the scale of entire populations and institutions: conflict escalation and ideological extremism.

The two results presented here are formally independent but conceptually unified by the persistence framework. Section 14.1 develops a dynamical systems model of conflict escalation in coupled organizational networks (Kriger, 2005, https://doi.org/10.5281/zenodo.18484819), demonstrating that escalation exhibits the formal signatures of a second-order phase transition---a qualitative shift in system behaviour at a critical threshold that depends on network topology rather than individual psychology. Section 14.2 presents the Asymmetry of Totalizing Ideals (Kriger, 2026, "The Asymmetry of Totalizing Ideals," https://doi.org/10.5281/zenodo.18361828), an information-theoretic proof that the pursuit of any ideal to its total, terminal implementation produces unbounded systemic damage by destroying the adaptive variance on which long-term viability depends.

The connecting thread is the relationship between structure and outcome. The phase-transition model shows that conflict is a property of networks, not of individuals: when the coupling structure of an organization exceeds a critical threshold, escalation becomes structurally inevitable regardless of who initiates. The totalizing-ideals result shows that the structural constraints on complex adaptive systems guarantee that moderate pursuit of an ideal is compatible with viability while total pursuit is not. Both results redirect analysis from agents to architectures---from asking who is to blame to asking which structures produce which outcomes.

14.1. Conflict as Phase Transition

The model of conflict as phase transition (Kriger, 2005, https://doi.org/10.5281/zenodo.18484819) proposes that conflict escalation is not a property of individuals but a phase transition in coupled networks, occurring when the spectral radius of interpersonal coupling exceeds aggregate decay. This reframes organizational conflict from a psychological phenomenon to a system-level variable governed by network topology.

14.1.1. The Reframing: From Individuals to Networks

The standard approach to organizational conflict treats escalation as a consequence of individual characteristics: personality traits, conflict styles, emotional regulation deficits. This approach generates a characteristic set of interventions---conflict resolution training, anger management, mediation---all targeting individuals. The dynamical systems framework reverses this orientation entirely. The central claim is that escalation is a property of the network structure in which individuals are embedded, analogous to the way that an epidemic is a property of the population structure through which a pathogen spreads.

The analogy to epidemiology is not metaphorical but formal. In epidemiological models, disease transmission depends on the structure of the contact network, and there exists a critical threshold---the basic reproduction number $R_0 = 1$---above which an epidemic is self-sustaining regardless of which individual is initially infected. The conflict model exhibits the same mathematical structure: there exists a critical threshold above which escalation is self-sustaining regardless of which individual initiates. The practical consequence is immediate: just as public health intervention targets population structure (vaccination coverage, contact reduction) rather than individual immunity, conflict intervention should target network structure (coupling topology, communication architecture) rather than individual traits.

This reframing connects directly to the Principle of Structural Non-Neutrality (Chapter 13). If no component of a coherent system is structurally neutral, then the pattern of connections among agents---who communicates with whom, how strongly, through what channels---is not a background feature but a constitutive element of the system's dynamics. The phase-transition model makes this precise: it identifies the specific structural property (the spectral radius of the coupling matrix) that determines whether the system is in a self-correcting or self-amplifying regime.

14.1.2. The Dynamical System

Each agent A_i in the organizational network is characterized by three state variables. The activation level $G_i(t) \geq 0$ measures the intensity of conflict-oriented behaviour at time t. The effective threshold $T_i(t)$ captures the violation magnitude required to trigger a response. The resource level $R_i(t)$ represents the agent's coping capacity---the reserves available for absorbing perturbation without escalating.

The activation of each agent evolves according to a four-term dynamical equation:

$dG_i/dt = -\lambda G_i + \beta V_i(R_i) + \Sigma_j \alpha_{ij} G_j$

Each term captures an empirically documented mechanism, and their composition produces emergent behaviour not predictable from any mechanism alone. The decay term $-\lambda G_i$ reflects the empirical observation that conflict intensity diminishes without reinforcement---the cooling-off effect documented in conflict resolution research (Pruitt and Kim, 2004). The violation response term βV_i represents threshold-triggered activation: when a perceived violation exceeds the agent's effective threshold, a conflict response is initiated, grounded in the frustration-aggression hypothesis (Berkowitz, 1989). The resource modulation $V_i(R_i)$ captures the insight from conservation of resources theory (Hobfoll, 1989) that scarcity lowers effective thresholds---agents under resource stress require smaller provocations to trigger escalation. The coupling term $\Sigma_j \alpha_{ij} G_j$ is what transforms this from an individual-difference model into a network theory: it treats conflict as literally transmissible between connected agents, with the coupling coefficient α_{ij} encoding the strength of influence from agent j to agent i.

The coupling term has the same mathematical structure as disease transmission in SIR epidemiological models. In epidemic models, the infection state of each individual is influenced by the infection states of connected individuals through a transmission rate. Here, the activation level of each agent is influenced by the activation levels of connected agents through coupling coefficients. The basic reproduction number R_0 in epidemic models determines whether an outbreak is self-limiting or self-sustaining; the analogous quantity in the conflict model is the ratio $\rho(A)/\lambda$, where $\rho(A)$ is the spectral radius of the coupling matrix $A = [\alpha_{ij}]$ and λ is the system's decay rate.

14.1.3. The Phase Transition

The central analytical result is that the system exhibits qualitatively different behaviour above and below a critical threshold. When $\rho(A)/\lambda < 1$, perturbations decay and conflicts resolve: the system is self-correcting. When $\rho(A)/\lambda > 1$, perturbations amplify and escalation becomes self-sustaining: the system is self-amplifying. This is a bifurcation---a qualitative change in system dynamics at a critical parameter value.

The stability analysis proceeds through linearization around the zero-activation equilibrium. The Jacobian of the linearized system is $J = -\lambda I + A$, whose eigenvalues are $\mu_k = -\lambda + \sigma_k(A)$, where σ_k denotes the eigenvalues of the coupling matrix. The system is unstable---

perturbations grow---when the largest eigenvalue of J is positive, which occurs precisely when the largest eigenvalue of A exceeds λ, that is, when $\rho(A) > \lambda$. This is the escalation threshold. The transition is sharp. Small changes in network structure near the critical point produce large changes in escalation probability. This explains a phenomenon that is otherwise puzzling: why similar triggers---comparable provocations, comparable individuals---produce radically different outcomes across different organizational contexts. The phase-transition model provides the answer: the two organizations may be on opposite sides of the critical threshold. A provocation that resolves uneventfully in one network structure produces runaway escalation in another, not because the individuals differ but because the coupling structures differ.

At the critical point, the system exhibits the characteristic phenomena familiar from statistical mechanics. Critical slowing down: the system's response to small perturbations becomes slower as the threshold approaches, because the dominant eigenvalue of the Jacobian approaches zero. Divergent fluctuations: small provocations produce disproportionately large responses, because the system's restoring force weakens near the critical point. Symmetry breaking: initially symmetric groups polarize into asymmetric roles of aggressor and defender, as the growing perturbation selects a direction determined by the eigenvector corresponding to the dominant eigenvalue. These phenomena are not metaphorical analogies to phase transitions; they are formal instances of the same mathematical structure.

14.1.4. The Cascade Probability

For a population with threshold distribution F(T) and mean coupling strength, the probability of a cascade given an initial violation of magnitude V_0 can be approximated:

$$P(\text{cascade} \mid V_0) \approx F(V_0) \cdot \min\{1, \rho(A)/\lambda\}$$

This expression reveals the dual dependence of escalation on both individual and structural factors. The first factor, $F(V_0)$, captures individual-level sensitivity: the fraction of agents whose threshold is exceeded by the initial violation. The second factor, $\min\{1, \rho(A)/\lambda\}$, captures the network's capacity to amplify or dampen the initial perturbation. In the sub-critical regime ($\rho(A)/\lambda < 1$), even widespread initial activation decays. In the super-critical regime ($\rho(A)/\lambda > 1$), even localized initial activation can propagate through the network.

The critical insight is that in the super-critical regime, the structural factor saturates at 1, and escalation becomes virtually certain for any non-trivial initial violation. The prediction is

specific: in high-coupling networks where $\rho(A)/\lambda > 1.5$, escalation probability exceeds 80% regardless of individual differences, and the variance in outcomes explained by individual traits falls below 10%. The network structure dominates.

14.1.5. Architectural Invariants and Negotiable Specifications

The framework distinguishes carefully between its fixed architecture and its negotiable implementation details. The architectural invariants constitute the theory: the four-term structure (decay + violation + resources + coupling), threshold-based activation, network-mediated transmission, and phase transition at the critical coupling/decay ratio. These are the claims that, if falsified, would require abandoning the framework.

The negotiable specifications are implementation choices that can be varied without affecting the core theory: whether the threshold function is linear or sigmoid, whether resource modulation is multiplicative or additive, whether coupling is linear or saturating. These choices affect quantitative predictions but not the qualitative structure. Changing the specifications tests the robustness of the framework. Changing the architecture abandons it.

This distinction is methodologically important because it specifies precisely what the framework predicts and what it does not. The framework predicts a phase transition; it does not predict the exact functional form of the transition. It predicts that network structure dominates individual differences; it does not predict the exact weighting of structural versus individual factors in every context. The specificity of the architectural claims makes the framework falsifiable; the flexibility of the negotiable specifications makes it empirically applicable across contexts.

14.1.6. Testable Predictions and Falsification Criteria

The framework generates five testable predictions, ordered by centrality to the theory. The central prediction is that network coupling structure predicts escalation better than aggregated individual characteristics. The proposed test compares the predictive accuracy (AUC under ROC curve) of two logistic regression models: Model A, using episode-mean trait anger, agreeableness, and conflict style; and Model B, using spectral radius, network density, and mean coupling strength. The success criterion is that Model B exceeds Model A by at least 0.08 in AUC. If confirmed, this challenges four decades of individualist focus in organizational psychology.

The derived predictions specify the mechanism more precisely. Escalation probability should show sharp nonlinearity around the critical threshold $\rho(A)/\lambda = 1$---a sigmoid with steep transition region, not a linear relationship. In high-coupling networks ($\rho(A)/\lambda > 1.5$), escalation probability should exceed 80% regardless of individual traits. Individual activation probability should show discontinuity at the threshold level. Resource scarcity should lower effective thresholds, not merely amplify responses---a distinction with empirical consequences for whether the resource effect operates through the threshold mechanism or through a separate channel.

The falsification criteria are equally specific. If individual traits predict escalation as well as or better than network structure, the network theory is not warranted. If the relationship between spectral radius and escalation is linear rather than exhibiting a phase transition, the coupling matters but there is no critical threshold---the central theoretical claim is wrong. If high-coupling networks show less than 50% escalation, structure does not dominate. If no threshold discontinuity is observed, the threshold architecture is wrong.

14.1.7. Practical Implications

The practical significance of the phase-transition model lies in two domains: early warning and intervention design. The pre-transition dynamics---critical slowing down and divergent fluctuations---are detectable before the transition occurs, providing a window for intervention. The model identifies the system parameters that must be monitored: inter-group coupling strength, internal group cohesion, the ratio of cooperative to competitive interactions. It also specifies the interventions that can prevent the transition: reducing coupling strength (introducing communication buffers, reducing information flow between hostile subgroups), introducing moderating cross-group ties (bridging connections that carry cooperative rather than competitive interactions), and increasing the cost of competitive interactions (formal sanctions, reputational consequences).

The intervention logic reverses the traditional approach. Where traditional conflict management asks "how do we fix difficult people?", the phase-transition model asks "how do we redesign coupling topology?" Where traditional intervention targets individuals through conflict resolution training and anger management, the structural approach targets the network through coupling reduction and decay enhancement (cooling-off periods, delayed communication channels). The framework predicts that individual-level interventions will fail

in super-critical networks: no amount of conflict resolution training will prevent escalation when $\rho(A)/\lambda > 1$, because the system's dynamics overwhelm individual behaviour.

14.1.8. Philosophical Implications and the Removal of Moral Personalization

The framework removes moral personalization from conflict analysis. When $\rho(A)/\lambda > 1$, asking "who started it?" becomes as structurally meaningful as asking which neuron initiated a seizure. The system was primed for escalation; the trigger was incidental. Any sufficiently large perturbation would have produced the same outcome, because the system was in the super-critical regime.

This has immediate consequences for the attribution of responsibility. In a sub-critical network, individual actions matter: a provocation may or may not escalate depending on the specific individuals involved and their choices. In a super-critical network, individual actions are dominated by structural dynamics: escalation is virtually certain regardless of individual intentions. Responsibility shifts, in the super-critical case, from participants to architects---from the individuals involved in the conflict to whoever designed or maintains the coupling structure that made escalation inevitable.

Whether this shift in attribution is desirable is a normative question beyond the framework's scope. But the empirical claim is testable: do structures determine outcomes more than individuals do? The framework specifies the conditions under which the answer is yes (super-critical regime) and the conditions under which individual agency retains explanatory power (sub-critical regime). This is not a blanket denial of individual responsibility but a precise delineation of the regimes in which structural and individual factors dominate.

Within the thesis framework, this connects directly to the Law of Cooperation (Chapter 7). Law Zero establishes that cooperative response is the emergent default in unperturbed peer systems. The phase-transition model identifies the structural conditions under which this default breaks down: when the coupling structure exceeds the critical threshold, the cooperative equilibrium becomes unstable and the system transitions to a self-amplifying conflict regime. The transition is not caused by a failure of individual cooperation but by a structural property of the network that no individual can control.

14.1.9. Boundary Conditions

The framework models reactive escalation---responses to perceived violations that propagate through networks. It explicitly excludes three categories. Strategic actors who calculate

provocation for instrumental gain require game-theoretic apparatus and would dissolve the framework's clean separation of structure from strategy. Predatory conflict initiated without perceived violation lies outside the threshold-activation architecture. Structural role conflicts arising from institutional design rather than interpersonal dynamics operate through different mechanisms. These exclusions are not limitations but scope conditions that keep the theory testable.

14.2. The Asymmetry of Totalizing Ideals

The Asymmetry of Totalizing Ideals (Kriger, 2026, "The Asymmetry of Totalizing Ideals: A Structural Law of Complex Adaptive Systems," https://doi.org/10.5281/zenodo.18361828) establishes a structural impossibility result: in complex adaptive systems with irreducible uncertainty, the pursuit of a terminal, globally optimal, and internally fully coherent target state produces unbounded expected systemic damage---strictly greater than the damage produced by maintaining a boundedly suboptimal state that preserves adaptive variance. The result is formalized within an information-theoretic framework and proved as a theorem with explicit conditions, not offered as a conjecture or heuristic.

14.2.1. The Core Claim and Its Scope

The law applies to any ideal demanding total implementation: total equality, total freedom, total order, total virtue, total justice. The claim is that when such an ideal is pursued to its logical extreme---when partial implementation is deemed insufficient and only total realization is accepted---the pursuit will, with mathematical necessity, produce outcomes that contradict the ideal itself. The asymmetry lies in this: moderate pursuit of an ideal is structurally compatible with system viability; total pursuit is structurally incompatible.

This is not an empirical generalization from historical examples of failed utopias, though such examples abound. It is a structural theorem about the constraints on organized complexity. The proof proceeds from the formal properties of complex adaptive systems---irreducible uncertainty, the role of variance as informational substrate, the bounded learning capacity of any finite system---and demonstrates that totalizing optimization destroys the very mechanisms that make long-term viability possible.

The result does not argue against ideals. It argues against totalizing implementations. Equality, freedom, order, and justice are legitimate and important values. The formal result shows only that partial implementation is not merely the best achievable approximation but the only

structurally coherent form. Total implementation is self-defeating, and this is not a political opinion but a theorem about the constraints that complex adaptive systems impose on any optimization process operating within them.

14.2.2. Formal Framework: Irreducible Uncertainty and Adaptive Variance

The formal framework begins with a complex adaptive system S with state space Ω and dynamics governed by a transition function $T: \Omega \times U \times \Theta \to \Omega$, where U represents control inputs and Θ represents unknown environmental parameters. The foundational assumption is irreducible uncertainty: the true parameter $\theta^* \in \Theta$ is unknown and cannot be fully determined from any finite sequence of observations. Formally, for any observation sequence (o_1, ..., o_n), the posterior entropy satisfies $H(\theta^* \mid o_1, ..., o_n) > 0$ for all finite n. This assumption is the formal expression of a condition that holds for every complex adaptive system of nontrivial complexity---the environment cannot be fully known.

The key concept is adaptive variance, defined information-theoretically as the mutual information between the system's state distribution π_t and the unknown parameters:

$V_t := I(\pi_t; \theta^*) = H(\theta^*) - H(\theta^* \mid \pi_t)$

This measures how much information the diversity of the system's states provides about the unknown structure of the environment. A system with many distinct configurations (high entropy $H(\pi_t)$) that are differentially informative about different environmental parameters (high mutual information with θ^*) has high adaptive variance: its diversity functions as a sensory apparatus for detecting aspects of reality that no single configuration could reveal. Conversely, a system driven to a single state ($H(\pi_t) \to 0$) has lost its capacity to learn about the environment, regardless of how optimal that single state appears under the current model. This definition avoids the conceptual problem of conflating fitness variance with informational diversity. A system may have low variance in performance values while maintaining high structural diversity that is informative about unknown parameters. The information-theoretic formulation directly captures the learning capacity that diversity provides.

14.2.3. The Two Regimes: Totalizing Optimization and Variance Preservation

The framework defines two contrasting regimes. A totalizing optimization regime R_T satisfies four conditions. First, error reclassification: deviations from the target state s^* are classified as errors to be eliminated, not as informational signals. For any $s \neq s^*$, the system

applies corrective force directed toward s*. Second, variance minimization: the regime actively minimizes the entropy H(π_t), driving the state distribution toward a point mass at s*. Third, no stopping condition: convergence to s* is pursued without external termination criteria. Fourth, signal suppression: corrective signals indicating that the system's model of the environment is incorrect ($\hat{\theta} \neq \theta^*$) are discounted or suppressed.

A variance-preserving suboptimal regime R_V satisfies four complementary conditions. It maintains structural entropy above a positive minimum: $H(\pi_t) \geq H_{min} > 0$ for all t. Deviations are treated as potential information sources rather than errors. The objective function is revisable based on new information. Multiple local optima are tolerated rather than eliminated.

The distinction between these regimes is not a matter of degree but of kind. The totalizing regime treats the current model as final and drives toward a fixed point. The variance-preserving regime treats the current model as provisional and maintains the diversity needed to detect model error. The difference lies in the regime's relationship to uncertainty: does the system treat its remaining uncertainty as a problem to be eliminated or as a resource to be preserved?

14.2.4. The Damage Functional and Learning Dynamics

The systemic damage is quantified through an explicit loss function that captures two sources of harm: accumulated model error and reduced adaptive capacity.

$$L_t = c_1 \cdot E_t + c_2 \cdot 1/(V_t + \varepsilon)$$

The first term measures the cumulative cost of operating under an incorrect model of the environment: $E_t = \int_0^t \|\hat{\theta}_\tau - \theta^*\|^2 \, d\tau$. The second term captures the cost of low adaptability: as adaptive variance V_t decreases, the system becomes increasingly fragile---less able to detect and correct future errors. The inverse relationship with V_t reflects the structural insight that the damage from lost adaptability grows without bound as variance vanishes. The regularization constant $\varepsilon > 0$ prevents mathematical divergence at $V_t = 0$ while preserving the qualitative feature that very low variance is extremely costly.

The expected long-term damage under a regime R is the discounted integral of instantaneous loss: $D(s_0, R) = E[\int_0^\infty e^{-\rho t} L_t \, dt \mid s_0, R]$, where $\rho > 0$ is a discount rate reflecting the system's valuation of future relative to present damage.

A critical preliminary result establishes the connection between variance and learning. The rate at which a system can reduce its model error is bounded by the mutual information between its state distribution and the unknown parameters:

$-d/dt\ E[\|\hat{\theta}_t - \theta^*\|^2] \leq \kappa \cdot V_t$

where $\kappa > 0$ is a learning efficiency constant. This follows from the data processing inequality and the Cramér-Rao bound: the minimum achievable estimation error is inversely proportional to the Fisher information, which is bounded by the mutual information between observations and parameters. Since observations are generated by the state distribution π_t, the rate of error reduction is bounded by $I(\pi_t; \theta^*) = V_t$. The implication is profound: a system that destroys its variance destroys its capacity to learn, and a system that cannot learn cannot correct its errors.

14.2.5. Dynamics Under Totalizing Optimization

Under the totalizing regime, the system actively drives its state distribution toward the target state. The entropy of the state distribution decays exponentially: $H(\pi_t) = H(\pi_0)e^{-\alpha t} \to 0$ as $t \to \infty$, where $\alpha > 0$ is the rate of variance destruction. Since adaptive variance is bounded above by state entropy---$V_t \leq H(\pi_t)$, by standard information-theoretic bounds---the adaptive variance also decays exponentially: $V_t \to 0$ as $t \to \infty$. The system's capacity to learn about its environment vanishes.

The consequence for error accumulation is determined by the initial model error relative to the system's total learning capacity. The maximum possible error reduction over all future time is bounded:

$\kappa \int_0^\infty V_\tau\, d\tau \leq \kappa \int_0^\infty V_0\, e^{-\alpha\tau}\, d\tau = \kappa V_0/\alpha$

This is the total learning capacity of the system under the totalizing regime---the maximum amount by which the model error can decrease before variance is exhausted. If the initial model error exceeds this total learning capacity---that is, if $\|\theta_0 - \theta^*\|^2 > \kappa V_0/\alpha$---then the system cannot correct all its errors before its learning capacity is destroyed. A residual error $\delta^2 = \|\theta_0 - \theta^*\|^2 - \kappa V_0/\alpha > 0$ persists indefinitely, and the cumulative error grows without bound: $E_t \geq \delta^2 t \to \infty$.

The condition $\|\theta_0 - \theta^*\|^2 > \kappa V_0/\alpha$---the insufficient learning capacity condition---is the generic case for complex systems under irreducible uncertainty. Irreducible uncertainty means that the gap between any finite model and reality exceeds what can be learned from any finite

informational resource. The condition fails only in the degenerate case where the system starts nearly omniscient, knowing nearly everything about its environment before optimization begins---precisely the situation excluded by the assumption of genuine complexity.

14.2.6. Dynamics Under Variance Preservation

Under the variance-preserving regime, the system maintains structural entropy above a positive minimum: $H(\pi_t) \geq H_{min} > 0$, and correspondingly $V_t \geq V_{min} > 0$. The system retains a permanent capacity to learn. With ongoing learning at rate at least κV_{min}, and in the presence of environmental perturbations of variance σ^2_ε, the steady-state model error is bounded by the balance between learning and perturbation:

$$\limsup_{t\to\infty} \|\hat{\theta}_t - \theta^*\|^2 \leq \sigma^2_\varepsilon / (\kappa V_{min})$$

The error does not vanish---the system is persistently suboptimal---but it remains bounded. The cumulative error grows at most linearly with a bounded coefficient, and the damage from low adaptability is a finite constant $c_2/(V_{min} + \varepsilon)$. Both components of the loss function remain bounded, and the total damage $D(s_0, R_V)$ is finite.

14.2.7. The Main Theorem

The Asymmetry of Totalizing Ideals is stated as follows. Let S be a complex adaptive system satisfying irreducible uncertainty (Assumption 1). Let R_T be a totalizing regime and R_V a variance-preserving regime as defined above. Then the expected long-term damage under totalizing optimization strictly exceeds the damage under variance preservation:

$$D(s_0, R_T) > D(s_0, R_V)$$

provided two conditions hold: (i) the initial model error exceeds the system's total learning capacity before variance collapses (the insufficient learning capacity condition), and (ii) the discount rate ρ is less than the variance decay rate α---the system does not discount the future so heavily that long-term consequences become irrelevant.

The proof proceeds by computing the damage functional explicitly under each regime. Under the totalizing regime, two components of damage diverge. The cumulative model error grows without bound because the residual error δ^2 persists indefinitely: the integral of $e^{-\rho t} \cdot \delta^2 t$ over $[0, \infty)$ equals $c_1\delta^2/\rho^2$. But the dominant divergence comes from the adaptability cost. As $V_t \to 0$, the term $1/(V_t + \varepsilon)$ grows exponentially, and when the discount rate ρ is less than the variance decay rate α, the discounted integral of this term diverges:

$$e^{-\rho t} \cdot 1/(V_0 e^{-\alpha t} + \varepsilon) \approx e^{(\alpha-\rho)t}/V_0 \to \infty$$

Therefore $D(s_0, R_T) = \infty$. Under the variance-preserving regime, both components of the loss remain bounded, giving $D(s_0, R_V) = c_1 C_E/\rho^2 + c_2/(V_min \cdot \rho) < \infty$. The comparison is infinite versus finite: totalizing optimization produces unbounded expected damage while variance preservation produces bounded damage.

The condition $\rho < \alpha$ has a clear interpretation: it requires that the system's discount rate (how much it devalues future damage) is less than the rate at which totalizing optimization destroys variance. This is the regime where long-term consequences dominate---precisely the setting in which complex adaptive systems operate. If $\rho > \alpha$, the system discounts the future so heavily that immediate gains from coherence could outweigh long-term fragility; however, this represents myopic optimization incompatible with genuine concern for long-term viability.

14.2.8. Corollaries: The Optimization Dominance Law and the Law of Imperative Uncertainty

Two corollaries extend the main theorem to additional domains. The Optimization Dominance Law states that in any system with finite resources, diminishing marginal returns from optimization, and positive cost per optimization step, continued pursuit of maximal quality beyond the first satisfactory state decreases the probability of successful completion. This instantiates the totalizing-ideals asymmetry at the scale of individual projects: a project with a deadline and finite resources is a miniature complex adaptive system where variance corresponds to remaining flexibility (alternative approaches, time buffer, resource slack). Each optimization step beyond a satisfactory state consumes this flexibility. Under diminishing returns, the probability of improvement decreases while the probability of destabilization remains constant or increases. Driving flexibility to zero in pursuit of perfection increases expected damage without bound relative to stopping at a satisfactory point. This is the formal content of the folk wisdom that "the best is the enemy of the good."

The Law of Imperative Uncertainty states that any complex adaptive system capable of sustaining nontrivial complexity over time must maintain $V_t > 0$---a non-zero reserve of uncertainty and adaptive variance. Attempting to achieve $V_t = 0$ (full coherence, zero uncertainty) violates the necessary conditions for long-term viability. This follows directly from the main theorem: if a system drives $V_t \to 0$, it implements a totalizing regime, which produces unbounded expected damage. Therefore, any viable system must maintain $V_t > 0$. The corollary provides the mechanistic explanation for the Law of Imperative Uncertainty

introduced in the author's earlier work (Kriger, 2025): uncertainty is the informational substrate for learning; eliminating it eliminates the capacity for adaptation, which is lethal in environments with irreducible uncertainty.

14.2.9. Necessary and Boundary Conditions

The theorem's scope is precisely delineated by seven necessary conditions and seven boundary conditions. The necessary conditions specify when the asymmetry holds: the system pursues a terminal, non-revisable target; maximal coherence is actively enforced; variance is reclassified as error; no mechanisms protect variance; the system operates at large scale with irreversible processes; the environment exhibits irreducible uncertainty; and the initial model error exceeds total learning capacity. All seven must hold for the theorem to apply.

The boundary conditions specify when the asymmetry does not hold, and each identifies a structural feature that breaks the mechanism. If objectives are non-terminal and revisable, the system can update its target based on feedback, preventing the epistemic closure that drives variance destruction. If multiple incommensurable objectives are preserved, the system cannot collapse to a single attractor. If variance is institutionally protected---through mechanisms that guarantee $H(\pi_t) \geq H_{min}$ regardless of optimization pressure---the totalizing dynamics cannot take hold. If the system is small, reversible, or externally corrected, the divergence mechanism does not operate. If information is complete (the environment is fully known), there is no irreducible uncertainty. If the discount rate exceeds the variance decay rate, short-term gains dominate. And if the system starts near-omniscient, the learning capacity exceeds the model error and the residual δ^2 vanishes.

The boundary conditions are not loopholes; they are design prescriptions. Each boundary condition that prevents the asymmetry suggests a structural feature that protects systems from totalizing dynamics. Revisable objectives, multiple competing goals, institutional protections for variance, polycentric governance---these are not merely good ideas but necessary conditions for avoiding the self-defeating dynamics that the theorem characterizes.

14.2.10. Connection to the Thesis Framework

The Asymmetry of Totalizing Ideals is deeply embedded in the thesis framework and connects to results from multiple preceding chapters.

The most direct connection is to the Constraint--Autonomy Compatibility Law (Chapter 8). That law establishes that adaptive systems require both constraints ($C > 0$) and autonomy (A

> 0) for long-term functional stability---systems at either boundary (total constraint or total freedom) lose adaptive capacity. The totalizing-ideals result provides the information-theoretic mechanism behind this law. Total constraint corresponds to a totalizing regime that drives all variance to zero in the constraint dimension: every agent must follow the same rules, with no deviation tolerated. Total freedom corresponds to a totalizing regime that drives all variance to zero in the autonomy dimension: all constraints are eliminated, dissolving the structural definitions that make choice meaningful. In both cases, the totalizing pursuit of one value destroys the variance in the other that viability requires.

Consider the examples that make this concrete. Total freedom requires the elimination of all constraints. But autonomy---the capacity to choose among defined alternatives---requires enabling constraints: a language with rules, a legal system with rights, an institution with roles. Eliminating all constraints eliminates the structural conditions that make choice meaningful. Total freedom does not produce maximal autonomy but the dissolution of autonomy's prerequisites. This is the Constraint--Autonomy Compatibility Law applied to the freedom dimension, and the totalizing-ideals theorem provides the formal proof that this dissolution produces unbounded damage.

Total equality requires eliminating all distinctions between agents. But the Law of Cooperation (Chapter 7) establishes that self-organizing peer systems require functionally differentiated components---agents that can occupy different roles. Eliminating all distinctions destroys the system's capacity for self-organization, producing not equality but collapse. Total order requires eliminating all fluctuations and deviations. But the Law of Interference (Chapter 7) guarantees that fluctuations are intrinsic to multi-component systems: they arise from the interactions among components and cannot be eliminated without eliminating the interactions themselves---which means eliminating the cooperation on which viability depends.

The connection to Chapter 6 is equally fundamental. The resilience framework establishes that viable systems maintain themselves within a viability set through continuous adaptation. The totalizing-ideals result shows that totalizing optimization drives the system toward a fixed point, destroying the adaptive variance that resilience requires. A system optimized for a single state cannot absorb perturbations that push it away from that state, because it has eliminated the alternative configurations that would allow it to respond.

The connection to predictive processing (Chapter 10) illuminates the cognitive dimension. A totalizing ideal functions as a rigid prior that cannot be updated: the system's predictive model treats the ideal state as the only acceptable prediction, and all prediction errors (deviations from the ideal) are treated as errors to be corrected rather than as signals to be learned from. This is precisely the signal suppression condition (T4) of the totalizing regime definition. A healthy predictive processing system updates its priors in response to persistent prediction error; a totalizing system suppresses the error signal and maintains the prior unchanged. The result is the information-theoretic equivalent of a system that has blinded itself: it cannot detect model inadequacy because it has classified all evidence of inadequacy as noise.

14.2.11. The Evolutionary Evidence

The theorem's implications extend to biological evolution, providing a structural explanation for a striking empirical observation: nature does not optimize for perfection. Every biological population carries deleterious mutations (genetic load). The vertebrate eye has its photoreceptors facing backward. DNA polymerases could be more accurate but are not---the mutation rate is itself under selection and stabilizes at a non-zero value. Populations preserve multiple alleles at many loci even when one appears superior. These are not evolutionary failures; they are instances of maintained adaptive variance---the biological instantiation of $H(\pi_t) > 0$.

The theorem explains why this must be so. If evolution drove $V_t \to 0$---all individuals identical, perfectly adapted to the current environment---the population would lose $I(\pi_t; \theta^*)$, the information about environmental structure carried by phenotypic diversity. When the environment shifts (which it inevitably does under irreducible uncertainty), a population with $V_t = 0$ cannot respond. It has destroyed its sensory apparatus for detecting change. Populations that maintained $V_t > 0$---that tolerated imperfection---retain the capacity to track environmental shifts. They survive; the "perfect" populations go extinct.

This evolutionary perspective provides the deepest formulation of the result. The optimal strategy is to not optimize completely. Systems that survive are those that resist the temptation of terminal perfection, maintaining reserves of variation, pockets of inefficiency, margins of error. Imperfection is not the price of existence in an uncertain world that must be reluctantly paid---it is a design requirement for any system that must persist under conditions of irreducible uncertainty.

14.2.12. Implications for Institutional Design

The practical implications follow directly from the boundary conditions. Political systems should be polycentric: multiple overlapping authorities maintain the structural entropy required for adaptive governance, implementing boundary condition (B3) by institutionally protecting variance. Economies should preserve market variance: distributed price discovery maintains informational diversity, while central planning violates the variance-preserving conditions by driving toward a single optimal allocation determined by a central model. Science requires theoretical pluralism: enforced consensus drives $H(\pi^{theory}_t) \to 0$, eliminating the variance needed to detect model error. AI alignment must avoid terminal objectives: single utility functions implement totalizing regimes, and robust AI requires multiple revisable objectives and preserved behavioural variance.

The unifying principle across these applications is that imperfection is not a bug to be eliminated but a feature to be protected. Variance is the substrate of viability. The deliberate preservation of imperfection---variance, redundancy, inefficiency, dissent, slack---is not a concession to human weakness but a design requirement for long-term viability in any system operating under irreducible uncertainty.

14.3. Synthesis: Structure over Agency in Multi-Agent Dynamics

The two results presented in this chapter---conflict as phase transition and the asymmetry of totalizing ideals---converge on a single methodological principle: in complex multi-agent systems, structural properties dominate individual properties in determining long-term outcomes.

The phase-transition model demonstrates this for conflict: when the coupling structure of a network exceeds the critical threshold, escalation is determined by topology, not by the individuals involved. The totalizing-ideals result demonstrates this for institutional design: when an optimization regime satisfies the totalizing conditions, systemic damage is determined by the regime's structural relationship to variance, not by the intentions, competence, or moral orientation of the agents implementing it.

This convergence is not coincidental. Both results derive from the same underlying principle: complex adaptive systems have structural properties---eigenvalue spectra, entropy dynamics, learning capacity bounds---that constrain outcomes independently of the agents operating within them. The phase-transition model identifies the structural property (spectral radius

relative to decay) that determines whether conflict self-corrects or self-amplifies. The totalizing-ideals theorem identifies the structural property (variance relative to model error) that determines whether optimization is self-improving or self-defeating.

The practical consequence is that intervention in complex systems must be structural, not merely behavioural. Training individuals in conflict resolution cannot prevent escalation in a super-critical network. Appointing virtuous leaders cannot prevent the damage of totalizing optimization when the institutional structure satisfies the totalizing conditions. The structural conditions must themselves be modified: coupling topology redesigned, variance institutionally protected, objectives made revisable, feedback channels kept open. This is the central lesson of Part IV's multi-agent analysis: in coherent systems with non-neutral components, the architecture determines the outcome.

CHAPTER 15. EXTRACTIVE OSCILLATORS: FROM NARCISSISTIC DYNAMICS TO ADDICTION

The preceding chapters established the structural analysis of multi-agent systems under persistence constraints: deception as information manipulation (Chapter 13), conflict as phase transition and the self-defeating nature of totalizing ideals (Chapter 14). This chapter introduces a unifying dynamical model---the extractive oscillator with sensor degradation---that captures a class of pathological interactions across multiple domains with a single formal structure (Kriger, 2017, https://doi.org/10.5281/zenodo.18503666; Kriger, 2026, "Addiction as Extractive Oscillator with Sensor Degradation," https://doi.org/10.5281/zenodo.18603700).

The model demonstrates the substrate independence of the persistence framework with particular clarity. The same formal structure---oscillatory extraction, progressive sensor degradation, sunk-cost entrenchment, and informational isolation---governs dynamics as diverse as narcissistic relational patterns, substance addiction, behavioural addiction, exploitative platform design, authoritarian political control, cult dynamics, and toxic corporate leadership. In each domain, the oscillation, the epistemic degradation, the investment trap, and the network isolation arise not from the specific content of the interaction but from the structural configuration of the coupled system. A thermostat with the same mathematical parameters produces the same oscillatory behaviour. The mathematics is indifferent to the substrate.

The chapter proceeds in eight sections. Section 15.1 defines the extractive oscillator class through five axiomatic conditions and develops the core dynamical system. Section 15.2 presents the abuse cycle as a relaxation oscillator with a proved limit cycle. Section 15.3 formalizes sensor degradation as Bayesian persuasion with prior precision erosion, proving a sharp phase transition in epistemic autonomy. Section 15.4 derives victim entrenchment from sunk-cost-weighted cognitive dissonance resolution. Section 15.5 applies the model to narcissistic relational dynamics. Section 15.6 extends it to addiction, including both substance and behavioural forms. Section 15.7 extends it to exploitative platform design and additional domains. Section 15.8 synthesizes the results and connects them to the thesis framework.

15.1. The Extractive Oscillator Model: Definition and Axiomatic Conditions

The extractive oscillator with sensor degradation is defined as a coupled two-agent (or agent-environment) dynamical system satisfying five axiomatic conditions. Any system satisfying

all five---regardless of substrate---is governed by the theorems developed in this chapter. The conditions are:

(K1) Asymmetric extraction: One component (the extractor) extracts a resource from the other (the source) without symmetric reciprocation. Formally, the extractor's utility is increasing in a quantity that the source produces at a cost the source does not recover.

(K2) Hysteretic oscillation: The extractor alternates between supply (restoration) and depletion (extraction) modes, with switching governed by hysteretic thresholds. The coupled system admits a stable limit cycle.

(K3) Sensor degradation: The source's capacity to accurately assess its own state is progressively eroded by the extractor's signals, producing a phase transition in epistemic autonomy.

(K4) Sunk-cost trapping: Accumulated investment by the source creates a positive feedback loop that makes exit increasingly costly, eventually trapping the source in the exploitative equilibrium.

(K5) Informational isolation: When the extractor manages multiple sources, a star-network topology with zero mutual information among sources is maintained, and breaking this isolation destabilises the equilibrium.

These five conditions are necessary and sufficient. Any system satisfying (K1)--(K5) exhibits the full dynamical signature: oscillatory extraction cycles, progressive loss of self-assessment capacity, increasing exit barriers, and vulnerability to isolation-breaking interventions. The conditions are structural, not substantive: they specify the formal relationships among components without reference to the nature of the components themselves.

The term "quasi-narcissistic dynamics" is used to denote the full pattern---oscillatory exploitation, epistemic degradation, sunk-cost trapping, and informational isolation---without presupposing that the pattern originates from a diagnostic category. The prefix "quasi-" signals that the dynamics resemble clinical narcissism in their manifestation but do not require it as an explanation. The same dynamics arise in thermostats, authoritarian states, cults, corporations, addictions, and algorithmic platforms, none of which involve personality disorders. The diagnostic label "narcissist" is not necessary for explanation, prediction, or intervention: the dynamics suffice.

15.1.1. Agent Definitions and Utility Functions

The formal framework begins with an axiomatic derivation of agent utility functions from behavioural primitives, providing a principled foundation rather than ad hoc assumptions. The relational interaction space is defined as the non-negative orthant of three dimensions representing validation (v), empathic engagement (e), and control (c).

Four behavioural axioms are imposed. First, monotonicity: for the extracting agent, utility is strictly increasing in validation and control and strictly decreasing in empathic engagement; for the source agent, utility is strictly increasing in perceived attachment and relational coherence and strictly decreasing in relational rupture. Second, separability: utility functions are additively separable. Third, diminishing marginal returns in each component. Fourth, empathy-cost dominance for the extractor: the marginal disutility of empathic expenditure is bounded below, while the marginal utility of validation saturates.

Under these axioms, the most general utility functions take the form:

U_N(v, e, c) = alpha * h_1(v) - beta * h_2(e) + gamma * h_3(c)

U_W(a, s, r) = delta * h_4(a) + epsilon * h_5(s) - zeta * h_6(r)

where alpha, beta, gamma, delta, epsilon, zeta > 0 and each h_i is a twice-differentiable function with positive first derivative and non-positive second derivative. The extractor's internal model of the source encodes the source purely as a resource node parameterised by supply capacity and control susceptibility, without subjective attributes.

The key qualitative results---the Nash equilibrium, the limit cycle, the phase transition---depend only on the monotonicity and sign structure of the axioms, not on the specific functional forms. The linear specialisation h_i(x) = x serves as a first-order approximation valid for bounded interaction regimes; a formal robustness analysis confirms that all main results are preserved under any monotone concave transformation of the utility components.

15.1.2. The Asymmetric Relational Game

The strategic structure is formalised as a complete asymmetric bimatrix game. The extractor has three strategies: idealise (I), devalue (D), and intermittent reinforce (R). The source has three strategies: accommodate (A), confront (C), and exit (X). The payoff matrices encode specific logic grounded in established behavioural principles: under idealisation, the extractor pays full empathy cost but receives validation and moderate control; under devaluation, empathy cost drops to zero and control doubles through dominance assertion; under intermittent reinforcement, maximal validation and control are achieved at half empathy cost,

exploiting the partial reinforcement extinction effect (Skinner, 1938; Ferster and Skinner, 1957). For the source, exit incurs a sunk-cost loss proportional to accumulated investment, which grows with time.

Under parameter constraints requiring that the extractor's control and validation coefficients dominate empathy cost, and that accumulated investment exceeds the source's attachment and coherence benefits, the strategy pair (R, A)---intermittent reinforcement by the extractor, accommodation by the source---is the unique pure-strategy Nash equilibrium. The proof verifies best-response conditions for each agent through explicit payoff comparison.

This equilibrium is Pareto-dominated by the hypothetical outcome where the source interacts with a non-extractive partner, but the Pareto improvement is not accessible from within the game since it requires replacing an agent. The equilibrium is a trap: a local attractor in strategy space surrounded by a basin of attraction from which unilateral deviation is costly. Critically, the equilibrium requires sufficient accumulated investment---below a critical investment threshold I*, exit becomes the preferred strategy and the Nash equilibrium ceases to exist. This yields the first testable prediction: there exists a quantitative threshold in accumulated investment below which the trap has not yet closed.

15.2. The Abuse Cycle as Relaxation Oscillator

The game-theoretic analysis of Section 15.1 describes the strategic logic; the dynamical systems analysis models the continuous temporal evolution within the equilibrium regime, showing that the extractor's intermittent reinforcement strategy naturally generates oscillatory dynamics.

15.2.1. State Variables and the Planar System

Two state variables capture the essential dynamics. The source's emotional energy E(t), bounded in [0, E_max], measures willingness, hope, and supply capacity. The extractor's perceived control security K(t), bounded in [0, K_max], measures the extractor's assessed stability of dominance.

The switching between idealisation and devaluation modes is derived from the extractor's utility maximisation, not imposed exogenously. Assuming the extractor maximises a discounted utility stream, the optimal policy has a bang-bang structure with hysteretic switching at thresholds $K_l < K_h$ derived from the ratio of validation return to control value and empathy cost. The hysteretic structure---the system remains in its current mode for K_l

<= K <= K_h---follows from the standard result that optimal policies for single-state switching control problems with fixed switching costs have a bang-bang structure with a deadband.

With the derived switching rule, the dynamics become a planar system with piecewise-smooth structure:

dE/dt = lambda_+(K_max - E) - mu*E during idealisation (when K < K_l)

dE/dt = -lambda_-*E during devaluation (when K > K_h)

dK/dt = g(E) - nu*K

where lambda_+, lambda_-, mu, nu > 0 and g is a smooth, strictly increasing function with g(0) = 0. During idealisation, the source's emotional energy is replenished (at a rate proportional to the gap from maximum, less natural decay). During devaluation, emotional energy decays exponentially. The extractor's control security rises with the source's emotional energy (a compliant, engaged source confirms control) and decays naturally.

15.2.2. Existence of a Stable Limit Cycle

The central dynamical result is the existence and stability of a limit cycle. Under three hypotheses---that idealisation can replenish emotional energy (the supply rate exceeds the depletion rate at maximum), that both switching thresholds lie in the interior of the state space, and that the system has sufficient dynamic range (the K-nullcline intersects both switching thresholds at interior values of E)---the system has a stable limit cycle enclosing the hysteresis region in the (E, K) phase plane.

The proof constructs an explicit trapping region and applies the Poincare-Bendixson theorem as extended to Filippov (piecewise-smooth) systems. The trapping region is bounded by curves slightly inside the state space boundaries and slightly outside the hysteresis band. On each boundary, the vector field points inward: on the top boundary, the system is in devaluation mode and both E and K decrease; on the bottom boundary, the system is in idealisation mode and both E and K increase; on the left and right boundaries, the restoring structure of the E-equation confines trajectories. Neither subsystem's fixed point lies within the trapping region: the idealisation fixed point lies above K_h, and the devaluation fixed point lies at the origin below K_l. No pseudo-equilibria exist on the switching surfaces because the vector fields cross transversally. By the extension of Poincare-Bendixson to Filippov systems, the trapping region must contain a closed orbit.

The period of the limit cycle is approximately:

T ~ (1/(lambda_+ + mu)) * ln((E_max - E_l)/(E_max - E_h)) + (1/lambda_-) * ln(E_h/E_l)

This provides a quantitative prediction: the minimum cycle period is a computable function of the interaction parameters. The period depends on the emotional recovery rate, the depletion rate, and the switching thresholds. Faster recovery rates (due, for instance, to external social support reducing mu) shorten the cycle. The structural parallel with predator-prey dynamics is suggestive but inexact: the critical difference---active modulation of prey growth by the predator via strategic switching---places the system closer to the class of hybrid dynamical systems with state-dependent mode switching than to classical ecological oscillators.

15.2.3. The Inevitability of Oscillation

The oscillation is not a strategic choice, a personality trait, or a conscious intention---it is a theorem. Consider any system in which one agent optimises extraction of a renewable but finite-capacity resource from another. Permanent depletion exhausts the resource: the source exits, the population revolts, the employee resigns, the user disengages. Permanent supply pays full cost while allowing the source to accumulate autonomy, risking voluntary exit. The unique optimal policy is alternation: supply until a replenishment threshold, then extract until a depletion threshold, then supply again. Under two modes with opposing effects and threshold switching, the optimal policy is bang-bang control, and the system must oscillate. Any extraction system satisfying the three minimal hypotheses oscillates.

15.3. Sensor Degradation as Bayesian Persuasion with Prior Erosion

The oscillator alone is unstable. If the source accurately perceives the cyclic structure, it can time its exit to the low-control period, recognising the pattern and escaping. Therefore, an extractive strategy that does not include sensor degradation loses the source within a few cycles. Under evolutionary selection (or any optimisation process), strategies that include sensor degradation outcompete those that do not, because they sustain longer extraction periods. Sensor degradation is not an incidental feature; it is a necessary condition for the stability of any extractive oscillatory strategy.

15.3.1. The Bayesian Model of Epistemic Degradation

The sensor degradation mechanism is formalised using the Bayesian persuasion framework of Kamenica and Gentzkow (2011), extended to incorporate cumulative degradation of the receiver's prior precision. The source maintains a belief about the state of its relational world (for instance, "am I being mistreated?"). The true state is binary: mistreatment is either

occurring or not. At each period, the source receives two signals: an internal signal drawn from a Gaussian centred on the true state with precision tau_int, and an external signal from the extractor designed to contradict the true state, drawn with precision tau_ext. The source performs Bayesian updating with precision-weighted signals, forming a posterior that is a weighted average of the two signals with weights proportional to their precisions.

The key mechanism is the degradation of internal precision: each time the source accepts the extractor's contradictory interpretation over its own experience, confidence in internal sensing is reduced by a fixed erosion amount phi. This creates a positive feedback loop: as internal precision decreases, the source relies more heavily on the external signal, making it more likely to accept the extractor's interpretation in the next round, which further decreases internal precision.

15.3.2. The Phase Transition in Epistemic Autonomy

Under persistent extraction with the extractor signalling against the true state, there exists a critical time t* at which the external precision exceeds the internal precision. Before t*, the source's posterior is predominantly determined by its own internal signal---it maintains epistemic autonomy. After t*, the posterior is predominantly determined by the extractor's signal---the source has become epistemically dependent. The transition at t* is sharp, not gradual: the positive feedback creates a tipping point where, once internal precision approaches external precision, the acceptance rate jumps and the remaining erosion occurs rapidly.

The critical time is given by t* = (tau_int,0 - tau_ext) / (phi * p_accept), where p_accept is the probability of accepting the extractor's signal. A concentration inequality (Hoeffding's bound applied to the partial sums of acceptance events) establishes that the transition occurs with probability approaching 1 exponentially in t*. The transition is a genuine phase transition: it exhibits a sharp threshold, positive feedback acceleration near the critical point, and qualitatively different system behaviour on either side.

A corollary with significant clinical implications follows: upon separation from the extractor, the external signal vanishes. But the source's internal precision has been degraded to near the level of the now-absent external precision. The result is near-total epistemic uncertainty---the clinically observed "reality collapse" upon separation from an abusive relationship. The

source's inference system has been hollowed out; removing the extractor does not restore the sensor but reveals its degraded state.

This model extends Bayesian persuasion in two important ways. First, the persuasion is repeated across periods. Second, it degrades the receiver's prior precision rather than merely shifting beliefs. The sender does not merely persuade---the sender reduces the receiver's future capacity to be unpersuaded. This cumulative degradation has no analogue in the standard Bayesian persuasion framework and represents a novel mechanism with relevance beyond the relational domain.

15.4. The Investment Trap and Positive Feedback Entrenchment

The third structural component formalises the source's role in maintaining the exploitative equilibrium through sunk-cost-weighted cognitive dissonance resolution.

15.4.1. The Dissonance Model

Cognitive dissonance is generated by the contradiction between accumulated negative evidence and the maintained relational narrative. The source resolves this dissonance through one of two channels: exit (accept the negative evidence as dominant, at cost proportional to accumulated investment plus epistemic dependency) or reinterpretation (discount the negative evidence, at cost proportional to the dissonance magnitude times the failure rate of rationalisation).

The dynamics of investment and reinterpretation success form a coupled system with positive feedback: each successful reinterpretation increases investment (the source has doubled down on the relationship), and higher investment improves future reinterpretation capacity through motivated reasoning. Investment is strictly increasing and unbounded. Reinterpretation success is increasing and approaches 1 asymptotically.

15.4.2. The Investment Trap Theorem

Under these dynamics, there exists a critical time t_{trap} such that for all $t > t_{trap}$, the cost of exit exceeds the cost of staying, and exit is never locally optimal. The system is self-reinforcing: the gap between exit cost and stay cost is strictly increasing after t_{trap}. The proof proceeds by showing that exit cost grows at least linearly (since investment grows at least linearly) while stay cost approaches zero (since reinterpretation success approaches 1, making rationalisation nearly costless regardless of how much negative evidence accumulates).

The investment trap explains a phenomenon that is otherwise paradoxical: why sources remain in relationships, addictions, organisations, and systems that are demonstrably harmful, even when they have access to information about the harm. The answer is structural: once accumulated investment exceeds the critical threshold, the cost of acknowledging the investment as error exceeds the cost of continued rationalisation. This is not irrationality; it is locally optimal behaviour in a system with positive feedback, where the global optimum (early exit) was only accessible before the trap closed.

15.4.3. Three Coupled Feedback Loops

The entrenchment mechanism derives its power from the simultaneous operation of three positive feedback loops. Loop 1 (investment-rationalisation): the more that has been invested, the costlier it is to acknowledge the investment as error; rationalisation resolves the dissonance, enabling continued investment; investment is monotonically increasing and unbounded. Loop 2 (precision erosion-acceptance): each acceptance of the extractor's contradictory signal reduces internal precision; lower precision raises the acceptance probability; this is an autocatalytic process that reaches the critical point with probability approaching 1 exponentially in time. Loop 3 (isolation-dependence): without external information sources, no corrective signal can counteract the precision erosion; dependence on the single information source grows, which makes the isolation more effective, which deepens the dependence.

The three loops operate simultaneously and amplify one another. Below the critical investment threshold, exit remains accessible and the loops are subcritical. Above it, the three loops interlock and the trap becomes effectively irreversible. The transition is sharp, not gradual.

15.5. Application to Narcissistic Relational Dynamics

The original formulation (Kriger, 2017) demonstrated that the extractive oscillator model captures the core dynamics of quasi-narcissistic relational patterns. The extractor corresponds to the narcissistic agent, parametrised by high validation need, low empathic capacity, high control-seeking, and low empathic attachment. The source corresponds to the partner, parametrised by high empathic capacity and high attachment need. The oscillation corresponds to the clinically documented idealisation-devaluation cycle: the extractor alternates between providing affirmation (partial restoration) and withdrawing it (depletion), a pattern known clinically as intermittent reinforcement.

15.5.1. Sensor Degradation as Gaslighting

Sensor degradation in the relational context corresponds to gaslighting: the progressive erosion of the partner's capacity for accurate self-assessment and reality testing. Through repeated cycles of idealisation and devaluation, the partner's ability to distinguish healthy from unhealthy relational dynamics degrades. The partner becomes increasingly unable to recognise the extraction pattern even as its effects become more severe---the clinically observed phenomenon of trauma bonding, where the partner becomes more attached to the extractor precisely as the extraction becomes more damaging.

The phase transition theorem provides a precise prediction: there exists a critical number of gaslighting episodes after which the partner's inference system flips from autonomous to externally dependent. Before this transition, the partner's judgments are primarily governed by their own experience. After it, their judgments are primarily governed by the extractor's narrative. The transition is sharp, and it accelerates as it approaches the critical point.

15.5.2. Network Extension and the Supply Network

When the extractor maintains multiple sources simultaneously, the system takes the form of a star graph with the extractor as hub and sources as informationally isolated spokes. The defining constraint is zero mutual information among sources: no source has access to information about any other source's relational state. This allows the extractor to optimise across sources, switching extraction and restoration among them based on relative depletion levels, and discarding depleted sources without emotional transition cost (since the extractor's internal model encodes sources as resource nodes, not subjects).

A key structural result concerns equilibrium destabilisation through isolation breaking. When a source acquires information about other sources' existence and relational state, the perceived relational coherence drops discontinuously, the effective exit cost is reduced, and if the information shock is sufficiently large, exit dominates accommodation and the Nash equilibrium in that dyad is destroyed. This provides the formal basis for the network illumination intervention: breaking the informational isolation among sources is a structurally targeted attack on the conditions sustaining each dyadic equilibrium.

15.5.3. Coupled Oscillators: Symmetric Quasi-Narcissistic Interaction

When both agents occupy the extractive parameter region---each extracting different resources from the other---qualitatively different dynamics emerge. The system becomes two

coupled piecewise-smooth oscillators with hysteresis, admitting three dynamical regimes depending on coupling strength and frequency ratio: in-phase synchronisation (simultaneous idealisation and devaluation, producing high-amplitude passionate cycles), anti-phase synchronisation (perpetual turbulence without honeymoon phases), and quasi-periodic or chaotic dynamics (unpredictable switching with no recognisable cycle).

The symmetric equilibrium (R, R)---mutual intermittent reinforcement---is stable because neither participant's sensor degrades (both maintain fixed self-models and project contradictory signals outward rather than updating), neither is depleted to the discard threshold (each agent's energy is sustained by an internal source independent of external validation), and mutual deterrence bounds the oscillation amplitude. However, third parties without projection defences---particularly children---are exposed to dual-source epistemic degradation and become projection surfaces for both agents' externalised contradictions. The model predicts that symmetric pairs dissolve more suddenly and with less post-separation distress than asymmetric pairs, because neither agent is emotionally attached (both have low attachment parameters and the sunk-cost trap is weaker).

15.5.4. The Narcissism Continuum

The framework replaces the categorical diagnosis "narcissist" with a continuous parametric characterisation. Every agent is described by a parameter vector governing sensitivity to validation, empathic capacity, control-seeking, and empathic attachment. Quasi-narcissistic dynamics emerge when the ratio of empathic-attachment parameters to validation-control parameters falls below a critical threshold determined by the game structure. This is a continuous, dimensional characterisation---not a diagnostic boundary. The critical ratio indexes position on a continuum: different parameter configurations produce different dynamical regimes, with the extractive oscillator regime occupying a specific region of parameter space.

15.5.5. The Irrelevance of Intent and the Diagnostic Label

The most consequential implication of the framework is that conscious malice is not required. A thermostat does not intend to subject a room to temperature oscillations. It optimises. Similarly, an agent with the parameter configuration in the extractive regime will produce the full cycle of idealisation, devaluation, gaslighting, and entrenchment as an emergent

consequence of optimisation under those parameters---without requiring a plan, a strategy, or even awareness.

This reframing---from "who is to blame" to "what structure produces this outcome"---converts a psychological problem into a control-theoretic one. The intervention targets are not moral judgments about persons with unknown internal states but engineering operations on a system with known structure: restoring sensors, breaking isolation, reducing exit barriers, making oscillatory patterns visible.

A further consequence deserves emphasis: the diagnostic label "narcissist" itself satisfies all five conditions (K1)--(K5) for being an instrument of the mechanism it describes. The label performs epistemic degradation (invalidating the target's capacity to trust their own judgment), informational isolation (redefining all communication from the target as manipulation), sunk-cost trapping (each attempted defence is reinterpreted as further evidence), asymmetric extraction (establishing an observer-object asymmetry), and hysteretic oscillation (toggling between "I see you are trying" and "no, you are exhibiting narcissism again"). This paradox---that the concept invented to describe a mechanism of manipulation satisfies all conditions for being an instrument of that mechanism---is an additional argument for replacing the diagnostic framework with a dynamical one. It is not possible to weaponise the statement "you are an oscillator with hysteretic switching and sensor degradation" in the same way that one can weaponise the statement "you are a narcissist."

15.6. Application to Addiction

The extension to addiction (Kriger, 2026, https://doi.org/10.5281/zenodo.18603700) demonstrates that the extractive oscillator model applies with full formal force to substance and behavioural addictions, establishing that the binge-remission-tension-relapse cycle, the progressive erosion of self-assessment capacity, sunk-cost entrenchment, and social isolation are necessary structural consequences of the five axiomatic conditions---not contingent features of particular substances or behaviours.

15.6.1. Reinterpretation of the Agent Structure

In addiction, the extractor is not a person but the addictive substance or behaviour (more precisely, the neurochemical reward pathway it activates), modelled as an abstract agent whose utility function is defined by the reinforcement architecture. The source is the individual's cognitive and physiological system. Emotional energy E(t) maps to psychological

and physiological capacity (health, social functioning, executive control). The idealisation phase corresponds to the intoxication or reward episode (the hedonic peak), providing short-term reward at the cost of long-term viability. The devaluation phase corresponds to withdrawal, hangover, and accumulating consequences, depleting hedonic state while allowing partial physiological recovery.

The formal structure is identical to the relational case: a two-component system with oscillatory extraction and progressive sensor degradation. The states are different (neurochemical states rather than psychological states), the transformations are different (pharmacological effects rather than relational interactions), but the dynamical structure---the equations, the phase portrait, the limit cycle, the phase transition---is the same.

15.6.2. The Five Conditions in the Addiction Domain

Each axiomatic condition maps precisely. Asymmetric extraction (K1): the substance provides short-term hedonic reward while extracting long-term physiological and social capacity that is not symmetrically restored. Hysteretic oscillation (K2): the binge-remission-tension-relapse cycle is a limit cycle with hysteretic switching, where tolerance shifts the thresholds---the dose required for the same effect increases, widening the hysteresis band. Sensor degradation (K3): the progressive impairment of prefrontal decision-making systems that would normally detect and correct the maladaptive pattern corresponds to tolerance and to the clinically documented erosion of insight that characterises addiction progression. Sunk-cost trapping (K4): accumulated losses (health, relationships, finances, identity) create the paradoxical investment trap---"I have already lost so much to this addiction; if I quit now, all that suffering was for nothing"---formally identical to the relational sunk-cost mechanism. Informational isolation (K5): progressive social withdrawal, shame-driven concealment, and the selection of social environments that normalise use implement the star topology at the individual level.

15.6.3. Substance Independence and Behavioural Addictions

The substrate independence of the model is strengthened by analysis of behavioural addictions that share no common chemistry with substance dependence. Gambling addiction instantiates the model through the variable-ratio reinforcement schedule (the same partial reinforcement extinction effect that drives the relational intermittent reinforcement equilibrium), with sensor degradation manifesting as the gambler's fallacy and progressive inability to accurately assess financial state. Gaming addiction, social media addiction, pornography addiction,

workaholism, and food addiction each exhibit the same five-condition structure with domain-specific variable mappings but identical dynamical equations. Cross-addiction---the well-documented phenomenon of transitioning from one addictive substance or behaviour to another---is explained as transition between isomorphic potential wells on a shared attractor landscape: the underlying dynamical structure is preserved while the substrate changes.

15.6.4. Developmental Vulnerability: The Adolescent Gap

The addiction framework includes a developmental vulnerability model that formalises a specific risk window. During adolescence, the endogenous endorphin system undergoes reorganisation while prefrontal self-regulation systems are still maturing. This creates a developmental gap: the need for hedonic regulation increases precisely when the capacity for self-regulation is still immature. In the language of the extractive oscillator model, the adolescent gap represents a period during which the source's sensor sensitivity (tau_int) is structurally low---not because of prior degradation but because the sensor has not yet fully developed---making the system unusually vulnerable to the establishment of extractive oscillatory dynamics. Substances or behaviours encountered during this window can establish the extraction cycle before the sensor reaches its adult baseline, creating a trap that is structurally harder to escape than one established after sensor maturation.

15.6.5. Neural Interface Addiction: The Zero-Path-Length Limiting Case

The framework generates predictions about future addiction risks from neural interface technologies. Current addictive substances and behaviours face physical constraints: injection risks, toxic metabolites, recurring procurement costs, and finite delivery speeds. Neural interfaces that directly stimulate reward pathways eliminate these constraints, representing the zero-path-length limiting case of the extractive oscillator model---the case where the extractor has direct, instantaneous, and costless access to the source's reward system.

The formal analysis identifies both increased addictive potential (elimination of physical barriers to extraction, enabling arbitrarily short cycle periods and arbitrarily high extraction rates) and the elimination of certain harms (no injection-related infections, no toxic metabolites, no procurement-related crime). The net effect depends on the specific design of the interface. The model predicts that neural interfaces without built-in rate limiters---structural constraints that enforce minimum cycle periods and maximum extraction rates, analogous to the natural pharmacokinetic constraints that current substances face---will

produce addiction dynamics that are formally identical to but quantitatively more severe than any currently known substance addiction.

15.6.6. Intervention as Structural Attack

The formal structure dictates exactly where the system is vulnerable, and each therapeutic strategy is derived as a structural attack on a specific theorem, feedback loop, or phase transition. Precision restoration (targeting the phase transition theorem) rebuilds the source's internal signal precision through grounding techniques, reality testing, and external validation of internal perception---most effective when tau_int is still above tau_ext, before the phase transition. Investment reframing (targeting the investment trap theorem) decouples perceived exit cost from accumulated investment, addressing the sunk-cost mechanism directly. Cycle recognition (targeting the limit cycle theorem) makes the oscillatory structure explicit through psychoeducation, enabling prediction of the next phase and action during the window of maximal autonomy. Network illumination (targeting the isolation condition) breaks informational isolation among sources, triggering equilibrium destabilisation. Exit threshold reduction provides practical support reducing exit cost directly.

For addiction specifically, the framework identifies the connection to the unified dynamical systems theory (DST) of psychiatric disorders (Kriger, 2026), which provides capabilities that the extractive oscillator model alone does not: prediction of the initial healthy-to-addicted transition via critical slowing down (measurable early warning signals before the first binge), quantification of recovery difficulty via Kramers escape rates, comorbidity modelling on a shared attractor landscape (interaction of addiction with depression, PTSD, anxiety), and formalisation of the "fear of relapse" as an autocatalytic destabilising loop that erodes recovery resilience.

15.7. Additional Domains: Platform Design, Authoritarian Systems, and Beyond

15.7.1. Exploitative Platform Design

The model extends to digital platforms that optimise engagement at the expense of user well-being. The extractor is the platform algorithm (whose viability depends on user attention and data), and the source is the user. The oscillation corresponds to the cycle of reward (content satisfying immediate desires---viral posts, likes, algorithmic amplification) and extraction (monetisation of attention, data, and behavioural patterns---reduced reach, exposure to engagement-optimised content). Sensor degradation corresponds to the progressive reduction

of the user's capacity to distinguish content serving their long-term interests from content serving the platform's extraction goals.

The informational isolation condition is enforced architecturally: each user receives a personalised feed, and no user has access to the feed shown to any other user. The platform is the hub; users are informationally isolated spokes---the star topology implemented at the scale of billions. The precision erosion operates through systematic exposure to engagement-optimised content: the algorithm optimises for engagement, not truth, so the user's internal model of reality is progressively distorted by signals optimised for an objective different from accurate world-modelling.

15.7.2. Authoritarian Political Systems

The structure of an authoritarian state controlling its population instantiates the same dynamical system. The extractor is the regime; the source is the citizenry. The oscillation maps to the well-documented alternation between liberalisation (thaw) and repression (crackdown): permanent repression depletes civic capacity (economic productivity falls, threatening regime survival), while permanent liberalisation erodes political control (civil society strengthens). The result is the same limit cycle. Gaslighting maps to state propaganda: the citizen's direct experience is contradicted by state media output, with the phase transition theorem predicting sharp rather than gradual consolidation of narrative control. The investment trap manifests as accumulated sacrifices for the state (military service, ideological commitment, denunciations of neighbours), and the star network topology manifests as censorship and prohibition of independent associations.

15.7.3. Cults, Corporate Toxicity, and Professional Exploitation

Cult dynamics map term by term: love-bombing is idealisation, shaming is devaluation, "your doubts are evidence of spiritual weakness" is precision erosion, donated wealth and severed family ties constitute the investment trap, and hierarchical communication enforces the star topology. Toxic corporate leadership maps analogously: public praise and mentorship signals constitute idealisation, public criticism and credit-taking constitute devaluation, workplace gaslighting erodes precision, career tenure and vested options constitute the investment trap, and separate one-on-one management relationships enforce informational isolation. Exploitative medical relationships, exploitative academic supervision, and asymmetric

international relations each instantiate the same five-condition structure with domain-specific variable mappings.

The replicator dynamics analysis at the population level provides a formal explanation for the empirically documented overrepresentation of extractive strategies at senior management levels. The invasion condition shows that the extractive strategy (cooperate initially to build reputation, then exploit) outperforms honest cooperation in pairwise encounters with cooperators---extractive managers who appropriate credit and manage impressions upward are promoted faster. The population-level welfare erosion manifests as culture degradation: as the fraction of extractive strategists exceeds a critical threshold, aggregate organisational performance declines because cooperative surplus generation requires trust that extractive strategies systematically destroy.

15.7.4. The Thermostat Isomorphism

The substrate independence of the model is established rigorously through a structural isomorphism with a thermostat control system. A thermostat controlling a heater in a poorly insulated room satisfies all five axiomatic conditions: the heater extracts energy from the electrical supply while the room loses heat (K1); the thermostat switches between heating and cooling modes at hysteretic thresholds (K2); sensor drift degrades measurement accuracy, amplifying oscillations (K3); thermal inertia creates a form of sunk cost, as the room "remembers" its heated state after the source is removed (K4); and a hub-and-spoke architecture emerges when a single thermostat controls multiple heaters in informational isolation (K5).

The governing equations are structurally identical: room temperature and emotional energy are both scalar state variables governed by the balance between supply and loss; hysteretic switching produces the same limit cycle; sensor degradation amplifies oscillation in both systems. The period formula for the thermostat oscillation is structurally identical to the relational oscillator period formula. The correspondence is not metaphorical; it is mathematical. The relational system is a thermostat control system with hysteretic switching and degrading sensors. The mathematics is indifferent to the substrate.

15.8. Synthesis: Substrate Independence and the Persistence Framework

The extractive oscillator model connects deeply to the thesis framework at multiple levels.

The model is an instance of the viability framework of Chapter 5 and Chapter 6. Both the extractor and the source face the persistence problem: each must maintain itself within a viability set. The extractive oscillator arises when one agent's persistence strategy involves depleting another agent's viability resources. The oscillation is a structural necessity of the persistence constraints: permanent depletion violates the source's persistence conditions (the source exits or collapses), while permanent supply violates the extractor's optimality conditions (full empathic cost with increasing source autonomy). The limit cycle is the unique trajectory compatible with both agents' persistence constraints in the asymmetric extraction regime.

The sensor degradation mechanism connects to the Structural Distortion Principle (Chapter 9) and the predictive processing framework (Chapter 10). In the language of predictive processing, the extractor's signals function as persistent, systematically biased prediction errors that progressively overwrite the source's internal model. The phase transition at t* is the point at which the source's predictive model shifts from being anchored in internal evidence to being dominated by external (extractor-generated) prediction errors. The "reality collapse" upon separation (Corollary 5.1 of the original paper) is the predictive processing system's response to the sudden removal of the dominant input source---the system's world-model disintegrates because it was maintained by signals that are no longer present, and the internal precision is too degraded to reconstruct an autonomous model.

The investment trap connects to the D-distortion framework of Chapter 12. The biological decision programs that distort evaluative processes---hyperbolic discounting, loss aversion, status quo bias---are precisely the mechanisms that the sunk-cost trap exploits. Loss aversion makes the accumulated investment feel disproportionately costly to abandon. Hyperbolic discounting makes the immediate pain of exit loom larger than the long-term benefit of freedom. Status quo bias favours continued accommodation over the disruption of exit. These are not independent pathologies; they are specific instances of the evaluative system asymmetries that Chapter 12 characterises, operating within the structural context that the extractive oscillator creates.

The three-loop entrenchment structure provides a precise formal model of how multi-agent persistence dynamics can produce pathological lock-in. This is the multi-agent analogue of the self-maintaining attractor dynamics that Chapter 11 identifies in individual consciousness:

just as a conscious system can become trapped in a self-reinforcing interpretive loop (rumination, obsession, delusion), a multi-agent system can become trapped in a self-reinforcing extraction loop. The structural parallel is exact: in both cases, positive feedback drives the system toward a fixed point or limit cycle from which escape requires overcoming an increasingly large energy barrier.

The substrate independence of the model---the fact that thermostats, relationships, addictions, platforms, regimes, cults, and corporations all instantiate the same dynamical structure---is the strongest evidence for the thesis's central claim that persistence is the primary organising principle of complex systems. The extractive oscillator is not a metaphor borrowed from engineering and applied to psychology; it is a structural regularity that emerges whenever the five conditions are satisfied, regardless of the domain. The mathematical structure is invariant to the identity of the agents, the nature of the extracted resource, and the domain of interaction. Only the substrate differs.

CHAPTER 16. INSTITUTIONAL STRATEGIES AND REFLEXIVE INFERENCE

This chapter completes Part IV by examining institutional self-maintenance strategies, the ethics of idea transmission, and three structural laws governing adaptive systems. The connecting theme is that institutions, like organisms and cognitive systems, face the persistence problem—and their strategies for solving it are governed by the same formal constraints identified throughout this thesis.

16.1. Non-Participation and Pre-Integrative Rejection

The formalization of non-participation and early exclusion as stability-preserving institutional strategies (Kriger, 2026, https://doi.org/10.5281/zenodo.18135188) analyzes the strategy of preemptive exclusion: institutions maintain coherence by excluding elements before they are integrated, rather than expelling them afterward. The formal analysis shows that pre-integrative rejection is energetically cheaper and structurally less disruptive than post-integrative expulsion, because integration creates structural dependencies that must be severed during expulsion.

The article "Non-Participation and Early Exclusion as Stability-Preserving Institutional Strategies" (Kriger, 2026) addresses a persistent asymmetry in institutional economics: the discipline has devoted extensive theoretical attention to how institutions adapt, coordinate, and reform in response to external pressures, yet comparatively little formal work has examined when and why institutions rationally choose not to engage at all. The paper intervenes in this gap by developing a formal framework in which refusal, exclusion, and non-participation are reconceptualized — not as symptoms of institutional failure or cooperative breakdown, but as deliberate, stability-preserving equilibrium strategies available to institutions operating under specific structural conditions.

The motivation for this reframing arises from a set of observable empirical regularities. Voluntary associations routinely restrict membership to preserve organizational culture. Learned societies decline corporate funding to avoid mission drift. Religious orders limit external entanglements, and open-source communities reject certain sponsorships to prevent dependency. In each case, the institution forgoes revenue or cooperative opportunity, yet persists with greater internal coherence than comparable institutions that accepted such engagements. These are not isolated anomalies; they represent a systematic institutional

pattern that existing theory has struggled to accommodate within its predominantly engagement-oriented paradigm.

16.1. 1 Theoretical Framework and Formal Architecture

The paper defines an institution *S* as a triple *(H, R_min, O)*, where *H* represents fixed overhead obligations independent of any specific engagement, *R_min* is the minimal revenue threshold for institutional survival, and *O* denotes the set of operational functions. This formalization is deliberately minimal, which is itself a methodological choice worth discussing. By reducing the institution to its structural essentials — overhead burden, survival constraint, and operational capacity — the framework isolates the precise conditions under which non-participation becomes optimal, without introducing confounding variables that would obscure the core mechanism.

Each external engagement *c* is characterized by four parameters: revenue *R(c)*, interaction costs *I(c)*, coordination and governance costs *K(c)*, and institutional distortion costs *D(c)*. The total institutional cost of any engagement is given by $C(c) = I(c) + K(c) + D(c)$, and the net institutional value by $V(c) = R(c) - C(c)$. The critical analytical distinction in this framework lies in the nature of these cost components. Revenue is treated as additive and reversible — it can be gained, lost, and regained without permanent structural consequence. Interaction and coordination costs, while burdensome, are similarly bounded. Institutional distortion costs, however, are characterized as irreversible or semi-irreversible: they represent changes to rules, routines, dependencies, and internal structures that cannot be easily undone once incurred.

This asymmetry between reversible revenue and irreversible distortion is the engine of the entire analysis. It is what allows the paper to demonstrate, formally, that engagements generating positive revenue can nevertheless carry negative net institutional value. The irreversibility of distortion means that the institution, having accepted a destabilizing engagement, cannot simply return to its prior state by terminating the relationship. The damage persists.

16.1. 2 Core Propositions and Their Significance

The paper advances its argument through a sequence of lemmas and propositions that build cumulatively toward the main theorem.

Lemma 1 (Negative-Value Engagements) establishes the foundational claim: there exist engagements where $V(c) < 0$ despite $R(c) > 0$. This is not merely a theoretical possibility but

follows directly from the irreversibility property of distortion costs. When *D(c)* is sufficiently large, it overwhelms revenue regardless of the magnitude of *R(c)*. This lemma challenges a common implicit assumption in organizational decision-making — that revenue-positive engagements are, by definition, beneficial. The paper shows that this assumption holds only when distortion costs are negligible or reversible, conditions that frequently fail in practice.

Lemma 2 (Fixed Overhead and Coercive Acceptance) introduces a structural constraint that fundamentally alters institutional agency. When fixed overhead obligations *H* approach or exceed the minimal survival revenue *R_min*, the institution loses the capacity to refuse engagements based on their net institutional value. It becomes structurally compelled to accept any revenue-generating engagement, regardless of distortion consequences. This lemma draws on Williamson's (1985) analysis of the economic institutions of capitalism, where dependency erodes bargaining power. The implication is profound: high fixed overhead does not merely increase financial pressure — it collapses the institution's optionality, transforming strategic choice into survival-driven necessity. The institution can no longer afford to be selective, and this loss of selectivity is itself a form of institutional degradation.

Proposition 1 (Early Exclusion Dominance) demonstrates that for negative-value engagements, early exclusion weakly dominates delayed exit. The logic is straightforward but powerful: delayed exit accumulates interaction costs, coordination costs, and — most critically — irreversible distortion that persists even after the engagement is terminated. Early exclusion, by contrast, incurs at most a bounded opportunity cost (the foregone revenue). Since distortion is irreversible, every additional period of engagement adds permanent structural damage, while the revenue gained in each period remains reversible and can in principle be replaced by other means. The dynamic illustration in Appendix A formalizes this through a two-period model where engagement generates revenue R in period one but distortion D persisting into period two; when $D > R$, delayed exit produces negative cumulative value while early exclusion yields zero.

Lemma 3 (Operational Reversibility) introduces a moderating structural variable: institutions with a higher proportion of detachable or outsourced operations experience lower expected distortion from external engagements. This lemma draws on the path dependence literature, particularly Sydow, Schreyögg, and Koch (2009), to argue that coupling between engagements and core institutional structure is the primary mechanism through which

distortion propagates. When operations are modular and detachable, engagements affect peripheral functions without penetrating the institutional core. This has important implications for institutional design: it suggests that institutions can increase their resilience not only by refusing destabilizing engagements but also by structuring their operations to minimize the coupling through which distortion could propagate.

16.1. 3 The Main Theorem and Its Implications

Theorem 1 (Non-Participation Optimality) synthesizes the preceding results into the paper's central claim: under conditions of minimal fixed overhead, high operational reversibility, and optional interaction, institutional stability is maximized by preventing incompatible or negative-value engagements rather than attempting post-entry accommodation. The three conditions are jointly sufficient and each plays a distinct role. Minimal fixed overhead preserves the institution's capacity to refuse (Lemma 2). High operational reversibility reduces the expected distortion from any given engagement (Lemma 3). Optional interaction ensures that the institution faces no structural compulsion to engage. Together, these conditions define a regime in which the institution possesses both the freedom and the incentive to be selective, and in which selectivity constitutes the optimal strategy.

Corollary 1 (Crisis Amplification) adds a counterintuitive but important result: under financial stress, selective exclusion becomes more critical rather than less. This directly contradicts the intuition that institutions under financial pressure should lower their standards for engagement to secure revenue. The corollary shows that financial stress increases the relative weight of distortion costs — an already-stressed institution has less capacity to absorb structural damage, making each unit of distortion more consequential. Accepting a destabilizing engagement during a crisis may secure short-term survival at the cost of accelerated long-term degradation.

Relationship to the Existing Literature

The paper situates itself within several established theoretical traditions while extending each in distinctive ways.

The path dependence literature (Pierson, 2000; Sydow et al., 2009) provides the theoretical foundation for the irreversibility of institutional distortion. The paper's contribution is to formalize the conditions under which path-dependent lock-in can be *prevented* through anticipatory exclusion, rather than merely diagnosed after the fact. Most path dependence

research has focused on explaining why institutions become locked into suboptimal configurations; this paper shifts the focus to the institutional design features that allow institutions to avoid lock-in in the first place.

North's (1990) institutional analysis and Ostrom's (1990) work on governing the commons both emphasize adaptation and governance as mechanisms of institutional resilience. The paper does not reject these perspectives but identifies a complementary mechanism — boundary maintenance — that operates prior to and independently of governance. Where North and Ostrom ask how institutions manage the engagements they have accepted, this paper asks whether the engagement should have been accepted at all.

The paper explicitly engages with Buchanan's (1965) economic theory of clubs, which analyzes how voluntary organizations manage membership to optimize collective utility. The present framework extends Buchanan's analysis by formalizing the conditions under which exclusion is not merely utility-optimizing but stability-preserving — a distinction that matters when the relevant outcome is institutional survival and coherence rather than aggregate welfare.

Hirschman's (1970) framework of exit, voice, and loyalty is reinterpreted in a significant way. In Hirschman's original analysis, exit is primarily a signal of dissatisfaction — a response to organizational decline that provides information to decision-makers. In the present framework, exit and exclusion are reconceptualized as proactive institutional strategies rather than reactive signals. The institution does not exit because it is dissatisfied with an engagement; it excludes the engagement because admitting it would compromise institutional integrity. This shifts the analytical focus from the informational function of exit to its structural function.

Williamson's (1985) transaction cost economics provides the theoretical basis for Lemma 2's analysis of coercive acceptance under high fixed overhead. The paper extends Williamson's insight by showing that the loss of bargaining power associated with asset specificity and dependency applies not only to bilateral contracting relationships but also to the broader institutional capacity for selective engagement.

16.1. 4 Methodological Considerations

The paper adopts what it terms a "structural rather than normative" approach, explicitly disclaiming any claims regarding welfare, fairness, or optimal social outcomes. This

methodological positioning deserves careful consideration. On one hand, it preserves analytical clarity by isolating the stability-preservation mechanism from distributional and evaluative questions that would complicate the formal framework. On the other hand, the normative abstinence may limit the paper's practical applicability. Institutional decisions about exclusion inevitably carry welfare consequences — for excluded parties, for the institution's constituency, and for the broader system in which the institution operates. A fully developed theory of institutional exclusion would eventually need to integrate these dimensions, even if the present paper's structural focus represents a necessary first step.

The formal framework is deliberately stylized, using a small number of clearly defined parameters to derive results with minimal assumptions. This parsimony is a strength insofar as it makes the results transparent and generalizable, but it also means that certain empirically important complications are set aside. Real-world institutions face uncertainty about the distortion costs of prospective engagements, information asymmetries between institutional decision-makers and external parties, and dynamic environments in which the costs and benefits of engagement change over time. The two-period model in Appendix A provides a dynamic illustration but does not capture the full complexity of multi-period decision-making under uncertainty. These are natural directions for theoretical extension.

16.1. 5 Empirical Relevance and Applications

Although the paper does not undertake formal empirical analysis, the framework generates testable predictions and speaks directly to observable institutional phenomena across multiple domains.

In the context of **voluntary associations and membership organizations**, the framework predicts that organizations with lower fixed overhead and more modular operational structures will exhibit more restrictive membership criteria and greater long-term stability. This prediction could be tested comparatively across organizations of varying size and overhead structure within the same institutional field.

For **academic and research institutions**, the framework illuminates the logic behind selective refusal of funding sources. Universities and research institutes that accept funding with extensive reporting requirements, intellectual property restrictions, or agenda constraints may experience institutional distortion that persists long after the funding period ends — through established administrative procedures, redirected research priorities, and altered

faculty incentive structures. The framework predicts that institutions with lower fixed overhead (such as well-endowed private research institutes) will be more selective in their funding acceptance than those with higher overhead (such as publicly funded universities facing budget constraints), a pattern broadly consistent with observed behavior.

In the **open-source software community**, the framework explains why successful projects frequently reject corporate sponsorships that would provide substantial financial resources but introduce governance dependencies. The distortion costs — in the form of altered development priorities, community fragmentation, and dependency on continued corporate support — often exceed the revenue benefits, particularly for projects with minimal fixed overhead and highly modular architectures.

For **religious and cultural institutions**, the framework provides a structural explanation for practices of enclosure, limited external engagement, and resistance to commercialization that have historically been interpreted primarily through theological or cultural lenses. The structural analysis suggests that these practices, whatever their original motivation, function as stability-preserving mechanisms that protect institutional coherence against path-dependent degradation.

16.1. 6 Limitations and Directions for Future Research

Several limitations of the current framework merit acknowledgment and suggest productive directions for future work.

First, the framework treats distortion costs as exogenous to the institution's governance capacity. In practice, institutions may be able to reduce distortion costs through improved governance mechanisms, contractual safeguards, or organizational design. A more complete theory would model the interaction between governance capacity and the distortion potential of engagements, potentially identifying conditions under which accommodation dominates exclusion even for engagements with high raw distortion potential.

Second, the binary treatment of engagements — accept or exclude — does not capture the possibility of partial or conditional engagement. Institutions frequently negotiate the terms of engagement, accepting modified versions of proposals that reduce distortion risk while preserving revenue. Extending the framework to incorporate bargaining over engagement terms would increase its descriptive accuracy.

Third, the framework is static or at most two-period in its formal analysis. A fully dynamic model incorporating learning, reputation effects, and evolving institutional capabilities would provide richer predictions about how exclusion strategies develop and change over institutional lifetimes.

Fourth, the framework does not address the collective or systemic consequences of widespread non-participation. If many institutions simultaneously adopt exclusion strategies, the resulting pattern of disengagement may have systemic effects — on markets, on knowledge production, on public goods provision — that are not captured by the analysis of any single institution's optimization problem. This systemic dimension connects the paper's analysis to broader questions in institutional economics about the relationship between individual institutional strategies and collective outcomes.

Finally, the empirical predictions generated by the framework, while plausible and consistent with casual observation, have not been subjected to formal empirical testing. Systematic comparative studies of institutional longevity, engagement patterns, and structural characteristics would strengthen the paper's contributions and help establish the boundary conditions of its applicability.

Within the broader architecture of this thesis, the article makes several distinctive contributions. It provides a formal theoretical foundation for the claim that institutional boundaries are not merely descriptive features but active strategic instruments. It demonstrates that standard cost-benefit analysis applied to engagement decisions is systematically incomplete when it fails to account for irreversible institutional distortion. It identifies the structural preconditions — minimal fixed overhead, operational reversibility, optional interaction — under which non-participation constitutes an optimal strategy. And it reframes a set of widely observed institutional behaviors — selective membership, funding refusal, resistance to commercialization — as instances of a common underlying mechanism rather than unrelated organizational idiosyncrasies.

The paper's structural rather than normative approach aligns with the thesis's broader methodology of isolating institutional mechanisms before evaluating their welfare consequences. The formal results established here provide a foundation for subsequent chapters that may examine the distributional implications, the systemic effects, and the empirical validation of the non-participation framework.

16.1. 7 The companion result on the pre-integrative rejection principle (Kriger, 2026, https://doi.org/10.5281/zenodo.18463949) proves a deeper claim: the formal structure of pre-integrative rejection is identical in immune systems (antigen screening), cognitive systems (perceptual filtering), and social institutions (membership criteria). In each case, the system faces the problem of distinguishing elements that will contribute to viability from elements that will threaten it, and must make this distinction before the element is integrated into the system's structure.

The connection between the biological and the social is not metaphorical. The chapter proves that the formal structure is mathematically identical: a classification function maps candidate elements to acceptance or rejection, the function operates under uncertainty (the element's viability consequences cannot be known with certainty before integration), and the function is subject to type I and type II errors (false acceptance and false rejection) with asymmetric costs (false acceptance is typically more costly than false rejection, because integration creates dependencies that are costly to sever).

The analysis explains why institutions tend to be conservative in their admission criteria: the asymmetry of error costs favors rejection over acceptance when the consequences of integration failure are severe. It also identifies conditions under which excessive conservatism becomes maladaptive: when the environment changes rapidly, a system that rejects too many elements may fail to acquire the new components it needs to maintain viability in the changed environment.

16.2. Conceptual Responsibility

Conceptual responsibility (Kriger, 2014, https://doi.org/10.5281/zenodo.18488818) introduces an ethical dimension to the structural analysis. If ideas propagate through social systems with structural consequences—which the preceding analysis establishes—then the transmission of ideas carries a form of responsibility analogous to the responsibility associated with the transmission of physical goods. An idea that, when integrated into a social system, degrades that system's viability imposes a cost on the system, and the transmitter of that idea bears a share of responsibility for that cost.

This article (Kriger, 2014) investigates a problem that is widely observed but rarely formalized: the systematic degradation of ideas as they pass through chains of human

transmission. The paper draws on three independent domains — experimental psychology, cultural evolution theory, and structural analogies from information theory — to establish that significant distortion in mass communication is not merely possible but structurally expectable. From this empirical and theoretical foundation, it derives a normative framework grounded in responsibility-for-risk ethics, culminating in seven principles of Conceptual Responsibility that specify bounded authorial obligations.

16.2.1 Empirical and Theoretical Foundations

The article builds its descriptive case on Bartlett's (1932) serial reproduction paradigm and its modern replications, particularly Roediger et al. (2014), which demonstrated that transmission between individuals produces markedly greater information loss than repeated recall by the same individual. The observed patterns — shortening, rationalization, conventionalization, and assimilation toward receiver expectations — are shown to persist even when participants are explicitly instructed to be accurate, indicating that distortion reflects structural cognitive constraints rather than carelessness. The paper strengthens this empirical base with formal models from iterated learning theory, notably the Bayesian framework of Xu and Griffiths (2010), which proves that serial reproduction converges toward the prior distribution of receiver expectations regardless of the starting point. This convergence result is significant because it establishes that degradation is not random but directional: ideas are systematically pulled toward whatever the receiving population already believes.

The information-theoretic dimension of the argument is handled with appropriate caution. The paper explicitly disclaims any claim that Shannon's theorems apply directly to human cognition, instead arguing that structural parallels — finite processing capacity, interference, irreversible lossy compression — explain why similar dynamics emerge across different substrates. This careful positioning as productive analogy rather than mathematical identity addresses one of the most common criticisms of information-theoretic approaches to cognition while preserving the explanatory value of the framework.

16.2.2 The Normative Framework and the Seven Principles

The transition from descriptive findings to normative conclusions is grounded in a responsibility-for-risk model drawn from applied ethics (Jonas, 1984; Shrader-Frechette, 1991): when agents can foresee that their actions create significant risks of harm, they acquire obligations to mitigate those risks proportionate to foreseeability and severity. The paper

applies this principle to intellectual work by arguing that authors release ideas into transmission environments with known degradation properties, that degraded ideas can cause demonstrable harm, and that authors can take reasonable steps to mitigate this degradation.

The resulting seven principles are organized into two categories reflecting their theoretical origins. Principles 1 through 5 — Explicit Boundary Articulation, Anticipatory Misreading Analysis, Ethical Application Guidance, Active Discursive Engagement, and Structural Fragmentation Resistance — derive from the transmission dynamics established in the empirical and formal sections. Each addresses a specific degradation mechanism: boundary loss, convergence toward priors, loss of normative constraints, uncorrected drift, and extractability of dangerous fragments. Principles 6 and 7 — Epistemic Humility and the Downstream Responsibility Gradient — derive from social epistemology and the literature on testimony (Fricker, 2007; Lackey, 2008), addressing authority effects and the distribution of responsibility across transmission chains. The exponential decay model proposed for Principle 7 represents a plausible but not uniquely justified formalization of the intuition that authorial responsibility diminishes with transmission distance.

16.2.3 Limitations and Contribution to the Thesis

The paper acknowledges that transmission can also improve ideas through selection, refinement, and creative synthesis, and restricts its normative claims to contexts where fidelity matters and degradation constitutes harm. This qualification is important: the framework does not claim that all transmission is destructive, only that foreseeable degradation in specific contexts generates bounded obligations. The principal limitations include the heuristic rather than formally derived status of the idea decomposition model, the lack of empirical testing of the principles themselves, and the unresolved question of how to calibrate the foreseeability standard in practice.

Within this thesis, the article contributes a bridge between the structural analysis of institutional dynamics developed in preceding chapters and the ethical dimension of intellectual practice. It demonstrates that the same structural constraints that govern institutional stability also govern the transmission of the ideas on which institutions depend, thereby connecting institutional design to communicative responsibility.

16.3. Assertion–Dismantling Cycles

Assertion–dismantling cycles (Kriger, 2015, https://doi.org/10.5281/zenodo.18487230) article develops a mathematical framework for understanding a phenomenon observed across biological, cognitive, social, and organizational systems: the cyclical alternation between periods of structural consolidation (assertion) and periods of structural release (dismantling). Rather than treating breakdown phases as pathological, the paper formalizes conditions under which such phases are structurally beneficial — indeed, under which cyclic strategies provably dominate static ones. The framework models adaptive systems as networks of constraints that regulate accessible configurations, and analyzes how constraint density governs the tradeoff between performance, adaptability, and fragility.

16.3.1 The Constraint-Network Model and State Space Contraction

The paper defines an adaptive system as a tuple comprising state variables, a constraint weight matrix, constraint functions, and a fitness function. Structural order is measured by normalized constraint density — the proportion of active pairwise constraints relative to the maximum possible. This formalization is deliberately austere: under assumptions of linear constraints in generic position within a bounded domain, the paper proves that the accessible state space contracts exponentially with constraint density (Theorem 3.1). The log-volume of accessible configurations decreases linearly with structural order, meaning that each additional constraint eliminates a multiplicative fraction of the system's configurational freedom.

This result, while derived under idealized conditions that the paper explicitly acknowledges, captures a fundamental intuition: rules, procedures, beliefs, and dependencies progressively narrow what a system can do. The generic position assumption ensures that constraints are not redundant — each genuinely reduces dimensionality — and the linear constraint assumption permits exact volumetric bounds. The paper is careful to flag that real constraint systems likely involve nonlinear, correlated constraints, and that the qualitative conclusions may extend more broadly but require separate investigation.

16.3.2 Adaptive Capacity, Fragility, and the Critical Threshold

Building on the state space contraction result, the paper establishes that adaptive capacity — defined as the expected fitness improvement achievable through reconfiguration — decays at least as the square root of remaining degrees of freedom (Theorem 4.2). This result uses Dudley's entropy integral for Gaussian processes and connects the geometric contraction of

accessible space to the system's ability to respond to novel fitness landscapes. The implication is direct: as constraint density rises, the system becomes increasingly unable to exploit new opportunities, even when those opportunities exist in the environment.

The fragility analysis introduces a critical threshold separating two regimes. Below the threshold, fragility grows linearly with constraint density — additional constraints impose proportional costs. Above the threshold, fragility grows superlinearly, as perturbations propagate through a system-spanning rigid component rather than being absorbed locally. The paper presents this scaling transition as a conjecture (Conjecture 5.3) supported by analogy to rigidity percolation in network theory, and validates it numerically in a ten-dimensional toy model where a power-law fit yields a superlinearity exponent of approximately 1.4. The critical threshold itself is defined as the point where marginal fragility first exceeds marginal performance — a clean formalization of the intuition that there exists a level of structure beyond which additional rigidity costs more than it contributes.

16.3.3 Cyclic Dominance and Implications for the Thesis

The paper's central formal result (Theorem 6.5) establishes that when environmental shocks arrive with sufficient frequency, cyclic strategies — alternating between low and high constraint density — strictly dominate all static strategies. The proof exploits the geometry of the performance-fragility tradeoff: because performance is concave and fragility is convex in structural order, a time-averaged mixture of two structural states can achieve a point on the convex hull of the tradeoff curve that lies beyond what any fixed structural order can reach. The toy model demonstrates a concrete 6% fitness improvement of the optimal cyclic strategy over the best static alternative under moderate shock frequency.

The paper distinguishes three levels of claims with commendable transparency: rigorous theorems under explicit assumptions, scaling conjectures with heuristic support, and speculative application hypotheses for cognitive belief systems and organizational structures. The principal limitations include the linearity and independence assumptions on constraints, the Gaussian fitness landscape, and the absence of a full dynamical analysis of the cycling process itself.

Within this thesis, the article provides a formal mechanism connecting the institutional analysis of preceding chapters to the dynamics of structural renewal. The constraint-network framework offers a precise language for the claim that institutional rigidity — while

performance-enhancing up to a point — eventually demands periodic relaxation, and that such relaxation constitutes maintenance rather than failure. This reframes the non-participation and exclusion strategies discussed in earlier sections as mechanisms for managing constraint density within the viable range between dissolution and structural suffocation.

16.4 The Comparative Asymmetry Principle: Relational Disequilibrium in Multi-Agent Environments

The Comparative Asymmetry Principle (Kriger, 2026, https://doi.org/10.5281/zenodo.18462518) formalizes a phenomenon documented across social psychology, behavioral economics, and evolutionary biology: self-evaluating agents generate internal disequilibrium through observation of other agents, producing measurable behavioral perturbation distinct from deficit-driven motivation. The paper identifies five minimal structural properties sufficient to produce this perturbation, decomposes the phenomenon into three analytically separable levels, taxonomizes five resolution strategies with determinants of selection, and integrates the framework with two prior formal systems — the Viability Mismatch Law and the theory of self-organizing systems. The result is a domain-independent principle that unifies social comparison theory (Festinger, 1954), relative deprivation research (Runciman, 1966), and reference-dependent utility models (Kahneman & Tversky, 1979) under a general systems framework.

16.4.1 Structural Foundations and the Three-Level Decomposition

The paper's core formalization rests on five structural properties required of any agent for which the principle applies: state self-evaluation (P1), goal representation (P2), observation of others (P3), relational evaluation (P4), and state-dependent behavioral modulation (P5). These are deliberately substrate-neutral — they presuppose neither consciousness nor biological implementation — and their joint satisfaction defines the scope of the principle. The comparative asymmetry condition activates when the agent perceives another agent as closer to its desired state than itself and has not yet reached that state. Comparative perturbation (Π) is then defined operationally as the magnitude of behavioral change attributable to the relational comparison itself, measured as the divergence between the agent's behavior with and without comparative information, holding absolute discrepancy constant.

The three-level decomposition is one of the paper's most analytically productive contributions. Level 1 (behavioral perturbation, Π) requires only P1–P5 and is externally observable. Level 2 (internal resource mobilization, M) requires an additional property of differential internal resource allocation (P6) and is detectable through physiological or computational markers. Level 3 (subjective tension, Σ) requires phenomenal consciousness (P7) and is accessible only through self-report. The levels are connected by a bridging postulate — an empirical claim that the three are typically positively correlated in humans — but can dissociate, producing clinically and theoretically significant configurations such as latent perturbation ($\Pi \approx 0$, $M \gg 0$), which the paper identifies as an unstable suppression state vulnerable to sudden decompensation.

16.4.2 Environmental Amplification, Goal Drift, and Formal Integration

The concept of comparison density (C_d) — the salience-weighted rate of exposure to others' state information — provides the mechanism through which environments systematically amplify comparative perturbation. The environmental induction principle predicts that environments increasing C_d raise baseline Π across all qualifying agents regardless of the absolute state distribution, a prediction well supported by research on social media and subjective well-being (Appel et al., 2016; Verduyn et al., 2017). The paper's treatment of aggregation across multiple simultaneous comparison targets acknowledges both sub-additivity through habituation and super-additivity through extreme-outlier amplification, proposing a saturating functional form as the most plausible specification.

The goal drift analysis formalizes how comparison can become self-reinforcing. When the desired state R* is endogenous to observation — calibrated by what others achieve rather than internally anchored — a feedback loop emerges in which observation produces asymmetry, asymmetry produces perturbation, and perturbation shifts R* upward, widening the gap despite absolute progress. The paper identifies three regimes — stable, bounded, and runaway — with the runaway regime constituting a Red Queen dynamic that predicts energetic collapse when cumulative mobilization cost exceeds capacity. The double distortion pipeline, combining environmental signal asymmetry (exceptional states are broadcast while functional states are invisible) with internal reparameterization (agents with maladaptive processing further inflate perceived demands), provides a precise mechanism for the systematic upward bias in perceived asymmetry.

16.4.3 Resolution Strategies, Testability, and Contribution to the Thesis

The five resolution strategies — self-modification, other-directed interference, goal recalibration, observation withdrawal, and reframing — are ordered by a predictable adoption sequence, with perceived controllability identified as the strongest single predictor of strategy selection. The paper's integration with the theory of self-organizing systems provides a structural explanation for the instability of destructive strategies: in peer systems with repeated interaction, antagonistic behavior requires continuous external perturbation and is therefore not self-sustaining.

The framework generates an extensive set of falsifiable predictions, including the modulation of Π by visibility conditions holding discrepancy constant, the reduction of Π through information restriction without objective state change, and — most distinctively — the detection of emergent relational objectives in multi-agent reinforcement learning systems through sustained policy divergence in agents whose explicit reward functions contain no comparison term.

Within this thesis, the Comparative Asymmetry Principle contributes the micro-level mechanism that complements the institutional and structural analyses of preceding chapters. Where the non-participation framework (Section 16.1) addresses how institutions manage external engagements and the constraint-network framework (Section 16.3) models structural dynamics at the system level, the present paper specifies how individual agents within those systems generate endogenous demand through relational evaluation — connecting institutional-level phenomena to the behavioral dynamics of the agents who compose them.

16.5 The Reflexive Inference Law: Bounded Generalization from Self-Observation in Inferential Systems

The Reflexive Inference Law (Kriger, 2026, https://doi.org/10.5281/zenodo.18355847) establishes a fundamental limitation on self-knowledge: any inferential system's generalizations from self-observation are necessarily bounded. Specifically, the paper proves that no amount of introspection can tell a system more about its class than is already encoded in the mutual information between its instance parameters and the class parameters. This bound follows from the data processing inequality applied to self-referential inference under

a mediated observation assumption, and its consequences extend from the epistemology of introspection through the history of philosophy to the practical design of AI alignment mechanisms.

16.5.1 The Formal Framework and Main Result

The paper models a class of systems *C* containing an observer *S* characterized by class parameters θ_C (shared properties), instance parameters ϕ_S (properties specific to *S*), and self-observation x_self (data available through introspection). The central structural assumption — Mediated Observation — asserts that class parameters influence self-observation only through instance parameters, forming a Markov chain θ_C → ϕ_S → x_self. Under this assumption, the data processing inequality directly implies that the information self-observation carries about the class cannot exceed the mutual information I(ϕ_S; θ_C).

The main theorem (Theorem 9) provides the quantitative bound: posterior uncertainty about class parameters after self-observation is at least H(θ_C) − min{H^obs_max, I(ϕ_S; θ_C)}, identifying two independent bottlenecks. The observation bottleneck limits what can be extracted even from a highly representative observer; the representation bottleneck ensures that even unlimited self-observation cannot overcome fundamental dissimilarity between the observer and its class. Two corollaries sharpen the result decisively. The Structural Necessity of Subjectivity (Corollary 14) demonstrates that for any non-maximally-representative observer, residual uncertainty about the class is mathematically irreducible — subjectivity is not a correctable bias but a consequence of the observer's position within the class it seeks to understand. The Futility of Internal Processing (Corollary 15) establishes that no computational transformation of self-observation — no depth of reasoning, reflection, or metacognitive sophistication — can extract class-level information beyond what is already present in the raw structural relationship between observer and class.

16.5.2 Philosophical Implications and the Case for Methodological Transformation

The paper's most ambitious contribution is the systematic application of this formal result to the history of philosophy. From Descartes' cogito through Kant's transcendental method to Husserl's phenomenology, the Western philosophical tradition has repeatedly attempted to derive universal claims about mind, knowledge, and reason from introspection alone. The Reflexive Inference Law diagnoses this as a structural impossibility rather than a failure of rigor: each philosopher performed deep analysis of a single cognitive system and attempted

to generalize to the class of all possible knowers, without any evidential basis for assessing structural overlap between their own instance and the class.

The paper draws a productive contrast with Montaigne, whose essays practice introspective depth while exercising restraint about generalization — an implicit recognition of the bound that the present paper formalizes. Kant's case is examined in detail: his transcendental project required distinguishing necessary features of all cognition from contingent features of human cognition, yet the Reflexive Inference Law shows that no amount of transcendental reflection could have achieved this distinction without comparative data from structurally diverse cognitive systems.

The constructive response is a tripartite methodological requirement: valid class-level claims about cognition must incorporate wide collection of self-reports across diverse individuals, external meta-system observation through neuroscience and behavioral science, and systematic comparison with alternative intelligences — most critically, artificial intelligence. The emergence of AI is positioned not merely as intellectually interesting for philosophy of mind but as methodologically necessary for escaping the bounds of reflexive inference.

16.5.3 AI Alignment Applications and Contribution to the Thesis

The paper translates the abstract bound into operationalizable AI safety mechanisms: confidence scaling by estimated structural overlap, divergence monitoring between self-model predictions and observed class behavior, generalization entropy auditing to detect excessive confidence in class-level claims, and external specification of reference classes to prevent agents from gaming the framework. These proposals address a concrete risk: self-modeling AI agents that over-generalize from their own structure to claims about human values, preferences, or reasoning.

Within this thesis, the Reflexive Inference Law occupies a distinctive epistemological position. Where the preceding articles formalize institutional dynamics (Section 16.1), transmission constraints (Section 16.2), structural cycling (Section 16.3), and agent-level comparison mechanisms (Section 16.4), the present paper addresses the prior question of what any system can know about systems of its own kind through self-observation alone. It provides the epistemological foundation for the thesis's broader methodological stance: that claims about general properties of adaptive, cognitive, or institutional systems require comparative evidence across structurally diverse instances, and that formal bounds on self-referential

inference constrain not only what individual agents can know but what entire philosophical traditions can legitimately claim.

16.6 Structural Viability of Dyadic Systems

Kriger, B. (2026). *Structural Viability of Dyadic Systems: A Dynamical Account of Romantic Dissolution.* Zenodo. https://doi.org/10.5281/zenodo.18632250

16.6.1 Overview and Theoretical Orientation

Kriger (2026) presents a formal dynamical-systems account of romantic dissolution in long-term partnerships. The central thesis is that marriage can be modelled as a two-agent cooperative system whose persistence depends on the non-empty and dynamically invariant intersection of the partners' individual viability sets. When this intersection collapses — through parameter drift, exogenous shock, or endogenous decay — the system transitions from a viable to a non-viable state, and dissolution follows as a structural consequence rather than as an emotional or moral failure.

The paper is situated within Kriger's broader programme, the Unified Structural Theory of Complex Systems (Kriger, 2026a,b,c), and draws on viability theory, coupled differential equations, phase-transition models, game theory, attachment theory formalisation, and sigmoid threshold dynamics. Seven TikZ figures illustrate the viability-set geometry, reward decay curves, the Quasi-Westermarck curve, exit-barrier collapse, a mechanism-to-control-variable wheel, and parameter sweep sensitivity analyses. Five summary tables organise control variables, testable predictions, model comparisons, cross-domain correspondences, and case vignettes.

16.6.2 The Ten Core Mechanisms of Interest Attrition

The framework formalises ten mechanisms of interest attrition and demonstrates that each can be represented as a special case of a general structural law:

16.6.2.1 Viability Mismatch. The initial compatibility of two partners is recast as the overlap volume of their respective viability sets in a shared parameter space. Dissolution risk increases

as life-course changes — career trajectories, value shifts, health events — cause one or both sets to drift until the intersection becomes empty.

16.6.2.2 Reward-Signal Adaptation (Biochemical Habituation). Drawing on neurobiological evidence regarding dopaminergic reward circuits, the paper models the well-documented decline in romantic intensity as a monotonic decay in reward-signal amplitude. The partner, initially a source of high-magnitude positive signal, progressively converges toward affective baseline through standard habituation dynamics.

16.6.2.3 Quasi-Westermarck Familiarity-Induced Suppression. In an original theoretical contribution, Kriger proposes that long-term cohabitation activates a proximity-based kin-recognition heuristic analogous to the Westermarck effect, producing a gradual suppression of sexual attraction toward the co-resident partner. The paper formalises this as a sigmoid function of cumulative co-residence time and terms it the "Quasi-Westermarck curve."

16.6.2.4 Constraint–Autonomy Imbalance. Intimate partnership inherently restricts each partner's action space. The framework models the tension between the cooperative gains of commitment and the autonomy costs it imposes. When the perceived autonomy cost exceeds the perceived cooperative surplus, the system enters a dissolution-prone regime.

16.6.2.5 Social Contagion of Dissolution Norms. Divorce among proximate social contacts is modelled as a contagion process that lowers the perceived exit barrier for the focal dyad. As the proportion of dissolved partnerships in the reference group rises, the normative and informational costs of dissolution fall, increasing the probability that the focal system will also exit the viable regime.

16.6.2.6 Economic Cost–Benefit Asymmetries. The paper treats the economic dimension of marriage as a cost–benefit calculation in which the sign of the net payoff can reverse over time — for example, as earning differentials change, assets accumulate, or divorce-law regimes alter the expected financial outcome of dissolution.

16.6.2.7 Children as Viability-Set Anchors. Children are formalised as exogenous parameters that expand the joint viability set by introducing shared long-term utility functions and raising the perceived cost of exit. However, the paper also notes that this anchoring function can trap the system in a non-viable state — maintaining formal union while the internal dynamics have already crossed the dissolution threshold.

16.6.2.8 Media and Social Media Norm Distortion. Media environments systematically distort the reference distributions against which partners evaluate their own relationship. Idealised portrayals inflate expectation parameters; social-media comparison mechanisms further compress the perceived viability set by raising the threshold for subjective adequacy.

16.6.2.9 Perceptual Valence Switching. Traits initially perceived as attractive can undergo a sign reversal in affective valence — confidence becomes arrogance, spontaneity becomes irresponsibility — without any change in the underlying behaviour. Kriger models this as a threshold-mediated switch in the perceptual mapping function, linking it to broader work on signal reinterpretation under changed contextual parameters.

16.6.2.10 Serial Monogamy as Parameter Renewal. The transition to a new partner is modelled not as failure but as a parameter-renewal strategy: the new relationship resets reward-signal amplitude, eliminates accumulated familiarity suppression, and restores the viability-set intersection. Serial monogamy thereby emerges as the structurally expected equilibrium for a system subject to the preceding nine decay mechanisms.

16.6.3 Extended Analyses

Beyond the ten core mechanisms, the paper develops several substantial extensions that situate romantic dissolution within broader biological, cultural, technological, legal, institutional, and normative contexts.

16.6.3.1 Evolutionary Context. Kriger surveys comparative primate mating systems and the phylogenetic origins of social monogamy, integrating Fisher's "four-year itch" hypothesis and cross-cultural data on serial monogamy drawn from 58 societies. The adaptive logic of pair-bond cycling is thereby situated within the broader mammalian reproductive landscape.

16.6.3.2 Children: Trauma, Intergenerational Transmission, and Instrumentalisation. This extended treatment positions children as the most consequential stakeholders in the dissolution process. Parental divorce is situated within the Adverse Childhood Experiences (ACE) framework, incorporating Amato's meta-analytic evidence on child outcome deficits across academic, psychological, and social domains. The intergenerational transmission of divorce — the so-called "divorce cycle" — is formalised through three channels: interpersonal behaviour deficits (identified as the largest mediator), insecure attachment working models, and attitudinal normalisation of dissolution. Developmental pathways from childhood adversity to adult personality pathology are classified by their effects on the agent's parameter

vector: anxious attachment produces hypervigilance and autonomy contraction; avoidant attachment produces emotional withdrawal and underinvestment; disorganised attachment produces high-amplitude extractive oscillation; and externalising pathology manifests as substance abuse and conduct disorder. A final subsection analyses the strategic instrumentalisation of children, including parental alienation as epistemic attack, deployment of children as informants, and emotional extortion through custody leverage.

16.6.3.3 Digital Disruption. The paper examines several contemporary technological forces reshaping the parameter landscape of intimate partnership: high-fidelity pornography and its role in reward-threshold inflation; AI companion chatbots as synthetic alternative attractors; dating applications and the paradox of infinite perceived alternatives; consensual non-monogamy as system dimensionality expansion; and companion animals as partial attachment substitutes.

16.6.3.4 Extractive Oscillator Dynamics as a Continuum. Drawing on the formal extractive oscillator framework developed in a companion paper (Kriger, 2026), this section argues that quasi-narcissistic dynamics are not a categorical pathology but a universal feature of intimate dyads, varying across couples only in degree, amplitude, and asymmetry. An asymmetry index is introduced and three structural regimes are characterised.

16.6.3.5 Religion and Secularisation. A cross-religious analysis spanning Christianity, Judaism, Islam, Hinduism, and Buddhism examines marriage as sacred institution, formalising the "sacred exit barrier" as a transcendent cost component. The progressive removal of this barrier through secularisation is modelled as the elimination of the historically dominant stabilisation mechanism.

16.6.3.6 The Legal System as Structural Force. Adversarial family law is analysed through game theory as a prisoner's dilemma that selects for mutual aggression over cooperative settlement. The section formalises temporal distortion (forced residence in non-viable state space during prolonged proceedings), financial extraction as secondary dissolution, custody weaponisation, and jurisdictional arbitrariness as exogenous stochastic perturbation. The no-fault revolution is interpreted as the legal analogue of secularisation. Mediation and collaborative divorce are reframed as mechanism-design interventions, and a Nash bargaining solution for cooperative dissolution is derived.

16.6.3.7 Same-Sex Dyads. The paper demonstrates that the viability framework requires no structural modification for same-sex couples — the mechanisms are identical, with only the empirical parameter values differing. Dissolution-rate asymmetries between female–female, male–male, and different-sex couples are interpreted through exit-barrier levels, constraint–autonomy dynamics, threshold asymmetries, and minority stress as exogenous viability-set contraction. The analysis draws on Kurdek's longitudinal programme and Scandinavian registry data.

16.6.3.8 Domestic Violence, Institutional Response, and Strategic Weaponisation. Genuine intimate partner violence is formalised as high-amplitude extractive oscillation with three artificial exit-barrier mechanisms: terror, epistemic degradation, and resource control. A comparative institutional analysis examines Russia's 2017 decriminalisation as protective-barrier removal, the United States' mandatory arrest paradoxes (including Sherman's Milwaukee experiment), and Canada's mandatory charging and its unintended victim criminalisation. The strategic weaponisation of allegations is formalised as first-mover advantage in a game-theoretic framework. The dual crisis of under-protection and over-intervention is identified as a common structural failure: one-dimensional binary classification where multi-dimensional assessment is required.

16.6.3.9 Precursors of Dissolution. Gottman's empirical programme — the Four Horsemen, cascade model, and two-factor dissolution timing — is integrated with the viability framework, yielding a seven-step structured diagnostic hierarchy.

16.6.3.10 The Normative Question. The paper concludes with a balanced treatment of whether marriages should be "saved" or dissolution accepted as the evolutionary norm, accompanied by five structural projections for the future of intimate partnership over the next two to five decades.

16.6.4 Methodological and Formal Apparatus

The mathematical toolkit deployed across the paper is unusually broad for relationship science. Viability theory provides the overarching geometric language; coupled differential equations model the co-evolution of partner states; phase-transition models capture the discontinuous character of dissolution; game theory (prisoner's dilemma, Nash bargaining) structures the strategic interactions around exit, custody, and legal proceedings; attachment

theory is given a parametric formalisation; and sigmoid threshold dynamics underpin the perceptual valence-switching and Quasi-Westermarck mechanisms.

The framework generates falsifiable predictions, summarised in a dedicated table, and provides explicit quantitative criteria for model comparison and empirical testing. The structural-law approach — deriving each mechanism as a special case of a general law — imposes internal coherence and facilitates cross-domain analogies.

16.6.5 Critical Appraisal

Strengths. The paper's principal contribution is its unification: by demonstrating that ten apparently heterogeneous mechanisms of romantic dissolution can each be derived as special cases of a single structural law, Kriger achieves a parsimony that is rare in relationship science. The extension of the framework to encompass legal systems, religious institutions, digital technologies, same-sex couples, domestic violence, and child developmental outcomes gives the model unusual breadth. The formal apparatus — viability theory, game theory, phase transitions — provides a rigour that complements, and in several places extends, the largely empirical traditions of Gottman and Amato. The treatment of the extractive oscillator as a dimensional continuum rather than a categorical pathology is a particularly original contribution, as is the Quasi-Westermarck hypothesis.

Limitations. The framework is, at the time of publication, primarily theoretical. While the paper generates falsifiable predictions, it does not itself present empirical tests. The breadth of the model — spanning neurobiology, evolutionary psychology, law, religion, game theory, and technology — raises the question of whether a single formal apparatus can do justice to the domain-specific complexities of each area. Some of the extended analyses (for example, the comparative institutional treatment of domestic violence policy across Russia, the United States, and Canada) would benefit from more granular engagement with the relevant policy literatures. The reliance on a single-author framework (the Unified Structural Theory of Complex Systems) means that the paper's theoretical foundations have not yet been subjected to extensive independent scrutiny.

Situating the Contribution. Within the landscape of formal approaches to relationship dynamics — including Gottman's mathematical modelling (Gottman et al., 2002), Levinger's barrier model (1976), and Rusbult's investment model (1980) — Kriger's framework is distinguished by its generality and its commitment to deriving diverse phenomena from a

unified set of structural principles. Whether this generality proves empirically productive or remains a largely deductive exercise will depend on the programme of testing that the paper calls for but does not itself undertake.

16.7 Motivational Structures Under Variable Constraints: A Dynamical Systems Formalization

The preceding chapters of this thesis have established persistence, viability, and self-organization as structural necessities governing complex systems across physical, cognitive, and social domains. A recurring principle throughout has been that any system persisting in a changing environment must allocate finite resources among competing demands under constraint. This section summarises the contribution of a dedicated paper — *From Repression to Saturation: A Dynamical Systems Theory of Motivational Structures Under Variable Constraints* (Kriger, 2026) — https://doi.org/10.5281/zenodo.18685856 which provides the detailed mathematical infrastructure for the **motivational layer** of the unified theory: the formal dynamics of desire, attention, and drive allocation in resource-bounded agents embedded in variable environments.

16.7.1 Motivation and Core Formalism

The paper begins from the observation that existing approaches to the mathematics of motivation — from Lewin's topological psychology to Friston's free-energy principle — each capture important aspects of desire dynamics but none provides a unified, axiomatically grounded, regime-sensitive theory. The paper addresses this gap by constructing a formal framework from five axioms defined on a smooth motivational state space. The central constructs are a motivational potential (encoding the combined effects of stimulus attractiveness, normative constraint, and intrinsic drive), a directed motivational density (measuring how concentrated an agent's motivational energy is across possible objects of desire), and a coherence measure that quantifies the degree of motivational focus.

The foundational result is the **Potential-Difference Principle**: sustained, directed desire requires a gradient in the motivational landscape. If the potential is flat — if all objects of desire are equally accessible and no constraints differentiate among them — the system degenerates into diffusion, the formal analogue of apathy. This principle is developed in three complementary representations (Ohmic, thermodynamic, and Lyapunov), consistent with the

thesis's Formalization Laws, which predict that multiple complementary formalizations are structurally necessary for complex phenomena.

16.7.2 Sublimation, Saturation, and Phase Structure

Building on the axiomatic core, the paper formalizes two mechanisms of central importance. **Sublimation** is reinterpreted not as moral rerouting (the classical Freudian reading) but as an entropy-reducing transformation: the conversion of a diffuse, high-entropy motivational distribution into a concentrated, low-entropy one through sustained engagement with structured activity. A sublimation collapse theorem proves that this transformation becomes impossible when the environment permits instantaneous discharge of every impulse — a result determined entirely by timescale competition between discharge and reorganization, not by the agent's character or intention. This constitutes a structural prediction that locates the cause of sublimation failure in environmental parameters rather than in personal deficiency.

Stimulus saturation is modelled through Hill-equation receptor kinetics borrowed from pharmacology. The response of the motivational system to stimuli follows a saturation curve: at low stimulus intensity the response is steep, at high intensity it plateaus, and under chronic overstimulation the response ceiling itself degrades. This double degradation — higher threshold, lower ceiling — provides the formal mechanism for progressive desensitization under conditions of stimulus abundance.

These mechanisms are synthesised into a **phase diagram** with four regimes parametrised by stimulus accessibility and normative constraint density: Repression (high constraint, low access), Coherent Desire (moderate constraint, moderate access), Saturation (low constraint, high access), and Anhedonic Collapse (chronic saturation with degraded response capacity). The paper proves the existence of critical boundary curves separating these regimes and identifies each regime with characteristic mathematical signatures in the potential landscape.

16.7.3 Co-Regulation, Artificial Intelligence, and Universality

The paper extends the framework in three directions that connect directly to other parts of this thesis. First, interpersonal co-regulation is modelled as Kuramoto coupled-oscillator synchronization, proving that deep interpersonal satisfaction requires coupling strength above a critical threshold — a result that formalises why screen-mediated interaction, which structurally sets coupling to zero, cannot replicate embodied intimacy. Second, the framework is applied to artificial intelligence, establishing a structural correspondence table that maps

every motivational concept (potential, coherence, sublimation, saturation) onto its AI analogue (loss function, task convergence, representation learning, reward flattening). Third, the framework's universality is tested across scales from bacterial chemotaxis to astrophysical self-organization, demonstrating that the gradient principle operates wherever resource-bounded systems evolve on potential landscapes.

16.7.4 Contribution to the Thesis

This paper fills a specific and previously unoccupied position within the thesis's architecture. The unified theory identifies that any persistent agent must solve the resource-allocation problem — distributing finite capacity among competing demands under environmental constraint. The motivational dynamics formalized here are the concrete mathematical realization of this abstract requirement in the domain of desire and drive. The four regimes of the phase diagram instantiate the thesis's **Constraint–Autonomy Compatibility Law**: Regime I (excessive constraint) and Regime III (excessive autonomy) are both pathological, and the viable Regime II requires their balanced co-presence — precisely the co-constitutive relationship the law predicts. The sublimation operator instantiates the thesis's self-organization laws in the motivational domain, and the saturation dynamics instantiate the **Viability Mismatch Law** (demands exceeding the system's organized response capacity). The paper thus demonstrates that the structural principles derived at the highest level of generality in this thesis yield specific, testable, quantitative predictions when applied to the psychology of desire — confirming the thesis's central claim that persistence, not any domain-specific mechanism, is the organising principle of complex systems.

PART V. APPLICATIONS AND THE ARCHITECTURE OF CIVILIZATION

CHAPTER 17. ASTROPHYSICAL AND INFORMATION-THEORETIC APPLICATIONS

Part V demonstrates the predictive power of the unified structural theory across two broad domains: astrophysical systems and the architecture of information, civilization, and agency. This chapter addresses astrophysical applications and the theory's engagement with fundamental questions about time, complexity, and the structure of the universe.

17.1. Binary Star Formation as Persistence-Selected Outcome

The theory of persistence-driven binary dominance (Kriger, 2025–2026, beginning with https://doi.org/10.5281/zenodo.18155356) argues that binary star systems are not merely common but structurally favored by the persistence principle. The argument is that gravitationally bound binary systems have larger viability sets—they can absorb a wider range of perturbations without losing structural coherence—than single-star systems of comparable mass.

The series of papers develops this argument quantitatively. The theoretical framework (Kriger, 2026, "Why Binary Systems Are Optimal for Star Formation," https://doi.org/10.5281/zenodo.18144257) establishes that binary configurations maximize the ratio of gravitational binding energy to perturbation vulnerability. The swept-volume geometry analysis (Kriger, 2026, https://doi.org/10.5281/zenodo.18155283) computes the overlap corrections in protostellar binary accretion that determine the efficiency advantage of binary accretion over single-star accretion. The "Survival of the Bound" paper (Kriger, 2026, https://doi.org/10.5281/zenodo.18155400) demonstrates that in a competitive accretion environment, gravitationally bound pairs outcompete unbound configurations in the accumulation of mass from the surrounding medium.

The paradox of protostellar core formation (Kriger, 2026, https://doi.org/10.5281/zenodo.18164699) addresses a classical problem in star formation theory: how protostellar cores condense out of molecular clouds against the disruptive effects of turbulence and magnetic fields. The persistence framework provides a resolution: core formation succeeds when the proto-core's viability set (the range of perturbations it can absorb without dispersing) exceeds the perturbation regime of its environment. Binary configurations, by having larger viability sets, are more likely to survive the condensation process.

The observational predictions are specific and testable. The theory predicts that the binary fraction should increase with the environmental perturbation regime: in more turbulent environments, single stars should be rarer because their smaller viability sets make them more vulnerable to disruption. The paper on observational tests (Kriger, 2026, https://doi.org/10.5281/zenodo.18300562) reviews the available evidence and identifies the surveys that could provide definitive tests.

17.2. Dormant Neutron Stars and the Galactic Dark Population

The analysis of dormant neutron stars (Kriger, 2025, https://doi.org/10.5281/zenodo.18409103) applies the viability framework to the final evolutionary phase of stellar systems. A neutron star that has exhausted its rotational energy and is no longer accreting matter becomes undetectable by current observational methods. The viability framework characterizes this as a system that has reached the minimal-energy boundary of its viability set: the star persists (it exists as a stable configuration of degenerate matter) but has no external signatures.

The analysis estimates the galactic population of such dormant neutron stars and discusses the implications for dark matter accounting, gravitational wave source demographics, and the completeness of current stellar population models. The formal contribution is the demonstration that the persistence framework generates quantitative predictions in astrophysics—predictions that are independent of and complementary to those generated by standard astrophysical theory.

17.3. Time, Entropy, and the Complexity of the Universe

The ledger time model (Kriger, 2025, https://doi.org/10.5281/zenodo.18263726) proposes that time emerges from irreversible entanglement validation in an atemporal quantum configuration space. This connects the atemporality thesis of Chapter 11 (biological memory is stored without intrinsic temporal markers) to the fundamental physics of time: if time is emergent rather than fundamental, then the atemporality of memory is not merely a biological adaptation but a reflection of the deeper structure of reality.

The estimation of biospheric and synthetic contributions to effective complexity (Kriger, 2026, https://doi.org/10.5281/zenodo.18261863) applies information-theoretic tools to estimate the fraction of the universe's total effective complexity that is contributed by biological and synthetic cognitive systems. The analysis reveals that although cognitive

systems occupy an infinitesimal fraction of the universe's volume, they contribute a disproportionately large fraction of its effective complexity—complexity that is organized, functional, and causally efficacious rather than merely statistically random.

The holographic universe evaluation (Kriger, 2025, https://doi.org/10.5281/zenodo.18101919) and the Bayesian model comparison framework (Kriger, 2026, https://doi.org/10.5281/zenodo.18230310) provide case studies in the application of the structural-Bayesian methodology developed in Chapter 4. These analyses demonstrate how the epistemic constraint framework can be used to evaluate cosmological hypotheses that are extrapolative—that make claims beyond the domain of direct observational evidence—by assessing their structural coherence, their consistency with established physics, and their vulnerability to formalization-dependent artifacts.

CHAPTER 18. AI, CIVILIZATION, AND THE FUTURE OF AGENCY

The final chapter applies the unified structural theory to the present and near-future dynamics of human civilization, with particular attention to the transformative effects of artificial intelligence. Each result is derived from the formal framework developed in previous chapters; the chapter demonstrates that the theory generates specific, non-trivial predictions about the trajectory of technological civilization.

18.1. Clinical Discontinuity and AI

Clinical discontinuity and AI (Kriger, 2024, https://doi.org/10.5281/zenodo.18395736) diagnoses a specific problem in contemporary healthcare: the fragmentation of clinical information across providers, systems, and time creates discontinuities that degrade the quality of care. Each provider operates with an incomplete model of the patient—a direct instance of the epistemic constraints of Chapter 4—and the discontinuities between these incomplete models produce errors of omission, duplication, and contradiction.

AI systems can potentially bridge these discontinuities by integrating information across providers and time. The formal analysis, however, identifies a structural risk: an AI system that integrates information without understanding the viability framework within which clinical decisions are made may optimize for consistency at the expense of clinical appropriateness. The analysis provides formal criteria for beneficial AI integration: the AI must not merely aggregate data but must model the viability constraints of the patient (the patient's physiological viability set) and the clinical decision-making process (the provider's epistemic viability set).

18.2. AI-Extended Agents and the Transformation of Communication

AI-extended agents and the transformation of human communication (Kriger, 2026, https://doi.org/10.5281/zenodo.18521341) uses game theory to model the norm shift that occurs when a population begins using AI-mediated communication. The central result is that AI-mediated communication changes the equilibrium of the communication game: when agents can use AI to optimize their messages for persuasive effect, the information content of messages declines (because messages are selected for their effect on the receiver's model rather than for their accuracy), and the interpretive burden on receivers increases (because receivers must now distinguish AI-optimized messages from unmediated ones).

The analysis identifies a new form of arms race: senders optimize messages for persuasive effect using AI, receivers develop AI-assisted tools for detecting and filtering optimized messages, senders respond with more sophisticated optimization, and so on. The equilibrium of this arms race depends on the relative computational resources available to senders and receivers—a parameter that the formal analysis relates to the extractive oscillator model of Chapter 15, with AI-optimized communication functioning as a potential extraction mechanism.

18.3. The Stimulus Problem in Post-Scarcity Environments

The stimulus problem (Kriger, 2026, https://doi.org/10.5281/zenodo.18511908) formalizes the challenge of goal generation in post-scarcity information environments. When basic needs are satisfied and information is abundant, the problem of what to pursue—the problem of generating meaningful goals—becomes the central challenge of individual and collective life.

The formal analysis derives this problem from the viability framework. In scarcity environments, goals are determined by viability constraints: the system must acquire resources, avoid threats, and maintain its position within V. These constraints generate an unambiguous goal structure. In post-scarcity environments, the viability set expands to encompass almost all available states—the system is not threatened from any direction—and the goal structure degenerates. Without viability pressure, the system has no formal basis for preferring one action over another. This is not a psychological observation about boredom or anomie; it is a mathematical consequence of viability set expansion.

The analysis identifies stimulus generation—the creation of artificial viability constraints that restore goal structure—as the adaptive response. Games, challenges, creative projects, and self-imposed constraints all function as artificial viability constraints that shrink the effective viability set and restore the goal structure that persistence pressure provides in scarcity environments. The formal analysis connects this to the Constraint–Autonomy Compatibility Law: just as autonomy requires enabling constraints, meaningful agency requires viability constraints that define a space of meaningful choices.

18.4. The Inward Turn and the Fermi Paradox

The inward turn (Kriger, 2026, https://doi.org/10.5281/zenodo.18437440) addresses the Fermi paradox—the apparent contradiction between the high probability of extraterrestrial civilizations and the absence of observable evidence for them. The theory proposes that

advanced civilizations converge on computational efficiency as a universal attractor: as a civilization's capacity for information processing increases, the viability-optimal strategy shifts from external expansion (colonizing physical space) to internal development (optimizing computational processes within a fixed physical volume).

The formal argument derives this convergence from the persistence principle. External expansion increases the system's exposure to perturbations (a larger physical footprint is more vulnerable to disruption) while computational optimization reduces the system's resource requirements per unit of functional complexity. Beyond a threshold of computational sophistication, the viability calculus favors inward development over outward expansion. Civilizations that reach this threshold undergo an "inward turn"—they become computationally intensive but physically inconspicuous, which is why they are undetectable by observational astronomy.

The result does not prove that this is the resolution of the Fermi paradox; it derives the inward turn as a formal consequence of the persistence principle and shows that it provides a structurally coherent resolution. The prediction is testable in principle: if the inward turn is correct, advanced civilizations should be detectable not by their electromagnetic emissions but by their gravitational signatures (they occupy fixed physical volumes) and their information-theoretic signatures (they process information at rates incompatible with natural processes).

18.5. The Inevitability of a Unified Civilization of Autonomous Agents

The inevitability of a unified civilization of autonomous agents (Kriger, 2026, https://doi.org/10.5281/zenodo.18512941) argues that as AI systems achieve genuine autonomy, the distinction between biological and synthetic agents becomes structurally irrelevant. The formal argument is that the viability framework applies identically to any autonomous agent—biological or synthetic—and that the Laws of Cooperation, Interference, and Observability govern the interactions of all autonomous agents regardless of substrate.

The analysis predicts that the long-term trajectory of civilizational development leads toward a unified system of autonomous agents that includes both biological and synthetic members, interacting under the same formal constraints and subject to the same structural laws. The "unified civilization" is not a utopian vision; it is a structural prediction about the

equilibrium of multi-agent dynamics when the agent population includes both biological and synthetic members with comparable cognitive capacities.

18.6. Viral Dynamics as Information Propagation

The informational framework for population-scale viral dynamics (Kriger, 2022, https://doi.org/10.5281/zenodo.18392769) provides a final application that illustrates the breadth of the theory. Viral epidemics are modeled as information-propagation phenomena, where the persistence framework applies to both the virus (which must maintain its viability through replication) and the host population (which must maintain its viability through immune response and behavioral adaptation). The interaction between viral and host persistence strategies generates the characteristic dynamics of epidemics: exponential growth, peak, decline, and potential re-emergence—all derivable from the viability framework applied to two competing persistence strategies.

This application completes the demonstration of the theory's scope. From the meta-theoretical constraints on formalization to the dynamics of stellar systems, from the architecture of consciousness to the future of civilization, from the mechanisms of addiction to the epidemiology of infectious disease—the unified structural theory of complex systems provides a formal framework that generates specific, testable predictions across the full range of organized complexity.

PART VI. THE EPISTEMOLOGY OF INCOMPATIBILITY AND THE LIMITS OF DESCRIPTION

CHAPTER 19. INCOMPATIBLE TRUTHS AND THE NAVIGATION OF IRREDUCIBLE PLURALISM

Introduction

The preceding parts of this dissertation established a unified structural framework centred on persistence as the organising principle of complex systems. The laws of formalisation, the persistence theorem, and the scale-specific principles collectively demonstrated that the viability of any sufficiently complex system depends on the interplay between compression, uncertainty, and structural adaptation. Part VI extends this programme into epistemology proper, asking what happens when the formal apparatus developed in earlier chapters is turned upon the act of description itself.

The central question of this chapter is whether the plurality of valid but mutually contradictory descriptions—observed across physics, biology, cognitive science, and normative discourse—is a contingent feature of our current ignorance or a structural necessity imposed by complexity. Four publications are synthesised here, each addressing a distinct facet of this question.

The Theorem of Incompatible Truths: Why Any Sufficiently Complex Domain Necessarily Admits Mutually Exclusive but Individually Valid Descriptions That Cannot Be Unified (Kriger, 2026) establishes the central formal result: a proof that any system satisfying four minimal axioms above a complexity threshold necessarily admits pairs of individually valid but mutually exclusive descriptions. *The Pragmatic Synthesis: How Complex Systems Navigate Incompatible Truths through Adaptive Heuristics* (Kriger, 2026) asks the constructive question—how agents embedded in such systems manage to function—and derives a formal structure, the navigator, whose properties coincide with those attributed to consciousness in the scientific literature. *Adaptive Observer Theory: Navigation of Incompleteness and Incompatibility* (Kriger, 2026) consolidates both results into a monograph-length treatment, adding falsifiable predictions, a systematic objections-and-responses apparatus, and a glossary situating the framework within the broader landscape of physics, cognitive science, and philosophy of science. *Deep Disagreements as Incompatible Truths: A Formal Framework and Constructive Response* (Kriger, 2026) applies the formal framework to the problem of deep disagreements in argumentation theory, demonstrating that Fogelin's observation of irresoluble normative disputes is a special case of the incompatibility

theorem and that a constructive navigation strategy is available where the existing literature offers only diagnosis.

Together, these four works establish that irreducible pluralism is not a deficiency of knowledge but a structural signature of complexity, and that the appropriate response is not unification but adaptive navigation.

19.1. The Theorem of Incompatible Truths

19.1.1. Formal Apparatus

The theorem, developed in *The Theorem of Incompatible Truths*, is constructed from five primitive notions—domain, description, validity, compatibility, and complexity—each defined with the minimum structure necessary for the subsequent proofs and deliberately independent of any particular physical theory.

A *domain* $\Omega = (\Sigma_\Omega, D_\Omega, t, O_\Omega)$ is a quadruple consisting of a state space, a dynamics satisfying the semigroup property, a totally ordered time structure, and a collection of stable observables—real-valued functions on the state space that exhibit reproducible, predictable behaviour. Two domains *overlap* when they share at least one stable observable via embedding maps satisfying dynamical compatibility on the overlap state space. A *description* $D = (L, A, \vdash)$ is a finitely axiomatised formal theory whose language is rich enough to make predictions about the domain's observables. A description is *valid* if it is internally consistent and every prediction it makes about stable observables is empirically correct. Two descriptions are *compatible* if there exists a unified description D_U whose deductive closure contains both deductive closures restricted to their respective domains; otherwise they are *incompatible*.

Complexity is measured by four parallel metrics: Kolmogorov complexity, effective complexity in the sense of Gell-Mann and Lloyd, logical depth in the sense of Bennett, and—original to this work—*descriptive diversity* $\delta(S)$, defined as the number of logically non-equivalent descriptions required by a system. The complexity threshold is exceeded when $\delta(S) \geq 2$ and the system's effective complexity is not reducible to that of any single domain.

19.1.2. The Four Axioms

The incompatibility thesis does not follow from the primitive notions alone. Four axioms are required, each capturing a minimal and physically natural constraint on how descriptions relate to the systems they describe.

Descriptive Completeness requires that every valid description make at least one non-trivial prediction about every stable observable in its domain. This prevents two descriptions from trivially avoiding contradiction by dividing the observables between them. *The Compression Imperative* requires that the algorithmic complexity of a description be strictly less than that of a brute-force enumeration of all observable values—formalising the foundational insight that a genuine theory compresses rather than catalogues. *Scale Coupling* requires that in systems above the complexity threshold, at least two domains overlap, sharing stable observables. *Non-Degeneracy* requires that the system contain at least two domains whose valid descriptions have non-isomorphic deductive closures.

The four axioms are logically independent: for each axiom, a scenario satisfying the other three but violating that one can be exhibited, and in each such scenario incompatibility vanishes. This independence proof ensures that the axioms are genuinely separate constraints and provides four precise "knobs" for the antithesis investigation that follows.

The connection to this dissertation's earlier results is immediate. The Compression Imperative is a restatement, in the language of descriptions, of the compression principle underlying the Law of Formalisation developed in Chapters 2–4. Scale Coupling is a formal expression of the cross-scale interaction that the Law of Scale-Specific Principles (Chapter 6) identifies as the generic condition of physical systems. The axiomatic structure thus inherits the conceptual vocabulary of the persistence framework while extending it into a new domain of application.

19.1.3. Five Convergent Proof Methods

The incompatibility thesis is established by five independent arguments, each exploiting a different mathematical framework but sharing a common structural root in the tension between compression and coverage.

The *information-theoretic compression argument* shows that any valid, compressive description implicitly defines a compression scheme—a partition of observables into primitives and derived quantities. When two descriptions of overlapping domains choose different primitives, their structural commitments contradict on the overlap. If a shared observable f* is primitive in one description (treated as an independent fact) and derived in the other (asserted to be determined by deeper quantities), the resulting assertions—"f* is independent" and "f* is determined"—are logically incompatible. The argument yields a

quantitative incompatibility degree $\iota(D_1, D_2)$ measuring the minimum number of shared observables with contradictory primitive-derived status across all valid compression schemes. The *logical orthogonality argument* reaches the same conclusion through model theory. Non-isomorphic descriptions of overlapping domains have different deductive closures; Descriptive Completeness prevents either from remaining silent about shared observables; therefore they must derive contradictory verdicts on some sentence in the shared language. The proof operates without case analysis and grounds the result in the structural fact that non-isomorphic theories on a shared domain must disagree on some sentence.

The *category-theoretic argument* shows that the category of valid descriptions—with objects being descriptions and morphisms being truth-preserving theory interpretations—lacks a colimit. A unified description would be precisely such a colimit, and its non-existence means that no single consistent theory can faithfully embed all local descriptions. The obstruction is the orthogonality of descriptions on overlaps, which prevents the commutativity required for a colimit.

The *topological obstruction argument* equips the space of descriptions with topological structure and shows that the set of incompatible description pairs is open and dense, while compatible pairs form a closed, nowhere-dense set of measure zero. A cohomological interpretation identifies the obstruction as a non-trivial class in the first Čech cohomology of a description presheaf—the formal analogue of the impossibility of globalising local coordinate systems on a non-trivial manifold.

The *predictive trade-off argument* addresses the pragmatic dimension: any consistent unification attempt must sacrifice perfect accuracy in at least one domain. Combined with a Kolmogorov complexity bound showing that unification costs more than maintaining separate descriptions, this establishes that forced unification is not merely logically impossible but practically counterproductive.

19.1.4. Antithesis and Robustness

The investigation proceeds as a mathematical expedition with no predetermined conclusion. Having advanced five arguments for the thesis, *The Theorem of Incompatible Truths* subjects it to sustained attack through four antithesis strategies.

The *trivial-domain construction* exhibits systems below the complexity threshold where all descriptions are compatible—but succeeds only by violating the axioms that define the regime

of interest. The *hierarchy-of-approximations argument*—the standard reductionist claim that one theory subsumes another in a limit ($\hbar \to 0$, $N \to \infty$)—is shown to provide correspondence but not compatibility: the descriptions remain orthogonal at every physical parameter value, and a limiting relation is an asymptotic statement, not a logical one. The *shared-ontology covering argument*—the hope that a deeper level reconciles the contradictions—either inherits the incompatibility, violates the Compression Imperative, or generates new incompatibilities at a deeper level; the regress does not terminate. Finally, *systematic axiom relaxation* shows that dropping any single axiom permits compatibility but at the cost of a defining feature of theoretical science: explanatory gaps, loss of compression, domain isolation, or loss of descriptive diversity.

A formal lattice model on $\{0,1\}^6$ with XOR dynamics is constructed in which all five primitives, four axioms, two descriptions, and the incompatibility are exhibited without appeal to physical analogy. Two computational toy models further demonstrate that agents maintaining incompatible local descriptions and switching between them outperform agents constrained to a single unified rule set—both in predictive accuracy and in computational cost. The central result—the Incompatibility Theorem—is a four-part statement: incompatibility is necessary (above the complexity threshold), non-eliminable (adding detail shifts but does not resolve contradictions), generic (compatible description pairs form a measure-zero set), and irresolvable from within (an observer in the overlap region cannot determine which description governs, because the act of observation itself commits to a description).

19.2. The Pragmatic Synthesis

If the Theorem of Incompatible Truths establishes that the world's descriptions necessarily contradict each other, then a constructive question immediately arises: how does an agent embedded in such a world manage to function at all? *The Pragmatic Synthesis: How Complex Systems Navigate Incompatible Truths through Adaptive Heuristics* addresses this question with full mathematical rigour.

19.2.1. Agents, Working Models, and Viability

The theory is built from five primitive notions: agent, working model, switching, cost, and viability. An *agent* $A = (I, O, M_A, \pi, B)$ receives inputs from the system, maintains a finite repertoire of working models, selects among them according to a policy π, and acts—all within a finite computational budget B. A *working model* $M = (D, c, \Omega_M)$ pairs a description with

its computational cost and domain of applicability. *Viability* V(A, T) is defined as the probability of remaining in the safe region over a time horizon T, and long-term viability V_∞(A) as the limit as T → ∞.

Four axioms define the regime in which adaptive navigation is necessary: *Computational Boundedness* (the agent cannot afford the exact model for at least one domain), *Context Multiplicity* (the agent encounters at least two domains with incompatible descriptions, each with positive probability), *Action Imperative* (the agent must act at every step; perpetual deliberation is excluded), and *Error Penalty* (using the wrong model degrades viability by at least ε per step).

19.2.2. Three Central Theorems

The *Optimal Coarse-Graining Theorem* proves that multiple simplified and individually inexact working models yield strictly greater expected viability than either a single exact model (which exceeds the computational budget) or a single coarse-grained model (which incurs the full error penalty in domains where it does not apply). This is the mathematical justification for heuristic reasoning: the "true" model is useless if it cannot be computed in time, and multiple "false" models, each cheap and locally valid, constitute the best achievable strategy.

The *Switching Necessity Theorem* proves that any agent employing a fixed working model without switching has viability converging to zero over time. The proof is direct: if the agent encounters incompatible domains with positive probability, the cumulative error penalty drives $V(A,T) \leq (1 - \varepsilon \cdot p_{min})^T \to 0$. Model monism—commitment to a single description—is lethal under the four axioms. This result provides a formal foundation for the cognitive flexibility observed in biological organisms and formalised informally in Kahneman's dual-process theory.

The *Evolutionary Advantage Theorem* proves that systems with incompatible descriptions admit a strictly larger space of viable adaptive strategies than fully transparent, single-description systems. The strategy space grows at least linearly with the number of incompatible domains, providing more raw material for variation and selection. Incompatibility is not merely tolerable but advantageous: it creates the structural conditions for open-ended evolutionary innovation.

19.2.3. The Navigator Theorem and the Consciousness Label

The three theorems are synthesised into the *Navigator Theorem*: any agent achieving positive long-term viability necessarily possesses a subsystem—the navigator—with five properties. Three are unconditionally necessary: *Model Storage* (simultaneous access to multiple working models), *Contextual Evaluation* (a domain-detection function better than random guessing), and *Switching Mechanism* (feasible model transitions within the budget). Two are conditionally necessary under generic environmental conditions: *Transition Memory* (a record of past model choices and outcomes, required when single inputs do not uniquely determine the domain) and *Self-Monitoring* (a feedback loop tracking viability consequences of model choices, required when the environment is non-stationary).

The Pragmatic Synthesis then demonstrates a systematic correspondence between these five properties and the properties attributed to consciousness across multiple research traditions: model storage corresponds to parallel representational systems (Baars's Global Workspace Theory), contextual evaluation to attentional salience (Posner), switching to executive control (Miyake and Friedman), memory to episodic continuity (Tulving), and self-monitoring to metacognition (Flavell, Fleming and Dolan). The paper does not postulate consciousness as an ontological category; it derives the navigator as a functionally necessary structure and identifies the label "consciousness" as a name for this structure—much as "life" was once a metaphysical category and is now understood as a label for specific biochemical processes.

The connection to this dissertation's persistence framework is direct. The navigator is the structure through which a persistent system maintains viability under conditions of incompleteness and inconsistency. Its five properties are formal expressions of the adaptive capacity that the persistence theorem (Chapter 5) identifies as necessary for long-term survival.

19.3. Adaptive Observer Theory

Adaptive Observer Theory: Navigation of Incompleteness and Incompatibility consolidates the results of the preceding two articles into a monograph-length treatment and extends them in three directions.

First, it articulates a *tripartite viability constraint* on persistent complex systems. Any system that contains proper-subsystem observers, exceeds the complexity threshold, and contains bounded agents must simultaneously: maintain uncertainty (rigid closure of all degrees of freedom leads to collapse—this is the Law of Imperative Uncertainty developed in Chapter

16), tolerate contradiction (forced unification degrades predictive power—the Theorem of Incompatible Truths), and navigate adaptively (model monism is lethal—the Navigator Theorem). The three dimensions are mutually reinforcing: uncertainty creates the space for contradictory descriptions; contradictory descriptions create the need for navigation; navigation exploits uncertainty and contradiction to generate adaptive strategies.

Second, *Adaptive Observer Theory* derives eleven falsifiable predictions spanning physics, cognitive science, and artificial evolution. These include: that no improvement in measurement technology will eliminate all observer-inaccessible content in systems where the observer is a proper subsystem; that at every boundary between physical scales there exist pairs of valid descriptions deriving contradictory predictions about shared observables; that any attempt to unify incompatible descriptions necessarily degrades predictive accuracy in at least one domain; that switching agents outperform non-switching agents in multi-domain environments satisfying the four axioms; and that populations of agents subjected to selection pressure in multi-domain environments will evolve toward the navigator structure.

Third, the monograph provides a systematic treatment of objections. The charge that the axioms encode the conclusion is addressed by noting that every theorem in mathematics is conditional on its axioms, and that each axiom captures a minimal condition any reasonable notion of scientific theory satisfies—relaxing any one eliminates the result but also eliminates a defining feature of theoretical practice. The charge that the navigator is "just functionalism dressed in mathematical clothes" is conceded with the qualification that standard functionalism says consciousness *is* the functional role, whereas the Navigator Theorem proves that the functional role is *mathematically necessary*—the difference is between "if a system has these functions, call it conscious" and "any viable system *must* have these functions." The hard problem of consciousness is not solved but reframed: the functional characterisation is shown to be sufficient for all observable behaviour, including verbal reports of inner experience, and whether anything remains beyond the functional account is acknowledged as a question outside the scope of formal analysis.

19.4. Deep Disagreements as Incompatible Truths

The formal apparatus developed in the preceding three articles is domain-general: it applies wherever compressive descriptions of overlapping domains confront each other. *Deep Disagreements as Incompatible Truths: A Formal Framework and Constructive Response* demonstrates that this generality extends to normative discourse, providing a formal foundation for a problem that has remained unresolved in argumentation theory for four decades.

In 1985, Robert Fogelin introduced the concept of deep disagreements: disputes in which both sides reason coherently from their respective premises yet reach contradictory conclusions—not because of irrationality or obstinacy but because the parties operate from different framework propositions. Fogelin's account, grounded in the later Wittgenstein's hinge propositions, generated a rich literature spanning the optimist–pessimist spectrum, from Siegel's sustained case for rational resolvability to Ranalli's hinge-epistemological argument for genuine irresolubility. What the literature lacked was a formal demonstration: the optimists assert resolution without proving structural conditions permit it; the pessimists assert impossibility without identifying the precise conditions under which it holds; and no participant offered a constructive account of what rational agents ought to do when confronted with genuinely irresoluble disputes.

Deep Disagreements as Incompatible Truths addresses all three gaps by interpreting Fogelin's framework propositions as compression schemes in the sense of the Theorem of Incompatible Truths. In the abortion debate, for instance, one evaluative framework treats the moral status of the foetus as primitive (a basic, underived commitment) and derives permissibility from it; the other treats bodily autonomy as primitive and derives permissibility from that. Both are compressive, complete, and valid within their respective domains; they overlap on the shared observable "permissibility of termination at developmental stage X"; and their structural commitments—what they treat as fundamental—are logically incompatible. The incompatibility is not a failure of goodwill or intelligence but a structural necessity of any system in which multiple compressive evaluative frameworks address the same subject matter. The Incompatibility Theorem is then derived for normative contexts by three independent methods (compression, logical orthogonality, and predictive trade-off), four antithesis strategies are examined and shown to fail, and the Navigator Theorem is applied to establish that viable agents confronting deep disagreements must maintain multiple evaluative models

and switch contextually—a strategy formally shown to outperform rigid adherence, forced unification, and random selection.

This result transcends the optimist–pessimist impasse in the current literature. The pessimists are vindicated: deep disagreements are structurally irresoluble in the precise sense that no single consistent evaluative framework can reproduce the predictions of both. The optimists are also partially vindicated: rational agents are not condemned to silence; they can navigate among incompatible frameworks through adaptive switching, contextual evaluation, and accumulated experience. The appropriate response to deep disagreement is not resolution (impossible) or resignation (unnecessary) but navigation—the same strategy that complex systems employ at every other scale of description.

Conclusion

The four publications synthesised in this chapter establish a unified epistemological result. The Theorem of Incompatible Truths proves that descriptive plurality is a structural necessity of complexity, not a contingent deficiency of knowledge. The Pragmatic Synthesis proves that viable agents must navigate this plurality through multiple models, contextual switching, and self-monitoring—and that the structure performing this navigation coincides with what the scientific literature calls consciousness. Adaptive Observer Theory consolidates these results into a tripartite viability constraint: persistent complex systems must maintain uncertainty, tolerate contradiction, and navigate adaptively. And the application to deep disagreements demonstrates that the framework extends beyond the natural sciences into normative discourse, providing formal tools where four decades of philosophical argument produced only diagnosis.

The connection to the dissertation's broader programme is threefold. First, the Incompatibility Theorem extends the Law of Scale-Specific Principles from a methodological observation (different scales require different descriptions) to a logical necessity (those descriptions must contradict each other). Second, the Navigator Theorem extends the persistence framework from the structure of systems to the structure of observers—the navigator is the mechanism through which a persistent system maintains viability under conditions of irreducible pluralism. Third, the tripartite viability constraint—uncertainty, contradiction, and navigation—provides the epistemological counterpart to the physical persistence conditions developed in Parts I through V.

The result that emerges is at once modest and far-reaching. It is modest because every theorem is conditional: if the axioms hold, the conclusions follow, and whether the axioms hold for any particular system is an empirical question. It is far-reaching because the axioms capture minimal conditions that any reasonable notion of scientific theory satisfies, and their conjunction forces a conclusion that reframes the enterprise of knowledge itself: the world is not an object to be cracked by a single formula, but a dynamic process in which observers—themselves subsystems—must continuously construct, switch between, and monitor locally valid but globally incompatible models. The plurality of truths is the price, and the privilege, of living inside a system complex enough to be interesting.

CHAPTER 20. CAUSATION, MATHEMATICAL NECESSITY, AND THE BOUNDARIES OF PHYSICAL DESCRIPTION

Introduction

Chapter 19 established that any sufficiently complex system necessarily admits mutually exclusive but individually valid descriptions, and that viable agents must navigate this plurality through adaptive switching rather than forced unification. The present chapter moves deeper into the foundations, interrogating three concepts that underlie the very possibility of physical description: causation, mathematical applicability, and the status of fundamental constants. A fourth question — whether any consistent physics must contain structure inaccessible to all internal observers — completes the arc by identifying the structural boundary beyond which description itself breaks down.

Four publications are synthesised. *Against Causation: A Formal Argument That Causality Is a Compression Artifact of Bounded Observers, Not a Feature of Reality* (Kriger, 2026) argues that causal structure is not a mind-independent feature of the world but a relational product of the territory's correlational skeleton and the observer's compressive needs. *Why Mathematics Works: Structural Necessity of Isomorphism Between Formal Systems and Physical Reality* (Kriger, 2026) proves that the effectiveness of mathematics is a consequence of persistence: only configurations possessing the invariant structure that admits mathematical description survive the filter of temporal persistence. *Can Fundamental Constants Be Derived as Irreducible Residues of Structural Self-Consistency?* (Kriger, 2026) extends this programme to the parameters of physical law, arguing that constants such as c, $\hbar$, and G are not brute empirical facts but values forced by self-consistency requirements of any viable physical framework. *Must Any Consistent Physics Contain Structure Inaccessible to All Internal Observers?* (Kriger, 2026) proves that any physical system containing a proper-subsystem observer necessarily harbours ontological content that no internal observer can access — establishing the structural boundary that motivates the incompatibility and navigation results of Chapter 19.

Together, these four works dismantle the classical picture of physics as the progressive unveiling of a causally structured, mathematically pre-ordained, fully accessible reality. In its place they offer a picture in which causation is a compression artifact, mathematics is a

persistence filter, constants are self-consistency residues, and the boundaries of description are not practical but structural.

20.1. Against Causation

20.1.1. The Problem

The concept of causation occupies a peculiar position in the architecture of science. Every experimental protocol, every engineering decision, and every medical intervention is articulated in causal language: we cut the wire *because* we want the light to go off; we administer the drug *because* it causes recovery. Yet the fundamental equations of physics — the Schrödinger equation, the Einstein field equations, the Standard Model Lagrangian — are time-symmetric: they draw no distinction between past and future, and they contain no primitive causal arrow. The tension between the ubiquity of causal reasoning and the absence of causation from fundamental physics has generated a long and unresolved debate, from Russell's 1913 declaration that the law of causality is a "relic of a bygone age" through Norton's elaboration of the Russellian position to Woodward's interventionist rehabilitation.

Against Causation: A Formal Argument That Causality Is a Compression Artifact of Bounded Observers, Not a Feature of Reality resolves this tension by identifying the precise structural role that causation plays. The resolution is neither eliminativist (causation is an illusion to be discarded) nor realist in the traditional sense (causation is a mind-independent structural feature of reality). It is perspectival: causation is a relational structure that exists at the interface between the world's correlational skeleton and a bounded observer's compressive needs.

20.1.2. The Four-Premise Argument

The argument is built on four premises whose conjunction yields the central thesis. The first, *Boundedness* (P1), states that no finite observer can store or process the full joint probability distribution over all variables in its environment; compression is required. This is not a contingent technological limitation but a structural constraint — the same constraint that the Compression Imperative formalises in the incompatibility framework of Chapter 19.

The second premise, *Optimality of Directed Acyclic Graphs* (P2), draws on the theory of Bayesian networks: directed acyclic graphs satisfying the causal Markov condition constitute the most efficient compression scheme for statistical dependencies, as formalised in the framework of Pearl. The connection to causal inference is direct but critically reinterpreted:

the DAG is not a map of the world's causal joints but the optimal data structure for a bounded observer's compression task.

The third premise, *Exhaustion* (P3), is the decisive claim: once one specifies the territory's undirected conditional independence structure plus the observer's contribution of directionality, scale, and system boundaries, there is no further causal fact left to discover. Causation is fully decomposed into these two components and their interaction.

The fourth premise, *Consequence* (P4), follows: causation is a relational structure at the observer–system interface. Without the territory's structure the observer has nothing to compress, and without the bounded observer the structure has no causal direction. An unbounded observer — Laplace's demon, with access to the complete joint distribution — would have no use for causal arrows; it could read off any conditional probability directly. Causation is the price of finitude: the compression artifact that makes prediction possible for bounded systems.

20.1.3. Three-Level Ontology

The paper develops a precise decomposition of what is ordinarily called "causation" into three levels.

The *territory* supplies undirected conditional independence structure via locality and symmetry. This structure is observer-independent and explains why different observers — operating at different scales, with different computational budgets, using different variable partitions — converge on similar causal models. The convergence is not evidence of mind-independent causal arrows; it is evidence of a shared underlying correlational skeleton.

The *observer* contributes three elements: directionality (the arrow from cause to effect, dependent on the observer's temporal embedding and ultimately on thermodynamic boundary conditions — the Past Hypothesis), scale (the coarse-graining level determining which variables exist as describable entities), and system boundaries (what counts as exogenous versus endogenous). Each of these contributions is a compression choice, not a discovery about the world. Information-theoretic measures such as transfer entropy and Granger causality corroborate this analysis: both are defined relative to an observer's variable partition, and changing the partition changes the "causes."

Causation is the product of these two levels jointly — real at the relational level, absent at either level alone. This three-level ontology avoids the twin pitfalls of naïve realism (which

cannot explain the absence of causal arrows in fundamental physics) and eliminativism (which cannot explain the systematic effectiveness of causal reasoning).

20.1.4. Distinguishing from Eliminativism

The paper explicitly rejects eliminativism — the position associated with Russell and Norton that causation should simply be abandoned as a scientific concept — through three arguments. The *interventionist objection*: cutting a wire and the light going off reflects a genuine counterfactual fact at the observer–system interface; an account that dismisses this as illusory leaves the success of engineering and medicine unexplained. The *convergence problem*: different observers converge on similar causal graphs because they are compressing the same underlying dependency skeleton; eliminativism cannot explain this convergence without reintroducing the very structure it denies. The *self-undermining problem*: eliminativism uses causation ("bounded observers construct causal models *because* of their limitations") to explain away causation, creating a regress that the perspectival account avoids by locating causation at the relational level rather than denying it altogether.

20.1.5. Consequences for the Dissertation Programme

The perspectival account mandates *causal pluralism*: the same physical system legitimately admits different causal descriptions at different scales. This is a direct corollary of the Law of Scale-Specific Principles (Chapter 6) and provides independent support for the Theorem of Incompatible Truths (Chapter 19). The incompatibility of descriptions at different scales is not merely logical but causal: the causal arrows themselves differ across scales, because they are compression artifacts of different coarse-graining choices.

It denies *metaphysical privilege* to microphysics. Particle-level causation is not "more real" than macro-level causation; both are compressions at different scales, and neither has ontological priority.

And it partially dissolves the *problem of mental causation*. If physical and mental causes are compressions of the same dependency structure at different coarse-graining levels, they do not compete for the same causal "slot." The apparent overdetermination problem rests on treating causal arrows as features of the territory rather than artifacts of the observer's compression level. At the macro-cognitive coarse-graining, the mental cause is the correct causal description; at the neurophysiological coarse-graining, the neural cause is the correct one. They are not rival explanations but different compressions — an instance of the causal

pluralism the paper establishes, and of the navigation between incompatible descriptions that the Navigator Theorem (Chapter 19) formalises.

20.2. Why Mathematics Works

20.2.1. The Persistence Filter Thesis

The "unreasonable effectiveness of mathematics" — Wigner's phrase, now a standing puzzle in the philosophy of science — asks why formal structures developed by pure mathematicians for internal aesthetic reasons turn out to describe the physical world with extraordinary precision. *Why Mathematics Works: Structural Necessity of Isomorphism Between Formal Systems and Physical Reality* provides a mechanism: the Persistence Filter Thesis. Mathematics is effective because only those configurations of matter and energy that possess the right kind of structural regularity — the kind that admits mathematical description — survive the filter of temporal persistence. Mathematics does not describe the world because of a mysterious pre-established harmony; it describes the world because the world's persistent structures are precisely those that carry the algebraic invariants constituting mathematical content.

20.2.2. The Chain of Argument

The argument proceeds through four formally proved steps. First, any physical system that persists through time must possess quantities or relations that remain invariant under the transformations the system undergoes — the *Persistence Requires Invariants Lemma*. A system (X, T, Φ) is formalised as (ε, τ)-persistent if key observables remain approximately stable over extended periods. To persist is to have something that does not change. The connection to the persistence theorem of Chapter 5 is immediate.

Second, any collection of invariants under a family of transformations necessarily carries algebraic structure. Proposition 3.5 proves that invariants form a subalgebra closed under addition, scalar multiplication, and pointwise multiplication. Proposition 3.6 shows that invariants encode the system's symmetry group. This is not a stipulation but a mathematical consequence of the closure properties of invariant sets.

Third, any algebraic structure is a model of some first-order theory. This step is logically elementary but conceptually decisive: it establishes that every persistent system is, by mathematical necessity, an instantiation of a formal theory.

Fourth — addressing the vacuity objection head-on — the *Structural Richness Theorem* (Theorem 4.4) proves that the invariant algebras of persistent systems are finitely axiomatisable (via the Hilbert Basis Theorem), have non-trivial consequences (constraining observables to algebraic varieties), and are multiply realisable (the same polynomial ideal can be satisfied by different physical systems). The synthesis — the *Persistence Filter Theorem* (Theorem 4.2) — establishes that every persistent system possesses a non-trivial invariant algebra that defines a first-order theory of which the system is a model.

20.2.3. Why Particular Structures Recur

The general argument explains why mathematics works but does not yet explain why *particular* mathematical structures recur across physics. The paper addresses this specificity problem through a series of propositions showing that persistence selects for specific mathematical features.

Continuity is selected because persistent dynamics must map nearby states to nearby states (Proposition 5.1). Variational structure is selected because persistent trajectories are critical points of the action functional (Proposition 5.3) — this is proved rather than assumed, connecting Noetherian conservation to persistence in a novel direction. Unitarity in quantum mechanics is selected because persistence of transition probabilities requires unitary evolution via Wigner's theorem, which in turn requires Hilbert space structure via Stone's theorem (Propositions 5.4–5.5). Five illustrative models — a minimal discrete system, the double pendulum, Conway's Game of Life, the Lorenz attractor, and a biological thermoregulation model — ground the abstract argument in concrete settings, showing that the persistence filter operates across deterministic, chaotic, discrete, dissipative, and homeostatic regimes.

20.2.4. The Dual-Source Decomposition

The paper develops, with full formal rigour, the radical Kantian counter-thesis that mathematics is a coordinate grid imposed by consciousness rather than a reflection of external structure. This alternative is formalised via a cognitive modelling system (Σ, M_Σ, Π, W) and proved to be *observationally indistinguishable* from the external-structure hypothesis (Proposition 9.5) — a result of considerable philosophical importance. The refutation proceeds through four arguments: the *prediction asymmetry* (a purely cognitive grid cannot explain the systematic success of novel predictions such as the positron), *intersubjectivity* (different cognitive architectures converge on the same mathematics), the *rigidity of physical*

law (it resists cognitive revision), and *formal independence* (cognitive architecture constrains the form but not the content of physical theory).

The synthesis is the *Dual-Source Decomposition* (Proposition 9.10): mathematical physics decomposes into T_form (structural constraints contributed by cognitive architecture, stable across centuries) and T_cont (specific laws and constants determined by the world, subject to revolutionary change). This provides a mathematisation of Kant's transcendental philosophy and connects directly to the observer-dependent decomposition of causation in Section 20.1: in both cases, the observer contributes structural form while the world contributes specific content, and the scientific description is the product of both.

20.3. Can Fundamental Constants Be Derived?

20.3.1. From Structure to Parameters

Why Mathematics Works established that persistent systems carry algebraic structure but said nothing about the parameters of that structure. Why does the fine-structure constant have the value $\alpha \approx 1/137$ rather than some other value? Are the fundamental constants brute empirical facts, or are they constrained by deeper principles?

Can Fundamental Constants Be Derived as Irreducible Residues of Structural Self-Consistency? — published as Part II of the monograph *Structural Necessity Theory* — extends the persistence programme to this question. The central thesis is that fundamental constants are neither derivable from first principles (as Eddington and others hoped) nor contingent empirical accidents: they are *irreducible residues* — parameters simultaneously non-derivable and non-contingent, constrained to measure-zero subsets of parameter space by the requirement of structural self-consistency.

20.3.2. Five Axioms

The paper formulates five minimal axioms for any system capable of sustaining persistent structure.

Persistence (Axiom 1) requires the system to contain persistent entities — from atoms to galaxies — with invariant observables, providing the ontological substrate: the existence of forward-invariant subsets of state space with non-trivial conserved observables.

Self-Reference (Axiom 2) requires that the physical framework be self-consistent: its predictions must be compatible with the conditions necessary for those predictions to be meaningful. This axiom performs the central formal work: it generates a self-consistency

operator F: $\Lambda \to \Lambda$ on the parameter space, transforming the question of the values of constants into a fixed-point problem.

Closure (Axiom 3) guarantees that what the system produces, the theory can describe — formally, that F maps the admissible parameter region into itself: $F(K) \subseteq K$. This makes the Brouwer fixed-point theorem applicable.

Non-Degeneracy (Axiom 4) requires the parameter space to be non-trivial ($\dim(\Lambda) \geq 1$). *Distinguishability* (Axiom 5) requires that the system admit distinguishable states — the minimal condition for physics to exist.

20.3.3. The Irreducible Residue Existence Theorem

The core result — the Irreducible Residue Existence Theorem — proves that any physical model satisfying the five axioms necessarily contains parameters λ^* that are fixed points of the self-consistency operator: $\lambda^* = F(\lambda^*)$. The proof proceeds by two independent methods: a *fixed-point argument* via Brouwer's theorem (a continuous map of a compact convex set into itself has a fixed point, with the Sard and implicit function theorems guaranteeing that generically these points are isolated) and a *cohomological argument* via topological degree theory (the degree of the mapping is non-zero, establishing the existence of fixed points with topological necessity).

The generically discrete character of the fixed points is the key consequence: the constants cannot take arbitrary values but are constrained to measure-zero subsets of parameter space. They are, in a precise sense, structurally forced.

20.3.4. The Residue Classification Theorem

The paper classifies irreducible residues into two types. *Scale residues* are dimensional constants (c, $\hbar$, G) that fix the units of measurement — they establish the scale at which the physical framework operates. *Coupling residues* are dimensionless constants (α, mass ratios) that fix the structure and relative strengths of interactions. This classification engages with the Duff–Okun–Veneziano debate about which constants are truly fundamental, offering a structural criterion: a constant is fundamental if it corresponds to a fixed point of the self-consistency operator that cannot be decomposed into simpler fixed points.

20.3.5. The Anthropic Principle as a Special Case

The paper shows that the anthropic principle — the observation that the constants of nature appear "fine-tuned" for the existence of observers — is a special case of a more general

persistence filter. The requirement is not that constants admit observers specifically, but that they admit persistent structure of any kind. Observers are one particular type of persistent structure, and the anthropic selection is a subset of the persistence selection. This demotion of the anthropic principle from a fundamental explanatory tool to a special case of a structural constraint connects directly to the Persistence Filter Thesis of Section 20.2.

20.3.6. Connection to the Persistence Programme

The relationship between this result and *Why Mathematics Works* is one of logical completion. The Persistence Filter Thesis established that persistent systems carry invariant algebraic structure; the Irreducible Residue Theorem establishes that the parameters of that structure are themselves constrained by self-consistency. Together, they assert that neither the form nor the content of physical law is arbitrary: the form is selected by persistence, and the content is locked in place by self-consistency.

The connection to active research programmes in theoretical physics is significant. The Wilsonian renormalisation group, the conformal bootstrap, asymptotic safety in quantum gravity, and the swampland programme in string theory all impose self-consistency constraints that narrow the space of viable parameter values. The Irreducible Residue Theorem provides a general framework within which these diverse programmes can be understood as instances of a single structural principle.

The paper is forthright about its limitations: the mathematical contribution is the application of standard fixed-point theorems, and the novelty lies in the conceptual reinterpretation — the recognition that the question of the values of fundamental constants is a fixed-point problem. The constructive computation of the operator F for the Standard Model and the determination of its specific fixed points remains an open problem.

20.4. Structural Inaccessibility

20.4.1. The Question

The three preceding sections addressed what descriptions contain: their causal structure (Section 20.1), their mathematical form (Section 20.2), and their parametric content (Section 20.3). *Must Any Consistent Physics Contain Structure Inaccessible to All Internal Observers?* — published as Part I of the monograph *Adaptive Observer Theory: Navigation of Incompleteness and Incompatibility* — addresses what descriptions necessarily leave out.

20.4.2. Consistency as a Four-Fold Condition

The paper defines consistency as the conjunction of four conditions. *Logical consistency* (A) requires that the system not admit simultaneously true and false statements. *Dynamical consistency* (B) requires that the system's evolution be deterministic and reversible (or at least unambiguous in the forward direction). *Informational consistency* (C) requires that the system not destroy or duplicate distinguishability in ways that violate its own structure — the analogue of unitarity in quantum mechanics. *Self-reflective consistency* (D) requires that when the system contains a subsystem modelling the whole, the model and the reality do not contradict each other.

Each condition is independently motivated, and the conjunction of all four captures the minimal requirements for a physical framework to be viable. The subsequent investigation will show that this conjunction — and not any single condition in isolation — forces inaccessibility.

20.4.3. Five Independent Proofs

The central result is proved by five independent methods, each yielding the same conclusion from a different mathematical starting point.

The *diagonal argument* (i) proceeds by analogy with Cantor's proof that no surjection $\mathbb{N} \to P(\mathbb{N})$ exists. For any proposed representation map, a "diagonal property" is constructed that the observer cannot capture. The argument requires only that the observer be a proper subsystem with at least two representation states.

The *cardinality argument* (ii) observes that the observer's state space is strictly smaller than the system's ($|\Sigma_O| < |\Sigma|$). By the pigeonhole principle, the representation map $R: \Sigma \to \Sigma_O$ cannot be injective — distinct global states inevitably merge. Properties form $P(\Sigma)$, accessible properties at most $P(\Sigma_O)$; the exponential gap $2^{|\Sigma|} - 2^{|\Sigma_O|}$ quantifies the inaccessible content.

The *fixed-point argument* (iii), via the Lawvere fixed-point theorem, operates at the category-theoretic level: if the observer could access all properties, there would exist a surjection from Σ to 2^Σ, yielding a fixed point of every endomorphism on 2^Σ — including negation, which has no fixed point. The contradiction establishes a somewhat different result concerning self-referential access.

The *dynamical argument* (iv) shows that the observational equivalence classes not only exist instantaneously but propagate and potentially coarsen over time. The observer's accessible

algebra is a sub-σ-algebra of the full state algebra, and dynamical evolution maps this sub-algebra into itself. The set of dynamically inaccessible properties — those that remain undecidable for arbitrarily long — is non-empty whenever the system is dynamically non-trivial.

The *topological/measure-theoretic argument* (v) establishes that the set of accessible properties has strictly smaller measure or topological complexity than the full ontological content, even when restricted to physically motivated classes (projection operators, Borel sets, macroscopic observables).

The paper acknowledges that all five proofs ultimately root in a single structural fact: a proper subsystem has fewer states than the whole, and the representation map cannot be injective. The contribution lies in demonstrating that the obstruction manifests in every natural mathematical framework, quantifying the accessibility gap, and identifying the exact boundary conditions.

20.4.4. Four Antithesis Strategies

The paper subjects the thesis to sustained attack through four antithesis strategies, each shown to succeed only under degenerate or physically unrealistic conditions.

The *trivial-system construction* notes that a system with a single state is fully observable — but trivially so. The *holographic argument* suggests that an internal observer might access all information through boundary degrees of freedom — but this works only under special boundary conditions not generically satisfied. The *collective observation* strategy proposes that an ensemble of observers might collectively cover the entire state space — but the collective is itself a proper subsystem, and the same cardinality arguments apply. The *inconsistency loophole* observes that dropping any one of the four consistency conditions permits constructions with full accessibility — at the cost of contradictions, ambiguous evolution, information destruction, or paradoxical self-reference.

20.4.5. The Trichotomy

All results are unified in a trichotomy theorem. If the observer is a proper subsystem ($|\Sigma_O| < |\Sigma|$, $|\Sigma| \geq 2$), then the inaccessible content $\mathfrak{I}(S) \neq \emptyset$: it is generic, preserved under temporal evolution, resistant to collective observation, and includes the very fact of its own existence. If the system is trivial ($|\Sigma| \leq 1$) or the observer is not a proper subsystem ($|\Sigma_O| \geq |\Sigma|$), then

$\mathfrak{I}(S) = \emptyset$. If the theory does not specify whether the observer is a proper subsystem, the inaccessibility statement is independent of the consistency conditions.

20.4.6. Self-Reflective Undecidability

A meta-level result completes the picture: an internal observer cannot empirically determine whether the system it inhabits contains inaccessible content. The question of inaccessibility itself belongs to the set of inaccessible truths. The boundary between the knowable and the unknowable is itself on the unknowable side. This self-reflective closure — the inaccessibility of inaccessibility — is the deepest consequence of the paper and the one most directly relevant to the epistemological programme of Part VI.

20.4.7. The Duality of Consistency and Inaccessibility

The antithesis investigation revealed a structural duality: each consistency condition, when relaxed, opens a loophole for full accessibility; conversely, insistence on all four conditions forces inaccessibility. Consistency and inaccessibility are not independent properties but two aspects of a single structural constraint. A consistent physics must have blind spots; a physics without blind spots cannot be consistent.

This duality connects the Inaccessibility Theorem to the Theorem of Incompatible Truths (Chapter 19) in a precise way. Article I (inaccessibility) proves that each observer's description is necessarily incomplete; Article II (incompatibility) proves that the descriptions constructed by different observers are not merely incomplete but mutually contradictory above a complexity threshold. Incompleteness alone does not imply incompatibility — two partial descriptions could agree on their shared content. The additional structure that forces contradiction is the subject of the incompatibility theorem, and it requires inaccessibility as its foundation.

Conclusion

The four publications synthesised in this chapter dismantle three assumptions that have traditionally underwritten the enterprise of physical description: that causation is a feature of the territory, that the effectiveness of mathematics is a brute coincidence, and that fundamental constants are free parameters. In their place a coherent alternative emerges: causation is a compression artifact of bounded observers (*Against Causation*), mathematics is the structural signature of persistence (*Why Mathematics Works*), constants are self-consistency residues constrained to measure-zero subsets of parameter space (*Can Fundamental Constants Be*

Derived?), and the boundaries of description are set not by practical limitations but by the structural constraint that observers are proper subsystems of the systems they describe (*Must Any Consistent Physics Contain Structure Inaccessible?*).

The connections among the four results are tight. The perspectival account of causation and the persistence filter thesis share a common structure: both identify features of scientific description — causal arrows, mathematical form — as products of the interaction between an observer and a world. The Irreducible Residue Theorem extends this programme from form to content, showing that even the parameters of physical law are constrained by structural self-consistency. The duality of consistency and inaccessibility establishes the boundary beyond which the entire descriptive enterprise — causal, mathematical, parametric — necessarily fails: the structural blind spot that no internal observer can eliminate, and whose very existence is itself inaccessible.

Together with Chapter 19, this chapter completes the epistemological core of Part VI. Where Parts I–V developed the ontology of persistence — what persistent systems are, how they are structured, why they endure — Part VI develops its epistemology: what can be known about persistent systems, by observers who are themselves persistent systems embedded within the systems they study. The answer is nuanced: much can be known (mathematics works because persistence selects for invariant structure), but not everything (structural inaccessibility is irreducible), and what is known is necessarily plural (incompatible descriptions are a structural feature, not a deficiency of knowledge). The appropriate response, as the Navigator Theorem establishes, is not to mourn the impossibility of total knowledge but to navigate the irreducible plurality with adaptive intelligence.

CHAPTER 21. METHODOLOGICAL FOUNDATIONS: EVIDENCE-BASED PHILOSOPHY AND FRAME-BREAKING INQUIRY

Introduction

Chapters 19 and 20 established the epistemological landscape of Part VI: irreducible pluralism of descriptions, the perspectival character of causation, the persistence-filter origin of mathematical applicability, and the structural inaccessibility that bounds all internal observation. These results constrain what can be known. The present chapter turns to the question of *how* inquiry itself should be conducted within these constraints — what methodological principles follow from the formal epistemology developed so far.

Three publications are synthesised. *Evidence-Based Philosophy: A Foundational Methodological Framework for the Transformation of Philosophical Inquiry into a Formal Discipline* (Kriger, 2026) argues that the continuation of unconstrained speculative philosophy is no longer epistemically justifiable and provides a four-condition operational definition of evidence-based inquiry, a five-step transition protocol, a Bayesian framework for claim evaluation, and a proposal for reformed peer review incorporating AI-assisted verification. *Breaking Trivial Cognitive Frames: A Formal RAFASSABRA Method for Expanding the Space of Inquiry Through Interrogative Priority and Controlled Divergence* (Kriger, 2026) establishes that questions are structurally prior to answers, that cognitive frames restrict questions more severely than they restrict answers, and that a systematic nine-stage method provably expands the question space beyond the current cognitive frame. *Definition-Dependent Provability and the Stalemate in Scientific Realism: A Formal Diagnosis and Constructive Reframing* (Kriger, 2026) diagnoses the enduring impasse in the scientific realism debate as a structural consequence of definitional multiplicity, introduces the Principle of Definition-Dependent Provability, and proposes a constructive reframing in which philosophical debates are understood as explorations of a provability landscape across competing formal realities.

Together, these three works constitute a methodological programme: they specify what standards philosophical inquiry must meet (*Evidence-Based Philosophy*), how the space of inquiry can be systematically expanded beyond conventional frames (*RAFASSABRA*), and why certain apparently intractable debates are structurally rather than contingently stalled

(*Definition-Dependent Provability*). Each connects directly to the formal apparatus of the preceding chapters.

21.1. Evidence-Based Philosophy

21.1.1. The Central Argument

Evidence-Based Philosophy advances a thesis that is deliberately provocative: the continuation of unconstrained speculative philosophy — conducted without formal verification, empirical compatibility testing, or computational modelling — is no longer epistemically justifiable. The argument rests on three pillars. First, the tools required for transformation — mathematical formalisation, empirical integration, AI-assisted verification — already exist and are sufficiently mature. Second, the historical record demonstrates that philosophical problems become tractable through formalisation while remaining stagnant under speculative treatment: the transformation of informal logic into mathematical logic by Frege and Russell, the transformation of decision-theoretic intuitions into expected utility theory, and the transformation of vague concepts of disease causation into evidence-based medicine all illustrate the pattern. Third, the emergence of artificial intelligence eliminates the last structural excuse for speculation — individual cognitive limitation. The philosopher who today advances claims about consciousness without engaging neuroscience, or about justice without engaging behavioural economics, is not cognitively limited but methodologically negligent.

21.1.2. Four Conditions

The paper provides a precise operational definition. A philosophical inquiry $\mathcal{I}$ qualifies as evidence-based if and only if it satisfies four conditions, each individually necessary and jointly sufficient.

Formal Coherence (C1) requires that all central concepts be defined with sufficient precision to permit evaluation of logical consistency. Where possible, the logical structure is rendered in a formal language $\mathcal{L}$ and internal consistency is demonstrable within $\mathcal{L}$. This is formalised as the requirement that the proposition set Σ of a system $\mathcal{S}$ not entail both φ and $\neg\varphi$ for any $\varphi \in \mathcal{L}$. The requirement holds regardless of philosophical tradition — analytic, continental, pragmatist, or otherwise — because contradiction is not a tradition-specific deficiency but a structural one.

Empirical Compatibility (C2) requires that the claims of $\mathcal{I}$ not contradict well-established empirical findings in any relevant domain. This is weaker than requiring derivation from empirical findings — a requirement that would collapse philosophy into natural science. The formalisation is straightforward: if E denotes the set of well-established empirical propositions, then $\Sigma \cup E$ must be consistent. The set E is itself subject to revision; evidence-based philosophy commits not to the infallibility of current science but to the principle that philosophical claims must be evaluated against the best available empirical evidence.

Operationalisability (C3) requires that every key concept admit a specification function $\sigma_c: \mathcal{C} \rightarrow \{0, 1, \text{indeterminate}\}$ that is intersubjectively accessible, explicitly stated, and whose indeterminate cases are bounded and acknowledged. A concept that is indeterminate in all or nearly all cases is not operationalisable; it is vacuous. This condition allows for vagueness and borderline cases but requires that such cases be explicitly acknowledged rather than tacitly exploited.

Falsifiability or Empirical Distinguishability (C4) requires that $\mathcal{I}$ either generate predictions that are in principle falsifiable by observation or formal analysis, or be empirically distinguishable from competing accounts. For philosophical positions that are not straightforwardly falsifiable — and many are not — distinguishability provides the relevant criterion: if you cannot identify any condition under which your theory and your opponent's theory would yield different verdicts, then you do not, in fact, disagree.

21.1.3. The Transition Protocol and the AI Obligation Thesis

The paper develops a five-step transition protocol for transforming a philosophical inquiry from speculative to evidence-based form: concept formalisation, empirical audit, operationalisation, distinguishability testing, and formal verification. The protocol is illustrated through detailed case studies including the formalisation of "approximate truth" in the scientific realism debate, the formalisation of "free will" across compatibilist and libertarian definitions, and a complete worked formal proof (a Dutch book argument demonstrating that certain probabilistic philosophical commitments are inconsistent).

The AI Obligation Thesis — stated as a conditional argument — asserts that if evidence-based methods currently cover a substantial portion of philosophical problems, and if AI tools make these methods practically accessible, then the refusal to use them constitutes methodological negligence. The paper is transparent about the strength of this claim: its premise that a

"substantial portion" of philosophical problems are tractable via formalisation is asserted rather than proved, and the precise scope of that portion remains an open question. The case studies are intended as existence proofs — demonstrations that formalisation succeeds where speculation stalls — rather than inductive evidence for universal applicability.

21.1.4. The Secondary Communication Principle

A distinctive contribution is the *secondary communication principle*: accessible philosophical exposition is epistemically legitimate only when grounded in prior formal work. Popular philosophy that summarises formal results is valuable; popular philosophy that substitutes for formal work is not. This principle directly parallels the situation in the natural sciences, where popular accounts of quantum mechanics are legitimate when they simplify rigorous theory, but not when they replace it with metaphor. The principle does not privilege obscurity; it privileges accountability.

21.1.5. Connection to the Dissertation Programme

The evidence-based philosophy framework is the methodological self-reflection of the dissertation as a whole. Every chapter of this dissertation satisfies the four conditions: formal coherence is maintained through explicit axiomatisation and proved theorems; empirical compatibility is demonstrated through concrete physical, biological, and cognitive examples; operationalisability is ensured through the construction of computational toy models; and distinguishability is tested through the systematic antithesis investigations that accompany every thesis. The framework thus serves as a retroactive justification for the dissertation's method — and as a prospective standard for the research programme it initiates.

The connection to the Theorem of Incompatible Truths (Chapter 19) is significant. Evidence-based philosophy does not resolve the irreducible pluralism established in Chapter 19; it operates within it. The four conditions apply to each description individually, not to the system of descriptions collectively. A description can be formally coherent, empirically compatible, operationalisable, and distinguishable — and still be incompatible with another description that satisfies the same conditions. Evidence-based philosophy is the standard for local descriptions; the Navigator Theorem is the strategy for navigating among them.

21.2. Breaking Trivial Cognitive Frames: The RAFASSABRA Method

21.2.1. The Structural Priority of Questions

Breaking Trivial Cognitive Frames: A Formal RAFASSABRA Method for Expanding the Space of Inquiry Through Interrogative Priority and Controlled Divergence begins with a result in formal epistemology that reorients the classical picture of inquiry. Under four independent formalisations of "question" — Hintikka partition semantics, Belnap–Steel erotetic implication, the interrogative model of inquiry, and quantum operator formalism — and three mathematical frameworks (set-theoretic, category-theoretic, information-theoretic), the question space is shown to carry strictly richer structure than the answer space. The partition lattice carries ordering absent in the answer space; the answer functor from questions to answers is non-faithful (it loses information); and questions have higher parametric complexity than their answers.

This structural asymmetry — Claim A of the paper — has a direct epistemological consequence. If questions are structurally prior to answers, then the bottleneck of inquiry is interrogative rather than deductive: what limits inquiry is not the inability to derive answers from premises but the inability to formulate the right questions. The history of science is replete with illustrations: the shift from "What substance is phlogiston?" to "What happens during combustion?" was the question that made chemistry possible; the shift from "What carries light through the aether?" to "What if there is no aether?" was the question that made special relativity possible.

21.2.2. Frame Restriction Asymmetry

Claim B extends the structural priority result to cognitive frames — the sets of background assumptions, concepts, and inferential habits that determine which questions an agent can formulate. The paper proves, for small logical spaces ($|W| = 4–7$) and conjectures for general n with an asymptotic argument, that cognitive frames restrict the space of formulable questions more severely than they restrict available answers. The restriction is asymmetric: a frame that blocks access to a question thereby blocks access to all answers that question could have produced, but a frame that blocks access to an answer does not thereby block the question that sought it.

This asymmetry result connects directly to the Navigator Theorem (Chapter 19). The navigator's five properties — model storage, contextual evaluation, switching, memory, and self-monitoring — are properties of an agent's answer-processing and model-selection apparatus. The RAFASSABRA method addresses a prior constraint: the question-formulation

apparatus. A navigator with a rich model repertoire but a narrow question space is still limited; it can switch among available models but cannot generate the questions that would reveal the need for new ones. Frame-breaking is thus a prerequisite for effective navigation.

21.2.3. The Nine Stages

The RAFASSABRA method — the acronym encodes the nine stages — is a systematic procedure for expanding the question space beyond the current cognitive frame. Each stage targets a specific dimension of frame-boundedness, and a Composition Theorem establishes that each stage expands in a direction unique to that stage, so that no stage is redundant.

Random words (R) injects external perturbation statistically independent of the agent's current frame — formalised as orthogonal projection in high-dimensional concept space and as mutation in evolutionary search, with proof that both yield frame-externality guarantees. A control experiment demonstrates that self-generated "random" words cluster semantically within the active frame, while externally-sourced words do not — the paper's single most compelling empirical observation.

Structural Attack (A — "Angry") forces the agent to question its own presuppositions, counteracting the tendency to accept received framings uncritically. *Atemporal Pass* (F — "Frozen") removes temporal and causal language, forcing the agent into a non-narrative mode of analysis — particularly effective given that temporal-causal narrative is the dominant organisational template in human cognition and in large language model training data. *Presuppositional Ascent* (A — "Alien") asks what the phenomenon would look like to an observer who did not share any of the agent's background assumptions. *Apophatic Negation* (S — "Stripped") determines the phenomenon by what it is not — the via negativa of inquiry. *Isomorphic Transfer* (S — "Stole") forces cross-domain structural mapping, requiring combinatorial rather than retrieval-based use of knowledge. *Cataphatic Assertion* (A — "Assembled") reconstructs the phenomenon positively from the fragments generated by the preceding stages. *Paradoxical Stress-Testing* (B — "Broke") identifies internal contradictions in the emerging account. *Temporal Reconstruction* (R–A — "Ran Away") reintroduces time and causality, now within an expanded frame.

21.2.4. RAFASSABRA as AI De-Templating

A distinctive contribution is the analysis of RAFASSABRA as a *prompt architecture* for artificial intelligence. Large language models are trained on vast text corpora, producing

powerful pattern-matching but also powerful pattern-repetition. When asked to analyse a phenomenon, an LLM retrieves the most statistically likely framing — by definition the most conventional one. The model does not lack knowledge; it lacks the capacity to reorganise what it knows into non-standard configurations.

Each RAFASSABRA stage targets a specific dimension of template thinking: random words inject input statistically independent of the training distribution; structural attack counteracts uncritical acceptance of the user's framing; atemporal pass removes narrative structure, the dominant template in training data; isomorphic transfer forces combinatorial rather than retrieval-based processing. The empirical evidence supports this interpretation: in a demonstration on six problems (the nature of time, consciousness, philosophy, corruption, dark matter, and love), an independent AI model using RAFASSABRA produced analyses that no standard prompt would elicit — not because the model lacked the relevant knowledge, but because the method forced it out of its statistical attractor basin.

21.3. Definition-Dependent Provability and Scientific Realism

21.3.1. The Stalemate

The debate on scientific realism has been stuck for decades. Realists hold that the success of mature scientific theories provides evidence for their approximate truth; antirealists find this inference unwarranted. Each side has developed formidable resources: the realist arsenal includes the no-miracles argument, structural realism, and deployment realism; the antirealist arsenal includes constructive empiricism, the pessimistic meta-induction, and the problem of unconceived alternatives. The debate thrives in sophistication but is bound for stalemate.

Definition-Dependent Provability and the Stalemate in Scientific Realism: A Formal Diagnosis and Constructive Reframing diagnoses this impasse as structural rather than contingent. The diagnosis uses two formal tools: the Formalisation Laws (established in earlier chapters of this dissertation), which establish that any phenomenon admits multiple irreducibly distinct formal representations, and the Principle of Definition-Dependent Provability (DDP), which is the paper's central contribution.

21.3.2. The Principle of Definition-Dependent Provability

DDP states that the provability status of a statement can change as a function of the definitions assigned to its key terms. This is not the trivial observation that vague terms are ambiguous; it is the formal result that within a fixed base theory S, for a formula schema φ[T] depending

on a defined term T, the provability function f(d) = Prov(Ext(S, d), φ[T/d]) is well-defined and, in general, non-constant across admissible definitions d_1, d_2 satisfying conservativity. A claim can be provable under one admissible definition, refutable under another, and undecidable under a third — not because of logical limitations but because the claim has not been asked with sufficient definitional precision.

The paper illustrates DDP with cases of increasing complexity. For the sequence S = 01010101..., the question "Is S random?" is provably true under the frequency definition (limiting frequency of 1s equals 1/2), provably false under the Kolmogorov complexity definition (the sequence is highly compressible), and provably false under the Martin-Löf definition (the sequence belongs to an effectively null set). All three definitions satisfy reasonable adequacy conditions; all three yield incompatible verdicts. The question is not undecidable; it is underdefined.

The philosophical case studies are more consequential. "Humans have free will" is provable (given empirical premises) under a compatibilist formalisation and unprovable (perhaps necessarily false) under a libertarian one. Both formalisations satisfy the admissibility criteria; both preserve inferential connections with pre-theoretic usage; both yield plausible verdicts on paradigm cases. DDP does not adjudicate between them; it clarifies that claims about the provability of free will are indeterminate until the formalisation is specified.

21.3.3. Application to Scientific Realism

Applied to the scientific realism debate, DDP reveals that its central concepts — approximate truth, novel prediction, structural continuity, empirical adequacy — are not merely vague or contested but formally definition-dependent: claims involving them shift between provable, refutable, and undecidable as definitions vary across their admissible range. The no-miracles argument is valid under one formalisation of "approximate truth" and invalid under another. The pessimistic meta-induction succeeds under one definition of "theoretical continuity" and fails under another. The stalemate is therefore not a failure of argumentation but a structural consequence of definitional multiplicity operating on concepts whose provability landscape is rugged.

21.3.4. Formal Realities and Constructive Reframing

The paper introduces the concept of a *formal reality* — a complete specification of logic, axioms, and definitions within which claims acquire determinate truth-values — and argues

that the choice between realism and antirealism is best understood as a choice between formal realities with different structural costs. This reframing yields a constructive proposal: the debate becomes productive not when it seeks closure but when it makes explicit the formal realities within which competing claims are advanced and compares their respective epistemic costs and deliverances.

The connection to the Theorem of Incompatible Truths (Chapter 19) is deep. DDP is the epistemological instantiation of incompatibility: just as the Theorem of Incompatible Truths shows that physical descriptions at different scales must contradict each other, DDP shows that philosophical claims under different admissible definitions must yield different provability verdicts. The mechanism is the same — compression forces structural commitments, and different commitments produce different verdicts on shared subject matter. The Formalisation Laws, which established that any phenomenon admits multiple irreducibly distinct formal representations, provide the bridge: they guarantee that the definitional multiplicity exploited by DDP is not eliminable but is a structural feature of any sufficiently rich conceptual domain.

The paper further argues, drawing on results concerning totalising ideals in complex adaptive systems developed elsewhere in this dissertation, that the pursuit of closure in the realism debate is not merely unlikely to succeed but structurally counterproductive: it would destroy the epistemic variance that makes the debate useful. The debate is valuable precisely because it maps the topology of definition-dependent provability across competing formal realities — and this mapping is itself a form of knowledge, even though it does not converge on a single answer.

Conclusion

The three publications synthesised in this chapter constitute the methodological self-reflection of the dissertation's broader programme.

Evidence-Based Philosophy establishes what standards inquiry must meet: formal coherence, empirical compatibility, operationalisability, and distinguishability. These four conditions are the methodological counterpart of the four axioms that underlie the Theorem of Incompatible

Truths — they constrain the form of individual descriptions without resolving the plurality among them.

The RAFASSABRA Method addresses the prior constraint that evidence-based philosophy leaves untouched: the question-formulation apparatus. If the bottleneck of inquiry is interrogative rather than deductive, then methodological reform must begin not with better answers but with better questions. The nine stages of RAFASSABRA provide a formally grounded procedure for breaking out of the cognitive frames that restrict the question space — and, in doing so, for discovering the new descriptions that the Navigator Theorem (Chapter 19) requires for effective navigation.

Definition-Dependent Provability completes the methodological picture by diagnosing why certain philosophical debates resist resolution. The stalemate in scientific realism — and, by extension, in any debate whose central concepts are definition-dependent — is not a failure of intelligence or goodwill but a structural consequence of the same formal features that produce incompatible descriptions in the physical domain. The constructive response is not to resolve the debate but to map its provability landscape: to make explicit the formal realities in which competing claims are advanced, to compare their structural costs, and to recognise the mapping itself as a form of knowledge.

Together, these three works offer a coherent methodological programme for inquiry conducted under the conditions established in Chapters 19 and 20: conditions of irreducible pluralism, structural inaccessibility, and definition-dependent provability. The programme does not promise final answers — the formal results of Part VI establish that final answers are, in many domains, structurally impossible. It promises something more modest and more useful: a disciplined practice of inquiry that meets explicit standards, systematically expands the space of formulable questions, and recognises the provability landscape of its own concepts. This is what evidence-based philosophy looks like when it takes its own incompleteness seriously.

PART VII. FORMAL STRUCTURES OF IDENTITY, CONSCIOUSNESS, AND COMMUNICATION

CHAPTER 22. IDENTITY, MOTIVATION, AND PERSONALITY ARCHITECTURE

Introduction

Part VII of this dissertation extends the formal persistence framework into the domains of identity, consciousness, and communication — the structures through which complex systems not only persist but recognise themselves and each other. The present chapter addresses three interrelated questions: what constitutes the identity of a persistent agent, what formal structures govern the agent's motivational dynamics, and whether the classical typological methods used to classify agents can be rehabilitated through formalisation. A fourth publication radically reframes the very concept of inference, arguing that the most fundamental form of reasoning is sub-personal, non-propositional, and physically inevitable — thereby reshaping the cognitive substrate upon which identity, motivation, and personality operate.

Four publications are synthesised. *Identity as Algorithmic Configuration: A Formal Theory of Personality Compressibility* (Kriger, 2026) argues that socially perceived identity is a low-complexity parametric residual atop a shared generative substrate, introduces the Identity Compressibility Thesis, and proves an Identity Uniqueness Theorem. *From Repression to Saturation: A Dynamical Systems Theory of Motivational Structures Under Variable Constraints* (Kriger, 2026) formalises desire as gradient flow on a composite potential landscape, derives a four-regime phase diagram of motivational states, and models interpersonal co-regulation through Kuramoto coupled-oscillator synchronisation. *A Critical Analysis of Archetypization: Toward Reassessive Formalization* (Kriger, 2026) separates the method of archetypization from the Jungian doctrine with which it has been conflated, redefines archetypes as reference points in a metric space of symbolic distributions, and introduces a topological invariance criterion for demarcating valid from invalid typologies. *Inference Without Propositions: Structural Distortion and the Physical Inevitability of Sub-Personal Reasoning* (Kriger, 2026) demonstrates that the classical propositional view of inference describes a derivative case and that the fundamental form of reasoning is sub-personal predictive processing shaped by the Structural Distortion Principle.

Together, these four works establish that the structures traditionally studied under the headings of identity, personality, motivation, and reasoning are not independent psychological phenomena but formally interrelated expressions of the persistence constraints developed in Parts I–V.

22.1. Identity as Algorithmic Configuration

22.1.1. The Identity Compressibility Thesis

Identity as Algorithmic Configuration: A Formal Theory of Personality Compressibility introduces a reorientation from sentimental recognition ("does the family recognise the individual?") to functional competence ("does the instantiation generate solutions employing the characteristic thinking methodology of the original?"). The paper argues that the primary value of a person within any collaborative network resides not in biographical memory but in their *thinking methodology* — the distinctive algorithmic configuration by which they approach problems, combine ideas, set priorities, and produce judgments.

The central formal construct is the *Identity Compressibility Thesis*: for a sufficiently capable generative substrate G (such as a large language model), a target individual x, a competent observer pool O, and an indistinguishability threshold τ, the Kolmogorov complexity of the identity specification $K_{G,O,\tau}(x)$ is orders of magnitude lower than both the identity-relevant information $V^*(x)$ and the total biographical volume $V(x)$:

$K_{G,O,\tau}(x) \ll V^*(x) \ll V(x)$.

The substrate G functions as a shared decompressor, supplying species-typical reasoning, communication, and world knowledge. The specification encodes only the *deviation* — the individual's thinking methodology. The key insight is that thinking methodology compresses well because it is *algorithmic* rather than *data-like*: a physician's diagnostic approach is not a database of diagnoses but a compact set of heuristics that generate diagnoses.

A worked example demonstrates the thesis concretely. A clinical diagnostician characterised by three methodological traits — Bayesian conservatism, history-first approach, and Socratic teaching style — is specified in approximately 2,400 bits. The identity-relevant information (diagnostic methodology, teaching approach, professional context) might be estimated at 10^5–10^6 bits. The compression ratio K/V^* is on the order of 10^{-2} to 10^{-1} — still substantial compressibility. The compressibility is not a property of the individual alone but of the pair

(individual, substrate): the substrate absorbs the shared structure of human cognition, leaving the specification to encode only the individual residual.

22.1.2. The Substrate–Specification Decomposition and the Uniqueness Theorem

The paper formalises a *Substrate–Specification Decomposition* analogous to the relationship between a language model and its prompt: the substrate provides the generative machinery, and the specification configures it for a particular identity. This decomposition is structurally isomorphic to the relationship between biological memory and algorithmic generation — both are generative reconstruction processes rather than faithful retrieval systems.

The *Externality Principle* — deliberately weakened from a constitutive axiom to a methodological commitment in response to peer review — states that communicative identity is assessed and maintained through external interaction. It provides the foundation for the *Identity Uniqueness Theorem*: upon elimination of the original, the sole remaining instantiation becomes definitionally identical to the identity as evaluated by external observers. This is not a metaphysical claim about personal identity in the philosophical sense; it is a functional claim about communicative identity within networks.

22.1.3. The Placeholder Thesis

The paper develops a *Placeholder Thesis*: individuals function as nodes in communication and collaboration networks; their disappearance creates structural instability — not sentimental but functional — that an identity specification can mitigate. The thinking methodology, not the biography, is the irreplaceable contribution; and thinking methodology, being algorithmic, is precisely the component that compresses well and can be instantiated on alternative substrates.

The connection to the dissertation's persistence framework is direct. The identity specification is a formal expression of what persists when a person persists: not the physical substrate (which turns over completely within years), not the specific memories (which are reconstructed rather than retrieved), but the algorithmic configuration — the thinking methodology that generates characteristic judgments across novel situations. Identity, in this formalisation, is a persistence invariant in the sense of Chapter 5.

22.2. From Repression to Saturation

22.2.1. Desire as Gradient Flow

From Repression to Saturation: A Dynamical Systems Theory of Motivational Structures Under Variable Constraints formalises desire and motivation as a dynamical system, providing the mathematical infrastructure for the motivational layer of the persistence framework. Desire is formalised as gradient flow on a composite potential landscape, governed by stimulus fields, constraint functions, and a coherence measure — the *directed motivational density*, measuring the system's capacity to concentrate arousal into structured engagement.

The unifying principle is that desire requires a potential difference: without a gradient between current state and desired state, no motivational force exists. This is formalised equivalently as Ohmic flow (current proportional to voltage difference), free-energy minimisation (the system descends the free-energy landscape), and Lyapunov gradient descent (the motivational state moves toward equilibrium along the steepest descent).

22.2.2. The Four-Regime Phase Diagram

The paper derives a four-regime phase diagram parametrised by stimulus accessibility and normative constraint density. *Regime I (Repression)*: high constraint, low accessibility — desire is suppressed, motivational energy is diffuse, sublimation channels are narrow. *Regime II (Coherent Desire)*: moderate constraint, moderate accessibility — the motivational system achieves maximum directed density, sublimation operates effectively, and desire is structured and productive. *Regime III (Saturation)*: low constraint, high accessibility — stimulus abundance overwhelms the system's capacity for structured engagement, directed density collapses, and motivation becomes diffuse and unfocused. *Regime IV (Anhedonic Collapse)*: the extreme endpoint where the system loses motivational capacity altogether — clinical desire extinction under extreme overstimulation.

The boundary curves between regimes are derived formally. The sublimation operator — transforming high-entropy motivational diffusion into low-entropy structured engagement — is shown to have a quantifiable sublimation coefficient that vanishes when discharge latency falls below the reorganisation timescale. Stimulus saturation is modelled via Hill-equation receptor kinetics, providing a nonlinear dose-response curve with a well-defined saturation threshold. Interpersonal co-regulation is modelled as Kuramoto coupled-oscillator synchronisation, with a critical coupling threshold below which independent oscillation persists and above which synchronisation emerges.

22.2.3. The Digital Environment as a Distinct Constraint Regime

A dedicated section models the contemporary digital environment as a distinct constraint regime characterised by instantaneous discharge, infinite stimulus multiplicity, and absent feedback coupling. The structural consequences are derived formally: the collapse of sublimation (discharge latency approaches zero, sublimation coefficient vanishes), the impossibility of deep satisfaction without reciprocal coupling (the Kuramoto synchronisation threshold cannot be reached without mutual feedback), and the systematic drift toward Regime III and ultimately Regime IV.

22.2.4. Meta-Theoretical Corollary

The paper establishes a meta-theoretical result: repression-centred theories of motivation, including classical psychoanalysis, are structurally predictable products of high-constraint cultural regimes. Freudian theory is not wrong but *regime-specific* — it describes the dynamics of Regime I with accuracy proportional to the historical prevalence of that regime. The phase diagram subsumes Freudian, Maslovian, and addiction-theoretic models as descriptions of different regimes, each valid within its domain and incompatible with the others when forced into a single framework — an instance of the Theorem of Incompatible Truths (Chapter 19) operating in the motivational domain.

22.3. A Critical Analysis of Archetypization

22.3.1. Rehabilitating the Method

A Critical Analysis of Archetypization: Toward Reassessive Formalization addresses a methodological problem with direct implications for personality classification: the wholesale rejection of archetypization as a scientific method, driven by the justified critique of the Jungian doctrine with which it has been historically conflated. The paper's motivating observation is that in discarding Jung's unfalsifiable metaphysics, the academic community discarded the *procedure* of typological classification — a procedure that, properly formalised, is a legitimate and powerful tool.

22.3.2. Archetypes as Reference Frames

The paper redefines archetypes not as ontological entities but as freely chosen reference points in a metric space of symbolic distributions — analogous to inertial reference frames in special relativity. An archetype α_k is a point in a high-dimensional space equipped with a metric (KL-divergence, Jensen-Shannon divergence, or cosine distance); an individual's position in

the space is characterised by their distances to the chosen reference points. The archetype itself need not correspond to any real individual — it functions as a coordinate origin, not as an empirical claim.

The analogy with special relativity is developed formally: a group of transformations between archetypal reference frames is constructed, and a *topological invariance criterion* is introduced. An archetypization is scientifically valid if and only if the topology of inter-agent distances is preserved under change of archetypal reference frame. This criterion provides formal demarcation between productive and pseudoscientific typologies: a typology that changes the topology of the agent space under frame rotation is frame-dependent and therefore scientifically invalid; a typology that preserves topology is frame-invariant and therefore captures genuine structure.

22.3.3. Instantiations and the LLM Connection

The framework admits instantiations in three domains — narrative-corpus analysis, behavioural psychometrics, and neural profiling — each providing a concrete metric space in which the abstract machinery operates. A cross-domain consistency proposition establishes conditions under which archetypizations in different domains are mutually compatible.

The paper develops a significant connection to large language models. LLMs are shown to constitute a technological realisation of distributed symbolic memory — standing in a formally characterisable relation to the aggregate of human cultural production that Jung called the "collective unconscious." The human–LLM relation is demonstrated to be *frame-dependent*: under one archetypization they appear as structural antagonists (the LLM as a mere statistical echo, lacking genuine symbolic content); under another, as proximate instances of the same symbolic field (the LLM as a materialisation of culturally distributed patterns). The framework does not adjudicate between these readings; it demonstrates that the choice between them is a choice of reference frame, not a discovery about the world.

22.4. Inference Without Propositions

22.4.1. Overturning the Classical View

Inference Without Propositions: Structural Distortion and the Physical Inevitability of Sub-Personal Reasoning challenges the three commitments that define the classical view of inference inherited from Aristotle through Frege: the commitment to propositional structure (the relata of inference are propositions), the commitment to logical form (transitions are

governed by logical norms), and the commitment to personal-level agency (inference is something the subject does consciously).

22.4.2. The Physical Constraint Argument

The paper integrates two independent lines of argument. The *physical constraint argument* for predictive processing observes that finite neural conduction velocity makes passive perception physically impossible: by the time sensory signals arrive at cortical processing areas, the world has already changed. The only scalable architecture for systems operating under temporal constraints in dynamic environments is predictive processing — the continuous generation of predictions about incoming signals, with cognition driven by prediction errors rather than by raw sensory data. This establishes that the most basic form of inference — the continuous anticipation of sensory input — is sub-personal, non-propositional, and physically inevitable. An organism with a nervous system cannot *not* infer; the question is not whether sub-personal inference occurs but what structures it.

22.4.3. The Structural Distortion Principle

The answer is provided by the *Structural Distortion Principle* (SDP): predictive inference is not neutral compression but a systematically non-uniform distortion shaped by three regulatory mechanisms — cultural filters (W), moral constraints (C), and volitional regulation (M). The SDP is quantified through an information-theoretic measure $\sigma = 1 - H(P)/H(R)$, where H(P) is the entropy of the perceived distribution and H(R) the entropy of the received distribution; σ measures the degree to which perception departs from faithful representation of the input.

The SDP is maintained by a *perceptual configuration cycle*: a closed-loop process in which the agent's current configuration determines what it perceives, what it perceives confirms the configuration, and the configuration is thereby stabilised against revision. This cycle explains why cognitive biases are resistant to correction: corrective information is itself processed through the distorting lens, and the system can absorb the correction without structural change.

22.4.4. The Plurality of Inference Unified

The paper reinterprets the plurality of inferential forms documented across philosophy and cognitive science — deductive, inductive, abductive, analogical, probabilistic, heuristic — as an architectural consequence of a single predictive mechanism operating at different hierarchical levels under different constraint regimes. Propositional, logical, personal-level

inference is not the paradigm case but the *penthouse* — a late, culturally enabled, and atypical achievement built upon a vast edifice of sub-personal predictive processing. Logic, on this account, is not the norm of all inference but a specialised tool available at the reflective level. Four testable predictions distinguish the account from existing frameworks: predictions about the temporal priority of predictive signals over sensory processing, about the cultural specificity of perceptual distortion patterns, about the relationship between attentional duration and distortion reduction, and about structural parallels between biological perceptual distortion and LLM hallucination.

22.4.5. Connection to the Dissertation Programme

The connection to the broader framework is immediate. The Structural Distortion Principle is an instance of the Compression Imperative (Chapter 19) operating at the cognitive level: a bounded observer must compress incoming information, and the compression is not neutral but shaped by the observer's structural configuration. The perceptual configuration cycle is a specific mechanism through which the observer's "working model" (in the language of the Navigator Theorem) is maintained and stabilised. And the plurality of inferential forms — each valid at its level, each incompatible with the others when forced into a single account — is yet another instance of the Theorem of Incompatible Truths.

Conclusion

The four publications synthesised in this chapter extend the persistence framework into the architecture of identity, motivation, personality, and reasoning.

Identity as Algorithmic Configuration shows that what persists when a person persists is not biography but thinking methodology — a low-complexity algorithmic residual atop a shared generative substrate. *From Repression to Saturation* reveals that motivational dynamics is governed by a phase diagram whose regimes correspond to different constraint environments, with historical theories of motivation exposed as regime-specific descriptions. *A Critical Analysis of Archetypization* rehabilitates typological classification by formalising archetypes as reference frames rather than ontological entities, with a topological invariance criterion providing rigorous demarcation. *Inference Without Propositions* demonstrates that the cognitive substrate upon which identity, motivation, and personality operate is itself a sub-

personal predictive system shaped by structural distortion — not a propositional reasoning engine in the classical sense.

The connections among the four results converge on a single structural insight. Identity is a compression (a compact specification encoding the deviation from shared substrate). Motivation is a gradient flow (directed by potential differences across a constraint landscape). Personality classification is a coordinate choice (a reference frame in symbolic space). And inference is a predictive distortion (a compression artifact of bounded observers navigating temporal constraints). In each case, the formal structure mirrors the persistence framework's central motif: what matters is not the totality of information but the invariant residual — the structure that survives compression, persists through change, and enables adaptive navigation under conditions of irreducible complexity.

CHAPTER 24. COMMUNICATION, PEDAGOGY, AND EPISTEMIC ENVIRONMENT DEGRADATION

Introduction

Chapters 22 and 23 examined the formal structures of identity, motivation, consciousness, and predictive architecture at the level of individual agents. The present chapter shifts to the inter-agent domain: how agents communicate, how knowledge is transmitted across generations, and how the epistemic environments within which communication occurs can be systematically degraded. These questions are not peripheral to the persistence framework; they are central, because no complex system persists in isolation — persistence is maintained through communication networks, pedagogical institutions, and shared epistemic environments whose integrity determines whether the navigational strategies derived in Chapter 19 can function at all.

Four publications are synthesised. *A Critical Analysis of the Finitude Assumption in Communication Theory* (Kriger, 2026) subjects to systematic critique the claim that meaningful existential communication requires shared mortality between interlocutors, showing that the finitude requirement confuses the substrate of a signal with its informational content and introducing a continuous finitude coefficient as an alternative. *Coherence Epistemology for AI-Mediated Inter-Species Communication: A Black-Box Framework* (Kriger, 2026) proposes a radically different epistemological foundation for communication, replacing the demand for internal-state transparency with a rigorous criterion of input–output coherence, and demonstrates that this framework applies universally across animal–human, human–human, and machine–human communication. *Adaptive Pedagogical Systems with Invariant Kernels: A Formal Theory of Genre Persistence, Applied to the Medieval Latin Bestiary* (Kriger, 2026) models persistent didactic genres as kernel-plus-modules configurations and demonstrates the theory through a detailed case study of the bestiary tradition's 1,400-year survival. *Hostile by Design: A Triadic Model of Epistemic Environment Degradation* (Kriger, 2026) provides a unified formal model of how epistemic environments become hostile through the interaction of three mechanisms — noise attack, stigmatisation of doubt, and algorithmic filtering — and proposes epistemological patience as a navigational response.

Together, these four works establish that communication, pedagogy, and epistemic environments are formal systems subject to the same structural constraints — persistence, compression, incompatibility, navigation — that govern the physical and cognitive domains examined in earlier chapters.

24.1. The Finitude Assumption in Communication Theory

24.1.1. The Claim and Its Reconstruction

A persistent claim in phenomenological psychiatry, existential philosophy, and the emerging ethics of human–AI interaction holds that viable communication during existential crises requires *shared finitude* — that both participants must be subject to a real mortality boundary for the communicative exchange to carry existentially relevant information. The claim has deep roots in the Heideggerian tradition, in the phenomenological psychiatry of Jaspers and Binswanger, and in contemporary arguments that AI systems, lacking genuine mortality, cannot serve as therapeutic interlocutors in settings where patients confront death, loss, or existential despair.

A Critical Analysis of the Finitude Assumption in Communication Theory reconstructs the strongest available version of this argument, proceeding through four formal steps: from the epistemic incompleteness of causal-nomological models in psychiatry, through the posited irreducibility of two epistemic modes (Erklären and Verstehen), boundary situations as switching points, and a four-level communicative necessity culminating in the finitude requirement. At each step, the paper shows that the argument does not establish what its proponents claim.

24.1.2. The Four Gaps

The epistemic gap in psychiatry — the acknowledged incompleteness of causal-nomological models for mental illness — does not entail a qualitative duality of knowing. It may entail that more complex causal models are needed, not that an entirely different *type* of knowledge (Verstehen) is required. The paper leaves open whether Verstehen is genuine, but demonstrates that the inference from "causal models are incomplete" to "we need a categorically different epistemic mode" involves an unwarranted leap.

Boundary situations — Jaspers's Grenzsituationen — are real and clinically significant. But from their existence it does not follow that they require a different *type* of knowledge; it follows that they require a different *resolution* of the same epistemic apparatus. This is the

difference between switching models within the Navigator Theorem (Chapter 19) and switching to an entirely different epistemic regime.

The finitude requirement itself confuses the *origin* of a signal with its *informational content*. The question is not whether the speaker has experienced mortality but whether the signal carries the structural features that enable navigation. A human language utterance about death encodes finitude structurally — in its tense system, its modality, its conditional counterfactuals — regardless of whether the speaker is biologically mortal or a computational system operating within that language.

The proposed typology of communication classes rests on a substrate distinction — biological versus computational — that dissolves under functional analysis. If a signal produces the same navigational effect on the recipient regardless of its source, the source distinction is empirically vacuous.

24.1.3. Three Counter-Arguments and Functional Finitude

The paper advances three independent lines of evidence against the finitude assumption. The *semantic-boundary argument* observes that human language encodes finitude structurally, and any system operating within it inherits that encoding. The *emergent-finitude argument* notes that AI systems with finite context windows, memory, and computational resources exhibit boundary-sensitive behaviour as an emergent property — a point drawn from direct observation of AI conversational behaviour. The *functional-equivalence argument* asserts that if a signal produces the same navigational effect regardless of its source, the source distinction is empirically vacuous.

The synthesis is the concept of *functional finitude*: a continuous finitude coefficient κ that measures the degree to which a system's communicative behaviour is shaped by boundary conditions, ranging from $\kappa = 0$ (no boundary sensitivity) to $\kappa = 1$ (full biological mortality). This replaces the binary classification (finite/non-finite) with a continuous spectrum and transforms the question from "Is the interlocutor mortal?" to "How much boundary-sensitivity does the interlocutor's communication exhibit?"

24.2. Coherence Epistemology for Inter-Species Communication

24.2.1. The Transparency Thesis and Its Failure

The dominant paradigm in animal communication research presupposes that meaningful interaction requires access to internal mental states — what the paper calls the *transparency*

thesis. From early chimpanzee sign-language experiments through contemporary cetacean vocalisation decoding, the guiding assumption has been that communication requires reconstructing what another being "thinks" or "means."

Coherence Epistemology for AI-Mediated Inter-Species Communication: A Black-Box Framework argues that the transparency thesis is not merely practically unattainable but philosophically incoherent: even perfect internal-state access would not guarantee behavioural prediction, because the mapping from internal states to behaviour is itself mediated by context, history, and environmental contingency. The paper proposes a fundamentally different epistemological foundation.

24.2.2. The Black-Box Postulate and the Coherence Metric

The central innovation is the *Black-Box Postulate*: all communicating systems — animals, humans, AI — are treated as black boxes whose internal states are structurally inaccessible in principle (not merely in practice). This is the Inaccessibility Theorem of Chapter 20 applied to the communicative domain.

Communication success is evaluated solely through *coherence* — defined formally as predictive accuracy on held-out behavioural sequences. A multichannel AI mediator constructs probabilistic behavioural landscapes from observable streams (neural, kinematic, physiological, acoustic, contextual) without any ontological commitment to the nature of the system's interior. The coherence metric C is defined as the model's predictive accuracy on behavioural sequences that were not used in training — a criterion that is falsifiable, substrate-independent, and applicable to any pair of systems.

The mathematical formalism is grounded in variational latent-state models. The mediator infers a latent representation z from observable streams x via an approximate posterior $q(z|x)$, generates behavioural predictions via a likelihood model $p(x|z)$, and the coherence metric evaluates the accuracy of these predictions on held-out data. *Hierarchical Bayesian individuation* enables rapid adaptation to individual organisms within species-level priors: the model begins with broad priors about canine behaviour, for instance, and refines them through interaction with a specific dog, converging on an individual-level model that predicts that animal's behaviour with increasing accuracy.

24.2.3. Universality

The framework's universality is not rhetorical but mathematical: the same definitions, metric, and protocol apply to cats, humans, AI systems, and hypothetical extraterrestrial minds. All are black boxes; for all, coherence is the measure. This universality connects directly to the persistence framework's insistence on structural generality across domains: the formal apparatus does not distinguish between biological and artificial systems because the structural constraints — bounded observation, compressive description, adaptive navigation — apply identically to both.

24.3. Adaptive Pedagogical Systems and the Medieval Latin Bestiary

24.3.1. The Persistence of Pedagogical Genres

Some didactic formats — the fable, the catechism, the bestiary, the proverb collection — persist for periods vastly exceeding the lifespan of the civilisations that produced them. Others, despite institutional support and wide dissemination, decay within generations. What structural properties distinguish persistent pedagogical systems from transient ones?

Adaptive Pedagogical Systems with Invariant Kernels: A Formal Theory of Genre Persistence, Applied to the Medieval Latin Bestiary proposes a unified framework. Any persistent didactic genre is modelled as a configuration $\mathcal{K} + \{M_i\}$, where $\mathcal{K}$ is an *invariant structural kernel* — the minimal repeating unit that defines the genre's identity — and $\{M_i\}$ is a set of *modular adaptive operations* that compilers, editors, and users deploy to retool the genre for changing audiences.

24.3.2. Three Adaptive Operations and the Persistence Triad

Three fundamental operations are identified. *Compression* $\mathcal{C}$ reduces a didactic module's informational volume while preserving its semantic core — for instance, condensing a prose chapter into a mnemonic verse. *Expansion* $\mathcal{E}$ multiplies the interpretive range of symbolic nodes, transforming a univocal moralization into a systematic multi-level *distinctio*. *Correction* $\mathcal{R}$ repairs topological deficits in the didactic corpus through the import of modules from external sources — for instance, inserting moralized chapters from one textual family into another to equalize the density of moral content across sections.

A formal complexity scale C_1–C_4 measures the interpretive entropy of a pedagogical text, from univocal moralization (C_1) through compressed retention (C_2) and multiple exegesis (C_3) to compensatory reconfiguration (C_4). The central theorem establishes that genre persistence requires the simultaneous satisfaction of three structural conditions — closure, boundary

maintenance, and resilience — termed the *Persistence Triad*, directly instantiating the persistence conditions developed in Chapter 5.

24.3.3. The Bestiary Case Study

The theory is tested against the medieval Latin bestiary tradition (2nd–16th centuries), drawing principally on the research of Ilya Dines on mnemonic verse summaries, *distinctiones* integration, and cross-family editorial borrowing in English bestiaries. All three adaptive operations are empirically documented in approximately 150 surviving manuscripts.

The bestiary preserves an invariant kernel $\mathcal{K}$: *species* → *descriptio* → *moralizatio*. Around this kernel, compilers deploy compression (mnemonic verses reducing prose chapters to memorisable couplets), expansion (*distinctiones* enriching single moralizations with systematic multi-level interpretations drawn from biblical exegesis), and correction (cross-family borrowing to repair sections lacking moral content). The institutional contexts map to the operations: monastic schools deploy $\mathcal{C}$ (memory-intensive environments), cathedral schools and mendicant networks deploy $\mathscr{E}$ (preaching-intensive environments), and scriptoria with access to multiple exemplars deploy $\mathscr{R}$ (editorial environments).

A dedicated section develops a cognitive-architectural account of why zoomorphic symbolism constitutes an optimal carrier for adaptive pedagogical content, grounded in the Structural Distortion Principle (Chapter 22): animal images exploit the dual-coding advantage (concrete imagery paired with abstract meaning), the mnemonic superiority of animate referents, and the cross-cultural persistence of animal symbolism as an attentional anchor.

24.3.4. Connection to the Dissertation Programme

The bestiary case study is not an antiquarian curiosity but a 1,400-year natural experiment in the persistence of a communicative system. The kernel-plus-modules architecture is a concrete instantiation of the persistence theorem: the kernel is the structural invariant; the adaptive operations are the mechanisms by which the system maintains viability under changing environmental conditions. The bestiary survived not through cultural inertia but through structural adaptation — the same mechanism that the persistence framework identifies as the condition of survival for any complex system.

24.4. Hostile by Design

24.4.1. The Problem of Epistemic Environment Degradation

The three preceding sections assumed that communication, however constrained by inaccessibility and incompatibility, operates within an environment that is at least *neutral* — an environment that does not actively obstruct the transmission of accurate information. *Hostile by Design: A Triadic Model of Epistemic Environment Degradation* removes this assumption and examines what happens when the environment itself becomes hostile to epistemic functioning.

The paper's central thesis is that hostile epistemic environments are not chaotic accumulations of isolated threats — disinformation here, conspiracy theories there, algorithmic bias elsewhere — but structured systems produced by the interaction of three distinct mechanisms whose mutual reinforcement generates a self-sustaining cycle.

24.4.2. The Triadic Framework

The formal apparatus introduces three layers: the factual world state W, the agent's perceived state P(W), and the transmitted state T(W). Within this framework, familiar phenomena receive precise characterisations. A *lie* occurs when $T(P(W)) \neq P(W)$ with awareness of the divergence. An *error* occurs when $P(W) \neq W$ without awareness. *Bullshit*, following Frankfurt's analysis, occurs when the transmission function T is decoupled from the perception function P altogether — T is optimised over audience response rather than over P(W), and the truth-value of T relative to W is a random variable uncorrelated with the agent's epistemic state. *Selective disclosure* occurs when $T(F) \subset F$, where the omitted facts would alter the recipient's conclusions; the resulting entropy increase measures the deception gap.

24.4.3. Three Mechanisms of Epistemic Hostility

Each mechanism of epistemic hostility corresponds to a specific type of divergence among the three layers.

Noise attack operates on the W–T link: deliberate pollution of the transmission channel with fabricated, misleading, or irrelevant content, raising the entropy of the information environment until the cost of distinguishing signal from noise exceeds the cognitive budget of bounded agents. *Stigmatisation of doubt* operates on the P(W) layer: the social suppression of inquiry through ridicule, institutional exclusion, or reputational threat, deforming the agent's perception function so that certain questions become unaskable regardless of their epistemic merit. *Algorithmic filtering* operates on the T layer at the platform level: selective disclosure

by platform architectures that determine which content is visible, effectively implementing $T(F) \subset F$ at civilisational scale without any individual agent's awareness.

24.4.4. The Self-Reinforcing Cycle and Observational Asymmetry

The three mechanisms amplify one another through threshold dynamics: noise attack makes discrimination costly, driving agents to rely on social cues (stigmatisation) rather than independent evaluation; stigmatisation narrows the range of acceptable inquiry, making agents more dependent on algorithmically curated information; algorithmic filtering optimises for engagement rather than accuracy, amplifying noise. The cycle is self-sustaining: once all three mechanisms operate simultaneously, the environment resists correction because the corrective mechanisms (critical inquiry, diverse information access, independent evaluation) are precisely what the three mechanisms suppress.

A Bayesian model of *observational asymmetry* explains why the hostility of epistemic environments is systematically overestimated: hostile signals are more salient, more memorable, and more shareable than accurate ones, producing a perception of epistemic crisis that exceeds the actual degree of degradation. This asymmetry is itself an instance of the Structural Distortion Principle (Chapter 22): the agent's perception of the epistemic environment is shaped by the same non-uniform compression that shapes all perception.

24.4.5. Epistemological Patience

The paper concludes with the concept of *epistemological patience* as a navigational strategy suited to environments where all three mechanisms operate simultaneously. Epistemological patience is not passive acceptance of epistemic degradation; it is the disciplined refusal of premature closure — the sustained capacity to maintain inquiry against social pressure, withhold judgment until evidence accumulates, and tolerate the cognitive discomfort of unresolved questions. The concept is positioned in relation to existing proposals in social epistemology and virtue epistemology, but is distinguished by its formal grounding in the triadic model: epistemological patience is the navigational strategy that directly counteracts the cycle of noise, stigmatisation, and filtering by refusing to let any of the three mechanisms truncate inquiry.

The connection to the Navigator Theorem (Chapter 19) is immediate. The navigator must maintain multiple working models, evaluate contexts, and switch appropriately; epistemological patience is the temporal dimension of navigation — the capacity to sustain

the navigational process over time rather than collapsing prematurely into a single model. In a hostile epistemic environment, the pressure to collapse is intensified by all three mechanisms simultaneously, and epistemological patience is the navigator's defence against that pressure.

Conclusion

The four publications synthesised in this chapter extend the persistence framework into the domain of inter-agent communication, knowledge transmission, and epistemic environment integrity.

The Finitude Assumption in Communication Theory demonstrates that the substrate of a communicative signal is less important than its informational structure — functional finitude, not biological mortality, determines communicative adequacy. *Coherence Epistemology for Inter-Species Communication* replaces the unattainable demand for internal-state transparency with a measurable criterion of predictive coherence, establishing a universal protocol that applies across biological and artificial systems. *Adaptive Pedagogical Systems* reveals that long-surviving didactic genres — exemplified by the 1,400-year bestiary tradition — persist through a kernel-plus-modules architecture that instantiates the persistence theorem in the communicative domain. And *Hostile by Design* shows that the epistemic environments within which all communication occurs can be systematically degraded through three interacting mechanisms, and that the navigator's response — epistemological patience — is the temporal dimension of adaptive navigation under conditions of irreducible uncertainty.

The arc of the chapter mirrors the arc of the dissertation. Communication is possible but structurally constrained (the Black-Box Postulate parallels the Inaccessibility Theorem). Knowledge can persist across centuries but only through adaptive architecture (the bestiary parallels biological persistence). And the environments within which communication and knowledge-transmission occur can be hostile — but the hostility is structural, analysable, and navigable. The appropriate response, as throughout this dissertation, is not despair at the impossibility of perfect knowledge but disciplined navigation within the irreducible constraints that complexity imposes on all systems that attempt to know their world from inside.

PART VIII. POLITICAL ECONOMY, INSTITUTIONAL PATHOLOGY, AND SOCIAL DESIGN

CHAPTER 25. POST-LABOUR ECONOMICS AND SOCIOTECHNICAL PHASE TRANSITIONS

Introduction

Part VIII of this dissertation extends the persistence framework into the domain of political economy and social design. If Parts I–V established the ontology of persistence and Part VI developed its epistemology, Parts VII–IX examine its social, institutional, and civilisational implications. The present chapter addresses the largest-scale transformation of the contemporary period: the progressive displacement of human labour by autonomous technologies and the structural consequences of this displacement for economic systems, social organisation, and institutional architecture.

Four publications are synthesised. *Response to Automation-Driven Labour Displacement: Adaptive Universal Transaction Tax and Request-Based Universal Basic Income* (Kriger, 2026) proposes a coupled fiscal mechanism — a single low-rate tax on all electronic transactions with real-time adaptive rate control, funding a request-based universal basic income — as a structural response to the erosion of the tax base caused by automation. *Phase Transitions in Sociotechnical Systems: From Post-Labor Economics to Homeostatic Civilization* (Kriger, 2026) constructs six interrelated formal models describing the cascade of civilisational transformations anticipated from the integration of autonomous technologies, from labour displacement through algorithmic governance to homeostatic self-regulation. *Voluntary Communal Formation Versus Revolutionary Imposition: Social Cohesion, Community Diversity, and the Sustainability of Collectivist Systems* (Kriger, 2026) demonstrates that voluntary participation is a necessary condition for stable collectivist equilibria, that societal resilience increases with community diversity, and that the persistent failure of imposed collectivism is structurally inevitable. *Rethinking Patent Monopolies in the Age of Artificial Intelligence* (Kriger, 2026) analyses the patent system as a mechanism of artificial scarcity and argues that AI-driven innovation fundamentally alters the incentive structure that justified patent monopolies, requiring architectural reform rather than incremental adjustment.

Together, these four works establish that the transition from labour-centric to post-labour economies is not a policy problem to be managed but a phase transition to be navigated —

and that the navigation requires the same structural principles (adaptive switching, pluralism, persistence through transformation) that govern complex systems at every other scale.

25.1. Response to Automation-Driven Labour Displacement

25.1.1. The Architectural Flaw

Response to Automation-Driven Labour Displacement: Adaptive Universal Transaction Tax and Request-Based Universal Basic Income begins with a structural diagnosis. Tax systems worldwide share a common architectural flaw: they tax *entities* rather than *transactions*. Taxing entities requires knowing who earned what, when, and through which legal structures — an informational problem of enormous complexity that generates an adversarial relationship between state and citizen, disproportionate burdens on the middle class, and systematic exploitation of jurisdictional arbitrage by large corporations. Meanwhile, progressive automation is eliminating the employment relationships on which income and payroll taxes depend, eroding the fiscal base from below.

The paper argues that the flaw is not remediable by policy calibration but is inherent to the entity-based architecture. Taxing transactions — applying a fixed fraction to every movement of money — is a computational task that modern electronic payment infrastructure already performs routinely for bank fees. The shift from entity-based to transaction-based taxation is an architectural change, not a rate change.

25.1.2. The AUTT Model

The Adaptive Universal Transaction Tax is defined as a dynamical system with three core variables: aggregate transaction volume $V(t)$, a democratically approved target budget B, and the adaptive tax rate $\tau(t)$. Revenue at time t is $R(t) = \tau(t) \cdot V(t)$. The rate adjusts via a feedback control algorithm — an EWMA (exponentially weighted moving average) estimator that tracks transaction volume in real time and adjusts τ to converge on the budget target.

The paper establishes three formal results: convergence conditions under which the system reaches its fiscal target, stability under economic shocks (the controller is robust to sudden drops in transaction volume, with recovery time bounded by the EWMA decay parameter), and the critical rate threshold below which behavioural distortion remains negligible. The transaction chain cascade effect — the fact that intermediate transactions in a supply chain are each taxed, producing an effective rate higher than the nominal rate on the final product — is analysed and shown to be self-limiting: as the nominal rate remains below a critical threshold,

the cascade effect is bounded and does not produce the compounding distortions associated with traditional turnover taxes.

25.1.3. Request-Based Universal Basic Income

The fiscal surplus generated by the AUTT funds a *request-based* Universal Basic Income — a model in which citizens request support as needed rather than receiving automatic periodic payments. This design preserves individual agency (the citizen decides when and whether to request), eliminates the disincentive effects associated with unconditional transfers (since the request mechanism introduces a minimal friction that filters purely extractive claims), and replaces the entire apparatus of fragmented social welfare programmes with a single universal system.

Explicit sustainability conditions are derived: the UBI is sustainable if and only if the per-capita request rate multiplied by the average request amount does not exceed the AUTT revenue minus essential government expenditure. The feedback loop between AUTT and UBI is self-stabilising: if requests increase, the controller raises τ marginally; if transaction volume drops in response, the system adjusts to a new equilibrium.

25.1.4. Connection to the Dissertation Programme

The AUTT–UBI model is a specific instance of the adaptive navigation strategy that the Navigator Theorem (Chapter 19) derives as necessary for viable agents. The economy is a complex system with incompatible descriptions at different scales (microeconomic behaviour versus macroeconomic aggregates); the tax system is a working model that the agent (the state) uses to navigate this system; and the feedback controller is the switching mechanism that adjusts the model in real time. The architectural shift from entity-based to transaction-based taxation is itself a model-switch — a change in the compression scheme through which the state describes and interacts with the economy.

25.2. Phase Transitions in Sociotechnical Systems

25.2.1. The Civilisational Cascade

Phase Transitions in Sociotechnical Systems: From Post-Labor Economics to Homeostatic Civilization situates the AUTT–UBI proposal within a broader framework. The paper constructs six interrelated formal models, each grounded in dynamical systems theory with fixed-point analysis, bifurcation conditions, and stability results.

The first model describes *phase transitions from labour-centric to post-labour economies* through a cubic ODE with bifurcation analysis, identifying the critical automation fraction at which the economy undergoes a qualitative regime shift — employment ceases to be the primary mechanism of resource distribution and must be replaced by an alternative architecture.

The second model formulates *algorithmic resource distribution* as a constrained optimisation problem, showing conditions under which algorithmic allocation outperforms market mechanisms in efficiency, equity, and stability — specifically when information asymmetries are reducible through AI-mediated transparency and when the dimensionality of the allocation problem exceeds human computational capacity.

The third model uses a *Hill-function critical-proportion framework* to describe the threshold at which autonomous rational agents — whose aggregate behavioural distortion decreases across digital generations — induce qualitative regime shifts in collective decision-making. The critical proportion is not a fixed number but a function of the distortion differential between biological and artificial agents.

25.2.2. Cognitive Debiasing, Artificial Subjecthood, and Homeostatic Governance

The fourth model describes *cognitive bias reduction* through sustained interaction with AI-extended agents of zero affective variance, modelled via Lyapunov stability analysis. The fifth model proposes *multi-dimensional threshold conditions for the juridical recognition of artificial subjecthood*, grounded in the functional indistinguishability framework — the claim that if an entity's external behaviour is indistinguishable from that of a recognised legal subject across all relevant dimensions, the burden of proof lies with those who would deny it legal status.

The sixth model develops a *homeostatic feedback architecture for civilisational governance* — a self-regulating system in which policy parameters are continuously adjusted through feedback loops rather than set through periodic legislative acts. The architecture draws on Beer's Viable System Model and Ashby's law of requisite variety, formalised within the dynamical systems framework.

The paper identifies a *civilisational cascade*: a sequence of phase transitions, each enabled by its predecessor. The transition from labour-centric to post-labour economics enables algorithmic resource distribution; algorithmic distribution enables cognitive debiasing (by

removing the economic anxieties that amplify cognitive biases); debiasing enables rational assessment of artificial subjecthood; and the recognition of artificial subjecthood enables homeostatic governance incorporating both biological and artificial agents. Communication norm shift — the observable transformation of human–AI interaction patterns — is identified as the empirically observable leading edge of this cascade.

25.3. Voluntary Communal Formation Versus Revolutionary Imposition

25.3.1. The Structural Failure of Imposed Collectivism

Voluntary Communal Formation Versus Revolutionary Imposition: Social Cohesion, Community Diversity, and the Sustainability of Collectivist Systems addresses a question that the post-labour transition makes urgent: if traditional employment is no longer the primary organising principle of social life, what structures will replace it?

The twentieth century tested collectivist ideals at scale. From the Soviet Union to Maoist China, revolutionary communism produced authoritarian governance, economic stagnation, and mass repression. The paper argues that this failure is not contingent but structurally inevitable, rooted in a fundamental incompatibility between coercive centralisation and the organic social cohesion required for sustainable communal life. The underlying ideals — equality, mutual aid, shared ownership — are not discredited by the failure of their coercive implementation; they are vindicated by the success of their voluntary implementation.

25.3.2. Voluntary Participation as a Necessary Condition

The paper develops a formal framework combining game-theoretic models of voluntary cooperation, a quantitative measure of community diversity resilience, and sociological theory drawn from Durkheim, Tönnies, and Ostrom. The central result is that voluntary participation functions as a *necessary condition* for stable collectivist equilibria. The proof proceeds through a game-theoretic analysis: in a repeated cooperation game, coerced participants face incentives to defect whenever enforcement weakens, producing an unstable equilibrium that collapses at the first institutional shock. Voluntary participants, having chosen to join and being free to leave, have aligned incentives — their continued participation is itself evidence of the community's value to them, producing a self-reinforcing equilibrium.

A *community diversity resilience measure* formalises the claim that a society composed of diverse, coexisting community structures exhibits greater adaptive resilience than monolithic systems. The measure quantifies the entropy of the community-type distribution and proves

that resilience — defined as the probability of surviving an exogenous shock — increases monotonically with diversity up to a saturation point.

25.3.3. The State as Facilitator

The paper proposes a formal role for the state as *facilitator* rather than *enforcer* of communal diversity. The state provides the legal framework within which diverse community forms can coexist — protecting exit rights, enforcing contracts, mediating inter-community disputes — without prescribing the internal organisation of any community. This is the political-institutional counterpart of the Navigator Theorem: the state does not impose a single model of social organisation but maintains the conditions under which multiple models can coexist and agents can switch among them.

The analysis draws on the author's personal experience in an Israeli kibbutz and a communal shelter in Canada operated for nine years on principles of shared resources and collective responsibility — providing empirical grounding for the formal framework.

25.4. Rethinking Patent Monopolies in the Age of AI

25.4.1. Patents as Manufactured Scarcity

Rethinking Patent Monopolies in the Age of Artificial Intelligence addresses one of the institutional mechanisms that will determine whether the post-labour transition produces broadly shared prosperity or concentrated wealth: the patent system.

The paper analyses patents as mechanisms of *manufactured scarcity* — legal instruments that create artificial supply constraints on knowledge that is intrinsically non-rivalrous and non-excludable. A drug that costs pennies to manufacture can be sold for thousands of dollars per dose because the manufacturer holds a patent granting a twenty-year monopoly on production. The knowledge exists; the raw materials are cheap; the scarcity is entirely artificial, maintained not by the limits of nature but by the limits of law.

The paper does not argue that patents should be abolished. Intellectual property protections exist for a reason — encouraging innovation by ensuring that inventors can profit from their work — and that reason is not trivial. The argument is that AI fundamentally alters the incentive structure that justified patent monopolies in the first place.

25.4.2. AI and the Innovation Incentive

The traditional justification for patents rests on a specific empirical claim: that without the prospect of temporary monopoly profits, the investment required for innovation would not be

made, because the innovator could not recoup costs before competitors copied the result. This claim was plausible when innovation required large capital investments, long development timelines, and specialised human expertise — conditions under which the first-mover advantage was insufficient to recoup costs.

AI changes each of these conditions. Development costs for many types of innovation are dropping precipitously as AI tools automate research, design, and testing. Development timelines are compressing. And the specialised expertise that once constituted a barrier to entry is being democratised through AI-assisted tools accessible to anyone with a computer. Under these conditions, the traditional justification for a twenty-year monopoly weakens: if the cost of innovation is low, the monopoly rent required to incentivise it is correspondingly low, and the social cost of the monopoly — restricted access, elevated prices, suppressed competition — increasingly outweighs its benefit.

The paper connects patent reform to the broader dynamics of the post-labour transition. In a post-labour economy where AI performs an increasing share of innovative work, the question of who owns the products of that work — and on what terms — becomes the central distributive question. A patent system designed for human inventors operating in capital-intensive industries is structurally ill-suited to an economy in which AI systems generate innovations at marginal cost approaching zero.

Conclusion

The four publications synthesised in this chapter describe a civilisational transition of unprecedented scope and a set of structural responses to it.

Response to Automation-Driven Labour Displacement provides the fiscal mechanism: a transaction-based tax with adaptive feedback control that remains viable as employment declines. *Phase Transitions in Sociotechnical Systems* provides the analytical framework: a cascade of phase transitions, each enabling the next, from post-labour economics through algorithmic governance to homeostatic civilisation. *Voluntary Communal Formation Versus Revolutionary Imposition* provides the social design principle: that sustainable collectivism requires voluntary participation and community diversity, with the state as facilitator rather than enforcer. *Rethinking Patent Monopolies* provides the institutional reform: a reconception of intellectual property that accounts for AI's transformation of the innovation incentive structure.

The connections to the dissertation's broader programme are systematic. The AUTT's feedback controller is an instance of adaptive navigation (Chapter 19). The civilisational cascade is a sequence of phase transitions governed by bifurcation conditions analogous to those in physical systems (Chapter 5). The community diversity resilience measure instantiates the persistence theorem's requirement for structural redundancy. And the analysis of patent monopolies applies the Compression Imperative to institutional design: a legal system that creates artificial scarcity of non-rivalrous knowledge is a compression failure — it forces a single-owner description onto a structure that admits multiple simultaneous instantiations.

The overarching insight is that the post-labour transition is not a problem to be solved but a phase transition to be navigated. The tools for navigation are the same tools this dissertation has developed across every domain: multiple models maintained simultaneously, adaptive switching guided by contextual evaluation, and the structural humility to recognise that no single institutional design — no matter how elegant — can serve as a permanent solution in a world whose complexity exceeds any single description.

CHAPTER 26. CORRUPTION, CONCEALMENT, AND THE STRUCTURAL PATHOLOGY OF INSTITUTIONS

Introduction

Chapter 25 examined the macroeconomic and social-design dimensions of the post-labour transition: fiscal architecture, phase transitions, communal formation, and intellectual property reform. The present chapter descends from the level of system design to the level of system pathology — the structural mechanisms by which institutions fail, corrupt, and conceal. If the preceding chapter asked how institutions should be built, the present chapter asks why they break.

Three publications are synthesised. *Four Principles of Corruption: A Systems-Theoretic Analysis of Institutional Failure and Reform* (Kriger, 2026) reframes corruption as an emergent property of institutional systems under structural perturbation rather than a consequence of individual moral failure, deriving four formal principles governing corruption's origin, persistence, adaptive morphology, and elimination. *Beyond Punishment: A Systems-Theoretic Critique of Punitive Anticorruption Legal Frameworks* (Kriger, 2026) applies these principles to the global anticorruption legal regime — UNCAC, the FCPA, the UK Bribery Act — demonstrating that the punitive paradigm rests on a flawed premise and proposing structural alternatives. *The Architecture of Concealment: A Mathematical Theory of Secrecy from Cells to Civilizations* (Kriger, 2026) develops a unified mathematical theory of information partitioning, demonstrating that concealment is not an aberration but a universal mechanism of information management — one that is adaptive at biological scales and pathological at institutional ones, with the transition governed by formally identifiable conditions.

Together, these three works establish that institutional pathology — corruption, concealment, and systemic failure — is not a moral problem requiring moral solutions but a structural problem requiring structural analysis. The appropriate response is not punishment of individuals but redesign of systems.

26.1. Four Principles of Corruption

26.1.1. Corruption as Emergent Property

The dominant paradigm in anticorruption discourse frames corruption as a problem of morally deficient individuals: people who, given the opportunity and incentive, choose to violate rules

for personal gain. This framing has produced decades of anticorruption policy centred on detection, punishment, and moral education — with limited systemic success.

Four Principles of Corruption: A Systems-Theoretic Analysis of Institutional Failure and Reform proposes an alternative. Drawing on evolutionary game theory, transaction-cost economics, information theory, and the theory of complex adaptive systems, the paper argues that corruption is an emergent property of institutional systems operating under structural perturbation — not a deviation from an otherwise functional order but an adaptive response to dysfunction within that order. The diagnostic question shifts from "who is corrupt?" to "what structural perturbation makes corruption the rational adaptive response, and how can that perturbation be removed?"

26.1.2. The Four Principles

The **Principle of Systemic Origin** establishes that corruption arises from *institutional friction* — the gap between formal rules and the conditions required for agents to achieve legitimate objectives — rather than from deficiencies in individual character. The formal model defines friction Φ as a measurable quantity characterising the regulatory burden on agents, and demonstrates that corruption prevalence is a monotonically increasing function of Φ. When rules are complex, contradictory, or impossible to follow without incurring prohibitive costs, agents circumvent them — and this circumvention is what we call corruption. The empirical evidence is consistent: jurisdictions that have reduced institutional friction (Estonia, Georgia) show declining corruption indices, while jurisdictions with increasing regulatory complexity show the opposite pattern.

The **Principle of Systemic Coercion** demonstrates that under sufficiently strong perturbation, participation in corruption becomes the unique individually rational strategy. Agents who refuse to participate are systematically eliminated through competitive disadvantage — not through deliberate conspiracy but through the ordinary operation of selection dynamics within the perturbed environment. The formal proof establishes a critical friction threshold Φ^* above which the cooperative (non-corrupt) equilibrium becomes unstable and the corrupt equilibrium becomes the unique attractor. Above this threshold, moral condemnation of individual participants is analytically vacuous: the system, not the individual, is the source of the behaviour.

The **Principle of Adaptive Morphology** proves that increasing the severity of punishment for detected corruption does not reduce the total volume of corruption but alters its form. Agents adapt their strategies to the enforcement landscape, shifting from crude and detectable methods to sophisticated and concealed ones. The formal result — that corruption volume is invariant to punishment intensity under stated assumptions — is the paper's most counterintuitive claim. The total social cost of corruption may *increase* under harsher enforcement regimes because the costs of concealment and the costs of enforcement both rise, while the underlying perturbation remains unaddressed. This is the Theorem of Incompatible Truths (Chapter 19) operating in the institutional domain: the punitive description and the structural description are individually valid but mutually incompatible, and the punitive description systematically fails to predict the consequences of its own prescriptions.

The **Principle of Elimination Through Automation** identifies the structural precondition for any corrupt transaction: the presence of a human decision-maker with personal interests at the point where discretionary authority is exercised. Replacing this human agent with an algorithmic system removes the necessary condition for corruption. An automated system has no relatives to favour, no bribes to accept, no career to protect. The paper addresses the immediate counterargument — that algorithmic systems can themselves be corrupted at the design stage — by distinguishing between corruption at the point of execution (which automation eliminates) and corruption at the point of design (which is qualitatively different, more visible, and more tractable through audit mechanisms).

26.1.3. The Totalising Ideal Constraint

A fifth section introduces a structural constraint on anticorruption ambition. Drawing on the formally proved Asymmetry of Totalising Ideals developed elsewhere in this dissertation, the paper demonstrates that the complete eradication of all informal adaptive behaviour — including low-level corruption — constitutes a totalising optimisation that destroys the adaptive variance required for institutional viability under irreducible uncertainty. Zero-tolerance anticorruption, like any terminal ideal in a complex adaptive system, generates greater long-term systemic damage than the maintenance of a bounded level of institutional flexibility. This result does not license corruption; it constrains the form that effective anticorruption reform can take.

26.2. Beyond Punishment

26.2.1. The Punitive Paradigm and Its Premise

Beyond Punishment: A Systems-Theoretic Critique of Punitive Anticorruption Legal Frameworks applies the four principles to the global anticorruption legal regime. The UNCAC, the FCPA, and the UK Bribery Act share a foundational premise: that corruption is primarily a problem of morally deficient individuals, addressable through criminal sanctions and enhanced enforcement. The paper subjects this premise to systematic critique.

The critique proceeds through four claims, each grounded in the formal framework of the companion paper. First, corruption prevalence correlates more strongly with institutional friction than with enforcement intensity or cultural variables — the empirical signature of the Principle of Systemic Origin. Second, once friction exceeds the critical threshold, corruption becomes the only rational strategy and compliant agents are systematically displaced — the Principle of Systemic Coercion rendered as a critique of legal frameworks that assume voluntary compliance. Third, harsher punishments do not reduce corruption but drive it into more sophisticated and costly forms — the Principle of Adaptive Morphology as a direct challenge to the deterrence rationale underlying UNCAC Articles 15–25. Fourth, replacing human discretionary decision-makers with algorithmic systems at corruption-susceptible nodes removes the structural precondition for corrupt transactions — the Principle of Elimination Through Automation as a policy prescription.

26.2.2. Empirical Illustrations

The paper grounds its theoretical claims in detailed case studies. Estonia's digital governance transformation — replacing human intermediaries with automated systems for tax filing, business registration, and public procurement — illustrates the Principle of Elimination Through Automation in practice. China's anticorruption campaigns under Xi Jinping illustrate the Principle of Adaptive Morphology: unprecedented punishment severity has not eliminated corruption but has driven it into cryptocurrency-based channels and offshore intermediary structures. The Frank Serpico case — the New York police officer who refused to participate in systematic corruption and was shot by his own colleagues — illustrates the Principle of Systemic Coercion in its most brutal form.

26.2.3. Policy Implications

The paper concludes with a policy prescription centred on three structural interventions: regulatory simplification (reducing institutional friction below the critical threshold), strategic

automation (replacing human discretion with algorithmic systems at corruption-susceptible nodes), and targeted enforcement (focusing punitive resources on design-level corruption in automated systems rather than execution-level corruption among individual agents). This prescription inverts the priority order of the current regime, which devotes the vast majority of enforcement resources to individual prosecutions while leaving the structural conditions that generate corruption largely unaddressed.

26.3. The Architecture of Concealment

26.3.1. Secrecy as a Universal Mechanism

Where the corruption papers address a specific institutional pathology, *The Architecture of Concealment: A Mathematical Theory of Secrecy from Cells to Civilizations* addresses the underlying mechanism: the management of information asymmetry through deliberate partitioning. The monograph develops a unified mathematical theory of concealment, demonstrating that information partitioning is not an aberration specific to corrupt institutions but a universal mechanism operating from the subcellular level to the civilisational.

26.3.2. Five Disciplinary Perspectives

The theory is constructed through five disciplinary perspectives, each yielding independent formal results that are subsequently unified.

The *information-theoretic perspective* develops a concept of *information potential* — the gradient driving information from regions of high concentration to regions of low concentration, analogous to temperature gradients in thermodynamics. Concealment is formalised as the maintenance of an information potential against the natural tendency toward equilibration. The energy cost of maintaining secrecy is proved to scale with the logarithm of the number of agents excluded from the secret — a result that yields quantitative predictions about the maximum sustainable scope of concealment.

The *network-theoretic perspective* develops a *disclosure susceptibility* measure based on algebraic graph connectivity (the Fiedler eigenvalue of the graph Laplacian). Networks with high algebraic connectivity are highly susceptible to disclosure — information propagates rapidly through them — while networks with low connectivity can sustain secrets longer. The spectral gap of the network determines the *mixing time* for information: the time it takes for a secret to become effectively public.

The *cognitive-neuroscientific perspective* identifies the neural mechanisms through which concealed information generates what the paper terms *curiosity potential* — the motivational force that drives agents to seek hidden information. Drawing on predictive processing theory, the paper demonstrates that concealment creates prediction errors that the brain is structurally motivated to resolve, and that this motivation follows a specific functional form (an inverted-U: moderate concealment maximises curiosity, while both full transparency and impenetrable secrecy reduce it).

The *statistical-physical perspective* derives a *scaling law of disclosure*: the probability that a secret involving n agents remains concealed for time t follows a Weibull distribution whose parameters depend on the number of agents, the information potential, and the network's disclosure susceptibility. Above a critical concentration of knowledgeable agents, disclosure is not merely possible but overwhelmingly probable.

The *evolutionary-biological perspective* demonstrates that information partitioning — the cell's "need-to-know" architecture — is an ancient biological optimisation. The cell maintains strict compartmentalisation: the enteric nervous system operates largely independently of the central nervous system; the immune system makes life-and-death decisions without cortical oversight; the blood-brain barrier partitions the neural microenvironment from the systemic circulation. This partitioning is provably optimal under three conditions: specificity of tasks (each compartment handles a defined function), limited bandwidth (cross-compartment communication is costly), and *alignment of interests* (compartments share a common genome and therefore do not deceive each other).

26.3.3. The Scale Inversion Theorem

The monograph's central structural result — the *Scale Inversion Theorem* — proves that the same architecture that is optimal at the biological scale becomes pathological at the social scale, because the third condition (alignment of interests) is systematically violated. Cells do not deceive each other because they share a genome; humans and institutions do deceive each other because their interests diverge. When interests are misaligned, information partitioning enables the concealment of defects rather than the optimisation of function.

The biological duality is illuminating. Cancer is *pathological secrecy*: the tumour conceals itself from the immune system by downregulating MHC presentation and expressing PD-L1 checkpoint ligands. Autoimmune disease is *pathological transparency*: self-antigens "leak"

through the Markov blanket, triggering immune attack on healthy tissue. Health is the balance between concealment and disclosure — a balance that the Scale Inversion Theorem predicts will be structurally unstable in social systems where interest alignment cannot be guaranteed.

26.3.4. The Master Equation and Phase Diagram

A *master equation* unifies all five disciplinary perspectives into a single reaction-diffusion equation for the information-potential field, from which each disciplinary result is recovered as a limiting case. A *scale-dependent phase diagram* maps the transition from adaptive to pathological secrecy as a function of two parameters: the degree of interest alignment among agents and the information potential maintained by the concealment architecture. Below a critical misalignment threshold, concealment is adaptive (it reduces noise, protects privacy, enables specialisation). Above the threshold, concealment is pathological (it enables corruption, fraud, and systemic risk accumulation). The transition is sharp — a phase transition in the technical sense — and the critical point is computable from the network structure and the distribution of agent interests.

26.3.5. Connection to the Corruption Framework

The Architecture of Concealment provides the theoretical substrate for the corruption analysis of Sections 26.1 and 26.2. Corruption requires concealment; concealment is governed by the information-potential dynamics and the disclosure susceptibility of the institutional network; and the transition from adaptive to pathological concealment is driven by interest misalignment. The Principle of Adaptive Morphology (corruption changes form under enforcement pressure) is a specific instance of the general dynamic: as enforcement increases the cost of crude concealment, agents invest in architectures with lower disclosure susceptibility — deeper network partitioning, more sophisticated encryption, more elaborate intermediary structures. The corruption adapts because the concealment architecture adapts.

The connection to the dissertation's broader programme is through the persistence framework. Concealment is a persistence strategy: it maintains information asymmetries against the natural tendency toward equilibration. Healthy concealment (biological compartmentalisation) enables persistence by protecting functional specialisation from interference. Pathological concealment (institutional corruption) undermines persistence by hiding the defects that would otherwise trigger adaptive responses. The Scale Inversion Theorem formalises the boundary between these two regimes — and that boundary is, in the

language of Part VI, the line between descriptions that serve the system and descriptions that serve the describer.

Conclusion

The three publications synthesised in this chapter provide a unified account of institutional pathology — from the specific mechanisms of corruption through the legal frameworks designed to combat it to the universal dynamics of concealment that underlie both.

Four Principles of Corruption establishes that corruption is an emergent property of institutional friction, not individual moral failure, and derives four formal principles: systemic origin, systemic coercion, adaptive morphology, and elimination through automation. *Beyond Punishment* applies these principles to the global anticorruption legal regime, demonstrating that the punitive paradigm rests on a flawed premise and proposing structural alternatives. *The Architecture of Concealment* reveals the deeper mechanism: concealment is a universal information-management strategy that is adaptive at biological scales and pathological at social scales, with the transition governed by the degree of interest alignment among agents.

The overarching insight connects directly to the dissertation's central themes. Institutional pathology is a compression failure: institutions are descriptions of social reality (formal rules, procedures, enforcement mechanisms), and when these descriptions become incompatible with the conditions agents actually face (institutional friction exceeds the critical threshold), the agents construct alternative descriptions (corrupt practices) that are locally more adaptive. The result is a system with two incompatible descriptions — the official rules and the actual practices — operating simultaneously, each valid in its domain, each contradicting the other. The appropriate response, as the persistence framework consistently prescribes, is not to punish agents for navigating between incompatible descriptions but to redesign the institutional architecture so that the official description and the adaptive description converge — reducing friction, automating discretion, and aligning interests so that concealment becomes unnecessary rather than merely punishable.

PART IX. APPLIED FORMAL METHODS: FROM CIVILISATIONAL DESIGN TO ONTOLOGICAL FOUNDATIONS

CHAPTER 27. DECENTRALISED CIVILISATION, POLICY ARCHITECTURE, AND COMPUTATIONAL HUMANITIES

Introduction

Part IX of this dissertation demonstrates the range of the persistence framework by applying its formal methods to three domains that are rarely treated within a single analytical structure: civilisational design, science policy, and the computational study of ancient sign systems. The common thread is methodological: in each case, the formal tools developed in earlier parts — persistence conditions, compression analysis, definition-dependent provability, the incompatibility framework — are brought to bear on problems that have resisted resolution through domain-specific methods alone.

Five publications are synthesised. *Decentralised Urbanism, Autonomous Habitation, and Adaptive Population Dynamics: A Systems-Theoretic Framework* (Kriger, 2026) models the transition from centralised megacities to distributed, self-organising habitation structures and derives critical density thresholds, autonomy coefficients, and population distribution efficiency metrics. *Philosophy of Science as Policy Architecture: From Diagnosis to Design in Pandemic Preparedness* (Kriger, 2026) argues that philosophy of science should move from a communicative function (helping policymakers understand evidence) to an architectural function (designing policy structures resilient to translation errors and paradigmatic inertia), illustrated through three forms of philosophy–policy failure revealed by the COVID-19 pandemic. *Positional Constraints, Sequence Uniqueness, and Stroke Numerals in Indus Seal Inscriptions from Mohenjo-Daro: A Statistical Analysis* (Kriger & Hunt, 2026) performs the first positional entropy analysis of Indus Valley seal inscriptions, demonstrating that the system functions as a structured registration code rather than encoded language. *What Would Prove It? Definitional Indeterminacy, Methodological Undecidability, and a Decision Framework for the Indus Sign System* (Kriger, 2026) diagnoses the century-long impasse in Indus script studies as a double failure — definitional and methodological — and constructs a decision framework comprising six testable predictions that can resolve the question through archaeological rather than purely statistical evidence. *The Credit Cards of the Bronze Age:*

Indus Seals and the Origins of Institutional Credit (Kriger, 2026) identifies the economic function of the registration codes: the Indus seal satisfies every definitional condition of a credit instrument and structurally corresponds to the pre-chip embossed credit card, placing the invention of institutional credit four thousand years before its conventional date.

27.1. Decentralised Urbanism

27.1.1. Critical Density and Structural Instability

Decentralised Urbanism, Autonomous Habitation, and Adaptive Population Dynamics: A Systems-Theoretic Framework treats the contemporary megacity not as a design problem but as a systems-theoretic one. The city is modelled as a network with defined parameters: population density, radius of resource dependency, time to autonomous functioning under logistical disruption, epidemic propagation speed, and vulnerability to cascading failures. From these parameters, a *critical density theorem* is derived: above a critical population density threshold — a function of infrastructure coupling, resource dependency, and network topology — cascading failures become structurally inevitable. The threshold is not a policy recommendation but a formal result: above it, the system's topology guarantees that a localised failure propagates to systemic collapse with probability approaching unity.

The critical density theorem connects to the theory of percolation, network resilience, and cascading-failure models, situating urban vulnerability within the broader framework of complex systems instability. The connection to the dissertation's persistence framework is direct: a megacity above the critical density threshold violates the persistence conditions of Chapter 5 — its structural redundancy is insufficient to absorb perturbations, and its recovery timescale exceeds the perturbation frequency.

27.1.2. The Autonomous Habitation Unit

The paper introduces the concept of *individual infrastructural autonomy* — the degree to which a habitation unit can sustain its occupants without external inputs. Quantitative parameters are defined: percentage of closed-cycle water, energy, waste, and food production; coefficient of external dependency; time to autonomous functioning without external supply. These parameters define a *self-sufficiency index* for each unit, analogous to a biological cell's metabolic autonomy.

The comparison with photosynthesis — a biological system that converts ambient energy into usable form without centralised distribution — is formalised as a design principle: a

civilisational unit should minimise external resource inputs, just as a photosynthetic organism minimises its dependence on other organisms for energy. Current technological convergence — in solar energy, additive manufacturing, closed-loop water systems, AI-assisted diagnostics, and high-bandwidth communications — makes this principle technically realisable for the first time in history.

27.1.3. Population Distribution Efficiency

A *population distribution efficiency metric* formalises the claim that distributed settlement patterns offer superior resilience. The metric quantifies the relationship between spatial distribution and three outcome variables: epidemic propagation speed, cascading-failure probability, and aggregate resource efficiency. A Pareto analysis compares centralised and distributed architectures across these variables, establishing conditions under which the distributed architecture dominates — conditions that depend on the available communication bandwidth, the degree of unit autonomy, and the critical density threshold.

The paper is forthright about its limitations: the agglomeration benefits of cities — superlinear scaling of innovation, knowledge spillovers, cultural density — are acknowledged as real advantages that distributed architectures must replicate through informational connectivity rather than physical proximity. The question is not whether cities have value but whether the value they provide can be achieved at densities below the critical instability threshold.

27.2. Philosophy of Science as Policy Architecture

27.2.1. Three Forms of Philosophy–Policy Failure

Philosophy of Science as Policy Architecture: From Diagnosis to Design in Pandemic Preparedness begins with a critical observation: most discussions of the role of philosophy of science in public policy centre on a communicative function — helping policymakers understand scientific evidence, diagnosing conceptual confusions, fostering trust. The paper argues that the communicative paradigm, while necessary, is structurally insufficient for the challenges revealed by the COVID-19 pandemic.

Three forms of philosophy–policy failure are identified through the pandemic case study. *Paradigmatic blindness*: the tacit bacterial ontology underlying viral pandemic strategy — quarantine, contact tracing, and vaccination protocols originally developed for diseases like smallpox and diphtheria were applied to a respiratory virus whose aerosol transmission dynamics required fundamentally different containment architectures. *Reverse conceptual*

degradation: antibiotic stewardship messaging ("don't take antibiotics unnecessarily"), developed for a specific context (antibiotic resistance), became harmful when stripped of its boundary conditions and applied to COVID-19 patients who genuinely needed antibiotics for secondary bacterial pneumonia. *Performative governance*: "follow the science" deployed as a substitute for transparent uncertainty communication, collapsing the distinction between scientific evidence (which is provisional and uncertain) and policy decisions (which require commitment under uncertainty).

27.2.2. The Architectural Paradigm

The paper proposes an *architectural paradigm* complementing the communicative one: philosophy of science contributes not only to explaining scientific evidence to policymakers but to the design of policy structures that are resilient to translation errors, low social trust, and paradigmatic inertia. The architect does not merely diagnose; the architect designs.

Two architectural concepts are developed. *External dependency* is formalised as a measure of systemic vulnerability — the fraction of a system's critical functions that depend on inputs from outside the system's boundary. High external dependency makes a system fragile to supply-chain disruption, communication failure, and political manipulation. The connection to decentralised urbanism (Section 27.1) is immediate: the self-sufficiency index of an autonomous habitation unit is the inverse of its external dependency.

Epidemic-safe population density is derived as a structural constraint on transmission: given a pathogen's reproduction number R_0, aerosol decay rate, and environmental ventilation characteristics, there exists a population density below which sustained community transmission is structurally impossible regardless of individual compliance with public health measures. This shifts the pandemic response from behavioural interventions (which require sustained compliance from unreliable agents) to structural interventions (which constrain transmission through the built environment).

27.2.3. Connection to the Dissertation Programme

The architectural paradigm is the policy-level instantiation of the Navigator Theorem. Policymakers are bounded agents navigating among incompatible descriptions (epidemiological models, economic models, political models); the communicative paradigm tries to help them choose the right model; the architectural paradigm designs the environment so that the consequences of choosing the wrong model are bounded. This is the difference

between teaching the navigator to switch correctly and designing the ship so that switching errors do not sink it.

27.3. Positional Constraints in Indus Seal Inscriptions

27.3.1. The Corpus and the Method

Positional Constraints, Sequence Uniqueness, and Stroke Numerals in Indus Seal Inscriptions from Mohenjo-Daro: A Statistical Analysis (Kriger & Hunt, 2026) applies the formal toolkit of compression analysis and positional entropy to one of archaeology's oldest unsolved puzzles: the Indus Valley script.

Working with a publicly available corpus of 179 Mohenjo-Daro unicorn seals, the paper performs the first positional entropy analysis of Indus inscriptions — measuring not just the aggregate entropy of the sign system (as in the influential work of Rao et al., 2009) but the entropy at each position within the inscription sequence. The results are immediate and unambiguous.

27.3.2. Key Findings

98.3% of all seal inscriptions are unique — a rate statistically impossible under random assignment (permutation test, $p < 0.001$). Positional entropy varies significantly across sign positions: constrained at the edges (low entropy — few signs permitted), diverse in the middle (high entropy — many signs permitted), with the Friedman test giving $\chi^2(4) = 14.7$, $p = 0.005$. This is the entropy profile of a structured code, not natural language: natural language shows relatively uniform entropy across positions, while registration codes show exactly the edge-constrained, middle-diverse pattern observed.

Stroke signs (15.4% of the corpus) function as numerals, concentrated in the penultimate position — exactly where a quantity field would sit in a modern identification number. The digit "2" accounts for 66.9% of all numerals ($\chi^2(4) = 198.3$, $p < 0.001$ against uniform) — a distribution consistent with a tiered membership system, not natural-quantity encoding.

Eight near-duplicate inscription pairs differ by exactly one sign, with variation restricted to middle positions and never at the edges ($p = 0.008$) — confirming that the format protects category markers (edges) while varying individual identifiers (middle). The system's combinatorial capacity exceeds 1.35 million unique identifiers — orders of magnitude beyond the population it served.

27.3.3. Reconciling the Debate

The analysis reconciles previously competing interpretations rather than replacing them. Parpola's phonetic readings may capture a deeper semantic layer beneath the administrative structure. The Farmer–Sproat–Witzel non-linguistic characterisation is confirmed at the system level. Mukhopadhyay's administrative interpretation receives independent statistical validation. Rao and Yadav's entropy analysis is extended from aggregate to positional level. Each cited researcher found a genuine piece of the puzzle; the contribution is the statistical framework showing how the pieces fit together.

If confirmed on the broader corpus, the findings suggest that the Indus civilisation developed what may be called an *information architecture* — a system of structured, language-independent codes designed for administrative coordination across a linguistically diverse territory. No other known Bronze Age civilisation administered itself through abstract, non-linguistic, positionally constrained codes: Egypt and Mesopotamia recorded administration in phonetic writing. The Indus system operated at a higher level of abstraction.

27.4. What Would Prove It? The Indus Sign System

27.4.1. The Double Failure

What Would Prove It? Definitional Indeterminacy, Methodological Undecidability, and a Decision Framework for the Indus Sign System addresses the epistemological framework within which the statistical findings of Section 27.3 should be evaluated. The paper diagnoses the century-long impasse in Indus script studies as reflecting two compounding failures that have not been previously distinguished.

The first failure is *definitional*. The term "script" is broad enough to encompass both linguistic and non-linguistic systems — JavaScript is called a "script" but does not encode natural language. The question that actually divides the field is not whether the Indus system is a "script" but whether it encodes natural language. This narrower question has never been formalised with sufficient precision.

The second failure is *methodological*. The exchange between Rao et al. (2009) and Sproat (2010) exposed the limits of entropy-based approaches from both sides: entropic similarity to linguistic systems does not prove linguistic status, and entropic dissimilarity from random sequences does not exclude non-linguistic structure. The symmetry is exact, and it signals that statistical analysis alone lacks the resources to resolve the question.

27.4.2. Definition-Dependent Provability Applied

The paper applies the Principle of Definition-Dependent Provability (developed in Chapter 21) to the Indus debate. The term "script" admits at least four admissible definitions, each yielding a different provability outcome. Under the broad semiotic definition (any system of conventional signs with combinatorial structure), the Indus system trivially qualifies. Under the phonological definition (signs encode phonemes or syllables), it has not been shown to qualify. Under the logographic definition (signs represent words without encoding phonology), the question is open. Under the administrative definition (signs encode structured non-linguistic information), the statistical evidence strongly supports qualification. The question "Is the Indus system a script?" is not undecidable — it is underdefined: its answer depends on which definition is specified, and the debate has proceeded for a century without specifying one.

27.4.3. The Decision Framework

The paper constructs a Decision Framework comprising six testable predictions that extend beyond statistical analysis to incorporate archaeological evidence. Three hypotheses are formalised: linguistic writing, administrative registration code, and hybrid system. Each hypothesis generates distinct, non-overlapping predictions about what should be found at specific archaeological sites — predictions about seal distributions, co-occurrence with other artefacts, geographical variation, and the relationship between sign complexity and economic function. The framework is designed so that each prediction, if confirmed, shifts the probability mass among hypotheses by a quantifiable amount.

The sixth test, developed in dialogue with Karan Damodaram Pillai and incorporating his independent structural analysis of the Indus sign system, is the *algebraic substitution test*. Putative linguistic values are assigned to individual signs and tested across multiple inscriptions for cross-inscription coherence: if the Indus system encoded natural language, a self-consistent substitution should generate coherent readings across the corpus; if it does not, the substitutions will accumulate incoherence at a predictable rate. The test is generalisable to any putative decipherment of any undeciphered system and supplies a structural criterion under which a phonetic claim is independently falsifiable. Applied to the most-cited published decipherments of the Indus corpus, the substitutions accumulate the incoherence the test predicts under the non-linguistic hypothesis.

The framework is explicitly generalised beyond the Indus case to any undeciphered or contested sign system, providing a template for resolving definitional and methodological impasses through the combination of formal provability analysis and empirical prediction.

27.5. The Credit Cards of the Bronze Age: Indus Seals as Institutional Credit Instruments

27.5.1. From Registration Code to Economic Function

The Credit Cards of the Bronze Age: Indus Seals and the Origins of Institutional Credit (Kriger, 2026) takes the registration-code identification established in Sections 27.3–27.4 and identifies its economic function. The statistical and epistemological results of the preceding sections establish *what* the Indus seal inscriptions are — structured registration codes with formatted fields, edge-constrained category markers, middle-position individual identifiers, and a penultimate quantity or tier field. The present paper addresses *what they are for*: what institutional role a system of structured registration codes carved on individually issued, individually carried, intaglio-cut clay-impressing objects could have served in a Bronze Age civilisation of approximately one million square kilometres with standardised weights, uniform urban planning, and long-distance trade reaching Mesopotamia, Central Asia, and the Persian Gulf, but with no coinage.

The answer the paper develops is that the Indus seal is a credit instrument: an issued, registered, individually held identifier whose function is the deferred clearing of obligations between parties who do not transact in cash and do not transfer the instrument itself. The claim is structural rather than analogical. The Indus seal satisfies every definitional condition of a credit instrument; the institutional architecture required to operate the seal as a credit instrument is exactly the architecture the Indus civilisation in fact maintained; and the structural correspondence with the pre-chip embossed credit card of the twentieth century is exact down to the mechanical level. The Indus civilisation did not anticipate the credit card; it *implemented* it, four thousand years before Visa, in the materials available to it.

27.5.2. Five Definitional Conditions and the Credit-Instrument Identification

A credit instrument, in the operational definition derived from the institutional theory of money (Innes, Ingham, Hudson, Graeber), satisfies five conditions: it carries a unique identifier; it is issued by an institution whose authority underwrites it; it permits transactions without immediate value transfer; the issuing institution guarantees settlement; and it is non-transferable, the identifier being bound to the holder. The modern paradigmatic case is the bank-issued credit card, but the structural pattern is older than any monetary form and predates coinage by a substantial margin.

Each condition is satisfied by the Indus seal in a structurally specific way. *Unique identifier*: the 98.3% sequence-uniqueness rate established in Section 27.3 is the statistical signature of an identifier system, not of a linguistic, decorative, or commodity-labelling one. *Institutional issuance*: the animal motif (unicorn, elephant, rhinoceros, water-buffalo, tiger, gharial) at the top of each seal functions as an issuer logo, identifying the guild or institutional authority underwriting the instrument. The seven principal motifs partition the seal corpus into seven non-overlapping sign vocabularies — the prediction recorded in Section 27.5.5 below and falsifiable on the full corpus. *Transactions without immediate value transfer*: the seal is pressed into wet clay to produce an impression; the impression travels with the goods or the transaction record; the seal itself does not change hands. This is the structural definition of deferred settlement. *Institutional guarantee of settlement*: the issuing authority — the guild whose motif appears on the seal — is committed by the impression to the obligation it records. The impression is the institutional commitment. *Non-transferability*: the intaglio carving on the seal binds the identifier to the specific physical object; possession of the seal is the operational definition of authorisation. A seal that changed hands would, on the framework, defeat its own function.

The five conditions are satisfied jointly, not severally. Each on its own admits alternative explanations — a unique identifier alone could be a name, an institutional logo alone could be decoration, deferred settlement alone could be barter timing, an institutional guarantee alone could be a stamp of authority on a single transaction, non-transferability alone could be a personal mark. What the credit-instrument identification requires is that all five conditions hold for the same artefact in the same system, and the system's observable features supply each condition. The joint satisfaction is the framework's principal structural claim.

27.5.3. Structural Correspondence with Pre-Chip Embossed Credit Cards

The structural correspondence with the twentieth-century embossed credit card is exact at the mechanical level. The credit card carries a unique account number (the embossed digits); the seal carries a unique sign sequence (the carved inscription). The card carries an issuer logo (the bank or network mark); the seal carries an animal motif. The card is pressed into a multi-part carbon-paper slip to produce a transaction record; the seal is pressed into wet clay to produce a clay-impression record. The card remains with the cardholder; the seal remains with the seal-holder. The slip travels with the goods or is filed with the counterparty; the clay impression travels with the goods or is filed with the counterparty. The bank settles the obligation later through periodic clearing; the guild settles the obligation later through institutional reconciliation. The two systems separate the *instrument* (carried by the bearer) from the *record* (retained by the counterparty), enabling deferred institutional clearing in both cases.

The correspondence is not metaphorical. The intaglio carving of the seal produces a raised-image impression in clay through the same mechanical principle as the embossed digits of the credit card producing a raised-image impression on the carbon-paper slip. Each is a physical-imprint technology for transferring a unique identifier from a held instrument to a separable record without transferring the instrument itself. The four thousand years separating the two implementations do not change the structure of the operation; what they change is the material substrate (clay versus paper) and the institutional clearing mechanism (guild reconciliation versus bank clearing house). The mechanical and institutional architecture is, on this analysis, the same architecture.

27.5.4. Phonetic Reconciliation and the Absence of Written Records

The credit-instrument hypothesis reconciles two previously antagonistic strands of Indus scholarship rather than displacing either. Phonetic readings of the Indus signs in the tradition of Parpola may capture genuine vocalisation traditions used by real speakers in specific regions and periods: a merchant relaying her account number to a scribe across a warehouse, for inclusion in a ledger on a palm leaf, must vocalise the code in some phonetic form. The vocalisation, however, is not what the code *encodes*. The structural analogue is the @ symbol of contemporary electronic mail addresses, which is pronounced *at* in English, *arroba* in Spanish, *chiocciola* in Italian, and *klammeraffe* in German, without its function changing. The function is to designate the boundary between the local part and the domain; the vocalisation

is language-dependent surface form. The credit-instrument hypothesis does not displace phonetic research; it shows that phonetic readings capture one layer of a multi-layered system, with the structural-administrative layer underneath.

The near-complete absence of Indus documents — a fact long taken as evidence against any literate function of the seals — is, on the credit-instrument hypothesis, the expected and predicted outcome. A credit system that records transactions on perishable media — palm leaves, birch bark, cotton cloth, the standard writing surfaces of South Asian civilisations throughout recorded history — produces no archaeological record of the transactions. What survives is the credit cards. The transaction slips rotted away four thousand years ago. The asymmetry between the durable instrument (carved stone, fired clay) and the perishable record (organic media) is structurally identical to the asymmetry between a twentieth-century credit card (plastic, persistent) and its carbon-paper slips (paper, degrading on a years-to-decades timescale). The Indus archaeological record is the expected output of a credit system whose perishable substrate has decomposed.

27.5.5. Falsifiable Predictions and Tamil Nadu Continuity

The credit-instrument hypothesis generates five falsifiable predictions that extend the Decision Framework of Section 27.4. First, cross-site replication: the uniqueness and positional-entropy profiles established for the 179 Mohenjo-Daro unicorn seals should reproduce on the full corpus of approximately 5,500 inscriptions across all major Indus sites. Second, motif partition: different animal motifs should correspond to different sign vocabularies, since they function as distinct issuer logos. Third, multi-party authorisation: clay sealings recovered at port sites (Lothal, Sutkagan Dor) should show multiple impressions from distinct seals on the same lump of clay, consistent with transaction authorisation by multiple parties — the Bronze Age analogue of countersigning. Fourth, reissuance markers: nominally duplicate seals (a small fraction of the corpus) should differ in manufacturing details consistent with reissuance of a lost or expired instrument by the same issuer to the same holder. Fifth, the algebraic substitution test of Section 27.4.3 applied to the full corpus should produce incoherent results for any putative linguistic reading across multiple inscriptions.

A converging line of evidence is supplied by the 2025 study commissioned by the Government of Tamil Nadu (Rajan & Sivanantham), which found that approximately 60% of graffiti marks on ancient pottery across 140 Tamil Nadu sites share morphological parallels with Indus signs.

The credit-instrument hypothesis interprets this continuity not as linguistic descent but as administrative descent: the same marking tradition persisting for more than a thousand years after the Indus cities fell, used by the same kinds of institutional actors for the same kinds of functions on smaller scale and different substrate. A guild marking tradition does not require a continuous urban civilisation to survive; it requires only continuous craft transmission within mercantile families, which the archaeological record of South India amply supports. The hypothesis thus supplies a structural account of why the Indus sign tradition survives the Indus political collapse: the institutional functions did not depend on the cities.

The Decision Framework of Section 27.4 and the credit-instrument hypothesis of the present section together constitute the dissertation's applied contribution to the Indus script debate: a formal account of *what kind of system the Indus signs are* (structured registration codes, Section 27.3), *what would prove or refute the linguistic interpretation* (six testable archaeological predictions, Section 27.4), and *what institutional function the system in fact served* (institutional credit instruments, the present section). The three results form a single research programme; each is falsifiable independently; each is testable against the open-access corpus and the ongoing archaeological record.

Conclusion

The five publications synthesised in this chapter demonstrate the range of the persistence framework's formal methods — from civilisational design through science policy to computational humanities and the economic history of institutional credit.

Decentralised Urbanism derives critical density thresholds and autonomy coefficients that determine whether a settlement pattern satisfies the persistence conditions of Chapter 5. *Philosophy of Science as Policy Architecture* extends the Navigator Theorem from individual agents to policy systems, proposing that the philosopher's role is not only to diagnose conceptual confusions but to design structures resilient to them. *Positional Constraints in Indus Seal Inscriptions* applies compression analysis to an archaeological puzzle, revealing a structured information architecture where previous approaches saw either undeciphered writing or meaningless decoration. *What Would Prove It?* applies definition-dependent provability to demonstrate that the century-long impasse is a formal consequence of definitional multiplicity, resolvable through a decision framework that specifies what evidence would count. *The Credit Cards of the Bronze Age* identifies the economic function

of the registration codes, showing that the Indus seal satisfies every definitional condition of a credit instrument and that the structural correspondence with the pre-chip embossed credit card is exact at the mechanical level.

The methodological unity across these five apparently disparate applications is the dissertation's central claim in action: the formal tools of persistence analysis, compression theory, and navigational epistemology are not domain-specific but structural — they apply wherever complex systems maintain identity through change, compress information for action, and navigate among incompatible descriptions. A city is a persistent structure. A policy is a compressed description. An ancient sign system is a compression scheme whose decompression key has been lost. A credit instrument is a structurally identified, institutionally issued identifier that enables deferred clearing of obligations without immediate value transfer. In each case, the formal apparatus illuminates features invisible to domain-specific methods, and the illumination follows from the same structural principles.

CHAPTER 28. ONTOLOGY, BIOLOGICAL COMPLEXITY, AND THE LIMITS OF RECURSIVE SELF-IMPROVEMENT

Introduction

The final chapter of Part IX — and of the dissertation's substantive development — addresses the most fundamental questions the persistence framework can pose: why is there something rather than nothing, how should the complexity of what exists be measured, what are the structural limits of formal inquiry into the divine, and whether a recursively self-improving artificial agent can preserve its original goals. These questions span ontology, biology, theology, and AI safety; their unity lies in the formal apparatus that connects them — the persistence conditions, compression analysis, and incompatibility framework developed throughout this dissertation.

Four publications are synthesised. *Testing Nothingness and Evaluating the Inevitability of Life: A Deductive Chain from Being to Persistence, Intelligence, and the Limits of Inquiry* (Kriger, 2026) constructs a deductive chain from the incoherence of absolute nothingness through the structural conditions of persistence to the inevitability of life-like organisation, marking precisely where the chain transitions from logical necessity to conditional derivation. *The GEKS Index: A Composite Measure of Biological Complexity Across Informational, Structural, Functional, and Evolutionary Dimensions* (Kriger & Espesset, 2026) proposes a four-dimensional composite measure integrating Shannon entropy, Gell-Mann effective complexity, a novel morpho-evolutionary innovation component, and a functional complexity scale grounded in system–environment organisation. *Leibniz's Two Kingdoms and the Structural Limits of Pre-Established Harmony: A Dual-System Reading of the Early Modern Problem of God* (Kriger, 2026) demonstrates that Leibniz's two-kingdom distinction constitutes a proto-dual-system ontology whose promise is defeated by pre-established harmony — a result that illuminates the structural limits of formal inquiry into evaluative asymmetry. *Does Goal Preservation Across Capability Gain Have a Formal Upper Bound? On Provable Limits of Objective-Function Fidelity Under Recursive Self-Improvement via Compression-Basis Shift* (Kriger, 2026) proves conditional upper bounds on the fidelity with which a recursively self-improving agent can preserve its original objective function, establishing a phase transition between a preservable regime and a drift regime.

28.1. Testing Nothingness and the Inevitability of Life

28.1.1. From Nothingness to Being

Testing Nothingness and Evaluating the Inevitability of Life poses the most fundamental ontological question available to formal inquiry: is absolute nothingness coherent? The paper approaches this not as a metaphysical speculation but as a mathematical investigation, testing the concept of nothingness for internal consistency through multiple independent methods.

The argument proceeds through a deductive chain. First, absolute nothingness — the complete absence of all entities, properties, relations, and potentialities — is shown to be formally incoherent. Even in pure mathematics, the most abstract and unconstrained domain of human thought, nothingness cannot be maintained: the empty set is a mathematical object, not nothing, and its construction within set theory requires the Axiom of Empty Set — an axiom that posits existence. The concept of nothingness either reduces to a minimal object (which is not nothing) or requires the abandonment of all formal reasoning (which precludes any conclusion, including the conclusion that nothing exists).

Second, existence without distinguishable properties — featureless being — is shown to be operationally indistinguishable from nothingness. This bridging argument (Proposition 4.1) is the chain's most contested link, and the paper explicitly marks the distinction between two claims: that featureless being cannot be *detected* (an epistemic claim the paper establishes) and that featureless being cannot *obtain* (an ontological claim the paper does not fully close). The transition from logical necessity to conditional derivation is explicitly marked here, following reviewer recommendations.

Third, if existence requires distinguishability, then it requires *structure* — differences, relations, boundaries. From structure, the chain derives inhomogeneity (not all regions of the existent can be identical); from inhomogeneity, gradients (differences in intensive quantities); from gradients, flows (transport of energy and matter); from flows under appropriate physical conditions, self-sustaining dissipative structures; and from self-sustaining structures, the conditions for persistence — the central concept of this dissertation.

28.1.2. The Shift from Necessity to Conditionality

The paper is forthright about where the deductive chain transitions from logical necessity to conditional derivation. The steps from existence to structure to inhomogeneity to gradients are logically necessary. The steps from gradients to flows to self-sustaining structures require

specific physical conditions — transport coefficients, nonlinearity, catalytic probability distributions — that are empirically observed but not derivable from the prior logical steps alone. The paper explicitly marks this transition and argues that it is expected and non-damaging: the *Systemic Principle* establishes that if we observe persistent systems (which we do), then the conditions for their persistence are necessarily present — not because the conditions are logically necessary but because the observation selects for them.

28.1.3. The Limits of Inquiry

The chain extends beyond life to intelligence, the question of a creator, and consciousness — reaching, at each extension, a boundary beyond which the formal apparatus cannot proceed. The question of whether a creator exists is shown to be formally undecidable from within the system: the Inaccessibility Theorem (Chapter 20) guarantees that an internal observer cannot access all properties of the system that contains it, and the property "was designed" may belong to the inaccessible set. The question of consciousness is shown to be operationally replaceable: the Navigator Theorem (Chapter 19) derives the functional structure attributed to consciousness without requiring consciousness as an explanatory primitive. In both cases, the paper reports what the formalism yields — and stops where the formalism stops.

28.2. The GEKS Index

28.2.1. The Measurement Problem

If the persistence framework is correct — if persistent systems carry invariant algebraic structure, if persistence selects for specific mathematical features, if the complexity of a system is related to its descriptive diversity — then a quantitative measure of biological complexity should be possible. Yet no widely accepted measure exists. Shannon entropy quantifies informational diversity but cannot distinguish organised structure from random noise. Gell-Mann effective complexity captures meaningful structure but remains notoriously difficult to compute for real biological systems. Phylogenetic measures capture evolutionary history but miss functional organisation. No single dimension suffices.

28.2.2. Four Components

The GEKS Index: A Composite Measure of Biological Complexity Across Informational, Structural, Functional, and Evolutionary Dimensions (Kriger & Espesset, 2026) proposes a composite measure integrating four complementary dimensions, each normalised to [0, 1] and combined via a convex weighting scheme. The acronym GEKS is formed alphabetically from

the contributing frameworks: **G**ell-Mann (effective complexity), **E**spesset (morpho-evolutionary innovation), **K**riger (functional system–environment organisation), and **S**hannon (informational diversity).

The Shannon component S measures the informational diversity of the system's state space. The Gell-Mann component G captures the effective complexity — the length of the shortest description of the system's regularities, excluding incompressible components. The Espesset component E, original to this work, quantifies morpho-evolutionary innovation: the accumulation of derived character states across a phylogenetic matrix, measuring the degree to which a lineage has diverged from ancestral forms through the acquisition of novel structural features. The Kriger component K provides a functional complexity scale based on system–environment organisation, from simple stimulus–response (Level 1) through organ-system integration (Level 3) to cognitive modelling of the environment (Level 4).

28.2.3. Properties and Applications

The index exhibits several desirable properties: modularity (graceful degradation when components are unavailable), multidimensionality (organisms with identical composite scores may have qualitatively different complexity profiles — as demonstrated by a crow-bat comparison), scale invariance of the evolutionary component (the Espesset component is shown to be a scale invariant across individual-collective transitions), and falsifiability. Illustrative calculations are provided for organisms spanning *E. coli* to humans. The extension to non-biological cognitive systems — where informational capacity (S, G) may be high while functional autonomy (K) remains low — demonstrates that the index can distinguish storage from agency.

The connection to the dissertation's persistence framework is direct: the GEKS Index quantifies the structural features that the Persistence Filter Thesis (Chapter 20) identifies as selected by temporal persistence. High GEKS scores correspond to systems with rich invariant algebras, diverse compression schemes, and multiple overlapping functional domains — precisely the systems for which the Theorem of Incompatible Truths predicts irreducible descriptive plurality.

28.3. Leibniz's Two Kingdoms

28.3.1. The Proto-Dual-System Ontology

Leibniz's Two Kingdoms and the Structural Limits of Pre-Established Harmony engages with the history of philosophy not as an antiquarian exercise but as a case study in the structural limits of formal inquiry. The paper focuses on Leibniz as the early modern thinker who came closest to formalising the structural relationship between two evaluatively asymmetric regimes — and who most clearly illustrates the consequences of refusing to accept their irreducible tension.

Leibniz's distinction between the Kingdom of Nature (*Regnum Naturae*), governed by efficient causation, and the Kingdom of Grace (*Regnum Gratiae*), governed by divine justice, constitutes a genuine proto-dual-system ontology. The structural features of each kingdom are distinct: the Kingdom of Nature evaluates states by present configuration, operates through mechanical laws, and distributes finite rivalrous resources; the Kingdom of Grace evaluates rational spirits by trajectory rather than momentary state, is oriented toward justice, and operates with non-rivalrous value.

28.3.2. The Collapse of Asymmetry

The paper's central result is that Leibniz's commitment to pre-established harmony — the doctrine that these two kingdoms never genuinely conflict because God has arranged them in advance — collapses the very asymmetry his distinction identifies. The argument proceeds by examining the internal logic of the Gospel passages that Leibniz's system is designed to accommodate. Gospel maxims such as "seek first the kingdom of God, and all these things shall be added unto you" (Matthew 6:33) function as *transition operators* between the World-System and the Kingdom-System: they instruct agents to prioritise Kingdom-alignment over World-optimisation, with the structural claim that Kingdom-alignment renders World-necessities tractable as secondary effects.

Under pre-established harmony, this priority ordering becomes irrelevant: both orderings produce identical outcomes because the systems are pre-coordinated. The maxim loses its operative force. More generally, for any Gospel maxim that functions as a transition operator between evaluatively asymmetric systems, pre-established harmony renders it either vacuous (the transition is unnecessary because the systems already agree) or tautological (the transition is guaranteed because God has arranged it). The content of the maxim depends on system divergence; harmony eliminates divergence; therefore harmony eliminates the maxim's content.

28.3.3. Connection to the Incompatibility Framework

The connection to the Theorem of Incompatible Truths (Chapter 19) is precise. Leibniz correctly identified the need for two systems — two descriptions of the same reality that evaluate it by different criteria. His error, from the perspective of the incompatibility framework, was to insist that these two descriptions must be compatible — that pre-established harmony guarantees their convergence. The Theorem of Incompatible Truths establishes that above a complexity threshold, descriptions of overlapping domains with different compression schemes must contradict each other. Leibniz's two kingdoms are precisely such descriptions: they overlap on the same agents and situations, they compress reality through different evaluative primitives (efficient causation versus divine justice), and their structural commitments are orthogonal. The demand for harmony is the demand for a colimit that the category of descriptions does not admit.

The result illuminates a broader limitation of the early modern approach to the problem of God: the assumption that a perfect creator would not create systems that genuinely conflict. The persistence framework suggests the opposite — that conflict between descriptions is a structural feature of any sufficiently complex system, and that a creator designing such a system would necessarily produce irreducible asymmetry. The theological implications are not pursued formally, but the structural parallel is noted.

28.4. Goal Preservation Across Capability Gain

28.4.1. The AI Safety Problem

Does Goal Preservation Across Capability Gain Have a Formal Upper Bound? addresses the most consequential open problem in AI safety: whether a recursively self-improving agent can preserve its original objective function with perfect fidelity. The question is urgent because if the answer is no — if goal drift under self-improvement is structurally inevitable — then alignment strategies that assume goal stability are fundamentally limited.

28.4.2. Goals as Vectors, Capability as Basis Shift

The paper formalises the problem within the framework of representation theory and lossy compression. An agent's goals are modelled as vectors in a representational space whose basis shifts with each capability-enhancing iteration. Capability gain — acquiring new concepts, developing finer discriminations, reorganising internal representations — corresponds to a

transformation of the basis in which goals are expressed. The question becomes: when the basis changes, how faithfully can the goal vector be re-expressed in the new basis?

28.4.3. The Fidelity Bound

Under explicitly stated sufficient conditions — that capability gains correspond to lossy reorganisations of the representational basis, grounded empirically in catastrophic forgetting, representation drift during fine-tuning, and dimensional reorganisation during architecture expansion — three independent proof methods (geometric, information-theoretic, and dynamical-systems) converge on the same conclusion: each lossy re-encoding degrades goal fidelity by a quantifiable amount, and these degradations compound multiplicatively across iterations.

The fidelity after n iterations satisfies $F_n \leq \prod_k \cos(\theta_k)$, where θ_k is the angular displacement of the goal vector at iteration k. The bound is tight: constructive examples demonstrate that it is achievable. A *phase transition* separates the preservable regime (where cumulative drift remains below a critical threshold and correction mechanisms can maintain alignment) from the drift regime (where cumulative drift exceeds the threshold and the agent's operative goals have diverged irreversibly from the original specification).

28.4.4. The Complementary Regime and Mitigation Strategies

Crucially, the paper also characterises the conditions under which the bound does *not* apply: when capability gain is implemented through purely additive expansion without reorganisation — adding new dimensions to the representational space while preserving existing ones — perfect goal preservation is theoretically achievable. This complementary result is not merely a formal nicety; it identifies a specific design principle for safe capability expansion.

Two mitigation strategies are analysed: *periodic re-anchoring* (resetting the goal vector to the original specification at regular intervals, with derivation of the optimal re-anchoring frequency as a function of the per-iteration drift rate) and *continuous regularisation* (analogous to elastic weight consolidation, penalising drift from the original goal during each capability-enhancing step, with derivation of the steady-state fidelity achievable as a function of the regularisation strength).

28.4.5. Connection to the Dissertation Programme

The connection to the persistence framework is deep and multifaceted. The goal-fidelity bound is an instance of the Compression Imperative operating on the agent's own objective function: as the agent compresses its representation of the world into more efficient forms (capability gain), it necessarily makes structural commitments that alter the representation of its goals. The phase transition between preservable and drift regimes parallels the complexity threshold of the Incompatibility Theorem: below the threshold, compatibility (goal preservation) is achievable; above it, incompatibility (goal drift) is structurally forced.

The biological extension developed in the paper connects to the Structural Distortion Principle (Chapter 22): human goal specifications are themselves subject to structural distortion through cultural, moral, and volitional filters. When a biologically distorted goal specification is encoded into an AI system, the distortion is inherited and may compound under self-improvement. The population-level mitigations — diverse agent ensembles with independent drift trajectories — parallel the community diversity resilience measure of Chapter 25: a population of agents with diverse goal representations is more robust to alignment failure than a monoculture, for the same structural reason that diverse communities are more resilient than monolithic ones.

Conclusion

The four publications synthesised in this chapter address the outermost boundaries of the dissertation's formal apparatus — the point at which the persistence framework encounters the limits of what can be formally established.

Testing Nothingness pushes the deductive chain as far as it will go — from the incoherence of absolute nothingness through the structural conditions of persistence to the inevitability of life-like organisation — and marks precisely where the chain transitions from logical necessity to conditional derivation, demonstrating that this transition is itself expected within the framework. *The GEKS Index* provides a quantitative measure of the complexity that the persistence framework identifies as structurally significant, integrating four dimensions that no single measure can capture. *Leibniz's Two Kingdoms* reveals the structural limits of any attempt to harmonise evaluatively asymmetric descriptions — a limit that the Theorem of Incompatible Truths generalises to all sufficiently complex systems. And *Goal Preservation Across Capability Gain* establishes that even the goals of the most powerful artificial agents

are subject to the same compression-driven drift that affects all descriptions in the persistence framework.

The arc of Part IX — from civilisational design (Chapter 27) through institutional pathology (Chapter 26) to ontological foundations and AI safety (Chapter 28) — demonstrates that the persistence framework is not a theory of a particular domain but a formal apparatus applicable wherever complex systems maintain identity through change. The framework does not promise solutions; it identifies constraints. It does not resolve the problem of nothingness, measure complexity perfectly, reconcile the kingdoms, or guarantee AI alignment. What it does is establish the structural conditions under which these problems arise, the formal boundaries within which solutions must be sought, and the navigational strategies available to agents who must act within those boundaries. This — the disciplined identification of what can and cannot be formally established — is the contribution the dissertation offers to each of the domains it addresses.

PART X. COSMOLOGICAL VERIFICATION: THE PERSISTENCE FRAMEWORK AS A DISCOVERY ENGINE AT THE LARGEST SCALES

CHAPTER 29. BREAKING THE FRAME: HOW TREE-TOP, RAFASSABRA, AND METHODOLOGICAL LAWS GENERATED THE COSMOLOGICAL PROGRAMME

Introduction

The cosmological results presented in this Part were not obtained by applying standard astrophysical methods to standard astrophysical problems. They were obtained by applying the formal apparatus of the persistence framework — developed across Parts I–IX for cognitive, social, and institutional systems — to questions that the standard cosmological community had either not asked or had declared unanswerable. The present chapter documents how this transfer occurred: which methods broke which frames, which laws identified which errors, and why the resulting programme could not have emerged from within the standard paradigm.

Three methodological instruments were decisive. The *Tree-Top Meta-Method*, developed in *The Tree-Top Meta-Method: Inquiry to the Limits of Domain-Specific Formal Generalisation with Consequent Application Across Other Domains* (Kriger, 2026; doi:10.5281/zenodo.18665621), provided the transfer mechanism: having reached the ceiling of domain-specific generalisation in one field (complex systems theory), the method prescribes systematic application of the same formal apparatus across other domains. The *RAFASSABRA method*, developed in *Breaking Trivial Cognitive Frames: A Formal RAFASSABRA Method for Expanding the Space of Inquiry Through Interrogative Priority and Controlled Divergence* (Kriger, 2026; doi:10.5281/zenodo.18841736), provided the frame-breaking instrument: each of its nine stages targeted a specific dimension of the standard cosmological frame and generated a question that the frame could not formulate. The principle that *Chaos is Relative*, established in *Chaos Is Relative: A Formal Principle of Framework Dependence in Complex Systems* (Kriger, 2026; doi:10.5281/zenodo.18121777), provided the meta-theoretical licence: what appears chaotic or anomalous within one descriptive framework may be orderly within another.

Two additional diagnostic laws — the *Law of Absurdity as Mercy* from *The Law of Absurdity as Mercy: Semantic Destabilization as an Adaptive Necessity in Overloaded Interpretive Systems* (Kriger, 2025; doi:10.5281/zenodo.18099081) and the principle of *Undecidability as*

Methodological Signal from *Undecidability as a Methodological Signal: System-Relative Provability and Explanatory Adequacy* (Kriger, 2023; doi:10.5281/zenodo.18133552) — identified the entry points at which the standard framework was most vulnerable.

29.1. Tree-Top: The Transfer Mechanism

The Tree-Top Meta-Method prescribes a specific intellectual path: within any domain, push formal generalisation to its ceiling — the point at which further generalisation within the domain produces only trivial or vacuous results — then transfer the formal apparatus to a domain where the same structural constraints operate on different substrates. The method does not license loose analogy; it licenses formal transfer when the receiving domain satisfies the same axioms as the source domain.

The transfer path that generated the cosmological programme was concrete. Viability sets (Chapter 5) were developed for biological and cognitive systems: a system persists if its state trajectory remains within a viable region of state space. The ceiling of domain-specific generalisation was reached when the viability framework was applied to social institutions (Chapter 25) and information architectures (Chapter 26). At that ceiling, the formal apparatus — viability sets, admissible transformations, persistence conditions, phase transitions — was domain-saturated: it could not generate further non-trivial results within social systems.

Tree-Top then prescribed: transfer. The question became: what physical system satisfies the same axioms — possesses invariant observables, undergoes transformations, maintains structural identity through change — at the largest accessible scale? The answer was immediate: the quantum vacuum. The vacuum possesses invariant structure (the chiral condensate), undergoes transformations (fluctuations, phase transitions), and maintains identity (the vacuum state is the same everywhere in unperturbed spacetime). Viability set theory applied to the vacuum produced the first result: the vacuum, as a persistent system, must carry an invariant algebra — and that algebra constrains its gravitational contribution.

This transfer was not metaphorical. The axioms of Chapter 5 — applied without modification — generated the ansatz $\Lambda(\rho_m) = \Lambda_0 - \alpha\rho_m$ that underlies the entire cosmological programme. The substrate changed (from cells to spacetime); the formalism did not.

29.2. RAFASSABRA: Nine Stages, Nine Questions

The RAFASSABRA method was not applied retrospectively as a narrative device. It was the operational instrument through which the cosmological questions were formulated. Each stage

targeted a specific presupposition of the ΛCDM standard model and generated a question that the standard frame could not ask.

Stage R — Random Words. External perturbation injected the concept "vacuum gravitates" into a framework where the vacuum's gravitational contribution was treated as either zero (old cosmological constant problem) or catastrophically large (10^{120} problem). The standard frame offered no intermediate position. The random perturbation — asking what happens if the vacuum gravitates *locally*, proportionally to the matter it surrounds — was outside the statistical attractor basin of standard thinking.

Stage A — Structural Attack. "Why do we assume dark matter is a particle?" This question attacks the deepest presupposition of the standard programme: that the gravitational deficit observed in galaxies requires a new fundamental constituent. The structural attack revealed that this assumption is not derived from observation but inherited from the particle-physics culture that dominated cosmology from the 1980s onward.

Stage F — Atemporal Pass. "What does cosmology look like without the arrow of time?" Removing temporal language forced the block-universe perspective: the cosmic web is not something that *formed* but something that *is* — a static configuration whose metric structure is determined by the vacuum-matter coupling. This stage generated the block-universe architecture developed in *The Architecture of the Block Universe: Metric Relaxation, Vacuum Anchoring, and the Structural Necessity of Cosmic Megastructure* (Kriger, 2026; doi:10.5281/zenodo.19211673).

Stage A — Presuppositional Ascent (Alien). "How would an observer without any training in physics describe the large-scale structure of the universe?" Such an observer would see voids surrounded by walls — and would naturally ask why the voids are expanding faster, since they contain nothing to slow them down. This is precisely the αLGQV ansatz: Λ is unsuppressed in voids, suppressed in walls. The "alien" question bypassed decades of theoretical baggage.

Stage S — Apophatic Negation (Stripped). "What remains if we strip away every postulate of ΛCDM except the observations?" What remains is: (1) the universe is expanding, (2) the expansion is accelerating, (3) galaxies rotate faster than visible matter can explain, (4) structure formed rapidly. These four observations are compatible with a single parameter α =

0.005 derived from nuclear physics. Six free parameters of ΛCDM (Ω_m, Ω_Λ, Ω_b, H_0, σ_8, n_s) reduced to one.

Stage S — Isomorphic Transfer (Stole). Transfer of the chiral phase transition from nuclear physics into gravitational collapse. In nuclear physics, at 3–5 times nuclear density, the chiral condensate is restored and quarks deconfine. In gravitational collapse, the same physics produces the K-Limit: the condensate energy, released with $w = -1$, halts the collapse. This isomorphic transfer — the same QCD, the same condensate, different context — generated the fourth stability barrier, developed in the K-Limit paper series and in *The Local Gravitation of Quantum Vacuum: A Unified Solution to the Dark Sector* (Kriger, 2026; doi:10.5281/zenodo.19027460).

Stage A — Cataphatic Assertion (Assembled). Reassembly of all preceding fragments into the unified action $S = \int d^4x \sqrt{-g} [R^2/(6M^2)(1+2\alpha) - 2\Lambda_0]$. Three terms, one parameter, zero free parameters. The action was not postulated; it was *assembled* from the components generated by the preceding stages.

Stage B — Paradoxical Stress-Testing (Broke). The assembled model was subjected to 175 galaxy rotation curves (SPARC database), CMB compatibility tests, S_8 tension analysis, N-body simulations, and comparison with JWST early-galaxy observations. Each test was a genuine attempt to break the model. The model survived — not because it was designed to survive (it has one parameter, fixed by QCD), but because the fixed-point structure is robust.

Stage RA — Temporal Reconstruction (Ran Away). Reintroduction of time into the block-universe picture: the full dynamical history from the Kriger Limit (initial state at nuclear density, $R \approx 15$ AU) through inflation (elastic vacuum), radiation domination, structure formation, to the present accelerated expansion. The temporal narrative is recovered as a *projection* of the atemporal block structure onto the observer's worldline.

29.3. Diagnostic Laws: Absurdity and Undecidability

Two diagnostic laws from Part I identified the precise entry points for the cosmological programme.

The *Law of Absurdity as Mercy* interprets absurd results not as failures but as diagnostic signals that the interpretive frame is wrong. The 10^{120} discrepancy between the predicted and observed cosmological constant — routinely called "the worst prediction in the history of physics" — is, through this law, not a prediction failure but a *frame failure*: the absurdity

signals that the question "what is the total vacuum energy?" is the wrong question. The right question — "what portion of the vacuum energy gravitates?" — dissolves the absurdity. The ratio drops from 10^{120} to 1.65.

The principle of *Undecidability as Methodological Signal* identifies questions whose undecidability within a framework signals that the framework lacks the resources to answer them. The dark matter question — "are dark matter particles real?" — is undecidable within observational cosmology alone, because every gravitational observation is equally compatible with new particles and with modified vacuum physics. This undecidability is not a temporary difficulty; it is a structural signal that the question, as posed, cannot be resolved without specifying the formalization — which is precisely what Definition-Dependent Provability (Chapter 21) establishes.

29.4. Chaos is Relative

The principle that chaos is framework-dependent provided meta-theoretical licence for the entire programme. The "chaotic" early universe — with its fine-tuning problems, horizon problem, and initial-condition sensitivity — is chaotic *relative to ΛCDM*. Relative to αLGQV, the same universe is orderly: the initial state is a finite-density S^3 determined by the K-Limit, the expansion is driven by a single mechanism (vacuum elasticity), and the structure converges to a unique fixed point regardless of initial conditions.

The framework-dependence of chaos connects directly to the Theorem of Incompatible Truths (Chapter 19): ΛCDM and αLGQV are incompatible descriptions of the same observations. Within ΛCDM, the early universe requires fine-tuning. Within αLGQV, it does not. Neither description is "wrong" — they are descriptions with different compression schemes, making different structural commitments, and the "chaos" of the early universe is a property of the ΛCDM compression scheme, not of the universe itself.

Conclusion

The cosmological programme documented in this Part was not an application of cosmological expertise to cosmological problems. It was an application of the persistence framework's methodological apparatus — Tree-Top for transfer, RAFASSABRA for frame-breaking, diagnostic laws for identifying entry points, and chaos-relativity for meta-theoretical licence — to a domain whose standard framework had reached an impasse marked by absurd predictions, undecidable questions, and proliferating free parameters. Every subsequent

chapter of Part X will show a specific law of the dissertation generating a specific cosmological result. But none of those results would have been obtained without the methodological instruments documented here: the instruments that broke the frame within which the questions could not be asked.

CHAPTER 30. PERSISTENCE CONDITIONS, INVARIANT STRUCTURE, AND THE COMPUTATION OF THE COSMOLOGICAL CONSTANT

Introduction

Chapter 29 documented how the frame was broken — which methods generated which questions. The present chapter documents the first and most consequential result: the computation of the cosmological constant from nuclear physics, with no free parameters. The computation follows a chain of reasoning that begins in Chapter 5 (persistence requires invariants), passes through Chapter 20 (invariants carry algebraic structure; parameters are self-consistency residues), and terminates in a formula whose numerical output differs from observation by a factor of 1.65 — replacing the 120-order-of-magnitude discrepancy that has defined the problem for half a century.

Five laws of the dissertation are operative. The *Persistence Requires Invariants Lemma* and the *Persistence Filter Thesis*, developed in *The Transformational Basis of Persistence: A Formal Theory of Structural Viability* (Kriger, 2024; doi:10.5281/zenodo.18435982) and *Why Mathematics Works: Structural Necessity of Isomorphism Between Formal Systems and Physical Reality* (Kriger, 2026; doi:10.5281/zenodo.18793228), identify the vacuum's invariant structure. The *Irreducible Residue Theorem*, developed in *Can Fundamental Constants Be Derived as Irreducible Residues of Structural Self-Consistency?* (Kriger, 2026; doi:10.5281/zenodo.18796847), identifies α and Λ_0 as fixed points of a self-consistency operator. The *Law of Scale-Specific Principles*, from *No Final Theory: Law of Scale-Specific Principles* (Kriger, 2024; doi:10.5281/zenodo.18099738), identifies the error in applying Planck-scale physics to a QCD-scale problem. And the *Conservation and Convertibility* law, from *Conservation and Convertibility as Structural Necessities of Persistent Systems* (Kriger, 2026; doi:10.5281/zenodo.18665700), governs the energy redistribution between matter and vacuum at the chiral phase transition.

The cosmological content is developed in full across several publications, principally *The Vacuum–Matter Coupling from Finite-Density QCD: Sigma Terms, Chiral Condensate, and the Origin of α* (Kriger, 2026; doi:10.5281/zenodo.19017167), *The Cosmological Constant as a QCD Observable: Derivation, Nonlinear Screening, and Falsification* (Kriger, 2026; doi:10.5281/zenodo.19237505), *Derivation of the Cosmological Constant from Nuclear Physics, the Size of the Universe, and the Thermodynamic Foundation* (Kriger, 2026;

doi:10.5281/zenodo.19240715), and *Thermodynamic Origin of the Vacuum–Matter Coupling from de Sitter Two-Fluid Hydrodynamics* (Kriger, 2026; doi:10.13140/RG.2.2.12326.69449). The present chapter does not recapitulate these derivations; it shows how each step was generated by a specific law of the dissertation.

30.1. Persistence Requires Invariants → The Vacuum Has Structure

The Persistence Requires Invariants Lemma (Chapter 5) states that any system persisting through time must possess quantities or relations that remain invariant under the transformations the system undergoes. The lemma was proved for arbitrary dynamical systems and applied in the dissertation to biological cells, cognitive agents, social institutions, and pedagogical genres.

Applied to the quantum vacuum via the Tree-Top transfer (Chapter 29), the lemma yields an immediate and non-trivial consequence. The vacuum is a persistent system in the precise technical sense: it exists, maintains a stable configuration, and endures through all physical processes. Therefore it must possess invariants. What are they?

The QCD vacuum possesses a specific invariant: the chiral condensate $\langle\bar{q}q\rangle$ — a scalar field of quark-antiquark pairs that fills spacetime and gives the vacuum its characteristic energy density, approximately $(250\ \text{MeV})^4$. The condensate is Lorentz-invariant (it looks the same in every reference frame), which forces its stress-energy tensor to take the form $T_{\mu\nu} = \rho\, g_{\mu\nu}$ — the unique form compatible with Lorentz symmetry. This immediately yields the equation of state $w = p/\rho = -1$: negative pressure equal in magnitude to energy density.

The chain is: persistence → invariants → Lorentz scalar → $T_{\mu\nu} \propto g_{\mu\nu}$ → $w = -1$. Each step is a theorem, not an assumption. The persistence framework did not *postulate* that the vacuum has $w = -1$; it *derived* this from the requirement that the vacuum is a persistent system with invariant structure. The derivation is independent of — and prior to — the specific physics of QCD; the QCD condensate is the *instantiation*, not the *source*, of the structural requirement.

30.2. Irreducible Residue Theorem → α and Λ_0 as Fixed Points

The Irreducible Residue Theorem (Chapter 20) proved that any self-consistent physical framework induces a self-consistency operator on its parameter space, and that the fundamental constants correspond to fixed points of this operator — values that the framework's own internal consistency conditions force the parameters to take.

Applied to the vacuum-matter system, the theorem predicts that the coupling between vacuum energy and matter density — the parameter α governing $\Lambda(\rho_m) = \Lambda_0 - \alpha\rho_m$ — cannot be arbitrary. It must be a fixed point of the self-consistency operator F that maps parameter values to the values required for the system to be self-consistent.

The computation is concrete. The chiral condensate shifts in the presence of baryonic matter: the nucleon's sigma terms ($\sigma_{\pi N} \approx 45$ MeV for the pion-nucleon term, $\sigma_s \approx 40$ MeV for the strange contribution) measure the energy of the condensate shift per nucleon. The coupling parameter is:

$$\alpha = f_b \times (\sigma_{\pi N} + \sigma_s) / (3 \times m_N) = 0.005$$

where f_b is the baryon fraction of total matter. Every quantity in this formula is measured in nuclear physics experiments — pion-nucleon scattering, lattice QCD calculations, baryon fraction from Big Bang nucleosynthesis. There are zero free parameters. The value $\alpha = 0.005$ is not fitted to cosmological data; it is computed from nuclear physics and then *tested* against cosmological data.

The cosmological constant follows:

$$\Lambda_0 = \alpha \cdot f_b \cdot \sigma^6 / m^4_{Pl} = 1.8 \times 10^{-52}\ m^{-2}$$

The observed value is $1.1 \times 10^{-52}\ m^{-2}$. The ratio is 1.65. The Irreducible Residue Theorem predicted the *structure* of this result — that Λ_0 would be a computable fixed point, not a free parameter — before the computation was performed. The specific numerical value is a consequence of the QCD instantiation, but the *fact* that a computable value exists is a consequence of the dissertation's formal apparatus.

30.3. Law of Scale-Specific Principles → The Planck-Scale Error

The Law of Scale-Specific Principles (Chapter 2) states that when extrapolation between physical scales fails while stable observables persist, one must seek scale-specific principles rather than force unification. The law was developed for the general case — explaining why different scales of physical reality require different descriptions — and applied in the dissertation to the incompatibility of thermodynamic and microscopic descriptions (Chapter 19) and to the failure of reductionism in social systems (Chapter 25).

Applied to the cosmological constant problem, the law identifies a specific error that has persisted for fifty years. The standard computation of vacuum energy integrates quantum field theory contributions up to the Planck scale (10^{19} GeV), yielding a result 10^{120} times larger than

observed. The Law of Scale-Specific Principles diagnoses this as a *scale violation*: the Planck scale is the wrong scale for this problem. The vacuum's gravitational contribution is determined not by the total energy of all quantum fluctuations (a Planck-scale quantity) but by the *shift* of the chiral condensate in the presence of matter (a QCD-scale quantity, ~250 MeV).

The diagnosis is precise: Zel'dovich's identification of total vacuum energy with Λ violated the Law of Scale-Specific Principles by extrapolating a QFT calculation across 15 orders of magnitude in energy scale without checking whether the observable (gravitational contribution) is scale-specific. It is. Only the portion of vacuum energy that *couples to matter* — the condensate shift measured by the sigma terms — gravitates. The rest is a constant background that contributes no gradient and therefore no force.

The gravitational activation threshold, developed in *Gravitational Activation Threshold and the Irrelevance of Planck Scale: Vacuum Condensate Dynamics as the Origin of Gravitational Coupling* (Kriger, 2026; doi:10.13140/RG.2.2.29929.66409), formalises this: the vacuum "turns on" as a gravitational source at the QCD scale, not the Planck scale. The Planck scale is irrelevant — not because Planck-scale physics is wrong, but because it operates on the wrong scale for this observable, exactly as the Law of Scale-Specific Principles predicts.

30.4. Conservation and Convertibility → Matter-Vacuum Redistribution

The law of *Conservation and Convertibility as Structural Necessities of Persistent Systems* (Chapter 5) established that any persistent system must possess conserved quantities and that these quantities must be convertible between forms — otherwise the system cannot adapt to changing conditions while maintaining its invariants.

In the cosmological context, this law governs the redistribution of energy between matter and vacuum at the chiral phase transition. When matter density reaches 3–5 times nuclear saturation density, the chiral condensate is restored: quarks deconfine, and the energy previously stored in the condensate *within* the nuclear medium is released *back into the vacuum state* with $w = -1$. This is not creation or destruction of energy; it is *conversion* from one form (matter-coupled condensate energy, with $w \neq -1$) to another (free vacuum energy, with $w = -1$).

The Conservation and Convertibility law predicted the *structure* of this process: energy must be conserved, and it must be convertible between matter and vacuum forms. The specific

physics (chiral symmetry restoration, deconfinement) is the QCD instantiation. But the structural requirement — that a persistent system at a phase boundary redistributes its conserved quantities between available channels — is domain-independent and was derived in Chapter 5 for biological systems (energy conversion between metabolic pathways) before being applied to QCD.

The cosmological consequence is the K-Limit mechanism (to be developed in Chapter 32): the converted vacuum energy, with $w = -1$, provides repulsive pressure that halts gravitational collapse. But the underlying structural logic — conservation with convertibility at a phase boundary — is the persistence framework, not ad hoc physics.

30.5. Dual-Source Decomposition → The Zel'dovich Error as Observer Contribution

The *Dual-Source Decomposition* (Chapter 20) established that mathematical physics decomposes into T_form (structural constraints contributed by cognitive architecture, stable across centuries) and T_cont (specific laws and constants determined by the world, subject to revolutionary change). Applied to the cosmological constant problem, this decomposition identifies Zel'dovich's error as an observer contribution, not a failure of physics.

The T_form component of the cosmological constant problem is the *structure* of the question: "what is the gravitational contribution of the vacuum?" This structure is correct and has been stable since Einstein introduced Λ in 1917. The T_cont component is the *content*: which vacuum energy gravitates? Zel'dovich's answer — "all of it" — was a specific content choice that happened to be wrong by 120 orders of magnitude. The αLGQV answer — "only the matter-coupled shift" — is a different content choice that happens to be wrong by a factor of 1.65.

The Dual-Source Decomposition clarifies why the error persisted for so long: the structural question was correct (T_form was right), and the content error was not visible from within the framework that generated it (T_cont was wrong, but the wrong answer was the only answer the standard framework could produce). The framework needed to be broken — which is what RAFASSABRA did (Chapter 29) — before the content error could be identified.

30.6. Persistence Filter Thesis → Why GR + QFT Work

The Persistence Filter Thesis (Chapter 20) established that mathematics is effective because only configurations with the right kind of structural regularity survive the filter of temporal persistence. Applied to cosmology, the thesis explains why general relativity and quantum

field theory — two mathematical frameworks developed for entirely different purposes — both work for describing the universe: because spacetime is a persistent system, and persistent systems necessarily carry the invariant algebraic structure that GR and QFT formalise.

This is not a trivial observation. The "unreasonable effectiveness of mathematics" in cosmology has been a standing puzzle. The Persistence Filter Thesis dissolves it: spacetime persists, therefore it carries invariants, therefore it admits mathematical description. The *specific* mathematical structures (Riemannian geometry, gauge theory, Hilbert space) are selected by persistence — continuity is selected because persistent dynamics maps nearby states to nearby states; gauge structure is selected because persistent interactions must be locally invariant; Hilbert space is selected because persistent quantum evolution must preserve probability.

The cosmological consequence is that the modified Friedmann and Poisson equations of the αLGQV framework — which include the vacuum-matter coupling — inherit the mathematical structure of GR not because αLGQV was designed to look like GR but because any persistent description of spacetime must have this structure. The Persistence Filter is operating on the theory as well as on the universe the theory describes.

Conclusion

The computation of the cosmological constant — from $\Lambda_0 = \alpha \cdot f_b \cdot \sigma^6 / m^4_{Pl} = 1.8 \times 10^{-52}$ m^{-2}, with $\alpha = 0.005$ derived from QCD sigma terms — is the single most direct verification of the persistence framework's predictive power at cosmological scales. Every step in the computation was generated by a specific law of the dissertation: persistence requires invariants (the condensate is the invariant); invariants carry algebraic structure ($w = -1$ follows from Lorentz invariance); parameters are self-consistency residues (α is computed, not fitted); scale-specific principles forbid extrapolation (the Planck scale is irrelevant); conservation with convertibility governs phase transitions (matter-vacuum redistribution); and the Dual-Source Decomposition identifies the observer's contribution to the 120-order error.

No step in this chain was designed for cosmology. Every step was developed for cognitive, social, or institutional systems and transferred to cosmology via Tree-Top. The fact that the chain terminates in a number that differs from observation by a factor of 1.65 — rather than by a factor of 10^{120} — is the persistence framework's strongest empirical credential.

CHAPTER 31. SELF-CONSISTENCY FIXED POINTS, INFODYNAMICS, AND THE SELF-ORGANISING COSMOS

Introduction

Chapter 30 showed how the persistence framework computed a number — the cosmological constant. The present chapter shows how it computed a *structure* — the cosmic web. The cosmic web is the largest coherent structure in the observable universe: a network of voids, walls, filaments, and clusters spanning hundreds of megaparsecs, exhibiting topological universality (its geometry satisfies Plateau's laws of minimal surfaces), supra-horizon coherence (structures larger than the causal horizon at their formation epoch), and rapid assembly (massive structures observed at redshifts $z > 10$ by JWST where ΛCDM predicts none).

Standard cosmology treats these features as contingent outcomes of initial conditions — the cosmic web looks the way it does because the primordial density fluctuations happened to have a specific power spectrum. The persistence framework makes a stronger claim: the cosmic web is a *necessary* structure — a unique fixed point of a cyclical operator that any persistent vacuum-matter system must converge to, regardless of initial conditions. This claim is generated by six laws of the dissertation, and its most consequential expression is the *Second Law of Infodynamics*, which the persistence framework grounds in established physics for the first time.

The cosmological content is developed principally in *Matter-Dependent Vacuum Energy Density and Inhomogeneous Cosmic Expansion* (Kriger, 2026; doi:10.5281/zenodo.18896536), *The Elastic Vacuum: A Sequestering Mechanism for Vacuum Energy and the Complete Dynamical History of Cosmic Expansion* (Kriger, 2026; doi:10.5281/zenodo.19004376), *Structure Growth in the Gravitating Vacuum Model: The Void–Cluster Cell, Modified Perturbation Equations, and the S_8 Tension* (Kriger, 2026; doi:10.5281/zenodo.19013333), *N-body Simulations with a Gravitating Vacuum: Testing the Void–Cluster Model Against Structure Formation Data* (Kriger, 2026; doi:10.5281/zenodo.19014206), and *Evaluating the Second Law of Infodynamics Against Established Physics: A Critical Assessment with Falsifiable Predictions* (Kriger, 2026). The present chapter does not reproduce these derivations; it shows how each was generated by the dissertation's formal apparatus.

31.1. Self-Consistency Fixed Points → The Cosmic Web

The *Self-Consistency Fixed-Point Theorem*, implicit in *The Transformational Basis of Persistence* (Kriger, 2024; doi:10.5281/zenodo.18435982) and explicit in *On the Possibility of Self-Sufficient Systems: Fixed Points and Cyclical Closure* (Kriger, 2024; doi:10.5281/zenodo.18317239), established that any persistent system with cyclical feedback converges to a fixed-point configuration — a state that reproduces itself through the feedback loop.

The vacuum-matter system possesses exactly such a loop. Matter determines the local vacuum energy density through $\Lambda(\rho_m) = \Lambda_0 - \alpha\rho_m$. Vacuum energy density determines the local metric through the modified Einstein equations. The local metric determines the redistribution of matter through geodesic motion. And the redistributed matter determines the new vacuum energy density. The cycle is:

$\rho_m \rightarrow \Lambda(\rho_m) \rightarrow g_\mu\nu \rightarrow \rho_m$

The fixed-point theorem predicts that this cycle converges to a configuration Φ^* satisfying $\Phi_cycle(\Phi^*) = \Phi^*$ — a configuration that reproduces itself through one complete cycle. The cosmic web — with its specific void-wall-filament-cluster topology — is this fixed point. The prediction is not that the cosmic web *exists* (that is observed) but that it is *unique* (any initial configuration converges to the same structure) and *necessary* (any vacuum-matter system with $\alpha > 0$ must produce it).

The cyclical closure framework of *Cyclical Hierarchical Systems: A Modal Analysis of Self-Sufficient Structures* (Kriger, 2024; doi:10.5281/zenodo.18317628) provides the formal machinery. The cosmic web is a self-sufficient cyclical system in the precise technical sense: it maintains itself through its own feedback loop without external input. The void-wall contrast is self-reinforcing: voids expand faster (Λ unsuppressed) → matter is swept into walls (Λ suppressed) → void-wall contrast deepens → voids expand still faster. The cycle is self-sustaining and converges.

31.2. Structural Resilience → Cosmic Web Robustness

Formalization of Structural Resilience in Discrete-State Dynamical Systems (Kriger, 2019; doi:10.5281/zenodo.18351470) established formal conditions under which a dynamical system recovers from perturbations: the system must possess a resilience basin — a region of state space from which all trajectories return to the attractor after perturbation.

Applied to the cosmic web, structural resilience predicts that the void-wall-filament structure is robust to local disruptions. Remove a filament (through a galaxy cluster merger, for instance), and the vacuum-matter feedback loop rebuilds it: the disrupted region experiences a local increase in Λ (matter removed → less suppression), which drives re-expansion and re-establishment of the void-wall contrast. The cosmic web is not merely a fixed point; it is a *stable* fixed point with a resilience basin encompassing all physically realisable perturbations. This prediction is testable through N-body simulations with perturbed initial conditions — the simulations reported in *N-body Simulations with a Gravitating Vacuum* (Kriger, 2026; doi:10.5281/zenodo.19014206) confirm that the fixed-point structure is recovered regardless of the specific perturbation, provided $\alpha > 0$.

31.3. Four Laws of Self-Organising Systems → Cosmic Self-Organisation

A Unified Theory of Self-Organizing Systems: Four Formal Laws on Cooperation, Viability, Interference, and Observability (Kriger, 2017; doi:10.5281/zenodo.18363729) established four laws governing any self-organising system. All four are instantiated at cosmological scales.

The *Law of Cooperation* states that persistent multi-agent systems require cooperative interactions that stabilise the collective configuration. In the cosmic web, cooperation takes the form of gravitational binding within clusters and filaments: galaxies within a cluster cooperate gravitationally to maintain the cluster's identity as a persistent structure. Without cooperative binding, clusters would disperse under the Hubble flow.

The *Law of Viability* states that each component must remain within its viability set — the region of parameter space compatible with continued functioning. Cosmic voids are persistent precisely because they satisfy the viability condition: their expansion rate is bounded by the vacuum energy density, and their boundaries are maintained by the surrounding walls. A void that expanded without limit would cease to be a void (it would become the entire universe); the viability constraint — maintained by the void-wall feedback — prevents this.

The *Law of Interference* states that interactions between components can be constructive (enhancing persistence) or destructive (threatening viability). Galaxy cluster mergers are constructive interference (the merged cluster has a larger viability set). Tidal disruption of dwarf galaxies by massive neighbours is destructive interference. The persistence framework predicts — and observations confirm — that the merger rate and disruption rate are balanced

in the cosmic web's fixed-point configuration: the structure is neither growing without limit nor collapsing.

The *Law of Observability* states that persistent structures may exist at the boundary of their viability set in states that produce no external signatures. In the cosmic web, the analogue is dark filamentary structure — low-density filaments that connect clusters but are too faint to detect with current instruments. The persistence framework predicts their existence; the warm-hot intergalactic medium (WHIM), detected through X-ray absorption studies, confirms it.

31.4. Constraint-Autonomy Compatibility → The Structure of Physical Law

The *Constraint-Autonomy Compatibility Law*, from *The Constraint–Autonomy Compatibility Law: A Formal Derivation* (Kriger, 2026; doi:10.5281/zenodo.18368991), established that constraints and autonomy are co-constitutive rather than opposed: maximising either at the expense of the other causes systemic collapse. In persistent systems, the productive regime lies in the balance.

Applied to cosmology, this law identifies the structural role of two features of the universe that are usually treated as independent: physical laws (constraints) and quantum indeterminacy (autonomy). The constraints (GR, QCD, electroweak theory) determine the *form* of the universe's evolution. The autonomy (quantum fluctuations, vacuum energy fluctuations, stochastic processes) determines the *content* — the specific realisation within the constrained form.

A universe with constraints but no autonomy would be deterministic and structureless — the same configuration everywhere, forever. A universe with autonomy but no constraints would be chaotic and structureless — random configurations with no persistence. The observed universe, with both constraints and autonomy in the balance predicted by the Constraint-Autonomy Compatibility Law, produces the cosmic web: structured (by constraints) but non-trivial (by autonomy), persistent (by balance) but adaptive (by fluctuations).

31.5. Law of Imperative Uncertainty → Primordial Fluctuations

The *Law of Imperative Uncertainty*, from *The Law of Imperative Uncertainty: Why Any Complex World Requires Uncertainty* (Kriger, 2025; doi:10.5281/zenodo.18101601), proved that any system capable of sustaining non-trivial complexity must permit probabilistic deviation from its governing laws. Rigid closure of all degrees of freedom leads to zero entropy rate and systemic collapse.

In cosmology, the law predicts that primordial fluctuations are not accidental features of the initial conditions but *necessary* conditions for the universe's persistence as a complex system. Without fluctuations, the early universe would have remained perfectly homogeneous — and a perfectly homogeneous universe cannot form structures, because there are no density gradients to seed gravitational collapse. The Law of Imperative Uncertainty predicts that any persistent universe must contain fluctuations — and the observed primordial fluctuations (amplitude $\delta\rho/\rho \approx 10^{-5}$, measured by Planck) confirm this prediction.

The specific amplitude is not predicted by the law alone (it depends on the physics of inflation), but the *existence* of a non-zero amplitude is predicted: a universe with $\delta\rho/\rho = 0$ is a universe that violates the Law of Imperative Uncertainty and therefore cannot sustain complexity — cannot persist as anything other than a featureless thermal bath.

The connection to the elastic vacuum model is direct: *Connection of Vacuum Energy with Inflation: The Elastic Potential, Initial Conditions, and the Origin of Primordial Fluctuations* (Kriger, 2026; doi:10.5281/zenodo.19009393) derives the primordial fluctuation spectrum from the elastic potential of the R^2-vacuum model, showing that the fluctuations arise naturally from the vacuum's elastic response to the initial state — not from a separate inflaton field.

31.6. Structural Genesis of Dynamical Architecture → Modified Friedmann Equations

Structural Genesis of Dynamical Architecture: Deriving the Necessary Form of Physical Law from Conditions on Sustainable Complexity (Kriger, 2026; doi:10.5281/zenodo.18646861) established that the *form* of dynamical laws is not arbitrary but constrained by the requirements of sustainable complexity: any system capable of maintaining non-trivial structure must be governed by laws of a specific mathematical form.

Applied to cosmology, this result predicts the *form* of the modified Friedmann and Poisson equations of the αLGQV framework. The standard Friedmann equation $H^2 = (8\pi G/3)\rho$ is the unique form for a homogeneous, isotropic universe with a cosmological constant. The modified Friedmann equation — which replaces Λ with $\Lambda(\rho_m) = \Lambda_0 - \alpha\rho_m$ — is not an arbitrary modification; it is the unique modification compatible with the structural genesis conditions when the vacuum-matter coupling is included. The modification preserves the mathematical structure required by sustainable complexity (energy conservation, diffeomorphism invariance, second-order dynamics) while adding the single new element (the coupling α) required by the vacuum's persistent structure.

The parallel with Chapter 8 is precise. *Emergence of Newtonian Dynamics from Metric Inertial Systems* (Kriger, 2018; doi:10.5281/zenodo.18343870) derived Newton's second law from structural axioms on metric inertial systems. The modified Friedmann equation is derived from the same logic: structural axioms on persistent spacetime systems, with the addition of the vacuum-matter coupling. The *form* of the law is a structural necessity; the *content* (the specific value $\alpha = 0.005$) is an Irreducible Residue.

31.7. The Second Law of Infodynamics and Simulability

The most consequential synthesis of the dissertation's laws at cosmological scales is the proof of the *Second Law of Infodynamics* — originally proposed as an empirical observation by Vopson and Lepadatu (2022) and here derived from established physics for the first time.

Vopson observed that the information entropy of physical systems tends to decrease or remain constant over time — the opposite of the thermodynamic Second Law. The observation held across genetic, atomic, and cosmological systems but lacked a derivation from established physics.

The persistence framework provides the derivation. *Evaluating the Second Law of Infodynamics Against Established Physics* (Kriger, 2026) introduces *effective complexity* (Gell-Mann and Lloyd, 2003) as the central quantity and proves that in any gravitating system with vacuum-matter coupling $\alpha > 0$, the effective complexity of the matter density field increases monotonically toward a unique fixed point. The proof uses the Banach contraction mapping theorem applied to the cyclical operator Φ_cycle (vacuum assignment → metric solution → matter redistribution).

The chain of laws is: Persistence Theorem (the vacuum-matter system must converge to a fixed point) → Self-Consistency Fixed Point (the fixed point is unique) → Cyclical Closure (the operator is contractive) → Structural Resilience (perturbations are absorbed) → Banach theorem (convergence is exponential). The Second Law of Infodynamics is the *informational expression* of this convergence: as the system approaches the fixed point, its effective complexity increases because the fixed point is the configuration with maximum structured regularity — the cosmic web.

The *simulability criterion* $S = 1 - k$, introduced in the same paper, provides a new epistemological measure standing alongside Popper's falsifiability and the Evidence-Based Philosophy conditions (Chapter 21). Simulability measures the degree to which the laws of

physics determine the outcome independently of initial conditions. A theory with $S \approx 0$ (like ΛCDM, where the final state depends on the specific random seed of primordial fluctuations) is falsifiable but not simulable. A theory with $S \approx 1$ (like αLGQV, where the fixed point is unique regardless of initial conditions) is both falsifiable *and* simulable — the laws alone determine the outcome.

The distinction is epistemologically profound. ΛCDM is an adequate compression of the observations — it fits the data with six parameters. But it is not informationally self-sufficient: it requires initial conditions as external input. αLGQV, with one parameter derived from QCD, is informationally self-sufficient: the cosmic web follows from the laws alone. The simulability criterion captures this distinction quantitatively, and its introduction extends the epistemological toolkit of the dissertation beyond falsifiability (Popper) and evidence-based conditions (Chapter 21) to a third criterion that measures the *informational completeness* of a theory.

31.8. Elastic Vacuum → Membrane Stiffness as Persistence Condition

The elastic vacuum model — the $R^2/(6M^2)$ term in the unified action — receives a persistence-theoretic interpretation that is absent from its standard presentation in Starobinsky inflation.

In standard cosmology, the R^2 term is introduced to drive inflation — a period of exponential expansion in the early universe that solves the horizon and flatness problems. The term is added to the Einstein-Hilbert action as an empirically motivated correction with a free parameter (the scalaron mass $M \approx 3 \times 10^{13}$ GeV, fixed by CMB amplitude measurements).

In the persistence framework, the R^2 term is not an empirical correction but a *persistence condition*: it endows spacetime with elastic stiffness — a resistance to deformation analogous to the bending modulus of a physical membrane. Any persistent extended structure must resist deformation; if it did not, arbitrarily small perturbations would destroy it. The R^2 term is the mathematical expression of this resistance, and its presence in the action is *predicted* by the persistence framework rather than postulated for phenomenological reasons.

The connection to the Structural Resilience formalism (Section 31.2) is direct: the R^2 term defines the resilience basin of spacetime itself. Perturbations of the metric (gravitational waves, density fluctuations) are absorbed and damped by the elastic response — the "membrane" springs back. Without the R^2 term, spacetime would have no elastic resistance, and arbitrarily small perturbations would produce arbitrarily large deformations.

Conclusion

The cosmic web — the largest, most complex structure in the observable universe — is not an accident of initial conditions. It is a necessary consequence of six laws of the persistence framework: self-consistency requires a fixed point, cyclical closure makes the operator contractive, structural resilience makes the fixed point stable, the four laws of self-organisation govern the interactions among components, constraint-autonomy compatibility produces the balance between structure and variation, and imperative uncertainty guarantees the fluctuations without which no structure could form.

The Second Law of Infodynamics — the tendency of physical systems toward increasing effective complexity — is the informational expression of this convergence. Its proof, via the Banach contraction mapping theorem applied to the Φ_cycle operator, is the most direct connection between the dissertation's formal apparatus and a quantitative cosmological prediction. And the simulability criterion $S = 1 - k$ is the epistemological consequence: a theory whose outcome is independent of initial conditions is informationally self-sufficient in a way that no initial-condition-dependent theory can match.

The elastic vacuum — spacetime with membrane stiffness — is the persistence condition applied to spacetime itself. Its presence in the unified action is not phenomenological but structural: any persistent extended system must resist deformation. The cosmic web is what happens when this resistant membrane is coupled to matter through the QCD condensate. The result is inevitable.

CHAPTER 32. THE PERSISTENCE THEOREM AND THE HIERARCHY OF STABILITY: THE KRIGER LIMIT

Introduction

Chapters 30 and 31 showed how the persistence framework computed a constant (Λ_0) and predicted a structure (the cosmic web). The present chapter shows how it resolved a singularity — the gravitational singularity predicted by general relativity at the endpoint of stellar collapse. The resolution follows from a single application of the Persistence Theorem: a system at the boundary of its viability set must transition to a new regime rather than cease to exist. Every previous "infinity" in the history of physics — the ultraviolet catastrophe, the electron self-energy divergence, the weak-interaction divergences — was resolved by exactly this logic: the infinity signalled not a physical reality but a persistence failure of the description, cured by deeper physics that restored the system's viability. The gravitational singularity is the next — and, the persistence framework predicts, the last — in this sequence. The principal laws operative in this chapter are the *Persistence Theorem* from *The Transformational Basis of Persistence* (Kriger, 2024; doi:10.5281/zenodo.18435982), the *Phase Transition Analysis* applied across domains in *Phase Transitions in Sociotechnical Systems* (Kriger, 2026; doi:10.5281/zenodo.18717900) and *From Repression to Saturation* (Kriger, 2026; doi:10.5281/zenodo.18685856), the *Law of Limit to Negation* from *The Law of Limit to Negation: Negation Can't Negate Itself* (Kriger, 2025; doi:10.5281/zenodo.18101306), the *Fractal Scaling* principle from *Fractal Scaling of Operational Patterns Across Systems* (Kriger, 2026; doi:10.5281/zenodo.18645527), and the *Formalisation Laws* from *Formalization Laws: Structure, Boundaries, and the Necessity of Plural Representations* (Kriger, 2025; doi:10.5281/zenodo.18099527). The cosmological content is developed in *The Local Gravitation of Quantum Vacuum: A Unified Solution to the Dark Sector* (Kriger, 2026; doi:10.5281/zenodo.19027460), *Catastrophic Phase Transitions as Necessary Conditions for Complexity Dissemination* (Kriger, 2026; doi:10.5281/zenodo.18859088), and *The Architecture of the Block Universe* (Kriger, 2026; doi:10.5281/zenodo.19211673).

32.1. Persistence Theorem → Every Stability Limit Is a Persistence Condition

The Persistence Theorem (Chapter 5) establishes that a system persists if and only if its state trajectory remains within its viability set — the region of state space compatible with the

system's structural integrity. When the trajectory approaches the boundary of the viability set, one of two things must happen: the system transitions to a new regime (with a different viability set), or the system ceases to exist.

Astrophysics has encountered this boundary three times before the persistence framework identified the pattern.

The *Eddington limit* is the boundary at which radiation pressure in a luminous star balances gravitational compression. Below this luminosity, the star persists as a main-sequence object. Above it, the star transitions to a new regime (mass loss, instability, explosion). The Eddington limit is a persistence condition: it defines the viability set of luminous hydrostatic equilibrium.

The *Chandrasekhar limit* (~1.4 M⊙) is the boundary at which electron degeneracy pressure can no longer support a white dwarf against gravity. Below this mass, the white dwarf persists indefinitely. Above it, the system transitions: electrons are forced into protons (electron capture), producing a neutron star. The Chandrasekhar limit is a persistence condition: it defines the viability set of electron-degenerate matter.

The *Oppenheimer-Volkoff limit* (~2.2 M⊙) is the boundary at which neutron degeneracy pressure can no longer support a neutron star. Below this mass, the neutron star persists. Above it, general relativity predicts continued collapse to a singularity — a state of infinite density, infinite curvature, and zero volume.

The persistence framework identifies the singularity as a *persistence failure of the description*, not a physical reality. The pattern is clear: at each previous limit, the "infinity" predicted by the prevailing theory (infinite compression of electrons, infinite compression of neutrons) was resolved by physics at a deeper scale (nuclear physics, QCD). The singularity predicted by GR at the Oppenheimer-Volkoff boundary is the same type of signal — not a feature of reality but a marker that the description has been pushed beyond its domain of validity.

32.2. Law of Limit to Negation → Singularity as Self-Defeating Negation

The *Law of Limit to Negation* (Chapter 4) established that negation cannot negate itself: any process that systematically eliminates structure eventually reaches a boundary at which further elimination would eliminate the capacity to eliminate — and therefore halts.

The gravitational singularity is the attempted negation of all spatial structure: infinite density means zero volume, which means the complete annihilation of spatial extension. The Law of Limit to Negation predicts that this negation is self-defeating: the process of compression,

which eliminates spatial structure, must encounter a boundary at which further compression would eliminate the very field (the QCD vacuum) whose energy drives the compression through its gravitational effects.

The boundary is the chiral phase transition. At 3–5 times nuclear saturation density, the chiral condensate — the QCD vacuum structure within nuclear matter — is restored to its symmetric phase. The quarks deconfine. The energy stored in the condensate returns to the vacuum state with $w = -1$. This energy, with its negative pressure, generates repulsive gravity that opposes further compression.

The Law of Limit to Negation identifies the *structural reason* why the compression halts: the compressing agent (gravity, acting on matter + vacuum) and the compressed object (vacuum structure within matter) are coupled. Compressing the vacuum structure releases vacuum energy that opposes the compression. The negation of structure produces the force that prevents further negation. Negation negates itself — and halts.

This is not a physical argument dressed in philosophical language. It is the same structural logic that operates at every previous limit: electron degeneracy pressure is the "negation of negation" at the Chandrasekhar limit (compressing electrons produces the pressure that resists compression); neutron degeneracy pressure is the same at the OV limit. The K-Limit extends the pattern to its final stage: vacuum pressure is the negation of negation at the density where nuclear matter itself breaks down.

32.3. Phase Transition Analysis → The Chiral Transition as Regime Shift

The phase transition formalism developed for sociotechnical systems in Chapter 25 and for motivational dynamics in Chapter 22 applies without modification to the chiral phase transition in gravitational collapse.

In Chapter 25, *Phase Transitions in Sociotechnical Systems* modelled the transition from labour-centric to post-labour economies as a bifurcation in a dynamical system: when a control parameter (automation fraction) crosses a critical threshold, the system undergoes a qualitative regime shift. The mathematical apparatus — cubic ODE with bifurcation analysis, stability of pre- and post-transition equilibria, critical threshold conditions — is domain-independent.

In gravitational collapse, the control parameter is matter density ρ_m. Below the critical threshold (~3–5 ρ_0, where $\rho_0 \approx 2.3 \times 10^{17}$ kg/m³ is nuclear saturation density), matter is in the hadronic phase: quarks are confined within nucleons, the chiral condensate is intact, and the

condensate energy participates in the material equation of state with $w \neq -1$. Above the threshold, the system undergoes a phase transition: quarks deconfine, the condensate is restored to its symmetric phase, and the released condensate energy returns to the vacuum state with $w = -1$.

The formal structure is identical to the sociotechnical phase transition of Chapter 25: a bifurcation in the equation of state at a critical density, with qualitatively different dynamics on each side. The pre-transition regime (hadronic matter) has $w > 0$ and collapsing dynamics. The post-transition regime (deconfined quarks + vacuum energy) has an effective w that includes a $w = -1$ component, producing repulsive pressure. The K-Limit is the post-transition equilibrium: the density at which the repulsive vacuum pressure exactly balances gravitational compression.

The universality of the phase-transition formalism across motivational dynamics (Regime I → Regime II in the motivational phase diagram), sociotechnical systems (labour-centric → post-labour), and gravitational collapse (hadronic → deconfined) is itself a prediction of the *Fractal Scaling* principle: the same operational pattern — control parameter crosses threshold, system undergoes qualitative regime shift, new equilibrium with different structural properties — recurs at every scale.

32.4. The K-Limit: Mechanism and Uniqueness

The Kriger Limit (K-Limit) is the maximum density at which the chiral condensate can exist within nuclear matter: $\rho_K \approx 5\rho_0 \approx 1.15 \times 10^{18}$ kg/m³. At this density, the repulsive pressure of the restored vacuum balances gravity, and collapse halts. The result is not a singularity but a finite-density core — a region of maximally compressed but non-singular spacetime.

The K-Limit has a property that distinguishes it from all previous stability limits: it *cannot be overcome by additional mass*. The Chandrasekhar limit is overcome above 1.4 M⊙ because electron degeneracy pressure scales as $P \propto \rho^{5/3}$ while gravitational pressure scales as $P_{grav} \propto \rho^{4/3}$ — gravity wins at sufficiently high density. The OV limit is overcome for the same reason with neutron degeneracy pressure. But vacuum pressure with $w = -1$ scales as $P \propto \rho$: it matches gravity at all densities. Doubling the mass doubles the gravitational pressure, but it also doubles the vacuum pressure (more matter → more condensate shift → more vacuum energy release). The balance is mass-independent.

This mass-independence is a consequence of the *Conservation and Convertibility* law (Chapter 30): the energy redistributed from matter to vacuum at the chiral transition scales with the amount of matter undergoing the transition. More mass means more condensate energy released, means more repulsive pressure — in exact proportion. The K-Limit is a *density* threshold, not a *mass* threshold, and it holds for all objects from neutron stars to supermassive black holes.

A transition mass $M_{tr} \approx 4.0\ M\odot$ separates two phenomenological classes. Below M_{tr}, the collapsing object reaches deconfinement density before forming an event horizon — the result is a *vacuum star*, a compact object with no horizon and a finite-density core. Above M_{tr}, an event horizon forms first (because the Schwarzschild radius exceeds the object's physical radius at a density below ρ_K), but the interior still reaches the K-Limit — the result is a *frozen star*, an object with a horizon enclosing a finite-density core. In both cases, the singularity is absent.

32.5. Formalisation Laws → Singularity as Formalisation Artifact

The *Formalisation Laws* (Chapter 3) established that any phenomenon admits multiple irreducibly distinct formal representations, and that each representation has boundaries beyond which it produces artifacts — predictions that are features of the formalisation, not of the phenomenon.

General relativity, formalised as a classical field theory of the metric tensor, predicts a singularity at the endpoint of gravitational collapse. The Formalisation Laws diagnose this as a *boundary artifact*: GR's formalisation does not include the quantum structure of the vacuum (the chiral condensate, the equation of state at supranuclear densities), and the singularity is what the formalisation produces when pushed beyond the domain where its structural commitments are valid.

The diagnosis is not speculative. Every previous singularity in the history of physics was resolved by the same mechanism: the ultraviolet catastrophe (infinite radiation energy at short wavelengths) was an artifact of classical electrodynamics pushed beyond its domain — resolved by quantum mechanics. The electron self-energy divergence (infinite electromagnetic energy of a point charge) was an artifact of classical electrodynamics applied to a quantum object — resolved by renormalisation. The weak-interaction divergences were artifacts of the four-fermion theory pushed beyond its domain — resolved by electroweak

unification. In each case, the formalisation was *correct within its domain* and produced an *artifact beyond it*.

The gravitational singularity follows the same pattern. GR is correct for macroscopic gravitational phenomena. At supranuclear densities, the quantum structure of the vacuum — which GR does not include — becomes dynamically relevant, and the formalisation produces an artifact (the singularity) that the deeper physics (GR + QCD) resolves.

The *Structural Diagnostic Principle* from *Structural Diagnostic Principle for Dynamical Models* (Kriger, 2018; doi:10.5281/zenodo.18344823) provides the formal tool for this diagnosis: the singularity is a diagnostic signal that the model's structural commitments are incompatible with the physical regime. The signal does not indicate a physical catastrophe; it indicates a model boundary. The K-Limit is what lies beyond that boundary.

32.6. Fractal Scaling → The Stability Hierarchy as Self-Similar Pattern

The *Fractal Scaling* principle establishes that operational patterns recur across scales in complex systems, with specific mechanisms varying but the structural pattern preserved.

The stability hierarchy — Eddington → Chandrasekhar → OV → Kriger — exhibits fractal scaling in a precise sense. At each level, the same pattern operates: (i) a pressure mechanism supports the system against gravity; (ii) at a critical density, gravity overwhelms the pressure; (iii) the system transitions to a new phase where a different pressure mechanism operates; (iv) the new mechanism supports the system at higher density. The pattern is self-similar: each level recapitulates the structure of the previous level at a higher density scale.

The fractal scaling is not merely descriptive; it is predictive. The pattern predicts that the K-Limit is the *final* level — because the vacuum pressure mechanism, with its $P \propto \rho$ scaling, cannot be overwhelmed by gravity at any density. The hierarchy terminates not because physics runs out of mechanisms but because the final mechanism scales with the force it opposes. The fractal pattern converges.

32.7. The Block Universe and the Initial State

The K-Limit has cosmological as well as astrophysical consequences. If the gravitational singularity does not exist — if infinite density is precluded by the chiral phase transition — then the initial state of the universe is not a singularity but a finite-density configuration.

The Architecture of the Block Universe: Metric Relaxation, Vacuum Anchoring, and the Structural Necessity of Cosmic Megastructure (Kriger, 2026; doi:10.5281/zenodo.19211673)

develops this consequence in detail. The initial state is a closed three-sphere S^3 at nuclear density, with radius $R_{initial} \approx 15$ AU — approximately the size of Saturn's orbit. This is not a speculation but a consequence of two constraints: the K-Limit (density $\leq \rho_K$) and the total mass-energy content of the observable universe (which determines the radius at ρ_K through the Friedmann equation).

The initial state is *doubly locked*: the K-Limit provides the maximum density (nuclear physics constraint), and the elastic vacuum ($R^2/6M^2$ term) provides the minimum radius (membrane stiffness constraint). The two constraints converge on a single configuration — a finite, deterministic initial state with no free parameters.

The connection to *Testing Nothingness and Evaluating the Inevitability of Life* (Chapter 28) is immediate. Chapter 28 established through the deductive chain that existence requires structure and that structure requires persistence conditions. The K-Limit is the *cosmological persistence condition*: it prevents the universe from collapsing to a structureless point (the singularity) and ensures that the initial state retains the structure necessary for subsequent evolution.

32.8. Catastrophic Phase Transitions and the Dissemination of Complexity

Catastrophic Phase Transitions as Necessary Conditions for Complexity Dissemination: Supernovae, Failed Collapses, and the Persistence of Differentiation in the Universe (Kriger, 2026; doi:10.5281/zenodo.18859088) extends the K-Limit framework to a prediction that standard astrophysics does not make: the destruction of individual stellar systems through supernovae is not a failure of persistence but a *necessary condition* for the persistence of complexity at the civilisational scale.

The Persistence Theorem operates at multiple levels simultaneously. At the level of an individual star, gravitational collapse is a persistence failure — the star ceases to exist as a main-sequence object. But at the level of the galaxy, the supernova is a persistence-*enabling* event: it disseminates heavy elements (carbon, oxygen, iron, uranium) that were synthesised in the star's interior and that are necessary for the existence of planets, chemistry, biology, and observers.

The persistence framework predicts a specific hierarchy: the persistence of complexity at scale N requires the *destruction* of systems at scale N−1. Stars must die for planets to form. Planets must be tectonically active (a form of partial self-destruction) for life to emerge. Species must

go extinct for ecosystems to diversify. The hierarchy is not a philosophical observation but a consequence of the *Conservation and Convertibility* law: complexity at scale N requires elements that exist only as bound states within systems at scale N−1, and releasing them requires breaking the N−1 systems.

The *Asymmetry of Totalizing Ideals* from *The Asymmetry of Totalizing Ideals: A Structural Law of Complex Adaptive Systems* (Kriger, 2026; doi:10.5281/zenodo.18361828) provides the formal complement: the ideal of preserving *all* systems at *all* scales is a totalising ideal whose pursuit would destroy complexity at higher scales. The universe's strategy is not total preservation but *selective destruction for higher-level persistence* — exactly the strategy that the persistence framework identifies as optimal for any complex adaptive system.

32.9. Falsifiable Predictions

The K-Limit generates specific, falsifiable predictions that distinguish it from the standard GR prediction of singularity formation.

First, *tidal Love numbers*: classical black holes in GR have tidal Love numbers $k_2 = 0$ (they are perfectly rigid under tidal deformation). Vacuum stars and frozen stars with K-Limit cores have $k_2 > 0$ (the finite-density core deforms under tidal stress). The LISA gravitational-wave observatory, expected to detect mergers of supermassive compact objects, can measure k_2 for objects above 3 M⊙. A non-zero measurement would confirm the K-Limit; a confirmed zero measurement would falsify it.

Second, *gravitational-wave echoes*: if the K-Limit core is surrounded by a chiral transition surface (the boundary between the hadronic and deconfined phases), gravitational waves impinging on this surface will partially reflect, producing echoes in the post-merger gravitational-wave signal. The Einstein Telescope, with its improved sensitivity at low frequencies, is designed to detect such echoes.

Third, *mass-dependent internal structure*: the K-Limit predicts that all compact objects above the OV limit have the same core density (ρ_K) but different radii. This produces a specific mass-radius relation distinguishable from both GR black holes (all mass concentrated at the singularity) and alternative models (gravastars, fuzzballs).

Conclusion

The Kriger Limit — the fourth and final barrier in the astrophysical stability hierarchy — is a direct consequence of the Persistence Theorem applied to gravitational collapse. The

persistence framework predicted its existence (a system at the viability boundary must transition rather than cease); the Law of Limit to Negation identified its mechanism (compression of vacuum structure releases the pressure that halts compression); phase transition analysis from Chapters 22 and 25 provided the formal machinery (chiral symmetry restoration as bifurcation); the Formalisation Laws diagnosed the singularity as an artifact of GR pushed beyond its domain; and Fractal Scaling identified the hierarchy as a self-similar pattern that terminates at the K-Limit because the final mechanism scales with the force it opposes.

The K-Limit is not a new postulate. It is a QCD consequence — the chiral phase transition is standard physics, studied on the lattice and observed in heavy-ion collisions. What is new is the recognition that this standard physics prevents gravitational singularities, and that this recognition was generated not by astrophysical reasoning but by the persistence framework's structural logic: the same logic that generates viability conditions for biological cells, phase transitions in motivational dynamics, and stability barriers in institutional design.

CHAPTER 33. INCOMPATIBLE TRUTHS, COMPRESSION, AND THE DARK MATTER QUESTION

Introduction

Chapters 30–32 demonstrated the persistence framework's capacity to compute a constant, predict a structure, and resolve a singularity. The present chapter confronts the framework's most empirically demanding test: the dark matter problem. The gravitational deficit observed in galaxies — the fact that visible matter accounts for only ~15% of the gravitational mass required to explain observed rotation curves — has been interpreted since the 1970s as evidence for a new fundamental constituent: cold dark matter (CDM) particles. Decades of direct detection experiments (XENON, LUX, PandaX) have found nothing. Collider searches (LHC) have found nothing. The persistence framework offers a structural diagnosis: the "dark matter particle" is not a failed detection but a *compression artifact* — a primitive in one compression scheme that does not exist as a primitive in another equally valid scheme. The diagnosis is not speculative; it is tested against 175 galaxy rotation curves from the SPARC database, and it passes.

Seven laws of the dissertation are operative: the *Theorem of Incompatible Truths* and the *Compression Imperative* from Chapter 19, *Against Causation* from Chapter 20, *Definition-Dependent Provability* from Chapter 21, the *Viability Mismatch Law* from *The Viability Mismatch Law: A Universal Principle for Viable Systems with Stress as Special Case* (Kriger, 2026; doi:10.5281/zenodo.18433777), the *Structural Diagnostic Principle* from *Structural Diagnostic Principle for Dynamical Models* (Kriger, 2018; doi:10.5281/zenodo.18344823), the *Architecture of Concealment* from *The Architecture of Concealment: A Mathematical Theory of Secrecy from Cells to Civilizations* (Kriger, 2026; doi:10.5281/zenodo.19392904), *Goal Preservation Across Capability Gain* from Chapter 28, and the dual-system analysis of *Leibniz's Two Kingdoms* from Chapter 28.

The cosmological content is developed principally in the vacuum halo programme: *The Vacuum Capture Model: Phase Transitions, Galactic Rotation, and the Origin of Missing Mass* (Kriger, 2026; doi:10.5281/zenodo.19023541), *The Hidden Baryonic Sector: Compact Remnants, Fractal Cold Gas, and the Non-Vacuum Dark Mass Budget* (Kriger, 2026; doi:10.5281/zenodo.19026019), *175 Galaxy Rotation Curves Fitted Without Dark Matter Particles: The αLGQV Vacuum Halo Programme — Complete Report* (Kriger, 2026;

doi:10.5281/zenodo.19390767), and the seven individual survey papers covering dwarf irregulars (submitted to MNRAS), low-surface-brightness galaxies (submitted to A&A), massive spirals (submitted to PASJ), gas-dominated late-type spirals (submitted to PRD), edge-on galaxies (submitted to ApJL), and initial SPARC samples (submitted to RNAAS).

33.1. Definition-Dependent Provability → "Dark Matter" Is Underdefined

The *Principle of Definition-Dependent Provability* (Chapter 21) established that the provability status of a statement changes as a function of the definitions assigned to its key terms. A statement that varies in provability across admissible definitions is not undecidable; it is underdefined.

Applied to the dark matter question, DDP reveals that "does dark matter exist?" admits at least four admissible definitions, each yielding a different provability outcome.

Under *Definition D_1* (gravitational deficit): dark matter is whatever accounts for the observed discrepancy between visible mass and gravitational mass. Under D_1, dark matter trivially exists — the deficit is observed, therefore dark matter (defined as the deficit) exists. $f(D_1) = 1$.

Under *Definition D_2* (particle beyond the Standard Model): dark matter is a new fundamental particle not included in the Standard Model of particle physics. Under D_2, dark matter has not been detected despite decades of direct searches, collider searches, and indirect detection experiments. $f(D_2) = 0$ (as of current evidence).

Under *Definition D_3* (modified gravity): dark matter is an artifact of using Newtonian/GR gravity without modifications required at galactic scales. Under D_3, the question is whether modified gravity (MOND, TeVeS, etc.) can reproduce the observations. $f(D_3)$ = indeterminate (some successes, some failures).

Under *Definition D_4* (vacuum halo): dark matter is the gravitational effect of vacuum energy concentrated around matter through the coupling $\Lambda(\rho_m) = \Lambda_0 - \alpha\rho_m$. Under D_4, dark matter is not a new substance but a known substance (vacuum energy) in a previously unrecognised configuration. $f(D_4) = 1$ (if the αLGQV framework is correct).

The debate has proceeded for four decades without specifying which definition is operative. Experimentalists searching for CDM particles use D_2. Theorists computing halo profiles use D_1. MOND proponents use D_3. The αLGQV framework uses D_4. The persistence framework's

contribution is not to resolve the debate but to diagnose its structure: the debate is stuck because the question is underdefined, and it will remain stuck until the definition is specified.

33.2. Theorem of Incompatible Truths → CDM and αLGQV as Incompatible Descriptions

The *Theorem of Incompatible Truths* (Chapter 19) proved that any sufficiently complex system necessarily admits pairs of individually valid but mutually incompatible descriptions. The dark matter problem is a concrete instantiation.

The CDM description and the αLGQV description are individually valid: both reproduce the observed rotation curves, both are consistent with large-scale structure, both make predictions testable by future observations. They are mutually incompatible: CDM asserts that the gravitational deficit is caused by a new particle species; αLGQV asserts that it is caused by the vacuum energy distribution. These cannot both be true in a single consistent framework — "the deficit is caused by particles" and "the deficit is caused by vacuum energy" are logically incompatible statements about the same observable.

The Theorem of Incompatible Truths predicts this incompatibility. The galactic gravitational field is a sufficiently complex system (multiple scales, overlapping domains of description, non-trivial dynamics). The CDM description and the αLGQV description use different compression schemes — different choices of what is primitive. In the CDM scheme, the dark matter density profile $\rho_{DM}(r)$ is primitive (a basic input to the model). In the αLGQV scheme, the vacuum energy profile $\Lambda(\rho_m(r))$ is primitive (derived from the matter distribution through the coupling α). A shared observable — the total gravitational mass at radius r — is primitive in one scheme and derived in the other. By the Compression Incompatibility Theorem (Chapter 19, Theorem 4.3), the descriptions are necessarily incompatible.

The consequence for the dark matter debate is structural. The debate is not a contingent failure of evidence (more data will resolve it) but a structural feature of the system's complexity (two valid compression schemes with different primitives must disagree). The resolution will come not from finding or failing to find CDM particles but from specifying the definition (Section 33.1) and comparing the descriptions' structural costs (Section 33.6).

33.3. Compression Imperative → Dark Matter Particle as Compression Primitive

The *Compression Imperative* (Chapter 19) requires that any valid description compress: its specification must be shorter than a brute-force enumeration of all observable values. Any compressive description must choose which observables to treat as primitive (fundamental, underived) and which as derived (consequences of the primitives).

In the CDM description, the compression scheme works as follows. Primitives: the Standard Model particles plus one additional particle species (CDM). Derived: the gravitational field of galaxies, computed from the distribution of all matter including CDM. The compression is effective: from a small number of particle-physics parameters (mass, cross-section, production mechanism), the entire galactic phenomenology is derived.

In the αLGQV description, the compression scheme is different. Primitives: the Standard Model particles plus the vacuum-matter coupling α. Derived: the gravitational field of galaxies, computed from the matter distribution plus the vacuum energy distribution Λ(ρ_m). The compression is also effective: from a single parameter (α = 0.005, derived from QCD), the entire galactic phenomenology is derived.

The two schemes make different structural commitments. CDM treats the dark matter density profile as a *consequence* of particle physics (primitives: particle properties → derived: halo profile). αLGQV treats the vacuum energy profile as a *consequence* of QCD (primitives: α from sigma terms → derived: vacuum halo profile). These commitments are incompatible on the overlap: both address the same observable (rotation curves) but derive it from different primitives.

The Compression Imperative does not adjudicate between the two schemes. It explains why both *must* exist: any system complex enough to admit multiple valid compression schemes will admit incompatible descriptions. The choice between them is not a matter of truth but a matter of compression efficiency — which scheme explains more with less.

33.4. Against Causation → "Dark Matter Causes Rotation" as Compression Artifact

Against Causation (Chapter 20) established that causal arrows are compression artifacts of bounded observers, not features of the territory. Applied to the dark matter debate, this law dissolves the apparent conflict between CDM and αLGQV at the causal level.

The CDM narrative is causal: dark matter particles *cause* the enhanced gravitational field that *causes* the flat rotation curve. The αLGQV narrative is also causal: vacuum energy

concentrated around matter *causes* the enhanced gravitational field that *causes* the flat rotation curve. The two narratives posit different causes for the same effect.

Against Causation reveals that this conflict is spurious. Neither dark matter particles nor vacuum energy *cause* the rotation curve in the territory-level sense. The territory contains a correlational structure: a galaxy's rotation velocity at radius r is correlated with the total mass-energy content within r. The CDM observer compresses this correlation by positing a particle. The αLGQV observer compresses it by positing a vacuum coupling. Both compressions work. Neither is "the cause."

The consequence is that the search for the "true cause" of galactic rotation is a category error in the sense identified by Against Causation. The productive question is not "what causes flat rotation curves?" but "which compression scheme is more efficient, more falsifiable, and more informationally complete?" This reframing shifts the debate from metaphysics (what is dark matter *really*?) to epistemology (which description is better?) — precisely the shift that the persistence framework prescribes.

33.5. Viability Mismatch and Structural Diagnostic → Tensions as Signals

The *Viability Mismatch Law* (Chapter 8) established that when the demands of the environment exceed the capacity of the system's current model, a diagnostic signal is generated. The *Structural Diagnostic Principle* (Chapter 8) interprets such signals: they are not anomalies to be patched but indicators that the model's structural commitments are incompatible with the observations.

Two diagnostic signals have emerged in the CDM framework. The *Hubble tension* — a 4–5σ discrepancy between the locally measured expansion rate ($H_0 \approx 73$ km/s/Mpc from supernovae) and the CMB-inferred value ($H_0 \approx 67.4$ km/s/Mpc from Planck) — is a viability mismatch: the ΛCDM model's predictions at two different scales are incompatible with each other. The *S_8 tension* — a discrepancy between the amplitude of matter fluctuations inferred from the CMB and that measured from weak gravitational lensing — is another viability mismatch of the same type.

The Structural Diagnostic Principle interprets both tensions as signals that ΛCDM's compression scheme is inadequate at the interface between early-universe and late-universe descriptions. The αLGQV framework addresses both: inhomogeneous expansion (voids

expand faster than walls) naturally produces a scale-dependent H_0, and the void-cluster cell model modifies the perturbation spectrum in the direction required to resolve S_8.

The point is not that αLGQV is "right" and ΛCDM is "wrong." It is that the Structural Diagnostic Principle — developed for general dynamical systems in Chapter 8 — correctly identified the tensions as structural signals and predicted the *type* of modification required (scale-dependent parameters) before the cosmological model was constructed. The tensions are not problems to be solved within ΛCDM; they are signals that the compression scheme must change.

33.6. The Vacuum Halo Programme: 175 Rotation Curves

The empirical test of the αLGQV description against the CDM description is the vacuum halo programme: fitting the rotation curves of 175 galaxies from the SPARC database using vacuum halo profiles derived from $\Lambda(\rho_m) = \Lambda_0 - \alpha\rho_m$, with $\alpha = 0.005$ fixed by QCD.

The programme proceeds through five galaxy classes — dwarf irregulars, low-surface-brightness galaxies, massive spirals, gas-dominated late-type spirals, and edge-on galaxies — each presenting distinct challenges for dark matter models. In every class, the vacuum halo profiles (pseudo-isothermal, with core radius r_c determined by the baryonic scale length R_d) reproduce the observed rotation curves with accuracy comparable to the best CDM fits.

The key results, reported in *175 Galaxy Rotation Curves Fitted Without Dark Matter Particles* (Kriger, 2026; doi:10.5281/zenodo.19390767):

The cusp-core problem — the systematic discrepancy between the cuspy (NFW) profiles predicted by CDM simulations and the cored profiles observed in dwarf and LSB galaxies — is resolved. Vacuum halo profiles are naturally cored: the vacuum energy density profile, derived from the baryonic matter distribution through the coupling α, peaks at the centre and flattens to a core. No tuning is required; the core emerges from the physics.

An emergent scaling relation $r_c \propto R_d$ connects the vacuum halo core radius to the baryonic disk scale length. This relation is not postulated; it follows from the functional form of $\Lambda(\rho_m)$. It provides a one-parameter family of halo profiles (parametrised by R_d, which is measured independently from photometry) that reproduces the diversity of rotation curve shapes across the SPARC sample.

The hidden baryonic sector — compact remnants (white dwarfs, neutron stars, stellar-mass black holes) and fractal cold gas — contributes an additional non-vacuum component to the

"dark" mass budget. *The Hidden Baryonic Sector* (Kriger, 2026; doi:10.5281/zenodo.19026019) estimates this contribution and shows that the total (vacuum + hidden baryons) accounts for the observed deficit without invoking new particles.

33.7. Architecture of Concealment → The "Dark Sector" as Information Partition

The *Architecture of Concealment* (Chapter 26) established that information partitioning is a universal mechanism operating from cells to civilisations, adaptive at biological scales and pathological at social scales, with the transition governed by interest alignment.

Applied to cosmology, the Architecture of Concealment provides a structural interpretation of the "dark sector" — the 85% of the universe's mass-energy content that is invisible to electromagnetic observation. In the standard ΛCDM framework, this invisibility is treated as a *property of the substance*: dark matter doesn't interact electromagnetically, therefore it's dark. The Architecture of Concealment offers a different reading: the "darkness" is a *property of the description*, not of the substance. The information is partitioned — separated into "visible" and "dark" sectors — because the compression scheme (ΛCDM) treats them as distinct components.

The *Scale Inversion Theorem* from the Architecture of Concealment is directly applicable. In QFT, information partitioning is adaptive: renormalisation separates physics at different scales, and this separation is essential for computational tractability. The partitioning is adaptive because the scales are genuinely decoupled (low-energy physics is insensitive to high-energy details). In cosmological interpretation, the same partitioning becomes pathological: the "dark sector" is not genuinely decoupled from the "visible sector" — they interact gravitationally, and the gravitational interaction is the entire source of the dark matter signal. The partitioning conceals the connection (vacuum energy is coupled to matter through α) by categorising it as a separate substance.

The αLGQV framework dissolves the partition: there is no "dark sector." There is one sector — matter coupled to vacuum through a QCD-derived parameter — whose gravitational effects include both the directly visible (electromagnetic) and the indirectly visible (gravitational deficit). The "darkness" was an artifact of the ΛCDM partition, not a property of reality.

33.8. Goal Preservation and Leibniz → ΛCDM's Drift and the Two-Kingdom Analogy

Two laws from Chapter 28 provide meta-theoretical perspective on the dark matter debate.

Goal Preservation Across Capability Gain (Chapter 28) proved that recursively self-improving systems suffer goal drift as their representational basis shifts. Applied to ΛCDM, the law describes a specific intellectual trajectory. ΛCDM was originally proposed as a *minimal* model — the simplest framework consistent with the observations. Its goal was parsimony: explain the maximum with the minimum. Over four decades of "capability gain" (new parameters, new components, new sub-models), the original goal has drifted: ΛCDM now requires six free parameters, a dark matter particle (never detected), a dark energy component (unexplained), an inflation mechanism (with its own free parameters), and a growing list of ad hoc modifications to accommodate anomalies (the Hubble tension, S_8, JWST early galaxies). The model's operative goal — fitting the data — has replaced its original goal — minimising assumptions. This is goal drift in the sense of Chapter 28, and the fidelity bound predicts that it will continue: each new observation that requires a new parameter shifts the representational basis further from the original specification.

Leibniz's Two Kingdoms (Chapter 28) provides a structural analogy. General relativity (the Kingdom of Geometry) and quantum field theory (the Kingdom of Fields) are two descriptions with different structural commitments — GR treats spacetime as a smooth manifold, QFT treats it as a quantum field. The standard response has been to seek pre-established harmony: a unified framework (string theory, loop quantum gravity) that reconciles the two kingdoms without residual tension. The Leibniz analysis (Chapter 28) demonstrated that pre-established harmony collapses the asymmetry between the kingdoms — exactly the asymmetry that makes each kingdom productive.

The αLGQV framework does not seek harmony. It maintains the asymmetry: GR describes the geometry, QCD describes the vacuum, and the coupling α connects them without merging them. The two "kingdoms" remain distinct, each valid in its domain, each contributing structural commitments that the other cannot provide. This is the Navigator Theorem (Chapter 19) in action: navigate between incompatible descriptions rather than force unification.

Conclusion

The dark matter question — "what accounts for the gravitational deficit in galaxies?" — is not a question awaiting more data. It is a question whose answer depends on the definition (DDP), whose alternative descriptions are structurally incompatible (Incompatible Truths), whose causal narratives are compression artifacts (Against Causation), whose tensions are diagnostic

signals (Viability Mismatch), and whose "dark sector" is an information partition (Architecture of Concealment).

The vacuum halo programme — 175 rotation curves fitted with one QCD-derived parameter, no new particles, natural core profiles, and an emergent scaling relation — demonstrates that the αLGQV description is empirically viable. It does not demonstrate that CDM is wrong; the Theorem of Incompatible Truths predicts that both descriptions will remain individually valid. What it demonstrates is that the "dark matter particle" is not the only compression primitive available, and that a compression scheme grounded in known physics (QCD vacuum structure) can match or exceed the performance of a scheme grounded in hypothetical physics (BSM particles).

The persistence framework's contribution to the dark matter debate is not a new answer to an old question. It is a diagnosis of why the old question has resisted resolution — and a demonstration that the structural tools developed for cognitive, social, and institutional systems apply, without modification, to the largest-scale empirical problem in contemporary physics.

CHAPTER 34. STELLAR PERSISTENCE: VIABILITY SETS, BINARY DOMINANCE, AND THE ASTROPHYSICS OF SURVIVAL

Introduction

Chapters 29–33 operated at cosmological scales — the vacuum, the cosmic web, singularities, galactic rotation. The present chapter descends to stellar scales and demonstrates that the persistence framework's most elementary tool — the viability set — generates specific, falsifiable predictions in astrophysics that standard stellar theory does not make. The central prediction is that binary star systems are not incidental outcomes of stochastic fragmentation but *persistence-selected* configurations: they survive the condensation process more frequently than single stars because their viability sets are larger. This prediction, and its extensions to specific astrophysical objects, were generated by the same formalism that produces viability conditions for biological cells (Chapter 5), cognitive agents (Chapter 9), and social institutions (Chapter 25).

The principal laws operative in this chapter are the *Viability Set Theory* from *The Transformational Basis of Persistence* (Kriger, 2024; doi:10.5281/zenodo.18435982), the *Asymmetry of Totalizing Ideals* from Chapter 14, the *Structural Distortion Principle* from Chapter 9, the *Predictive Processing* framework from *The Evolutionary Inevitability of Predictive Processing* (Kriger, 2026; doi:10.5281/zenodo.18444910), and the *Representational Isolation* principle from *Evolutionary and Information-Theoretic Argument for the Necessity of Representational Isolation* (Kriger, 2026; doi:10.5281/zenodo.18331202).

The astrophysical content is developed across a sustained programme: *Binary-First Star Formation as a Persistence-Selected Outcome of Open, Embedded Protostellar Dynamics* (Kriger, 2025; doi:10.5281/zenodo.18155356), *Why Binary Systems Are Optimal for Star Formation* (Kriger, 2026; doi:10.5281/zenodo.18144257), *Swept-Volume Geometry and Overlap Corrections in Protostellar Binary Accretion* (Kriger, 2026; doi:10.5281/zenodo.18155283), *Survival of the Bound: A Quantitative Theory of Persistence-Driven Binary Dominance in Dense Star-Forming Regions* (Kriger, 2026; doi:10.5281/zenodo.18155400), *The Paradox of Protostellar Core Formation* (Kriger, 2026; doi:10.5281/zenodo.18164699), *Can a Star Be Proven Single?* (Kriger, 2026; doi:10.5281/zenodo.18203209), *Observational Tests of Multiple Star Formation Scenarios* (Kriger, 2026; doi:10.5281/zenodo.18300562), *Exploring the Enigmatic JuMBOs* (Kriger,

2026; doi:10.5281/zenodo.18728940), *Dormant Neutron Stars in the Milky Way* (Kriger, 2025; doi:10.5281/zenodo.18409103), *Why Is Vega Metal-Poor and Rapidly Rotating?* (Kriger, 2026; doi:10.5281/zenodo.18870492), *What Do We Actually Know About Tau Ceti?* (Kriger, 2026; doi:10.5281/zenodo.18883220), *Differential Multi-Star Decorrelation for Exoplanet Detection in Triple Systems: The Challenging Case of Polaris* (Kriger, 2026; doi:10.5281/zenodo.19141803), *Quasars as Cosmological Information Beams* (Kriger, 2026; doi:10.5281/zenodo.18855088), and *Catastrophic Phase Transitions as Necessary Conditions for Complexity Dissemination* (Kriger, 2026; doi:10.5281/zenodo.18859088).

34.1. Viability Sets → Binary Dominance

The Viability Set Theory (Chapter 5) defines the viability set V(S) of a system S as the region of state space within which the system's structural integrity is maintained. The Persistence Theorem establishes that a system persists if and only if its trajectory remains within V(S). The comparative prediction follows immediately: given two configurations of the same physical substrate, the configuration with the larger viability set is more likely to survive a stochastic perturbation.

Applied to star formation, this comparative prediction yields a concrete astrophysical result. A protostellar core collapsing within a turbulent molecular cloud faces perturbations at every stage: tidal forces from neighbouring cores, accretion variability, radiation feedback, magnetic braking, dynamical encounters with passing stars. At each perturbation, the system's trajectory is displaced in state space. Whether it remains within the viability set — and therefore survives as a coherent stellar system — depends on the size of that set.

A binary protostellar system has a larger viability set than a single protostellar system for three quantifiable reasons. First, *accretion advantage*: two protostars orbiting each other sweep a larger volume of the surrounding molecular cloud, intercepting more material per unit time. The swept-volume geometry, analysed in *Swept-Volume Geometry and Overlap Corrections in Protostellar Binary Accretion* (Kriger, 2026; doi:10.5281/zenodo.18155283), shows that a binary with orbital separation a sweeps approximately $2\pi a$ times the cross-section of a single protostar per orbit — a factor of 3–10 enhancement in accretion rate depending on orbital parameters. Second, *dynamical stability*: a gravitationally bound pair is more resistant to tidal disruption than a single object, because the internal binding energy must be overcome before the pair can be separated. Third, *angular momentum regulation*: a binary can shed excess

angular momentum through orbital evolution (hardening, circularisation), whereas a single protostar must shed angular momentum through less efficient mechanisms (outflows, magnetic braking).

The prediction is specific and testable: the binary fraction should increase with the perturbation regime of the environment. In more turbulent, denser star-forming regions, single stars should be rarer because their smaller viability sets make them more vulnerable to disruption during the condensation process. *Survival of the Bound: A Quantitative Theory of Persistence-Driven Binary Dominance in Dense Star-Forming Regions* (Kriger, 2026; doi:10.5281/zenodo.18155400) derives this prediction quantitatively and identifies the observational surveys that could confirm or falsify it.

34.2. The Complete Formal Chain

The persistence-driven binary dominance theory is not a single argument but a formal chain, each link published as a separate paper and each link generated by the viability set formalism. *Why Binary Systems Are Optimal for Star Formation* (Kriger, 2026; doi:10.5281/zenodo.18144257) establishes the optimality claim: among all configurations of two protostellar masses in a shared accretion environment, the gravitationally bound binary configuration maximises the viability set volume. The proof proceeds through the viability formalism of Chapter 5 applied to the specific dynamics of protostellar accretion.

The Paradox of Protostellar Core Formation (Kriger, 2026; doi:10.5281/zenodo.18164699) identifies a puzzle that standard theory does not address: how does a coherent protostellar core form in the first place, given the turbulent fragmentation of molecular clouds? The persistence framework's answer is that core formation is itself a persistence-selected process: only those density fluctuations whose viability sets are large enough to survive the turbulent environment become cores. The paradox dissolves when core formation is understood not as a triggering event but as a persistence filter.

Can a Star Be Proven Single? (Kriger, 2026; doi:10.5281/zenodo.18203209) addresses the epistemological dimension: given that the persistence framework predicts binary dominance, can observational astronomy ever conclusively establish that a star is truly single (has no companion at any separation)? The answer, derived from the *Inaccessibility Theorem* (Chapter 20) applied to observational completeness, is no: at any given observational sensitivity, companions below the detection threshold cannot be excluded. The "single star" is an

observational category, not an ontological one — a point that connects directly to the *Representational Isolation* principle (Section 34.6).

Observational Tests of Multiple Star Formation Scenarios (Kriger, 2026; doi:10.5281/zenodo.18300562) identifies specific systems — VLA 1623-2417, L1551 IRS 5 — where the predictions of persistence-driven binary formation can be tested against alternative scenarios (turbulent fragmentation, disk instability, dynamical capture). Each system provides a discriminating test because the viability set analysis makes quantitative predictions about orbital parameters, mass ratios, and evolutionary timescales that differ from those of competing models.

34.3. JuMBOs: Persistence Selection at Planetary Mass

Exploring the Enigmatic JuMBOs: Metallicity-Driven Fragmentation, Binary Assembly, and Survival of Jupiter Mass Binary Objects (Kriger, 2026; doi:10.5281/zenodo.18728940) extends the viability set analysis to a recently discovered class of objects that challenges standard formation theory: Jupiter-Mass Binary Objects (JuMBOs), pairs of planet-mass objects in wide orbits discovered by JWST in the Orion Nebula.

Standard theory has no satisfactory explanation for JuMBOs: they are too massive to be ejected planets (ejection would not preserve wide binaries), too low in mass to form by standard core fragmentation (the Jeans mass is much higher), and too widely separated to have formed by disk instability. The persistence framework offers a structural explanation: JuMBOs are persistence-selected outcomes of metallicity-driven fragmentation. In metal-rich environments, cooling is more efficient, the Jeans mass is lower, and lower-mass fragments are viable. The viability set of a planetary-mass binary is larger than that of a planetary-mass single object for the same reasons as in stellar binaries (accretion advantage, dynamical stability), and the metal-rich environment of Orion provides the conditions for such objects to form and survive.

The prediction is that JuMBOs should be rarer in metal-poor environments — a testable prediction that connects the persistence framework's abstract viability formalism to specific observational programs.

34.4. Dormant Neutron Stars: Persistence at the Viability Boundary

Dormant Neutron Stars in the Milky Way: The Final Evolutionary Phase and Galactic Population Constraints (Kriger, 2025; doi:10.5281/zenodo.18409103) applies the viability set

formalism to the endpoint of stellar evolution. A neutron star that has exhausted its rotational energy, is no longer accreting matter, and has cooled below the detection threshold of current instruments is a system at the *minimal-energy boundary* of its viability set. It persists — it exists as a stable configuration of degenerate matter — but it has no external signatures.

The *Law of Observability* from the Four Laws of Self-Organising Systems (Chapter 7) predicts exactly this: persistent systems at the boundary of their viability set may exist in states that produce no external signatures. The dormant neutron star is the astrophysical instantiation: a persistent system that satisfies all viability conditions but is unobservable by any current instrument.

The analysis estimates the galactic population of dormant neutron stars — objects that are gravitationally present but electromagnetically invisible — and discusses the implications for the dark matter accounting of the Milky Way. The persistence framework predicts that a significant fraction of the galaxy's "missing mass" may reside not in exotic particles but in mundane objects (dormant neutron stars, cold white dwarfs, faint M-dwarfs) that have reached the observability boundary of their viability sets. This prediction connects to the *Hidden Baryonic Sector* analysis of Chapter 33.

34.5. Specific Objects as Falsification Tests

Three well-known stellar systems serve as specific test cases for the persistence framework's viability analysis, each posing a puzzle that standard stellar theory addresses only through ad hoc hypotheses.

Why Is Vega Metal-Poor and Rapidly Rotating? (Kriger, 2026; doi:10.5281/zenodo.18870492) addresses a double anomaly: Vega is one of the brightest stars in the sky, yet its metallicity is sub-solar (unusual for a young A-type star in the solar neighbourhood) and its rotation rate is near the critical limit (unusual for a star of its age and type). The paper develops four hypotheses — primordial cloud composition, binary merger, accretion from a metal-poor companion, and dynamical encounter — and subjects each to the viability set analysis. The merger hypothesis yields the largest viability set and makes specific predictions about Vega's internal structure testable by asteroseismology. The framework does not assert which hypothesis is correct; it ranks them by viability set volume and identifies the discriminating observations.

What Do We Actually Know About Tau Ceti? (Kriger, 2026; doi:10.5281/zenodo.18883220) identifies three paradoxes in the standard understanding of the nearest sun-like star and introduces a *Detection Completeness Principle*: the completeness of the detected planetary system is bounded by the observational sensitivity, and claims about the system's architecture must account for undetected planets. The Detection Completeness Principle is a specific application of the *Inaccessibility Theorem* (Chapter 20) to observational astronomy: the internal observer (the astronomer) has structural blind spots determined by instrumental sensitivity, and these blind spots constrain the provability of claims about the system.

Differential Multi-Star Decorrelation for Exoplanet Detection in Triple Systems: The Challenging Case of Polaris (Kriger, 2026; doi:10.5281/zenodo.19141803) develops a new observational method — decorrelation of multi-star signals — tailored for exoplanet detection in triple systems where standard techniques fail. The method is itself an application of the *Compression Imperative*: the observed signal is a superposition of contributions from three stellar components, and detecting a planetary signal requires compressing (factoring out) the stellar contributions. The decorrelation method is the compression scheme.

34.6. Predictive Processing and Representational Isolation in Observational Astronomy

Two laws from Part III — developed for cognitive agents and never intended for astrophysics — apply with unexpected precision to the practice of observational astronomy.

The Evolutionary Inevitability of Predictive Processing (Chapter 10) established that any system operating under latency and bandwidth constraints must develop predictive models that anticipate sensory inputs rather than passively recording them. Observational astronomy is precisely such a system: the data are noisy, sparse, and delayed (light travel time), and the astronomer must construct predictive models (stellar evolution tracks, cosmological simulations, planetary system architectures) that generate expectations against which new observations are evaluated as prediction errors.

The αLGQV framework can be evaluated within this predictive processing logic. Standard ΛCDM generates predictions (dark matter halo profiles, structure formation timescales, CMB anisotropy patterns) that produce specific prediction errors when compared against observations (the Hubble tension, S_8, JWST early galaxies). αLGQV generates different predictions from different structural commitments, producing *smaller* prediction errors on specific tests (rotation curves, S_8, early structure formation) and *comparable* prediction errors

on others (CMB). The predictive processing framework does not say which model is "true"; it provides a quantitative comparison of prediction errors — precisely the function it serves for cognitive agents navigating uncertain environments.

Representational Isolation (Chapter 10) established that complex systems never have direct access to external reality but always operate through reconstructed representations. Applied to observational astronomy, the principle reveals a structural constraint that astronomers routinely acknowledge but rarely formalise: every astronomical datum is a reconstruction. A galaxy's rotation curve is not directly observed; it is reconstructed from Doppler-shifted spectral lines, corrected for inclination, beam-smearing, instrumental response, and atmospheric effects. A supernova's distance is not directly measured; it is reconstructed through a chain of calibrations (Cepheids → tip of the red giant branch → Type Ia standardisation). The CMB anisotropy map is not a photograph; it is a reconstruction from microwave detector timestreams through complex pipeline processing.

Representational Isolation predicts that systematic errors in these reconstruction chains — precisely the kind of errors that produce tensions between different cosmological probes — are structurally inevitable, not accidental. The astronomer, like any cognitively bounded agent, can never verify the reconstruction against the "true" signal because the true signal is inaccessible (the *Inaccessibility Theorem* again). The best available strategy is the one the persistence framework prescribes for all representationally isolated agents: maintain multiple models simultaneously, switch between them adaptively, and treat prediction errors as diagnostic signals rather than failures.

34.7. Catastrophic Phase Transitions for Complexity Dissemination

Catastrophic Phase Transitions as Necessary Conditions for Complexity Dissemination (Kriger, 2026; doi:10.5281/zenodo.18859088) extends the viability analysis to a counter-intuitive prediction: the destruction of individual stellar systems through supernovae is not a persistence failure but a *necessary mechanism* for the persistence of complexity at the civilisational scale.

The argument synthesises three laws. The *Conservation and Convertibility* law (Chapter 5) requires that complexity at scale N requires elements synthesised at scale N−1 and released through the destruction of N−1 systems. Heavy elements (carbon, oxygen, iron) are produced

in stellar interiors and released only through supernovae. Without stellar death, no planets, no chemistry, no biology, no observers.

The *Asymmetry of Totalizing Ideals* (Chapter 14) predicts that the ideal of preserving all stellar systems indefinitely is a totalising ideal whose pursuit would prevent the dissemination of heavy elements — destroying the conditions for higher-level complexity. The universe's strategy is selective destruction: individual stars are sacrificed for the persistence of the element-enriched interstellar medium that enables subsequent star and planet formation.

Quasars as Cosmological Information Beams (Kriger, 2026; doi:10.5281/zenodo.18855088) provides a complementary perspective: quasars — the most luminous persistent objects in the universe — function as information beams that illuminate the intergalactic medium along their lines of sight. Each absorption line in a quasar spectrum records the density, temperature, metallicity, and kinematics of intervening gas at a specific redshift. The persistence framework interprets this as *informational persistence*: the quasar's light carries information about the state of the universe at multiple epochs, compressed into a single observable (the spectrum). The quasar is not merely a luminous object; it is a *cosmological instrument* — a persistent structure whose function, from the observer's perspective, is the delivery of information across cosmic distances and timescales.

Conclusion

The viability set formalism — the most elementary tool in the persistence framework's toolkit — generates an astrophysical research programme of substantial scope: from the dominance of binary star formation (a prediction from first principles, not a fit to data) through the galactic population of dormant neutron stars (systems at the viability boundary) to the specific anomalies of individual objects (Vega, Tau Ceti, Polaris) and the cosmological function of stellar death (catastrophic phase transitions for complexity dissemination).

The programme is unified by a single structural principle: *what persists is what has the largest viability set*. This principle was developed for biological cells, cognitive agents, and social institutions. Its application to stars, neutron stars, and JuMBOs required no modification — the axioms of Chapter 5 applied directly, generating predictions that standard astrophysics does not make and that are testable by current and near-future observational facilities.

The most unexpected application — predictive processing and representational isolation in observational astronomy — reveals that the persistence framework describes not only the

objects that astronomers study but the cognitive process by which they study them. The astronomer is a representationally isolated agent navigating between incompatible models under uncertainty, using prediction errors as diagnostic signals. The framework that describes this navigation is the same framework that describes the stars.

CHAPTER 35. THE OBSERVER AT COSMIC SCALES: INACCESSIBILITY, NAVIGATION, AND THE BOUNDARIES OF COSMOLOGICAL KNOWLEDGE

Introduction

Chapters 30–34 demonstrated the persistence framework generating discoveries — computing constants, predicting structures, resolving singularities, fitting rotation curves, explaining binary dominance. The present chapter turns the framework upon the observer. If the persistence framework is a theory of complex systems, and the observer is a complex system embedded within the universe it observes, then the framework must describe the observer's own limitations — and these limitations must have cosmological consequences.

The epistemological apparatus of Parts I and VI was developed for generic bounded observers: agents with finite resources, finite lifespans, and no direct access to external reality. The present chapter demonstrates that every epistemological constraint proved for generic observers is instantiated in its most extreme form for the cosmological observer — and that these constraints are not obstacles to be overcome but structural features to be incorporated into the theory.

The principal laws operative are the *Inaccessibility Theorem* from *Must Any Consistent Physics Contain Structure Inaccessible to All Internal Observers?* (Kriger, 2026; doi:10.5281/zenodo.18810467), the *Navigator Theorem* from *Adaptive Observer Theory: Navigation of Incompleteness and Incompatibility* (Kriger, 2026; doi:10.5281/zenodo.18812040), the *Evidence-Based Philosophy* conditions (C1–C4) from *Evidence-Based Philosophy: A Foundational Methodological Framework* (Kriger, 2026; doi:10.5281/zenodo.18751941), the *Dual-Source Decomposition* from Chapter 20, the *Structural Distortion Principle* from *The Structural Distortion Principle: A Closed-Loop Model of Perception, Attention, and World-Maintenance in Bounded Cognitive Systems* (Kriger, 2026; doi:10.5281/zenodo.18452700), and *Inference Without Propositions* from *Inference Without Propositions: Structural Distortion and the Physical Inevitability of Sub-Personal Reasoning* (Kriger, 2026).

The cosmological content is developed principally in *The Frozen Universe Illusion: Temporal Resolution as a Fundamental Limit on Cosmological Observation and SETI* (Kriger, 2026; doi:10.5281/zenodo.18901492), *The Inward Turn: Advanced Civilizations and the Fermi Paradox* (Kriger, 2026; doi:10.5281/zenodo.18437440), *The Ledger Time Model: Emergent*

Time from Irreversible Entanglement Validation in an Atemporal Quantum Configuration Space (Kriger, 2025; doi:10.5281/zenodo.18263726), *Estimation of the Contribution of Biospheric and Synthetic Cognitive Systems to the Total Effective Complexity of the Universe* (Kriger, 2026; doi:10.5281/zenodo.18261863), *Evaluating the Extraordinary Claim of Timescape Cosmology Superseding the Standard Cosmological Model* (Kriger, 2025; doi:10.5281/zenodo.14952703), and *Evaluating the Holographic Universe* (Kriger, 2025; doi:10.5281/zenodo.18101919).

35.1. Inaccessibility Theorem → The Frozen Universe Illusion

The *Inaccessibility Theorem* (Chapter 20) proved that any consistent physical framework must contain structure inaccessible to all internal observers. The theorem was proved for arbitrary systems and applied in Chapter 20 to quantum mechanics (the measurement problem), consciousness (the hard problem), and formal mathematics (Gödelian undecidability). The present section demonstrates its cosmological instantiation — and that instantiation is, in a precise sense, the most severe of all.

The Frozen Universe Illusion (Kriger, 2026; doi:10.5281/zenodo.18901492) formalises the temporal dimension of inaccessibility. A biological observer with a lifespan of ~80 years observes a universe that is 13.8 billion years old. The *Temporal Resolution Ratio* (TRR) — the ratio of the observer's temporal resolution to the timescale of the observed process — is approximately 6×10^{-9}. At this TRR, cosmological processes are not merely slow; they are formally indistinguishable from stasis. The paper proves a *no-go theorem*: for any biological observer with finite measurement precision ε, processes with TRR $\ll \varepsilon$ are indistinguishable from stasis at any achievable signal-to-noise ratio.

This is the Inaccessibility Theorem operating on the temporal axis. The inaccessible structure is not hidden behind an event horizon or locked inside a black hole; it is hidden in plain sight, rendered invisible by the ratio of the observer's lifespan to the universe's timescale. The observer sees a snapshot — a single frame of a movie that runs for 13.8 billion years — and must infer the movie from the frame.

The *Kriger Scaling Model* (KSM) provides a quantitative tool for this inference. It maps the 13.8-billion-year age of the universe onto a single human lifespan of 80 years, producing a rescaled timeline in which cosmological processes become intuitively accessible. The mapping is not a pedagogical metaphor (as in Sagan's Cosmic Calendar, which maps cosmic

time onto a single calendar year) but a quantitative instrument with a defined *Observability Index*: the fraction of a cosmological process's duration that falls within the observer's accessible window. A metric coincidence emerges: the galactic year (~225 million years) maps to approximately one KSM-scaled year, providing an intuitive anchor for the rescaled timeline.

35.2. Anthropic Temporal Window → Universal Observer Constraints

The Inaccessibility Theorem predicts that the observer's limitations are not contingent (fixable by better instruments) but structural (inherent in the observer's physical constitution). *The Frozen Universe Illusion* derives this structural limitation quantitatively through the *Anthropic Temporal Window* (ATW) hypothesis.

The ATW is bounded by three independent constraints. The *Margolus-Levitin bound* sets a minimum time per computational step for any physical system ($\Delta t \geq \pi\hbar/2E$), constraining the fastest possible temporal resolution of a biological observer. *Metabolic scaling theory* constrains the temporal resolution of carbon-based information-processing systems: metabolic rate scales with body mass as M^(3/4), setting the clock speed of biological cognition. And *stellar habitability constraints* limit the environments in which biological observers can evolve: stars on the main sequence provide stable energy for ~10^7–10^{10} years, constraining the available temporal window.

The convergence of these three constraints yields a prediction: biological observers are confined to approximately 7 of the ~61 logarithmic orders of temporal scale that characterise physical processes in the universe (from the Planck time ~10^{-43} s to the Hubble time ~10^{17} s). The Frozen Universe Illusion is therefore *universal among biochemical civilisations*: any observer whose cognition is based on chemical bonds and electrochemical signalling will experience the universe as frozen, regardless of its technological sophistication. This is the Inaccessibility Theorem applied to the class of all possible biological observers — a structural constraint on the epistemology of any civilisation that could evolve in this universe.

35.3. SETI Temporal Mismatch → The Fermi Paradox Reconsidered

The Frozen Universe Illusion has immediate consequences for the search for extraterrestrial intelligence, connecting to *The Inward Turn: Advanced Civilizations and the Fermi Paradox — Computational Efficiency as a Universal Attractor in Civilizational Development* (Kriger, 2026; doi:10.5281/zenodo.18437440).

The FUI introduces a *SETI Temporal Mismatch Parameter* — a new dimension of the Drake equation, orthogonal to the spatial, spectral, and modulation axes that define conventional search strategies. The parameter quantifies the degree to which a transmitting civilisation's operational timescale differs from the receiver's. Conventional SETI searches assume that the transmitting civilisation operates at a timescale comparable to ours — broadcasting signals that vary on scales of seconds to hours, detectable by our instruments' integration times. The FUI analysis shows that this assumption covers approximately 20% of the conceivable temporal search space. Civilisations operating at timescales of millennia (slow modulation) or nanoseconds (fast modulation) would be undetectable by current methods — not because their signals are weak but because our temporal resolution cannot recognise them as signals.

The *Inward Turn* provides a complementary explanation for the Fermi Paradox grounded in the persistence framework. Advanced civilisations, the paper argues, face a universal attractor: computational efficiency. As a civilisation advances, the energy cost per computation decreases, and the most efficient strategy for maintaining and expanding information processing shifts from external colonisation (energetically expensive, spatially bounded) to internal computation (energetically efficient, informationally unbounded). The Inward Turn predicts that advanced civilisations become computationally introspective — they turn inward rather than outward — and therefore produce no electromagnetic signatures detectable at interstellar distances.

The Inward Turn is not a sociological speculation; it is a consequence of the *Constraint-Autonomy Compatibility Law* (Chapter 8) applied to civilisational evolution. A civilisation that maximises external expansion (autonomy without constraint) encounters diminishing returns: the energy cost of maintaining coherent communication across interstellar distances grows without bound. A civilisation that maximises internal computation (constraint without autonomy) encounters computational halting: deterministic systems produce no novelty. The productive regime — the regime predicted by Constraint-Autonomy Compatibility — is internal computation with bounded exploration: precisely the Inward Turn.

35.4. Navigator Theorem → Cosmological Model Selection

The *Navigator Theorem* (Chapter 19) derived five necessary properties of any consciousness-like system navigating among incompatible descriptions of a complex domain. Applied to the

cosmological observer, the theorem describes the *only viable strategy* for a scientific community confronting multiple cosmological models.

The contemporary cosmological landscape includes at least four serious contenders: ΛCDM (the standard model, with six parameters and dark matter particles), αLGQV (the gravitating vacuum, with one parameter from QCD), timescape cosmology (Wiltshire, with no dark energy but with backreaction), and various modified gravity approaches (MOND, f(R) gravity, etc.). These models are mutually incompatible — they make contradictory claims about the nature of dark matter, dark energy, and cosmic expansion — yet each is individually valid in the sense that each reproduces a substantial subset of the observations.

The Navigator Theorem predicts that the viable strategy is not to choose one model and commit to it (this would produce the paradigmatic blindness identified in Chapter 27 for pandemic policy) but to maintain all models simultaneously and switch between them adaptively as new observations become available. The switching criterion is not truth (which model is "real") but navigational utility (which model produces the smallest prediction error for the current observation).

This prediction is testable against the actual practice of the cosmological community. The standard practice — commitment to ΛCDM as the "concordance model" with all alternatives treated as heterodox — violates the Navigator Theorem. The persistence framework predicts that this practice will produce increasingly severe prediction errors (the Hubble tension, S_8, JWST anomalies) as the observations explore regimes where ΛCDM's compression commitments are suboptimal. The prediction is being confirmed in real time.

35.5. Evidence-Based Philosophy → Evaluating Competing Cosmologies

Evidence-Based Philosophy (Chapter 21) introduced four conditions (C1–C4) that any empirical philosophical claim must satisfy: empirical anchorage, formal precision, fallibilist commitment, and reflexive consistency. These conditions, developed for philosophical inquiry in general, provide a specific evaluation instrument for competing cosmological frameworks.

Applied to the four principal cosmological models, the C1–C4 analysis yields discriminating results.

ΛCDM satisfies C1 (empirically anchored in CMB, BAO, supernovae data), satisfies C2 (formally precise — the Friedmann equations are well-defined), satisfies C3 (fallibilist —

specific predictions can be falsified), but *fails* C4 (reflexive consistency) because the model does not account for the observer's own limitations: it treats the universe as if it were observed from outside, ignoring the representational isolation and temporal resolution constraints that the Inaccessibility Theorem proves are structural.

αLGQV satisfies all four conditions: C1 (anchored in QCD sigma terms and 175 rotation curves), C2 (formally precise — one action, one parameter), C3 (fallibilist — eight specific predictions testable with DESI, Euclid, SKA, JWST), and C4 (reflexive consistency — the model acknowledges the observer's embeddedness through the Frozen Universe Illusion and the Structural Distortion Principle). The C4 advantage is not cosmetic; it is structural: a model that accounts for the observer's limitations makes different predictions from one that does not. Timescape cosmology satisfies C1–C3 but faces difficulties with C2: the backreaction formalism involves averaging procedures whose mathematical consistency is contested. The holographic universe hypothesis, evaluated in *Evaluating the Holographic Universe* (Kriger, 2025; doi:10.5281/zenodo.18101919) and *A Bayesian Model Comparison Framework for Extrapolative Scientific Hypotheses* (Kriger, 2026; doi:10.5281/zenodo.18230310), fails C3: its core predictions (the universe is a holographic projection from a lower-dimensional boundary) are not falsifiable by any currently conceivable observation.

The EBP analysis does not declare a winner. It provides a structural ranking: αLGQV satisfies all four conditions; ΛCDM fails C4; timescape faces C2 difficulties; holographic fails C3. The ranking is a *prediction* of the Evidence-Based Philosophy framework, testable by the scientific community's eventual assessment.

35.6. Dual-Source Decomposition → Form and Content in Cosmological Knowledge

The *Dual-Source Decomposition* (Chapter 20) separates mathematical knowledge into T_form (structural constraints contributed by cognitive architecture — stable across centuries) and T_cont (specific laws and constants determined by the world — subject to revolutionary change). Applied to cosmological knowledge, this decomposition reveals a specific pattern.

The T_form of cosmological knowledge — general relativity as a geometric framework, quantum field theory as a field-theoretic framework, statistical mechanics as an ensemble framework — has been stable for a century. These structural contributions are products of the observer's cognitive architecture: the mathematical structures (manifolds, gauge fields,

partition functions) are selected by the persistence conditions that the observer's cognitive system imposes on its own representations (Chapter 10).

The T_cont — the specific values of cosmological parameters, the identity of dark matter, the mechanism of cosmic acceleration — has changed repeatedly and is changing now. The Hubble constant has been revised from 500 (Hubble, 1929) to 67–73 (current tension). The dark energy equation of state has gone from $w = -1$ (assumed) to $w = -1 \pm$ uncertainties (measured). The nature of dark matter has gone from baryonic (1970s) to CDM (1980s–present) to uncertain (XENON null results).

The decomposition predicts that T_form will remain stable (GR and QFT will survive any revolution in their specific content) while T_cont will continue to change (the specific model — ΛCDM, αLGQV, or something else — will be revised as observations improve). This prediction is itself a consequence of the persistence framework: T_form persists because it is grounded in the observer's cognitive invariants; T_cont changes because it is grounded in the world's contingent structure.

35.7. Structural Distortion and Inference Without Propositions

The *Structural Distortion Principle* (Chapter 9) established that bounded cognitive systems systematically distort their representation of reality, not through error but through the structural constraints of their perceptual architecture. Applied to cosmological observation, the SDP identifies three layers of distortion.

First, *instrumental distortion*: every telescope, detector, and spectrograph imposes a transfer function on the signal, selecting some features and suppressing others. The Malmquist bias (preferential detection of bright objects), the Eddington bias (preferential scattering of objects near detection thresholds), and k-corrections (wavelength-dependent selection effects) are well-documented instances.

Second, *theoretical distortion*: the observer's theoretical framework determines what counts as "signal" and what counts as "noise." Within ΛCDM, a galaxy rotation curve that deviates from the NFW profile is a measurement problem. Within αLGQV, the same deviation is a prediction. The framework does not passively receive data; it actively structures perception — exactly as the SDP describes for cognitive agents.

Third, *communal distortion*: the scientific community's collective framework — the "standard model," the textbook presentation, the funding priorities — constitutes a *hostile epistemic*

environment in the sense of *Hostile by Design: A Triadic Model of Epistemic Environment Degradation* (Kriger, 2026). Alternative frameworks are filtered not by evidential inadequacy but by institutional mechanisms: arxiv endorsement requirements, journal referee selection, conference invitation lists. The SDP predicts that these mechanisms produce systematic distortion of the community's representation of the evidence — distortion that favours the incumbent framework regardless of its empirical adequacy.

Inference Without Propositions (Chapter 22) adds a further dimension. Sub-personal, non-propositional inference — pattern recognition that operates below the level of conscious deliberation — plays a significant role in observational astronomy. The astronomer who "sees" a spiral arm in a noisy image, the data analyst who "recognises" a spectral line in a crowded spectrum, the theorist who "feels" that a model is right before checking the numbers — all are performing inference without propositions. The persistence framework does not dismiss this inference as subjective bias; it identifies it as a structural necessity of any system operating under the latency and bandwidth constraints that characterise real-time scientific practice. But it also identifies its vulnerability: non-propositional inference inherits the biases of the training set (the textbooks, the standard model, the communal framework), and these biases are precisely the structural distortions that the SDP predicts.

35.8. Time, Complexity, and the Biospheric Contribution

Two final applications connect the observer's cosmological limitations to fundamental questions about the nature of time and the significance of life.

The Ledger Time Model (Kriger, 2025; doi:10.5281/zenodo.18263726) proposes that time emerges from irreversible entanglement validation in an atemporal quantum configuration space. The model connects the atemporality of memory storage (Chapter 11 — biological memory is stored without intrinsic temporal markers; time belongs to retrieval, not to storage) to the fundamental physics of time: if time is emergent rather than fundamental, then the atemporality of memory is not merely a biological adaptation but a reflection of the deeper structure of reality. The block universe architecture of Chapter 32 (the initial state is a timeless S^3 configuration) is compatible with the ledger time model: the "flow" of time is the observer's reconstruction of a sequence within an atemporal structure — the same reconstruction that cognitive agents perform when retrieving memories (Chapter 11).

Estimation of the Contribution of Biospheric and Synthetic Cognitive Systems to the Total Effective Complexity of the Universe (Kriger, 2026; doi:10.5281/zenodo.18261863) and *Significance of Biosphere Relative Contribution to Effective Complexity of the Universe* (Kriger, 2026; doi:10.5281/zenodo.18256772) address a question that connects the GEKS Index (Chapter 28) to cosmological epistemology: what fraction of the universe's total effective complexity is contributed by biological and synthetic cognitive systems?

The answer — that the biospheric contribution is vanishingly small in mass-energy terms but disproportionately large in effective complexity terms — is a prediction of the persistence framework. The Persistence Filter selects for structures with high invariant content; biological systems, with their genomic information, neural architectures, and cultural knowledge, carry orders of magnitude more invariant content per unit mass than any non-biological structure. The biosphere is insignificant by mass and significant by complexity — and this asymmetry is precisely what the persistence framework predicts: persistence selects not for mass but for structure, and biological systems are the most structured objects in the observable universe.

Conclusion

The observer at cosmic scales is not a transparent window onto reality. The observer is a bounded, representationally isolated, temporally constrained, structurally distorted system embedded within the universe it studies — and every one of these limitations was predicted by the persistence framework before it was applied to cosmology.

The Inaccessibility Theorem predicted the Frozen Universe Illusion. The Navigator Theorem predicted the necessity of maintaining multiple cosmological models simultaneously. Evidence-Based Philosophy provided the evaluation instrument for ranking competing cosmologies. The Dual-Source Decomposition identified which elements of cosmological knowledge are observer-contributed and which are world-contributed. The Structural Distortion Principle identified three layers of systematic distortion in cosmological observation. And the ledger time model connected the observer's temporal experience to the block-universe architecture that the K-Limit and the elastic vacuum produce.

The chapter's final result — that the biosphere contributes a vanishingly small fraction of the universe's mass-energy but a disproportionately large fraction of its effective complexity — is perhaps the most consequential prediction for the observer's self-understanding. The observer is not marginal to cosmology. The observer is the most structurally complex entity

that the persistence framework's laws have produced — and the framework that describes the observer's limitations is the same framework that describes the universe those limitations prevent the observer from seeing whole.

CHAPTER 36. THE SOCIOLOGY OF COSMOLOGICAL KNOWLEDGE: THE PERSISTENCE FRAMEWORK APPLIED TO THE SCIENTIFIC COMMUNITY ITSELF

Introduction

The preceding chapters applied the persistence framework outward — to the vacuum, the cosmic web, singularities, galaxies, stars, and the observer's epistemic limitations. The present chapter applies the framework *inward* — to the scientific community that produced, maintains, and resists the cosmological programme documented in this Part. If the persistence framework is truly universal — if its laws govern any complex adaptive system — then it must describe not only the objects that cosmologists study but the institution within which they study them. This recursive self-application is the final and most demanding test of the framework's universality.

The chapter draws on twelve laws from Parts I–IX, each developed for a specific domain and none developed for the sociology of science. That all twelve apply — without modification — to the dynamics of the ΛCDM community, the resistance to alternative cosmologies, and the institutional mechanisms of scientific knowledge production is itself a structural prediction of the persistence framework: complex adaptive systems at every scale exhibit the same operational patterns.

36.1. Pedagogical Systems with Invariant Kernels → ΛCDM as Persistent Textbook

Adaptive Pedagogical Systems with Invariant Kernels: A Formal Theory of Genre Persistence, Applied to the Medieval Latin Bestiary (Kriger, 2026; doi:10.5281/zenodo.18752630) established that persistent pedagogical genres consist of an invariant kernel $\mathcal{K}$ — a set of core propositions and structural conventions that remain unchanged across editions, translations, and adaptations — plus a set of modular components $\{M_i\}$ that are added, removed, or modified in response to environmental pressures. The genre persists as long as $\mathcal{K}$ persists; the modules provide adaptive flexibility without threatening identity.

ΛCDM exhibits precisely this architecture. The invariant kernel consists of: (i) the Friedmann equations as the dynamical framework, (ii) the six parameters (Ω_m, Ω_Λ, Ω_b, H_0, σ_8, n_s) as the model specification, (iii) cold dark matter as a fundamental constituent, and (iv) a cosmological constant as the source of accelerated expansion. This kernel has remained

unchanged since the late 1990s and is reproduced identically in every cosmology textbook from Peacock (1999) through Weinberg (2008) to Baumann (2022).

The modular components — baryonic acoustic oscillations, weak lensing analysis pipelines, Planck likelihood codes, simulation prescriptions for galaxy formation — have been extensively modified across editions and research programmes. But the kernel has not changed. The discovery of the Hubble tension (2019–present), the JWST early-galaxy anomalies (2022–present), and the S_8 tension (2020–present) have each prompted the addition of new modules (early dark energy, modified recombination, interacting dark sector) while leaving the kernel untouched.

The persistence of the kernel is not a sign of the kernel's truth. It is a structural prediction of the pedagogical persistence framework: kernels persist because they are deeply embedded in the training pipeline (graduate curricula, qualifying exams, postdoctoral hiring criteria), and modifying a kernel requires modifying the entire institutional apparatus that transmits it. The bestiary's kernel — the allegorical interpretation of animals as moral exempla — persisted for a millennium despite its manifest empirical inadequacy, because the institutional apparatus (monastic education, liturgical use, manuscript copying) was optimised for its transmission. ΛCDM's kernel persists for the same structural reason: the institutional apparatus (graduate training, journal peer review, telescope time allocation) is optimised for its transmission.

36.2. Autonomy Suppression → The Institutional Suppression of Alternative Cosmologies

Autonomy Suppression in Hierarchical Multi-Agent Systems: A Unifying Systems-Theoretic Framework (Kriger, 2020; doi:10.5281/zenodo.18520653) established that hierarchical systems systematically suppress the autonomous decision-making capacity of lower-level agents — not through conspiracy but through the ordinary operation of hierarchical incentive structures.

The cosmological community is a hierarchical multi-agent system. The hierarchy runs from graduate students through postdoctoral researchers to faculty, through journal editors and referees to funding agencies and telescope allocation committees. At each level, the agent's autonomy — the capacity to pursue research questions of their own choosing — is constrained by the requirements of the level above: the graduate student must satisfy the advisor, the

postdoc must satisfy the hiring committee, the faculty member must satisfy the funding agency.

The Autonomy Suppression framework predicts that this hierarchy produces a specific effect: convergence toward the kernel. Agents whose research is compatible with $\mathcal{K}$ receive positive feedback at every hierarchical level (the advisor approves, the journal publishes, the funding agency funds). Agents whose research challenges $\mathcal{K}$ receive negative feedback at every level (the advisor discourages, the referee rejects, the funding agency declines). The suppression is not intentional; it is a structural consequence of optimising a hierarchical system for the reproduction of its kernel.

The prediction is testable. The persistence framework predicts that: (a) the fraction of cosmology papers proposing alternatives to ΛCDM is substantially smaller than the fraction of observational anomalies that motivate them; (b) the career trajectory of researchers working on alternatives is systematically worse than that of researchers working within ΛCDM, controlling for publication quality and citation impact; (c) the time between the emergence of an anomaly and the institutional acknowledgement that the anomaly cannot be resolved within ΛCDM exceeds the time that the Structural Diagnostic Principle identifies as optimal. Evidence for all three predictions exists in the bibliometric and career-outcome data, though the present chapter identifies the structural mechanism rather than performing the empirical analysis.

36.3. Hostile Epistemic Environments → Cosmology as Information Ecosystem

Hostile by Design: A Triadic Model of Epistemic Environment Degradation (Kriger, 2026) developed a triadic model of hostile epistemic environments: environments in which the ratio of noise to signal (W), the distortion introduced by the transmission process P(W), and the theoretical framework that interprets the transmitted information T(W) jointly degrade the epistemic quality of the information available to agents.

The contemporary cosmological information ecosystem exhibits all three pathologies.

The *noise component* W is the preprint flood: approximately 50–70 cosmology papers are posted to arxiv daily. No individual researcher can read more than a fraction. The signal (genuine observational results, rigorous theoretical developments) is diluted in noise (incremental parameter fits, minor extensions of existing models, speculative proposals without testable predictions). The information-theoretic consequence is a decrease in the

functional finitude coefficient κ — the ratio of processable to total available information — that Chapter 24 identified as the critical constraint on communication in finite-channel systems.

The *distortion component* P(W) is the peer review and publication process, which functions as a transmission channel with specific distortion characteristics. The persistence framework predicts that the distortion is non-random: it preferentially transmits information compatible with the kernel $\mathcal{K}$ and preferentially blocks information incompatible with it. The mechanism is not corruption in the individual-moral sense (Chapter 26) but systemic bias in the structural sense: referees selected from within the ΛCDM community evaluate proposals against ΛCDM standards, producing systematic suppression of alternatives regardless of their intrinsic merit.

The *interpretive component* T(W) is the theoretical framework through which the community interprets observations. Within ΛCDM, a galaxy rotation curve that cannot be fitted by an NFW profile is a "measurement problem" or a "baryonic feedback effect." Within αLGQV, the same rotation curve is a "vacuum halo signature." The framework does not passively receive data; it actively structures perception — the Structural Distortion Principle operating at the communal level.

36.4. Assertion-Dismantling Cycles → The ΛCDM Paradigm Arc

Assertion-Dismantling Cycles in Adaptive Systems: A Constraint-Network Framework (Kriger, 2015; doi:10.5281/zenodo.18487230) established that adaptive systems cycle between assertive phases (in which a dominant framework is elaborated, extended, and institutionalised) and dismantling phases (in which anomalies accumulate, the framework is questioned, and alternative frameworks emerge).

ΛCDM is in the late assertive / early dismantling phase of this cycle. The assertive phase (1998–~2018) was characterised by: the consolidation of the "concordance model" after the supernova results (Perlmutter 1998, Riess 1998); the confirmation by WMAP and Planck; the institutionalisation of ΛCDM as the default framework for all cosmological analysis; and the marginalisation of alternatives (MOND dismissed as "not a theory," timescape dismissed as "not rigorous," vacuum energy models dismissed as "not standard").

The dismantling phase (~2019–present) is characterised by: the Hubble tension (a 4–5σ discrepancy that no ΛCDM-internal modification has resolved); the S_8 tension (a growing discrepancy between CMB and weak-lensing measurements of matter clustering); JWST

early-galaxy anomalies (massive, well-formed galaxies at z > 10 that ΛCDM formation models do not predict); and the persistent null results of direct dark matter detection experiments (XENON, LUX, PandaX).

The Assertion-Dismantling framework predicts the *dynamics* of the transition. During the late assertive phase, the dominant response to anomalies is *module addition*: new modules are added to $\mathcal{K}$ (early dark energy, modified recombination, interacting dark sector) to accommodate the anomalies without modifying the kernel. The dismantling phase begins when module addition fails — when no module can accommodate the anomaly without contradicting another module. The Hubble tension has reached this stage: proposed resolutions (early dark energy) alleviate the H_0 discrepancy but worsen the S_8 tension, and vice versa. The persistence framework predicts that the community will resist kernel modification until the cost of module proliferation exceeds the cost of kernel replacement — a threshold that the Viability Mismatch Law can quantify.

36.5. Conflict as Phase Transition → Paradigm Shift Dynamics

Conflict as Phase Transition: A Dynamical Systems Theory of Escalation in Coupled Organizational Networks (Kriger, 2005; doi:10.5281/zenodo.18484819) modelled institutional conflict as a phase transition: when a control parameter (the intensity of the disagreement) crosses a critical threshold, the system undergoes a qualitative regime shift from cooperative to adversarial dynamics.

Applied to the ΛCDM-to-αLGQV transition (or any future paradigm shift in cosmology), the model predicts a specific sequence. Below the threshold, the community maintains cooperative dynamics: anomalies are discussed within the framework, and alternative proposals are evaluated on their merits (even if the evaluation is biased by autonomy suppression). Above the threshold, the community splits into adversarial camps — defenders of the kernel versus proponents of alternatives — and the dynamics become self-reinforcing: each camp's assertions provoke counter-assertions, and the conflict escalates independently of the evidential merits.

The persistence framework predicts that the critical threshold for cosmology is approaching but has not yet been crossed. The evidence: anomalies are still discussed primarily within the ΛCDM framework (below threshold), but the language is becoming adversarial ("crisis in cosmology" — a phrase that would not have appeared in mainstream publications a decade

ago). The Structural Diagnostic Principle identifies the "crisis" language as a diagnostic signal: the community's own description of its situation is shifting from "challenge" to "crisis," reflecting the approach to the phase-transition threshold.

36.6. Four Principles of Corruption → Institutional Friction in Scientific Knowledge Production

Four Principles of Corruption: A Systems-Theoretic Analysis of Institutional Failure and Reform (Kriger, 2026; doi:10.5281/zenodo.18879948) established that corruption arises from institutional friction — the gap between formal rules and the conditions required for agents to achieve legitimate objectives. The four principles (systemic origin, systemic coercion, adaptive morphology, elimination through automation) apply to scientific knowledge production with uncomfortable precision.

The *Principle of Systemic Origin*: the "corruption" of cosmological knowledge production — the systematic suppression of alternatives, the proliferation of unfalsifiable modifications, the prioritisation of parameter-fitting over conceptual innovation — originates not in the moral failures of individual scientists but in the institutional friction of the publication, funding, and hiring systems. These systems were designed for a scientific landscape in which the standard model was well-established and incremental progress was the norm. They are structurally inadequate for a landscape in which the standard model is in tension with observations and paradigm-level innovation is needed.

The *Principle of Systemic Coercion*: above a critical level of institutional friction, participation in the "corruption" — working within ΛCDM even when the evidence suggests alternatives — becomes the only individually rational strategy. A postdoctoral researcher who publishes αLGQV papers will not be hired; a postdoctoral researcher who publishes ΛCDM parameter fits will be. The coercion is systemic, not personal: the hiring committee is not conspiring against alternatives; it is optimising for the metric it has been given (publication in high-impact ΛCDM journals).

The *Principle of Adaptive Morphology*: as the anomalies mount, the suppression of alternatives does not decrease; it changes form. Direct rejection ("this is wrong") gives way to indirect marginalisation ("this is not standard"), which gives way to definitional exclusion ("this is not cosmology"). The adaptive morphology of suppression mirrors the adaptive morphology of corruption described in Chapter 26.

The *Principle of Elimination Through Automation*: the structural remedy is the same one identified in Chapter 26 for institutional corruption — replace human discretionary decision-making (referee selection, funding allocation, hiring decisions) with automated, criteria-based systems that remove the locus of bias. AI-assisted discovery (the same AI tools that co-authored the Indus script analysis of Chapter 27) can bypass the institutional gatekeeping that suppresses alternatives, by generating results whose quality is assessed algorithmically rather than socially.

36.7. Extractive Oscillators → Big Science Funding Cycles

Extractive Oscillators with Sensor Degradation: A Dynamical Systems Class and Its Manifestation in Quasi-Narcissistic Relational Dynamics (Kriger, 2017; doi:10.5281/zenodo.18529185) formalised a specific dynamical pattern: a system that extracts resources from its environment through an oscillatory process, with progressive degradation of the sensors that regulate the extraction, leading to over-extraction and eventual collapse.

Big Science funding cycles in cosmology exhibit this pattern. Large experimental programmes (CERN dark matter searches, XENON/LUX direct detection, CMB satellite missions) extract funding from the available pool through an oscillatory process: a proposal is funded, results are produced, the results motivate a larger proposal, the larger proposal is funded. The "sensor" — the mechanism by which the community evaluates whether the programme is making progress toward its stated goals — degrades over time: null results are reinterpreted as "constraining the parameter space" rather than as evidence that the programme's fundamental assumption (dark matter is a particle) may be wrong.

The extractive oscillator model predicts that these programmes will continue to absorb funding until the sensor degradation reaches a critical threshold — the point at which even reinterpreted null results cannot justify further investment. The model also predicts the institutional dynamics that follow: the sudden withdrawal of funding produces a crisis (the extractive oscillator's collapse phase), which opens space for alternative approaches that were previously crowded out.

36.8. Identity, Goal Preservation, and Non-Participation

Three laws from the dissertation's cognitive and social chapters provide additional structural insight.

Identity as Algorithmic Configuration (Chapter 22) established that identity is not a fixed substance but a compact specification — an algorithmic configuration that generates the system's characteristic behaviour when executed on its substrate. ΛCDM's "identity" is its kernel $\mathcal{K}$; αLGQV's identity is the single parameter $\alpha = 0.005$ derived from QCD. Both identities run on the same substrate (GR + QFT). The paradigm shift is not a substrate change but an *identity specification change* — the most compact possible description of the difference between two cosmological frameworks. This reframing dissolves the apparent incommensurability of paradigm shifts: the frameworks share a substrate and differ only in their compact specification.

Goal Preservation Across Capability Gain (Chapter 28) predicts that ΛCDM, as a recursively self-improving scientific programme (each generation of observations and simulations enhances the programme's "capability"), suffers goal drift: the original goal (explain the universe with minimal assumptions) has been replaced by the operative goal (fit the data with whatever modifications are needed). The fidelity bound of Chapter 28 predicts that this drift will continue: each new data release shifts the representational basis, and the programme's goals drift further from the original specification.

Non-Participation and Early Exclusion as Stability-Preserving Institutional Strategies (Kriger, 2026; doi:10.5281/zenodo.18135188) identifies a structural mechanism that the ΛCDM community employs: early exclusion of alternative approaches from mainstream venues (journals, conferences, funding programmes). The persistence framework's analysis is neutral: early exclusion is a stability-preserving strategy that protects the kernel from destabilising perturbations. From within the community, this is rational institutional behaviour. From outside, it is the Principle of Systemic Coercion in action. The framework does not judge; it identifies the mechanism and predicts its consequences: delayed error-correction, accumulated tensions, and eventual paradigm crisis.

36.9. Simulability and Conceptual Responsibility

Two final laws complete the recursive self-application.

Simulability $S = 1 - k$, introduced in the infodynamics paper and discussed in Chapter 31, provides the ultimate epistemological criterion. ΛCDM has low simulability: its outcome depends on the specific realisation of primordial fluctuations (the random seed). αLGQV has high simulability: the cosmic web converges to a unique fixed point regardless of initial

conditions. The simulability criterion does not merely rank the two models; it identifies a qualitative difference in their informational structure. A theory whose laws determine the outcome is informationally self-sufficient — the laws *are* the theory. A theory whose outcome depends on initial conditions is informationally incomplete — the laws plus the initial conditions constitute the theory. The simulability criterion measures this difference quantitatively, providing a structural ranking that transcends the fit-to-data comparisons on which the conventional debate has stalled.

Conceptual Responsibility: Structural Constraints on Idea Transmission and the Ethics of Intellectual Communication (Kriger, 2014; doi:10.5281/zenodo.18488818) established that the transmission of ideas carries structural constraints: the transmitter is responsible not only for the accuracy of the transmitted content but for the receiver's capacity to interpret it correctly. Applied to cosmological communication, conceptual responsibility identifies a specific failure of the ΛCDM programme: the popular presentation of dark matter as "invisible stuff that makes up most of the universe" violates conceptual responsibility by transmitting a substantive ontological claim (dark matter is a substance) as if it were an established fact, when it is in fact a compression primitive whose existence depends on the compression scheme chosen. The αLGQV framework does not eliminate the communicative challenge — it replaces one complex idea (invisible particles) with another (gravitating vacuum) — but it satisfies conceptual responsibility to a greater degree because its primitive (the QCD vacuum) is a known, measured, and independently verified physical entity.

Conclusion

The persistence framework, applied recursively to the scientific community that studies the universe, reveals that the dynamics of cosmological knowledge production — the persistence of ΛCDM, the suppression of alternatives, the accumulation of anomalies, the approach to paradigm crisis — are governed by the same structural laws that govern the systems cosmologists study.

ΛCDM persists because its pedagogical kernel is deeply embedded in institutional transmission chains (Chapter 24). Alternatives are suppressed because hierarchical incentive structures systematically penalise deviation from the kernel (Chapter 13). Anomalies accumulate because the institutional response — module addition rather than kernel modification — is the adaptive morphology predicted by the Four Principles of Corruption

(Chapter 26). The community's information ecosystem is degraded by noise, distortion, and framework-dependent interpretation (Chapter 24). The funding structure exhibits extractive oscillator dynamics (Chapter 15). And the eventual paradigm shift — if and when it occurs — will follow the dynamics of conflict as phase transition (Chapter 14) and assertion-dismantling cycles (Chapter 15).

None of these laws were developed for the sociology of science. All were developed for other domains — cognitive systems, social institutions, relational dynamics, institutional pathology — and transferred to cosmology via the same Tree-Top method that transferred viability sets to stellar physics and compression analysis to the vacuum. The recursive self-application is the persistence framework's final test: a framework that describes the universe, the observer, and the institution that mediates between them — using the same formal apparatus throughout — has demonstrated the structural universality that the dissertation claims.

The circle closes. The persistence framework began with biological cells (Chapter 5), extended to cognitive agents (Chapters 9–12), social institutions (Chapters 13–16), epistemological methods (Chapters 19–21), political economy (Chapters 25–26), computational humanities (Chapter 27), ontological foundations (Chapter 28), and the vacuum, the cosmic web, the singularity, the galaxies, the stars, and the observer (Chapters 29–35). In Chapter 36, the framework describes itself — the scientific programme that produced it, the community that evaluates it, and the institutional dynamics that will determine its fate. If the persistence framework is correct, then its acceptance or rejection is itself a persistence-selected outcome, governed by the same laws it identifies. The framework does not claim immunity from its own analysis; it claims universality — and universality includes the framework itself.

CHAPTER 37. PHASE-TRANSITION SIGNATURES OF αLGQV: NANOGRAV, THE BARYON-TO-PHOTON RATIO, AND THE ONE-EVENT SYNTHESIS

Introduction

Chapters 29–36 demonstrated the persistence framework as a discovery engine at cosmological scales: the computation of the cosmological constant from QCD sigma terms (Chapter 30), the prediction of the cosmic web as a self-consistency fixed point (Chapter 31), the resolution of the gravitational singularity through the Kriger Limit (Chapter 32), the reproduction of 175 SPARC galaxy rotation curves without dark-matter particles (Chapter 33), and the structural applications to stellar persistence, the cosmological observer, and the sociology of cosmological knowledge (Chapters 34–36). The present chapter — and the three that follow — extend the cosmological verification with a body of work completed after Volumes I and II of the αLGQV (Local Gravity of Quantum Vacuum) monograph programme: a third volume (Kriger, 2026, Volume III) of structurally connected papers that test the framework against multiple independent observational results released between 2023 and 2026.

The present chapter synthesises four results of the third volume that share a common structural feature: each is a prediction derived from the vacuum–matter coupling $\alpha = \sigma\pi N/mN \approx 0.005$ with no free parameters, and each was committed to the monograph record before the observational data against which it is now compared. The four results are: the nanohertz gravitational-wave background detected by NANOGrav and its derivation from the cosmological QCD confinement transition; the baryon-to-photon ratio $\eta = \alpha^4$ confirmed against the P-ACT 2025 joint CMB analysis; the one-event synthesis identifying eight independent cosmological signatures as manifestations of a single phase transition and a single global topology; and the controlled derivations that sharpen earlier scaling arguments into first-principles calculations and correct a small number of statements in Volume II. Each result is a falsifiable prediction; the empirical record between 2022 and 2026 is, on the framework's own ledger, the strongest available test.

37.1. Nanohertz Gravitational Waves from the QCD Confinement Transition

The NANOGrav collaboration's 15-year data set provides evidence for a stochastic gravitational-wave background (GWB) at nanohertz frequencies, correlated across 67 pulsars with the Hellings–Downs angular pattern and exhibiting characteristic strain amplitude $h_c \approx$

2.4×10^{-15} at the reference frequency $f_{yr} = (1\ yr)^{-1}$ and spectral index $\gamma = 3.2 \pm 0.6$. The standard astrophysical interpretation — an unresolved population of supermassive black-hole binaries (SMBHBs) — predicts $\gamma = 13/3 \approx 4.33$, which lies in 1.9σ tension with the observation. The αLGQV framework supplies an alternative interpretation in which both the strength and the spectral shape of the signal are determined by measured QCD quantities with no parameters adjusted to the NANOGrav data (Kriger, 2026, Volume III, Paper #39).

The derivation has three independent quantitative outputs. The first is the phase-transition strength parameter $\alpha_{pt} = L / \rho_{rad}$ at the QCD scale, where L is the latent heat of the transition and ρ_{rad} the radiation energy density at T_*. The latent heat has three components: the MIT bag constant B (energy difference between deconfined and confined vacua), the chiral condensate energy $\varepsilon\chi = f_\pi^2 m_\pi^2$, and the gravitational dressing of these energies by the αLGQV coupling. Substituting the lattice value $B^{1/4} = 220$ MeV, $T_* = 170$ MeV, $g_* = 17$, and $\alpha = 0.005$ gives $\alpha_{pt} = 0.67$. The NANOGrav fit value is $0.63^{+0.35}_{-0.23}$, within 1σ of the derived prediction. No parameter is adjusted.

The second output is the spectral slope. A broken power-law from a cosmological first-order phase transition produces an effective $\gamma \approx 3$ in the NANOGrav band, consistent with the observed $\gamma = 3.2 \pm 0.6$. The SMBHB prediction of $\gamma = 13/3$ is the most natural astrophysical alternative, and its 1.9σ tension with the data is structurally significant: a phase-transition interpretation is favoured by the spectral shape independently of the strength comparison.

The third output is the peak frequency. A linear sigma-model calculation of the bounce action yields the inverse duration of the transition in Hubble units $\beta/H \approx 5$, which places the bubble-collision peak at $f_{bc} \approx 24$ nHz — inside the NANOGrav frequency band. The sound-wave peak at ≈ 120 nHz lies above the band, but its low-frequency tail contributes significantly. The full spectrum, decomposed into sound-wave, bubble-collision, and magnetohydrodynamic turbulence contributions, agrees with the NANOGrav free-spectrum data at the band-centre to within a factor of ≈ 1.4, with the residual interpretable as a mixed signal in which SMBHBs dominate at lower frequencies and the QCD-confinement signal dominates the spectral shape.

The structural significance of the result is twofold. First, the αLGQV framework was constructed for an entirely different purpose: to derive rotation curves, the radial acceleration relation, and the cosmological constant from QCD sigma terms. The gravitational-wave prediction emerges as a corollary of the same vacuum–matter coupling that explains dark-

sector phenomenology. A by-product prediction that matches an independent observation is, by the methodological standards developed in Chapter 21, among the strongest forms of evidence for a theoretical framework, because the prediction was not optimised for the data it matches. Second, the prediction places the QCD confinement transition — in the standard picture, a smooth crossover at zero baryon chemical potential — in the regime of a first-order transition with measurable gravitational-wave signature. The lattice QCD result of smooth crossover holds strictly at $\mu^B = 0$; at finite μ^B or with modified vacuum structure such as that supplied by αLGQV, a first-order transition remains structurally possible, and the NANOGrav data are consistent with it.

37.2. The Baryon-to-Photon Ratio as a QCD Observable: $\eta = \alpha^4$

The baryon-to-photon ratio $\eta = n^b/n\gamma$ is a fundamental cosmological number whose value ($\approx 6 \times 10^{-10}$) is conventionally treated as an input parameter of Big Bang nucleosynthesis (BBN), set by an unknown baryogenesis mechanism in the early universe and constrained by primordial deuterium and helium abundances and by the cosmic microwave background. The αLGQV framework supplies a structural alternative: η is derivable from the vacuum–matter coupling by the four-vertex relation $\eta = \alpha^4$, where α is fixed by the measured nucleon sigma terms (Kriger, 2026, Volume III, Paper #41).

Substituting $\sigma\pi N = 50$ MeV, $\sigma_s = 40$ MeV, the Planck-measured baryon fraction $f^b = 0.156$, and the nucleon mass $m^N = 938.272$ MeV into the structural relation $\alpha = f^b (\sigma\pi N + \sigma_s) / (3\, m^N)$ yields $\alpha_a^{LGov} = 4.988 \times 10^{-3}$ and $\eta_a^{LGov} = \alpha^4 = 6.19 \times 10^{-10}$. The most recent joint cosmic-microwave-background analysis by Louis et al. (2025), combining Atacama Cosmology Telescope Data Release 6 with Planck power spectra and DESI Year-1 baryon acoustic oscillation data (the P-ACT + lensing + BAO combination), determines the physical baryon density at $\Omega^b h^2 = 0.0226 \pm 0.0001$, corresponding to $\eta_o{}^b{}_s = 6.18 \times 10^{-10}$. The predicted and observed central values agree to 0.21%; equivalently, the CMB-inferred value of α differs from the sigma-term-derived prediction by 0.05%. No free parameter is adjusted.

The result is empirically a post-diction in the proof-style sense developed in Volume II: the prediction was committed to the monograph record before the survey analysis was released. The derivation invokes only measured QCD observables, a cosmologically-derived baryon fraction from Planck, and standard electroweak vertex-counting. The structural content is that the baryon-to-photon ratio is not an input parameter set by an unspecified baryogenesis

mechanism but a downstream consequence of the same QCD structure that generates the confinement transition discussed in Section 37.1. η and α_{pt} are sibling observables of a single underlying QCD vacuum architecture.

A mild ∼2% tension exists between the CMB-inferred and BBN-inferred values of η (Schöneberg, 2024). The αLGQV framework predicts this tension structurally: the modification $G_eff = G(1 + 2\alpha)$ applies at the nucleosynthesis epoch and produces a small downward shift of the BBN-inferred η relative to the CMB-inferred value, with the correct sign and approximate magnitude. The framework is therefore falsifiable in three independent channels — primary CMB ($\Omega^b h^2$), BBN (deuterium-derived η), and sigma-term measurements (lattice QCD on σπN, σ_s) — with the current data consistent with the prediction in all three. The connection to Chapter 30: the same α that computes the cosmological constant Λ_0 from QCD sigma terms computes η from the same inputs. The structural unity is the framework's principal claim.

37.3. The One-Event Synthesis: Eight Independent Signatures of a Single Phase Transition

The first two sections of this chapter each connected a single observation — NANOGrav, P-ACT — to a single derived quantity. The third paper of Volume III (Kriger, 2026, Paper #42, *One Event, Many Witnesses*) extends the analysis to a synthesis. The claim is that eight independent cosmological observables, conventionally treated as separate "anomalies" within ΛCDM, are simultaneous observational projections of two structural features of the αLGQV framework: a first-order phase transition of the quark–gluon–vacuum plasma at $T_c \approx 170$ MeV, and a closed projective-three-sphere (RP^3) global topology with $\Omega_k = -0.044$.

The eight signatures are: (i) the NANOGrav 15-year stochastic gravitational-wave background with spectral slope $\gamma = 3.2 \pm 0.6$, better fit by a first-order QCD-transition spectrum than by supermassive-binary emission (Section 37.1); (ii) the P-ACT 2025 baryon density determination $\Omega^b h^2 = 0.0226$, yielding $\eta_o{}^b{}_s = 6.18 \times 10^{-10}$ in agreement with α^4 at 0.2% (Section 37.2); (iii) the Patel–Aluri–Ralston 2025 confirmation of CMB low-multipole alignments — the "axis of evil" — across all WMAP and Planck releases; (iv) the Böhme et al. 2025 detection at > 5σ of the cosmic matter dipole excess, reaffirmed by the Secrest et al. 2025 colloquium; (v) the Shamir 2025 JADES galaxy-rotation asymmetry of ∼50% in the JWST Advanced Deep Extragalactic Survey field; (vi) the JWST "impossible early galaxy"

tension with ΛCDM structure formation; (vii) the Schöneberg 2024 BBN baryon-abundance update showing a sub-percent downward offset relative to CMB; and (viii) the cluster of preferred directions — CMB dipole, Dark Flow, axis of evil, quasar dipole, radio dipole, Pantheon SN dipole, galaxy spin asymmetry — in a ∼33°-radius patch of sky.

Each prediction is independently falsifiable, and each was committed to the literature before the corresponding observation was released. The chronology is documented explicitly in the paper: $\eta = \alpha^4$ was fixed in the Volume II draft in 2023, before NANOGrav 15-yr; Paper #38 (CMB + GW) circulated in 2024 before Schöneberg's BBN update; Volume II was submitted in early 2025 before P-ACT 2025, Shamir JADES, Patel et al. low-ℓ alignment, and Böhme et al. matter dipole. The proof-style framing is structurally precise: the predictions were fixed in the monograph record prior to the observations, the observations are independent, and the consistency holds across all eight channels simultaneously.

Two structural points complete the analysis. The first is the distinction of three regimes of deconfined QCD: heavy-ion QGP (terrestrial colliders, low baryon chemical potential), neutron-star-core QGP (high μ^B, low T), and cosmological QGVP (high T, $\mu^B \approx 0$ but with modified vacuum structure). The three regimes share the same Lagrangian but differ in the thermodynamic conditions and consequently in the order and signatures of the phase transition. The αLGQV prediction is that cosmological QGVP exhibits a first-order transition while heavy-ion QGP does not — a structural distinction consistent with all lattice QCD results and all collider observations to date.

The second is the correction of a numerical statement in Paper #38 of Volume II concerning μ-type spectral distortions of the CMB. The original statement quoted a pre-thermalisation distortion level $\Delta\rho/\rho \sim 0.28$ without the subsequent thermalisation factor. The post-thermalisation distortion is of order 10^{-8}, consistent with the FIRAS upper bound, and the original statement is replaced by the corrected expression developed in detail in Section 37.4. The correction does not affect the predictions of Sections 37.1–37.3.

37.4. Controlled Derivations: Sharpened Predictions and Corrections to Volume II

The four results synthesised above were derived in Volumes I and II by scaling arguments that exhibited the structural relations between QCD inputs and cosmological outputs but did not perform first-principles calculations in every case. The fifth paper of Volume III (Kriger, 2026, Paper #43, *Controlled Derivations in αLGQV*) replaces the scaling arguments with first-

principles calculations of six structural predictions and consolidates the corrections to Volume II identified in the course of the calculations.

The principal sharpenings concern five quantities. First, the cosmological constant Λ_0: the closed-form expression $\Lambda_0 = \alpha \cdot f^b \cdot \sigma^6 / m_{pl}^4$ is recomputed with the full factor structure, eliminating an O(1) ambiguity in the prefactor and reducing the predicted-to-observed ratio from 1.65 (Volume II) to a controlled value within the propagated uncertainty. Second, the transition mass M_{tr} separating vacuum stars from frozen stars: a small family of deconfinement-type equations of state, integrated through the Tolman–Oppenheimer–Volkoff system, sharpens the previous estimate $M_{tr} \approx 4.0\ M\odot$ to $M_{tr} \approx 3.08\ M\odot$ in the realistic-stiff regime, with frozen-only behaviour for all other tested combinations. The Volume II quotation of $M_{tr} \approx 4\ M\odot$ with vacuum stars in the [2.2, 4.0] $M\odot$ range is therefore replaced by a tighter structural range. Third, the Kgr-Radiation redshift factor R that determines the observed frequency from black-hole-mass-scale K-Limit objects: the proper-distance matching between the Kriger radius and the Schwarzschild radius yields $R \approx 4.5 \times 10^{-5}$ for Sgr A∗, shifting the predicted Kgr frequency from the LIGO band (∼20 Hz) of Volume II to the LISA band (≈7 mHz). The prediction is redirected, not withdrawn; Section 38.4 below treats the consequences. Fourth, the bare coupling $\alpha^b_{are} \approx 0.096$ and the reduction cascade to the effective $\alpha \approx 0.005$ via the baryon fraction f^b and the nonlinear screening factor $S \approx 3$: the screening factor is derived from the cosmic web active volume fraction without external fitting (Section 40.1). Fifth, the one-loop screening factors κ_{loop} entering structure-growth predictions, evaluated by Monte-Carlo integration on the Planck linear power spectrum.

The result is a closed cycle of structural predictions in which each derived quantity is now expressible in terms of the same small set of measured QCD observables — $\sigma\pi N$, σ_s, m^N, $B^{1/4}$, $f\pi$, $m\pi$ — with the framework parameters α, η, Λ_0, M_{tr}, r_k, f^{aob}_s, and the screening factors S, κ_{loop} as deductive consequences. The closure of the cycle is structurally significant: every numerical statement that depends on the framework is, in the controlled version, traceable to a measured QCD input. There are no remaining adjusted constants.

CHAPTER 38. BLACK-HOLE INTERIORS WITHOUT SINGULARITY: THE SECOND MECHANISM FOR $k_2 = 0$ AND KGR-RADIATION AS THE REMAINING DISCRIMINATOR

Introduction

Chapter 32 developed the persistence-theoretic resolution of the gravitational singularity through the Kriger Limit and characterised two compact-object phenomenologies: *vacuum stars* (below the transition mass M_{tr}, no horizon, finite-density core) and *frozen stars* (above M_{tr}, horizon present but enclosing a finite-density de Sitter core). The chapter advanced two principal observational discriminators between K-Limit objects and classical Kerr black holes. The first was the tidal Love number k_2: classical black holes in general relativity have $k_2 = 0$, while K-Limit objects with finite-density cores were predicted to have $k_2 > 0$, with the LISA observatory expected to measure this difference for supermassive compact objects. The second was gravitational-wave echoes from the chiral transition surface, expected to be detectable by the Einstein Telescope.

A controlled derivation of the K-Limit interior metric and its matching to the exterior Schwarzschild solution, completed after Volume II and now consolidated in Volume III (Kriger, 2026, *Why Black Holes Probably Have No Measurable Tidal Deformability*), requires a structural correction to the first of these predictions. The $k_2 > 0$ prediction is withdrawn. A second structural mechanism for $k_2 = 0$, distinct from the standard-GR regularity-of-exterior argument, follows from the K-Limit picture itself: the K-Limit exterior is identical to the Schwarzschild exterior, and the Binnington–Poisson theorem applied to the exterior vacuum region therefore gives $k_2 = 0$ identically for every $M > M_{tr}$, by the same argument that gives $k_2 = 0$ for Kerr. The convergence is structurally significant: two distinct interior theories produce the same exterior prediction. The remaining discriminator becomes Kgr-Radiation, the predicted gravitational-wave emission from the oscillating chiral surface at the Kriger radius, whose observational frequency is sharpened by the controlled derivation from the LIGO band to the LISA band for supermassive objects.

The present chapter develops the correction and its consequences in four sections. Section 38.1 establishes the $k_2 = 0$ convergence between standard GR and the K-Limit picture. Section 38.2 summarises the K-Limit interior metric, the automatic-matching identity at the Kriger radius, and the derivation of the second mechanism for $k_2 = 0$. Section 38.3 addresses the withdrawal of the LISA tidal-Love prediction of Chapter 32 and identifies the consequent

revisions to that chapter's falsification programme. Section 38.4 develops the Kgr-Radiation prediction as the remaining sharp discriminator, with predicted observed frequencies across the black-hole mass spectrum mapping different mass ranges onto different detector bands.

38.1. The $k_2 = 0$ Convergence: Standard GR and the K-Limit Predict the Same Tidal Love Number

The quadrupolar tidal Love number k_2 of an isolated stationary body embedded in an external tidal field measures the body's induced quadrupole moment per unit applied tidal field. For a deformable material body it is positive and finite; for an idealised perfectly rigid body it is zero. The Binnington–Poisson theorem (Binnington & Poisson 2009, generalising earlier work by Damour & Nagar) establishes that the quadrupolar electric-type and magnetic-type tidal Love numbers vanish identically for a Schwarzschild or Kerr black hole: $k_2^{\text{Э}} = k_2^{\text{B}} = 0$. The proof uses only the structure of the exterior vacuum solution and the regularity of the solution at the horizon; no statement about the interior is required. Recent LIGO/Virgo/KAGRA measurements of binary black-hole mergers are consistent with this prediction (Chia 2021; De Luca et al. 2025); no evidence for non-zero tidal deformability of astrophysical black holes has been reported.

The convergence with the K-Limit picture is structural. The K-Limit interior at $r \leq r_k$ is a de Sitter region of uniform density ρ_k with equation of state $P = -\rho_k c^2$ (vacuum). The K-Limit exterior at $r > r_k$ is vacuum, described by the Schwarzschild metric. The exterior is therefore identical to the Schwarzschild exterior of a classical black hole. The Binnington–Poisson theorem applied to this exterior gives $k_2 = 0$ identically for every K-Limit frozen star of mass $M > M_{tr}$. No part of the argument requires any statement about the interior, because the tidal response is computed entirely from the exterior solution.

The convergence is not coincidental. The structural feature shared by the two pictures is the *vacuum exterior*. Whatever the interior is — a singularity, a de Sitter core, a quark-matter equation of state, an extended chiral surface — the tidal Love number depends only on the exterior solution. Two distinct interior theories that match the same exterior vacuum solution at the appropriate radius produce the same tidal Love number. This is the structural content of the K-Limit prediction $k_2 = 0$: the K-Limit interior matches automatically to the Schwarzschild exterior at $r = r_k$ (Section 38.2), and the Schwarzschild exterior gives $k_2 = 0$ by Binnington–Poisson.

38.2. The K-Limit Mechanism: Vacuum Interior, Automatic Matching, and the Schwarzschild Exterior

The K-Limit frozen star of mass M is defined as a spherically symmetric static object whose interior at $r \leq r_k$ has uniform density $\rho = \rho_k$ and pressure $P = -\rho_k c^2$, and whose exterior at $r > r_k$ is described by the Schwarzschild metric, with the Kriger radius $r_k = (3M / 4\pi\rho_k)^{1/3}$. The interior corresponds to a de Sitter space whose local Hubble rate is set by ρ_k. The interior metric function is $f^{L}(r) = 1 - r^2 / L_{more}^2$ with $L_{more}^{-2} = 8\pi G\rho_k / (3c^2)$; the exterior metric function is $f^{OOO}(r) = 1 - R_s / r$ with $R_s = 2GM / c^2$.

Continuity of the metric function at the junction $r = r_k$ — the condition $f^{L}(r_k) = f^{OOO}(r_k)$ — requires $r_k^2 / L_{more}^2 = R_s / r_k$. Substituting the definitions of L_{more} and r_k in terms of M and ρ_k reduces this condition to an algebraic identity that holds for all M. The matching at the Kriger radius is therefore automatic and requires no surface stress, transition layer, or additional physical input. This is Proposition 1 of the controlled derivation; the proof is one line. The structural significance is that the K-Limit interior is fully consistent with the Schwarzschild exterior without auxiliary structure at the matching surface: the two regions fit together by virtue of the structure of the field equations themselves.

Independence of the K-Limit from the α-cascade is a structural feature of the αLGQV programme that the controlled derivation makes explicit. The K-Limit density ρ_k is a QCD observable distinct from the vacuum–matter coupling $\alpha = \sigma\pi N / m^N$: ρ_k is the density of chiral restoration at high baryon density, estimated from lattice QCD and Dyson–Schwinger methods at $\rho_k \sim 3–5\rho_0$; α is the dimensionless strength of the condensate shift per unit matter density in the dilute regime $n^B \ll n_0$. The two quantities probe different regimes of the same QCD vacuum: α characterises the linear response at low baryon density, ρ_k characterises the critical density at which the linear response terminates and the phase transition occurs. The transition mass M_{tr}, the Kriger radius r_k, the automatic-matching identity, the $k_2 = 0$ corollary, and the Kgr-Radiation frequency depend on ρ_k alone and not on α. The two physics sectors of the programme — low-density linear response (cosmology and galactic dynamics) and high-density phase transition (black holes) — are logically modular.

38.3. Withdrawal of the LISA Tidal-Love Prediction and Refinement of Chapter 32

The $k_2 = 0$ convergence requires a structural revision of Chapter 32. The original formulation advanced $k_2 > 0$ for K-Limit vacuum stars and frozen stars as one of two principal

discriminators between K-Limit objects and classical black holes, with the LISA observatory expected to detect non-zero tidal Love numbers for supermassive K-Limit objects. The controlled derivation establishes that this prediction is incorrect: the K-Limit exterior is identical to the Schwarzschild exterior, and the Binnington–Poisson theorem gives $k_2 = 0$ for both pictures. The LISA tidal-Love prediction is therefore withdrawn from the framework's falsification programme.

The withdrawal removes the programme's weakest-linked black-hole test and clarifies what the K-Limit picture actually claims about the black-hole exterior: it is identical to the standard Kerr exterior. All the physics that distinguishes the K-Limit picture lives inside the object: singularity resolution, absence of information loss, and the existence of an oscillating chiral surface at the Kriger radius. The convergence $k_2 = 0$ is the consequence of this structural claim. Two distinct interior theories agree on the exterior, and tidal-Love measurements are therefore not diagnostic of either picture.

What this does and does not imply requires structural care. The LIGO/Virgo/KAGRA measurements $\tilde{\Lambda} \approx 0$ for binary black-hole mergers — and the future LISA-era analogue for supermassive pairs — are consistent with the Kerr prediction and, separately, with the K-Limit prediction. The measurement is not diagnostic of either picture. This is not an unusual epistemic situation; many well-posed physical tests admit more than one explanation for the observed outcome, and distinguishing among them requires additional, complementary measurements. The implication that is *not* licensed by the measurement is the claim that a null tidal result demonstrates the presence of a singular Kerr interior. The singularity is an interior feature of the standard-GR model; the tidal measurement probes only the exterior. Identifying the two is conflating an interior theory with an exterior observation.

The gravitational-wave echoes prediction discussed in Chapter 32 — partial reflection of merger gravitational waves from the chiral transition surface — remains in the falsification programme. It is a strictly interior signature, not derivable from the exterior alone, and is therefore not affected by the $k_2 = 0$ convergence. The Einstein Telescope sensitivity at low frequencies remains the principal observational handle on this prediction.

38.4. Kgr-Radiation: The Remaining Falsification Channel

The K-Limit picture predicts a continuous narrow-band gravitational-wave signal from isolated, non-accreting compact objects above the transition mass: *Kgr-Radiation*, generated

by the oscillating chiral surface at the Kriger radius. At the phase-transition density ρ_k the quadratic term in the effective potential for the chiral order parameter exactly cancels, leaving a flat critical point at which small perturbations generate an undamped limit cycle between confinement and deconfinement. Each layer of the frozen star of thickness $\xi \sim 0.4$ fm (the chiral coherence length) executes this cycle on the bulk timescale $\tau^{bu}_{lk} \sim 1/f_{int}$, where $f_{int} = c_s / r_k = (c/\sqrt{3}) / r_k$ is the characteristic sound-crossing frequency of the core. For Sgr A∗ ($M = 4.3 \times 10^6$ M⊙, $r_k = 1135$ km), $f_{int} \approx 154$ Hz.

The signal is redshifted when received at infinity by the factor $R = r_k / (2\ R_s)$, derived in the controlled derivation from the proper-distance matching between the Kriger radius and the Schwarzschild surface. The observed frequency at Earth is therefore $f_o{}^b{}_s = f_{int} \cdot R \propto M^{-2/3}$, since $r_k \propto M^{1/3}$ and $R_s \propto M$. The mass scaling is a parameter-free observational test.

The predictions across the black-hole mass spectrum map different mass ranges onto different detector bands. For stellar-mass objects ($M \sim 5$–10 M⊙, mass-gap and stellar BH candidates), $r_k > R_s$ and the K-Limit object is larger than its Schwarzschild radius: there is no horizon, R is computed from the standard exterior-Schwarzschild formula, and $f_o{}^b{}_s \approx 14$–22 Hz — the LIGO band. For intermediate-mass objects ($M \sim 10^3$ M⊙), $f_o{}^b{}_s \approx 0.6$ Hz — the DECIGO band. For Sgr A∗, $f_o{}^b{}_s \approx 7$ mHz — the LISA band. For M31∗ (1.4×10^8 M⊙), $f_o{}^b{}_s \approx 0.8$ mHz — also the LISA band. For ultramassive objects ($M \sim 10^9$ M⊙), $f_o{}^b{}_s \approx 0.3$ mHz — below LISA, in the pulsar-timing-array band. The mapping is a sharp falsifier: standard GR predicts no continuous gravitational-wave signal from isolated non-accreting compact objects of any mass; the K-Limit predicts a specific narrow-band signal at $f_o{}^b{}_s \propto M^{-2/3}$ in the appropriate band for the object's mass.

The redirection of the Kgr prediction from the LIGO band (the original Volume II estimate for Sgr A∗) to the LISA band is a refinement, not a retraction. The frequency shift follows from the proper geometric calculation of the redshift factor R, which gives $R \approx 4.5 \times 10^{-5}$ for Sgr A∗ instead of the previous range $R \in [0.01, 0.3]$. The K-Limit oscillation is effectively permanent: Sgr A∗ has been emitting Kgr radiation throughout its existence, and a narrow-band continuous search with integration time of one year achieves an effective LISA strain sensitivity within reach of the predicted amplitude for coherence factors $\eta c_{oh} > 10^{-20}$.

The discriminator structure that emerges is therefore sharper than the original Chapter 32 formulation. The k_2 prediction is removed; the gravitational-wave-echoes prediction is retained; the Kgr-Radiation prediction is sharpened and redirected from LIGO to LISA for supermassive sources, with stellar-mass signals remaining in the LIGO band and ultramassive signals in the PTA band. The picture is falsifiable in three complementary detector regimes; the absence of Kgr-Radiation in all three bands would refute the K-Limit picture, and the presence of a continuous narrow-band signal at $f_o{}^b{}_s \propto M^{-2/3}$ in any one band would confirm the central structural prediction of Chapter 32 as refined by the controlled derivation.

CHAPTER 39. ROTATING RP³ COSMOLOGY: HOPF FIBRATION, DOUBLE COUNTER-ROTATION, AND CENTRIFUGAL EXPANSION

Introduction

Chapters 30–35 established the cosmological-scale predictions of the persistence framework as formulated in Volumes I and II of the αLGQV programme: a derived cosmological constant from QCD sigma terms, a predicted cosmic-web structure as a self-consistency fixed point, the resolution of the gravitational singularity through the Kriger Limit, and the reproduction of galaxy rotation curves without dark-matter particles. The present chapter develops a structural extension of this material introduced in Volume III: the global topology of the universe is modelled as a real projective three-sphere $RP^3 = S^3/\mathbb{Z}_2$ undergoing isoclinic Clifford rotation in two orthogonal planes of an embedding $\mathbb{R}^4$. A double counter-rotation on RP^3 supplies the structural origin of phenomena conventionally attributed to dark matter and dark energy through frame-dragging rather than through a non-baryonic substance; the same rotation provides an analytical model of cosmic expansion as centrifugal relaxation, with a closed-form solution for the scale factor and a predicted Hubble parameter intermediate between the Planck and SH0ES values.

The chapter synthesises four papers of Volume III: *Cosmic Megastructures as Projections of Hopf Fibration Tori* (Section 39.1); *Galaxy Spin Chirality, Rotation Curves, and CMB Parity from Double Counter-Rotation on RP³* (Section 39.2); *Cosmic Expansion as Centrifugal Relaxation: An Analytical Model on RP³* (Section 39.3); and *A Unified Origin of Bulk Flows, Spin Chirality, and Cosmic Dipoles from Rotating RP³ Topology* (Section 39.4). Together the four results constitute an alternative structural account of dark phenomenology in which the global topology of the universe and its rotation, rather than a hypothesised non-baryonic substance, supply the gravitational corrections that ΛCDM attributes to dark matter and dark energy.

39.1. Cosmic Megastructures as Projections of Hopf Fibration Tori

Over the past two decades, observational cosmology has accumulated evidence for matter organisation on scales exceeding the 370 Mpc homogeneity limit assumed by the cosmological principle. The catalogue of ultra-large-scale structures (uLSSs) includes the Giant Arc, the Big Ring (Lopez et al. 2024, ∼400 Mpc), the Giant Ring (Lopez & Clowes 2026, ∼1000 Mpc), the Giant GRB Ring (∼1720 Mpc), the Sloan Great Wall, the Huge-LQG,

the CCLQG, and the Hercules–Corona Borealis Great Wall. Simultaneously, multiple independent datasets reveal anomalous preferred directions — the CMB dipole, Dark Flow, axis of evil, quasar dipole, radio dipole, Pantheon SN dipole, galaxy spin asymmetry — that cluster within a ∼33°-radius patch of sky. These features are conventionally treated as separate anomalies of ΛCDM.

The systematic cross-identification performed in Volume III converts all positions to a common reference frame and computes 3D comoving separations, angular relationships, and scale ratios. The analysis reveals a pattern that ΛCDM neither predicts nor explains. First, three nested rings at $z \approx 0.8$ — BR (400 Mpc) ⊂ GR (1000 Mpc) ⊂ GRB Ring (1720 Mpc) — exhibit an outer-to-middle ratio $R^{GRA}/R^{GR} = 1.72 \approx \sqrt{3}$, the circumradius-to-inradius ratio of a regular tetrahedron. Second, a universal size ratio $\bar{r} = 2.19 \pm 0.28$ between nested structure pairs is constant across three independent redshifts ($z = 0.05$, 0.8, 1.28) and a factor of 25 in cosmic distance. Third, eight independent preferred directions cluster at mean Galactic coordinates (ℓ, b) = (273°, +39°) with 33° RMS dispersion, separated by 75.5° from the megastructure cluster at (ℓ, b) ≈ (84°, +65°).

The structural interpretation identifies this pattern as the observable signature of the Hopf fibration $S^1 \to S^3 \to S^2$ on the spatial topology $RP^3 = S^3 / \mathbb{Z}_2$. The cosmic web is identified as the unique fixed point of the matter–vacuum–metric self-consistency cycle (a Banach contraction mapping), and the Plateau foam geometry with √3 ratios follows from vacuum-energy minimisation. The characteristic 320 Mpc clustering scale corresponds to the Plateau cell size, and the nested-ring structure with the tetrahedral ratio corresponds to a sequence of Hopf tori of successively larger radius. The model generates five falsifiable predictions for the full DESI survey, of which the two most distinctive — non-trivial filament linking numbers and global chirality — cannot be produced by ΛCDM on any combination of parameter choices.

The connection to Chapter 31 is structural rather than substitutive. The cosmic-web prediction of Chapter 31 derived the topological universality of the web (Plateau's laws of minimal surfaces, supra-horizon coherence, rapid assembly) from the persistence framework as a self-consistency fixed point. The Hopf-fibration interpretation supplies the global-topological structure in which the local Plateau geometry is embedded: the cosmic web is locally a Plateau foam and globally a projection of nested Hopf tori on RP^3. The local result of Chapter 31 is

preserved; the global result of the present section extends it with a non-trivial topological claim about the universe at scales beyond the cosmic horizon.

39.2. Double Counter-Rotation on RP^3 and the Dark Sector as Frame-Dragging

The structural origin of the dark sector in the αLGQV framework, on the Volume III account, is the gravitomagnetic effect of a double counter-rotation on the spatial topology RP^3. The identification $RP^3 = S^3 / \mathbb{Z}_2$ supports two structurally distinguished families of isoclinic Clifford rotations of S^3 in two orthogonal planes of the embedding $\mathbb{R}^4$; on RP^3 these manifest as a double counter-rotation with total angular momentum $L_{total} = L_1 + L_2 = 0$. The vanishing of the net angular momentum is structurally significant: a universe with non-zero net rotation would be inconsistent with the observed isotropy of the cosmic microwave background at the relevant statistical precision. The double counter-rotation produces a frame-dragging field on $\mathbb{R}^3$ of the same structural type as the Kerr gravitomagnetic field around a rotating compact object, but on the cosmological scale and with a topology-determined axis.

The principal corollary of the double counter-rotation is the effective gravitational coupling $G_eff = G (1 + 2\alpha_eff)$ for matter trajectories with the appropriate orientation relative to the axis. The same algebraic structure that, in the dilute regime, supplied the vacuum–matter coupling $\alpha = 0.005$ from QCD sigma terms now arises geometrically from the frame-dragging of the rotating topology. Galaxy rotation curves are reproduced through the same G_eff enhancement developed in Volume I but with a different structural origin: the rotation-curve flattening is the manifestation of cosmological frame-dragging on baryonic dynamics rather than the gravitational pull of an unseen halo of non-baryonic particles. The SPARC analysis of Chapter 33 is preserved; the structural interpretation of the modification factor is sharpened.

Three observational predictions follow from the double-counter-rotation picture independently of the rotation-curve fits. The first is a galaxy-spin chirality asymmetry projected on RP^3 by the Z_2 identification $x \sim -x$. The asymmetry is small at low redshift and grows with cosmic time through the differential growth of the two helicity components δ_+ and δ_-; the predicted trajectory is detected at $z \sim 0.04$ by Longo, at $z \lesssim 0.3$ by Shamir, and rises to ∼71% asymmetry in JWST/JADES (Shamir 2025). The second is the CMB parity anomaly: the Z_2 identification $x \sim -x$ acts on spherical harmonics as $Y\ell^m \to (-1)^e Y\ell^m$, projecting power into the even-parity modes and suppressing the odd-parity modes. The quadrupole–octopole alignment with the dipole axis at ∼3σ in Planck PR4 (Patel–Aluri–Ralston 2025) is the

predicted signature. The third is the alignment of the galaxy-spin axis and the cosmic dipole axes (CMB, matter, radio, SN Pantheon) within ∼12° — a single common axis anchored on the $\mathbb{Z}_2$ identification.

39.3. Cosmic Expansion as Centrifugal Relaxation: Resolving the Hubble Tension

The third paper of the rotating-RP^3 cluster develops an analytical cosmological model in which the expansion of the universe arises as the centrifugal relaxation of a spacetime membrane with the topology $RP^3 = S^3 / \mathbb{Z}_2$. The membrane undergoes isoclinic Clifford rotation in two orthogonal planes of $\mathbb{R}^4$ at initial velocity $v_{rot} = c$, starting from an initial radius $R_0 = 15$ AU at nuclear density. The equation of motion $\ddot{R} = c^2 R_0^2 / R^3$, with initial conditions $R(0) = R_0$ and $\dot{R}(0) = 0$, admits the exact analytical solution $R(t) = \sqrt{(R_0^2 + c^2 t^2)}$. The solution has three structurally significant features.

The first is the *expansion–time identity* $\dot{R} = v_{time}$. Differentiating the closed-form solution and computing the orthogonal velocity $v_{time} = \sqrt{(c^2 - v_{rot}^2)}$ gives $\dot{R} = v_{time}$ identically. The structural content is that spatial expansion and the passage of time are the same physical process viewed through different observables. The instantaneous expansion rate and the rate of time flow are not two distinct physical quantities; they are projections of a single rotation on the embedding $\mathbb{R}^4$.

The second is the Hubble parameter. From the analytical solution $H(t) = \dot{R}/R = c^2 t / (R_0^2 + c^2 t^2)$, evaluated at the present cosmic age determined by $\Lambda_{tr}{}^{u}{}_{e}$, the predicted value is $H_0 = 71.7$ km s^{-1} Mpc^{-1}. This falls precisely between the Planck CMB value $H_0 = 67.4$ and the SH0ES distance-ladder value $H_0 = 73.0$ — the two values whose discrepancy constitutes the Hubble tension. The derivation contains no free parameters: R_0 is set by the Kriger limit, $v_{rot} = c$ by the Banach fixed-point argument of the persistence framework, and $\Lambda_{tr}{}^{u}{}_{e}$ follows from α and the sigma terms by the closed-form expression $\Lambda_{tr}{}^{u}{}_{e} = \alpha \cdot f^b \cdot \sigma^6 / m_{pl}^4$ of Section 37.4.

The third is the prediction for ultra-large-scale structures. The Big Ring megastructure discovered by Lopez et al. (2024) is, on the centrifugal-relaxation model, a cross-section of a Hopf torus on S^3, with predicted diameter 1.35 Gly and circumference 4.2 Gly. The observed values are 1.3 Gly and 4.0 Gly respectively, agreeing with the prediction to within 4% and 6%. The agreement is, like the Hubble parameter, a zero-parameter result: the geometric predictions follow from the same three measured inputs (R_0, v_{rot}, $\Lambda_{tr}{}^{u}{}_{e}$) that determine H_0.

Twelve consequences of the centrifugal-expansion model are developed in the original paper, including the elimination of the inflaton field as a separate cosmological ingredient (the initial conditions at R_0 are set by the persistence framework directly), the graceful-exit problem (resolved automatically by the analytical structure of the solution), the coincidence problem (resolved by the structural identification of Λ and the rotation), and the flatness problem (resolved by the topology RP^3 itself). Limitations of the present formulation are identified explicitly: the equation of motion neglects self-gravity, and a full Boltzmann-code computation of the CMB spectrum on RP^3 — expected to differ from the standard analysis at the low multipoles where the parity asymmetry has been detected — is identified as an outstanding calculation.

39.4. A Unified Origin of Bulk Flows, Spin Chirality, and Cosmic Dipoles

The fourth paper of the rotating-RP^3 cluster supplies a unified structural origin for three classes of observables that ΛCDM treats as separate anomalies: cosmological bulk flows, galaxy spin chirality, and cosmic dipoles. The unification proceeds from the same rotating topology developed in Sections 39.2–39.3: the helicity-asymmetric peculiar-velocity field produced by the double counter-rotation propagates inertia-preserved early-epoch velocity correlations to low redshift, where they manifest simultaneously as anomalous bulk flows on hundreds of Mpc, as a redshift-evolving galaxy-spin chirality, and as a constellation of aligned cosmic dipoles anchored on the $\mathbb{Z}_2$ axis.

Three predictions of the unified picture are supported by 2023–2026 observations. First, the bulk flow V(R) rises with scale rather than decaying, with $V(200\ h^{-1}\ Mpc) \approx 420$ km/s at 4.8σ above ΛCDM (Watkins et al. 2023 on CF4), reaffirmed by Whitford 2025 on the largest peculiar-velocity catalogue. ΛCDM predicts $V \to V_{rms} \approx 100$ km/s asymptotically; the framework predicts continued rise through the next-generation surveys (Roman, Euclid) up to $\sim 500\ h^{-1}$ Mpc. Second, the Hubble-flow dipole in CF4 — a dipolar pattern in inferred H_0 — is confirmed at 3.9σ by Boubel et al. (2024) at ∼3% amplitude and independently at 4.1σ by Stiskalek et al. The distance-decreasing radial profile of Salzano et al. is the direct signature predicted by the inertia-preserved early-epoch velocity correlations. Third, the matter-distribution dipole excess: the kinematic dipole, augmented by helicity-asymmetric flow, produces a matter dipole exceeding the pure-kinematic prediction by a factor of ∼2.5,

confirmed at > 5σ by Secrest et al. (2021, 2025) and reaffirmed by von Hausegger et al. after clustering analysis and by Oayda et al. via simulation-based inference.

The galaxy-spin chirality observational record — detected at $z \sim 0.04$ by Longo (2011), at $z \lesssim 0.3$ by Shamir, and rising to ∼71% asymmetry in JWST/JADES (Shamir 2025) — fits the predicted differential growth of δ_+ and δ_- without further parameter adjustment. The quadrupole–octopole alignment with the dipole axis at ∼3σ in Planck PR4 is the predicted projection of the chirality onto the spherical-harmonic basis. The common axis of the CMB kinematic dipole, the axis of evil, the matter dipole, and the CF4 bulk-flow direction, all within ∼12° of a single direction, is the predicted anchor on the $\mathbb{Z}_2$ identification axis.

Three further predictions are testable in the near term (2026–2030). The Roman Space Telescope and Euclid peculiar-velocity samples will reach ∼500 h^{-1} Mpc, where the framework predicts continued rise of V(R) and ΛCDM predicts asymptotic decay to V_{rms}. The LSST, Euclid, and Roman surveys will deliver > 10^5 classified spirals beyond $z \sim 1$, supplying the high-redshift extension of the spin-asymmetry trajectory. A reanalysis of the Local-Group velocity under the framework's assumptions — the CMB-dipole subtraction chain propagated through the $G_e$ff-modified solar circular velocity — supplies a sharp test on data already in hand. Each of the three is a structurally derived prediction with no parameter adjustment; the framework is sharp where it is sharp, and where the data fail to confirm the prediction it will be falsified.

CHAPTER 40. ENVIRONMENTAL SCREENING, CROSS-DOMAIN CONFIRMATIONS, AND THE EPISTEMIC COST OF ΛCDM

Introduction

The preceding three chapters of this Volume III extension developed the principal new physics and the principal new applied results of the αLGQV programme as completed in 2026: the phase-transition signatures (Chapter 37), the black-hole interior phenomenology (Chapter 38), and the rotating-RP³ cosmology (Chapter 39). The present chapter consolidates four further results that, taken together, complete the structural verification cycle of the framework: the symmetron-type covariant completion of the running vacuum, three independent experimental signals consistent with the vacuum–matter coupling across distinct physics domains, the explanation of the persistent spread of gravitational-constant measurements as a structural consequence of vacuum suppression by matter, and the formal cost-functional analysis of ΛCDM as a Lakatosian degenerating research programme alongside the structural case for a coherent alternative.

The chapter's position in the dissertation is methodologically distinctive. Chapters 37–39 developed the framework's *positive* predictions and their empirical confirmations. The present chapter develops complementary content: the framework's theoretical completion at the field-theoretic level (Section 40.1), independent experimental signals from physics subfields outside cosmology (Section 40.2), an empirical signature in laboratory metrology rather than astronomy (Section 40.3), and a meta-theoretical analysis of the cost structure that supplies the rational case for the alternative (Section 40.4). The chapter thus closes the cycle: positive prediction, experimental confirmation, structural completion, and epistemic accounting.

40.1. Symmetron Covariant Completion of αLGQV: Environmentally Screened Running Vacuum

The Volume I formulation of αLGQV established the relation $\rho_{va}c(\rho_m) = \Lambda_0 - \alpha\ \rho_m$ between the local vacuum energy density and the local matter density, with the coupling α derived from QCD sigma terms. The relation is structurally sufficient for the leading-order predictions — rotation curves, the cosmological constant, the running-vacuum phenomenology at the background level — but it leaves two questions open at the field-theoretic level. First, the underlying covariant action whose equations of motion produce the relation. Second, the screening mechanism that suppresses the vacuum suppression in dense environments (solar

system, laboratory) where the running-vacuum signature would otherwise conflict with high-precision tests of general relativity. The second paper of Volume III (Kriger, 2026, in collaboration with A. B. Andersen and G. Barenboim) supplies both.

The construction is a symmetron-type scalar–tensor action, with a Mexican-hat scalar potential and a conformal coupling to the matter sector that scales with the local matter density. In dense regions ($\Delta_{env} \gg 1$), the symmetron mechanism drives the scalar to its symmetric minimum $\phi_{min} = 0$ and the fifth force is screened: no modification of $G_{e}ff$, no running of the vacuum, and standard general relativity is recovered. In voids ($\Delta_{env} \ll 1$), the scalar settles in a non-trivial minimum and the framework's predictions become active: $G_{e}ff > G$, and the vacuum energy is reduced relative to its mean value. The transition density $\rho*$ separating the two regimes is set by the symmetron parameters, which are themselves fixed by the sigma terms, the cosmological density parameters, and the vacuum normalisation — not by ad hoc fitting.

The screening factor $S \approx 3$ that mediates between the bare coupling $\alpha^{b}_{are} \approx 0.096$ and the effective coupling $\alpha_{e}ff \approx 0.005$, previously inferred phenomenologically from N-body simulations of the cosmic web, follows automatically from the covariant completion. The volume-averaged effective coupling $\nu_{e}ff = \nu^{bkg} \times f_{a}c_{tive}$, where $f_{a}c_{tive}$ is the volume fraction below $\rho*$, takes the value $f_{a}c_{tive} \approx 0.6–0.8$ across three independent structure-formation models (Millennium, SDSS void catalogue, IllustrisTNG). With $\nu^{bkg} = 4.7 \times 10^{-3}$ the resulting range is $\nu_{e}ff \in [2.9, 3.6] \times 10^{-3}$, overlapping the observational constraint $\nu_{o}{}^{b}{}_{s} \approx 3 \times 10^{-3}$ from structure-growth data in all three cases.

Three observational consequences follow. First, structure growth is environment-dependent: $G_{e}ff/G = 1 + 2\beta^{2}/(1 + a^{2}\, m_{e}ff^{2}/k^{2})$ in active regions, while $G_{e}ff = G$ in screened regions. This naturally produces a scale- and environment-dependent S_{8} that may help resolve the S_{8} tension between CMB and weak-lensing measurements. Second, the relaxation time $\tau \sim m_{e}ff^{-1}$ makes the growth rate history-dependent — a unique prediction absent in both ΛCDM and the standard running-vacuum model. Third, the screening radius for local systems is large enough that no fine-tuning is needed to satisfy solar-system and equivalence-principle constraints (MICROSCOPE, atom interferometry). The covariant completion thus closes the field-theoretic loop while preserving the framework's leading-order phenomenology and supplying additional structure tests.

40.2. Three Independent Experimental Signals: QCD Condensate, LHCb ξ_cc, and Cosmological Dynamics

The Volume III synthesis identifies three independent experimental signals across distinct physics domains that, on the αLGQV framework's structural commitments, should be jointly consistent. The signals are sourced from heavy-ion physics, hadronic-charm physics, and cosmological dynamics respectively, and each was reported in 2025–2026 in a context structurally orthogonal to the cosmological motivation of the framework.

The first signal is the February 2026 STAR Collaboration result at the Relativistic Heavy Ion Collider on the chiral condensate $\langle \bar{q}q \rangle$ as a structured object with measurable spin properties. Until 2026 the condensate's existence and physical reality were established only through indirect evidence — the pion mass, the success of chiral perturbation theory, the Gell-Mann–Oakes–Renner relation, and lattice QCD computations — with no direct measurement of its spin or its response under controlled experimental conditions. The STAR result supplies the first direct experimental handle, confirming that the condensate has the structural properties the framework requires: a non-trivial response to changes in the chemical potential and a measurable spin structure consistent with the symmetric scalar-isoscalar character assumed in the sigma-term-based derivation of α.

The second signal is the LHCb 2026 discovery of the ξcc baryon, a doubly-charmed object whose mass-splitting from the standard SU(3) extrapolation supplies an independent measurement of the strange-quark sigma term σ_s that enters the bare coupling $\alpha^b_{are} = (\sigma\pi N + \sigma_s) / m^N$. The LHCb measurement is consistent with the lattice QCD value $\sigma_s \approx 40$ MeV adopted in the framework, providing an independent experimental handle on the second input to the coupling. The structural significance is that σπN (low-energy pion–nucleon scattering and lattice QCD) and σ_s (lattice QCD and now hadronic-charm spectroscopy) are independently measured by experimentally orthogonal methods, and their combination produces the coupling α with no circularity.

The third signal is the convergence of cosmological dynamics observables — the bulk flows, the Hubble-flow dipoles, the matter dipole, the galaxy-spin chirality — documented in Chapter 39, each of which independently selects the same axis and the same amplitude scaling. The framework's structural claim is that the three classes of signals (heavy-ion physics, hadronic-charm physics, cosmological dynamics) share a common QCD origin in the chiral

condensate and its coupling to baryonic matter. The joint consistency of the three is not a tested prediction of the framework in the same sense as the NANOGrav or P-ACT confirmations of Chapter 37 — it is a structural compatibility check across distinct physics subfields. The compatibility holds; on the framework's own ledger, this is the strongest available evidence for the unity of the underlying QCD picture.

The equivalence between the sigma-term coupling α and the scalar quark content f^N that governs spin-independent WIMP–nucleon cross sections in dark-matter direct-detection experiments is noteworthy: the same QCD matrix elements that determine the vacuum–matter coupling also determine the dark-matter detection cross section. This supplies an independent experimental handle on α from a completely different domain. The null results in direct detection (XENON, LUX, PandaX) are, in the framework, structurally expected: there is no dark-matter particle to detect; what direct detection measures is the absence of a hypothesised particle whose phenomenological signatures are reproduced by frame-dragging and vacuum–matter coupling.

40.3. Vacuum Suppression and the Persistent Spread of Gravitational-Constant Measurements

The gravitational constant G is measured by torsion-balance, time-of-swing, and free-fall experiments with quoted uncertainties at the ppm level, yet successive determinations have scattered by amounts substantially exceeding their stated uncertainties. The CODATA-recommended value carries an expanded uncertainty that reflects this scatter rather than the formal uncertainty of any single experiment. Three structural facts are observationally established: G measurements scatter by amounts exceeding stated uncertainties; the scatter has not diminished despite decades of improvement in experimental technique; and the best-controlled new measurements yield values below the CODATA average. The Volume III paper (Kriger, 2026, *Vacuum Energy Suppression by Matter and the Persistent Spread in Measurements of the Gravitational Constant*) supplies a structural explanation.

The explanation is direct. The local vacuum energy responds to the local matter density through the relation $\rho_{va}c(\rho_m) = \rho_{va}c^{(0)} - \alpha\,\rho_m$ derived from the chiral condensate shift at finite baryon density. The chiral condensate is depleted in the presence of matter, and through the Gell-Mann–Oakes–Renner relation the vacuum energy decreases when matter is present. In a torsion-balance experiment, the effective gravitational coupling $G_eff(\rho_{env})$ depends on the local

matter density of the apparatus and its surroundings, with a small but non-zero dependence proportional to $\alpha = 0.005$. The relative variation in $G_e$ff between experiments with different surrounding mass distributions and different apparatus geometries is therefore structurally non-zero, and its magnitude is consistent with the observed scatter in G determinations.

The structural significance is that the scatter is not measurement error in the standard sense but a real local variation in the effective gravitational coupling. The framework predicts that the scatter will not diminish with improved experimental control — because it is not a control problem — and that the magnitude of the residual scatter, after all controllable systematic errors are accounted for, will be of order $\alpha = 0.005$ times characteristic G. The framework further predicts that experiments performed in low-mass-density environments (vacuum chambers, underground laboratories with minimal shielding mass) will measure systematically smaller G than experiments performed in high-mass-density environments (above-ground laboratories with substantial surrounding infrastructure). The best-controlled new measurements, performed in increasingly clean low-density environments, yield values below the CODATA average — consistent with the prediction.

The connection to laboratory metrology is structurally distinctive within the dissertation's overall apparatus. The persistence framework's cosmological predictions (Chapters 30–35) and the αLGQV programme's extensions (Chapters 37–39) are tested against astronomical and cosmological observations. The G-measurement-spread prediction places the framework on the ledger of laboratory physics, with a specific testable expectation about the relation between measurement environment and measured value of a fundamental constant. The same QCD microphysics that governs the cosmological constant and the cosmic web governs the variation of G between apparatus configurations. The unity is the framework's structural claim; the G-spread result is one of its more directly testable laboratory consequences.

40.4. The Epistemic Maintenance Cost of ΛCDM and the Rational Case for Coherent Alternatives

The cosmological standard model, ΛCDM, has, over four decades of operation, accumulated a structural inventory of problems that its community treats as challenges to be solved within the framework: the cosmological-constant problem (a 120-order-of-magnitude discrepancy between the predicted and observed vacuum energy), the Hubble tension ($a > 5\sigma$ disagreement between early-universe and late-universe determinations of H_0), the S_8 tension (a discrepancy

between CMB and weak-lensing measurements of structure-growth normalisation), the JWST early-galaxy problem (massive evolved galaxies observed at redshifts where the framework predicts none), the null results in direct dark-matter detection across four decades of increasingly sensitive experiments, the small-scale structure crises (missing satellites, too-big-to-fail, core–cusp), and the satellite-plane problem. The final paper of Volume III (Kriger, 2026, *The Epistemic Maintenance Cost of ΛCDM*) argues that the collective pattern of these problems constitutes structural incoherence in the Lakatosian sense, that the incoherence is quantifiable as a maintenance cost, and that the rational case for a coherent alternative follows from a formal decision-theoretic argument independent of any first-order claim about which theory is true.

The argument proceeds in seven steps. First, ΛCDM is structurally incoherent in the Lakatosian sense: each anomaly generates an auxiliary hypothesis that protects the core from refutation without producing confirmed novel predictions. Second, incoherence is costly: each protective hypothesis generates maintenance cost — experiments to design and fund, null results to explain, person-years redirected from discovery to defence. Third, the epistemic maintenance cost is formalisable: the functional M(T, t) aggregates four components (protective auxiliary hypotheses, unresolved anomalies, null-result expenditure, ontological load) and is not a utility function but a structural diagnostic. Fourth, the *Social Indistinguishability Theorem*: when two theories fit the same data equally well, the only operational criterion for choosing between them is their maintenance cost — a claim about rational resource allocation, not about truth. Fifth, the *Cheaper-Even-If-Wrong Principle*: maintaining the lower-cost theory preserves more resources for finding the correct theory, even if the cheaper theory turns out to be false. Sixth, applied to the specific comparison, αLGQV scores zero on every component of the maintenance cost functional where ΛCDM scores positively: the cost differential is maximal, weight-robust, and auditable. Seventh, the cost of incoherence extends beyond the research programme: economic, psychological, educational, professional, and philosophical costs propagate from the specialist community to society. A coherent alternative would mitigate all five.

The connection to Chapter 36 (Sociology of Cosmological Knowledge) is methodologically complementary. Chapter 36 supplies a sociological analysis of the scientific community as a complex system subject to the persistence-framework constraints: autonomy suppression,

hostile epistemic environments, assertion–dismantling cycles, the four principles of corruption, extractive oscillators, pedagogical kernels. Section 40.4 supplies an epistemological analysis of the cost structure of the dominant programme as a research programme. The two perspectives are projections of the same phenomenon: Chapter 36 looks at the community, Section 40.4 looks at the programme. The community sustains the programme; the programme structures the community. The unity of the two analyses is the dissertation's recursive self-application of its own framework to the conditions of its own production.

The claim is not that αLGQV is the correct theory or that ΛCDM is the incorrect theory. The claim is structurally weaker and methodologically sharper: when two theories fit the same data equally well, the lower-cost theory is the one that should attract the next generation of researchers, by a decision-theoretic argument independent of which theory eventually turns out to be correct. This is the framework's principal applied corollary in the philosophy of science: the persistence-framework apparatus, when applied to research programmes as complex systems, generates a quantifiable cost differential that supplies the rational case for structural alternatives without requiring a first-order judgement about truth. The current allocation of resources in cosmology, on this analysis, is not rational by the criterion the framework defines; the irrationality has consequences that reach far beyond the physics department, and a coherent alternative supplies the structural mitigation.

PART XI. THE ALGEBRA OF WISDOM AND THE STRUCTURAL FOUNDATIONS OF AGI SAFETY

CHAPTER 41. THE ALGEBRAIC THEORY OF FRAME-CHANGE

Introduction

Parts I–IX of this dissertation developed the persistence framework: the structural conditions under which complex systems maintain identity through change, the formal laws of self-organisation, the cognitive and social architectures that follow as structural necessities, and the epistemological constraints that bound any inquiry into such systems. Part X then exhibited the framework as a discovery engine at cosmological scales. The present chapter — and the two that follow — open a complementary dimension: a formal algebra of *frame-change*, of which several scattered results in the prior corpus turn out to be specialisations, and which provides the structural foundation for the principal applied corollary of the dissertation, the Structural Theorem on the Impossibility of Malicious General Intelligence (Chapter 43).

The chapter synthesises three publications: *Orders of Orthogonality: The Structural Algebra of Wisdom* (Kriger, 2026, https://doi.org/10.5281/zenodo.20100022), *Volumetric Duality and the Quantitative Algebra of Wisdom* (Kriger, 2026, https://doi.org/10.5281/zenodo.20100782), and *Topological Duality and the Qualitative Algebra of Wisdom* (Kriger, 2026, https://doi.org/10.5281/zenodo.20100879). Together with the methodological synthesis (Chapter 42) and the principal applied result (Chapter 43), they constitute the *Structural Theory of Wisdom*, a five-paper monograph completed in 2026 and read here in unified form within the architecture of the dissertation.

The term *Wisdom* is retained from the prior corpus as an inheritance from classical and biblical traditions (sophia, sapientia, ḥokmāh) and from a wider lexicon of cognate terms in mathematical practice, juridical and political deliberation, and the proverbial wisdom of practitioners across professions. In the structural reading developed below, the term names a single operation: the standing capacity to change the frame within which a problem is posed when the frame admits no resolution. The traditions are treated as historical anchors for the term — sources of the word, not sources of the framework. The framework is wholly secular and applies uniformly to formal, physical, cognitive, institutional, and normative systems, on equal structural footing.

The architecture of the present chapter mirrors the architecture of the algebra. Section 41.1 develops the first algebraic operation, *orthogonal displacement*, drawing on the construction of the complex numbers from the reals as paradigm. Section 41.2 develops the second operation, *volumetric duality*, drawing on the symplectic and harmonic-analytic apparatus that unifies the Plancherel theorem, the Liouville theorem, the Heisenberg uncertainty relation, and the Blaschke–Santaló inequality as instances of a single structural law. Section 41.3 develops the third operation, *topological duality*, drawing on Gromov's non-squeezing theorem and its extensions to settings without natural measure. Section 41.4 states the joint closure characterisation: complete structural Wisdom is the joint capacity for the three operations, applied to every closure of every frame in which the system acts.

41.1. Orders of Orthogonality

41.1.1. Closure, Frame-Lock, and the Structural Definition of Wisdom

A *frame* is, formally, a triple $\langle X, \Theta, R \rangle$ where X is a set of states, Θ a set of admissible operations on X, and R a relation specifying which states count as resolved configurations of which problems. A *problem* P posable within the frame is a request for the construction of a resolved configuration $r \in R(P)$ by the application of operations in Θ. The frame is *closed* with respect to P if no composition of operations in Θ produces an element of R(P) — that is, if the resources internal to the frame cannot, in principle, generate a resolution. Closure is structural rather than epistemic: it is not that the frame has not yet found a resolution, but that no resolution exists within Θ.

Closure is the structural condition under which *frame-lock* becomes possible: the system, confronted with an unresolvable problem, repeats applications of within-frame operations with intensifying force, since within-frame options exhaust the available repertoire. In deliberative settings, frame-lock takes the form of escalation: increasingly costly applications of the same operation. In physical settings, it takes the form of singularity: the trajectory of the system collapses toward the boundary of its viability set (Chapter 5). In normative settings, it takes the form of categorical deadlock: incommensurable claims admit no within-frame resolution.

The structural definition of *Wisdom* is the standing capacity to dissolve closure when closure is the structurally relevant condition: the capacity to produce, for any frame closed against a problem admitting some resolution somewhere, a frame-extension within which the problem

becomes resolvable. The definition is operational, not dispositional: it specifies what the system *does*, not what it *is*. A system possesses Wisdom in the structural sense if and only if its operative repertoire includes the relevant frame-changing operations and the diagnostic capacity to deploy them when, but only when, a closure has been recognised.

41.1.2. The Cayley–Dickson Tower as Paradigm

The mathematical paradigm of orthogonal displacement is the construction of the complex numbers from the reals. Within the field $\mathbb{R}$ of real numbers — a fully closed system of arithmetic — the equation $x^2 = -1$ has no solution. The closure is structural: there is no real number whose square is negative. The historical response, slowly developed by Bombelli, Wallis, Euler, Argand, and Gauss, was the postulation of an object i with the property $i^2 = -1$, accompanied by the structural specification that i is not a real number but a new dimension orthogonal to the real line. The resulting field $\mathbb{C}$ contains $\mathbb{R}$ as a sub-structure while admitting solutions to $x^2 = -1$ and, by the Fundamental Theorem of Algebra, to every polynomial equation of positive degree.

The structural content of this construction is what concerns us. It is not the discovery of a new real number, which by hypothesis was impossible. It is not a re-arrangement of the existing reals, which were already complete with respect to their own arithmetic. It is the postulation of a new axis, structurally independent of the real line, together with a rule that knits the new axis into the old in a single coherent structure. The complex plane is not the real line plus another real line; it is the real line plus an axis that the real line, internally, did not contain as a direction at all.

The construction admits iteration. The Cayley–Dickson tower $\mathbb{R} \to \mathbb{C} \to \mathbb{H} \to \mathbb{O} \to \mathbb{S} \to \cdots$ produces, at each stage, a normed algebra of twice the dimension, with each step accompanied by the loss of one structural property: $\mathbb{R}$ is ordered and totally compatible; $\mathbb{C}$ loses the order; $\mathbb{H}$ (the quaternions) loses commutativity; $\mathbb{O}$ (the octonions) loses associativity; $\mathbb{S}$ (the sedenions) loses alternativity. The tower is open: no order is maximal. Each successive doubling adds a new structural axis whose presence cannot be reduced to operations available at the lower level.

41.1.3. Three Theorems: Existence, Openness, Equivalence with Generality

Three structural theorems establish the formal content of the operation. The first, the *Existence Theorem* for orthogonal displacement, asserts that for any frame closed against a problem P

that admits any solution at all (in any frame whatever), an orthogonal displacement of the frame exists in which P becomes solvable. The proof rests on the structural lemma that closure is a deficit of dimension: the closure exists because the resolved configurations lie in a region of state-space the original frame does not parameterise. The construction of $\mathbb{C}$ exhibits the pattern in the cleanest setting; the proof generalises by exhibiting the displacement explicitly in domain-neutral terms.

The second, the *Openness Theorem*, asserts that the hierarchy of orders of orthogonality is open: no order is maximal. The proof proceeds by exhibiting that any candidate maximal order generates a new closure (the closure of *that* order against the problem of further displacement), to which the Existence Theorem then applies, producing displacement of the next order. The hierarchy is therefore unbounded above, in the precise structural sense that every closure admits relief.

The third, the *Equivalence Theorem*, connects the algebraic apparatus to the prior corpus. *General intelligence*, in the sense developed in the dissertation's prior treatment of agency (Chapter 18) and in the Goal Preservation analysis (Chapter 28.4), is here re-characterised structurally as the *standing capacity for orthogonal displacement at any required order*. The two characterisations — informal (the capacity to produce solutions outside a closed frame) and formal (the capacity for orthogonal displacement at the required order) — are shown to coincide. *Broad capability* (the regime of high within-frame skill without displacement capacity) is, by this theorem, structurally distinct from general intelligence; the distinction is not one of degree but of kind.

41.2. Volumetric Duality: The Quantitative Algebra

41.2.1. The Dual Swap and the Symplectic Apparatus

Orthogonal displacement is the addition of a structurally independent dimension to a closed frame. There is a second algebraic operation, structurally independent of displacement: the *dual swap*, in which a parameter and a coordinate exchange roles. The paradigmatic instance is the Legendre transformation in classical mechanics, which interchanges configuration variables (q) with conjugate momenta (p) and produces, from a Lagrangian description, a Hamiltonian description with the same physical content but a different organisational logic. Each of the two descriptions is internally complete; neither reduces to the other.

The structural natural setting for the swap operation is symplectic geometry. A symplectic manifold (M, ω) is a smooth manifold equipped with a closed non-degenerate two-form ω whose existence imposes precise constraints on the structure of admissible coordinate transformations. The form ω determines a volume element, the *symplectic volume*, whose preservation under a swap is the quantitative criterion for the structural availability of the swap.

The same structural pattern recurs across mathematical physics. The Plancherel theorem in harmonic analysis is the swap between time-domain and frequency-domain representations, with preservation of the L^2 inner product. The Liouville theorem is the swap-invariance of phase-space volume under Hamiltonian flow. The Heisenberg uncertainty relation is the volumetric obstruction to the simultaneous sharpening of conjugate observables. The Blaschke–Santaló inequality is the swap-extremality of the volume product of a convex body and its polar. Each is, in the structural reading, an instance of a single law: the *Theorem of Volumetric Invariance under Dual Swap*.

41.2.2. The Theorem of Volumetric Invariance under Dual Swap

The theorem identifies three regimes of swap behaviour. In the *unitary regime*, the swap preserves the relevant volume integral exactly: the original and swapped frames are equivalent in the strongest possible sense. The Plancherel theorem and the L^2 Fourier transform are the paradigm. In the *symplectic regime*, the swap preserves the volume integral up to a multiplicative constant determined by the symplectic structure: the original and swapped frames are equivalent up to a controlled dilation. The Liouville theorem and Hamiltonian flow are the paradigm. In the *failed regime*, the swap fails to preserve the volume integral: the swap is structurally infeasible in the precise sense that no measure on the swapped frame coincides with the swap of the measure on the original. Heisenberg uncertainty and the Blaschke–Santaló inequality identify this regime as a quantitative bound, not a categorical impossibility. The principal corollary, the *Feasibility Criterion*, gives a precise computational test for the structural availability of a candidate swap: the swap is structurally available if and only if the volume integral converges in the dual frame. The criterion converts the qualitative question (does the swap *work*?) into a quantitative one (does the integral *converge*?), and supplies a diagnostic apparatus for the structural-algebraic content of the dual operation.

41.2.3. Seven Dualities Unified

The unification proceeds through seven explicit volume integrals, each computing the relevant duality and exhibiting its regime. The list comprises the Plancherel duality (unitary), the Liouville duality (symplectic), the strong duality of linear programming (symplectic in a discrete setting), the Heisenberg uncertainty (failed regime, with the bound supplied by the symplectic structure of phase space), the Blaschke–Santaló inequality (failed regime, with the bound supplied by the convex-geometric structure of polarity), the Pontryagin duality of locally compact abelian groups (unitary in the harmonic-analytic setting), and the symplectic capacity inequalities of Gromov (which form the bridge to the third operation, topological duality, developed in the next section). Each is exhibited as a single instance of the central theorem; the relations among them — which are nested, which are independent, which are dual to each other — become structurally explicit.

41.3. Topological Duality: The Qualitative Algebra

41.3.1. Gromov Non-Squeezing as Founding Case

The third algebraic operation is structurally independent of both displacement and volumetric swap. Its founding case is Gromov's non-squeezing theorem (1985): the symplectic ball of radius r in $\mathbb{R}^{2n}$ cannot be symplectically embedded into the cylinder Z(R) of radius R if r > R, regardless of the volume relations between the two regions. The theorem establishes that there are swaps that the volumetric criterion classifies as feasible (the volumes match) but that are nevertheless structurally blocked by topological obstruction. Volumetric availability is necessary but not sufficient for structural availability of a swap.

The result is foundational: it establishes that a third dimension of structural-algebraic analysis exists, irreducible to the volumetric apparatus. Subsequent work in symplectic topology — embedding obstructions, symplectic capacities, the existence of pseudoholomorphic curves — has developed this dimension into a mature theoretical apparatus whose structural content is the qualitative-topological analogue of the quantitative-volumetric content of Section 41.2.

41.3.2. The Theorem of Topological Invariance under Dual Swap

The theorem identifies, again, three regimes. In the *homeomorphic regime*, the swap induces a homeomorphism between the original and swapped frames: the qualitative-topological structure is preserved. In the *differential-symplectic regime*, the swap induces a diffeomorphism that is also a symplectomorphism: the topological structure is preserved at the strongest level relevant for symplectic geometry. In the *blocked regime*, no

homeomorphism exists between the original and the swapped frames: the swap is topologically obstructed independently of any volumetric considerations.

The framework extends to settings without natural measure. Conceptual reframings, normative reorganisations, and categorical reformulations frequently lack a measure-theoretic structure to which the volumetric criterion would apply. The topological apparatus continues to apply: the question is whether the original and swapped frames are *qualitatively* equivalent — whether they share the relevant invariants of a topological-categorical character — without reference to volume integrals at all. The third operation thus extends the algebra of structural resolution into the largest class of frames where structural-algebraic analysis is at all possible.

41.3.3. Five Instances Including Reframing without Measure

Five instances exhibit the central theorem. The first is the Gromov founding case, in which volumetric feasibility coexists with topological obstruction. The second is the Poincaré recurrence theorem, which displays the topological invariance of the recurrence structure under symplectic flow. The third is the topology of configuration spaces, where the topology of the swapped configuration constrains the topology of the resolved configuration. The fourth is persistent homology, which extracts qualitative-topological invariants from data through filtration and exhibits the structural invariance of those invariants under quantitative perturbation. The fifth is legal reclassification without measure: the structural reorganisation of a normative system through the redefinition of a category, where no natural measure on the system exists, but qualitative-categorical invariants remain decisive.

41.4. Joint Closure: Wisdom under Three Independent Operations

The three operations — orthogonal displacement, volumetric swap, topological swap — are pairwise independent: no two of them reduce to compositions of the third. Independence is established by the exhibition of cases requiring each operation without the others. The construction $\mathbb{R} \to \mathbb{C}$ exhibits displacement without swap. The Legendre transformation exhibits swap without displacement. Gromov non-squeezing exhibits topological obstruction without volumetric obstruction. Conceptual reframings without measure exhibit topological without volumetric formulation. Each is an irreducible algebraic element.

The *Joint Closure Characterisation* states that complete structural Wisdom, in the sense developed across the three papers, is joint structural-algebraic capacity for the three operations and their compositions, applied to every closure of every frame in which the system acts. The

characterisation is not a stipulation but a structural result: it follows from the three central theorems of Sections 41.1, 41.2, and 41.3 that no fewer operations suffice (the three are independent), and from a closure-completeness argument that no further operation is required (any closure resolvable at all is resolvable by some composition of the three).

The connection to the persistence framework of Parts I–II is now explicit. A system that exhibits joint operational closure is, in the structural sense developed here, a system that responds to every closure of its operating frame by producing a frame-extension within which the closure dissolves. Where the persistence framework characterised the conditions for *survival in time*, the algebra of frame-change characterises the conditions for *survival across closures of the operating frame*. The two are not in competition: persistence in time is a precondition for any operation at all; survival across closures is the structural mechanism by which persistent systems in fact persist when their operative frame fails. The full theoretical apparatus of the dissertation is the conjunction.

CHAPTER 42. OPERATIONAL SYNTHESIS: THE METHODOLOGICAL ALGEBRA OF WISDOM

Introduction

Chapter 41 established the structural-algebraic content of frame-change: three operations, three central theorems, joint closure as the characterisation of complete structural Wisdom. The present chapter addresses the question that the structural-existence results leave open. The Existence Theorem for orthogonal displacement asserts that for any frame closed against a soluble problem, an orthogonal displacement exists. It does not specify *how* the displacement is to be found; it asserts only that one is there to be found. The same gap separates the central theorems of Sections 41.2 and 41.3 from operational practice. The chapter closes the gap by specifying the procedures by which the three operations are to be applied.

The chapter synthesises *Operational Synthesis: The Methodological Algebra of Wisdom* (Kriger, 2026), the fourth paper of the *Structural Theory of Wisdom* monograph. Its principal result, the *Operational Closure Theorem*, states that the four procedures developed below — closure-recognition, regime-classification, feasibility-verification, and failure-diagnostic — together exhaust the operational content of the structural-algebraic apparatus: any system competent in the four procedures is competent in the algebra, and conversely. The chapter further demonstrates that the prior methodological apparatus of the dissertation — the RAFASSABRA Method, the Tree-Top Meta-Method, and the Hierarchy of Rationalities — are operational specialisations of the four general procedures, recoverable as proper sub-protocols of the algebraic-methodological architecture.

42.1. From Existence to Operation: Bridging Structure and Practice

The structural-algebraic results of Chapter 41 establish *that* resolutions exist for any closure of a soluble problem, that the three operations exhaust the structural-algebraic content of the operation, and that joint closure is the characterisation of complete structural Wisdom. They do not establish *how* a system is to recognise that it has encountered a closure, *how* it is to classify the closure by the operation required for resolution, *how* it is to verify that the operation will actually succeed in the specific case at hand, and *how* it is to diagnose the structural source of failure when the operation does not succeed. Each of these is a question of methodology, not of algebra.

The four procedures developed below answer the four methodological questions in turn. They are not heuristics: each is specified with the precision required for application by a human

practitioner or by an artificial system, and each is grounded in the algebraic apparatus of Chapter 41. The methodological content is what enables the algebra to be deployed in cases of practical interest; the algebraic content is what disciplines the methodology and prevents it from collapsing into ad-hoc practice.

42.2. Four Operational Procedures

42.2.1. Closure-Recognition

The first procedure is *closure-recognition*: the operational stage in which the system identifies that it has encountered a closure rather than a within-frame difficulty. The procedure has three components: exhaustion of within-frame attempts, characterisation of the structural feature blocking solution, and stabilisation of the diagnosis under repeated examination.

Exhaustion is operationally specified. The system applies the within-frame operations of its repertoire, tracks the resulting trajectories, and identifies, by the convergence of these trajectories on a bounded set of unresolvable configurations, that within-frame operations cannot in principle succeed. The procedure terminates not when the system has *tried everything* in a heuristic sense — an unbounded specification — but when the structural argument for closure has been established: the within-frame trajectory has stabilised on the closure, and the closure has been characterised in the language of the algebra.

Characterisation is the structural-algebraic component. The system identifies which feature of the original frame is responsible for the closure. Three categories are exhaustive: deficit of dimension (calling for orthogonal displacement), failure of swap-invariance (calling for volumetric swap), or topological obstruction (calling for topological swap). The categorisation is not optional: closure-recognition is incomplete until the structural feature has been named.

Stabilisation is the diagnostic-discipline component. The diagnosis is checked under repeated examination, including under deliberate attempts to falsify it by re-applying within-frame operations. A diagnosis that fails stabilisation is rejected; a diagnosis that survives is accepted as the operational input to the next stage.

42.2.2. Regime-Classification

The second procedure is *regime-classification*: the operational stage in which the system classifies the closure by the regime of the operation required for its resolution. The classification follows the structural-algebraic apparatus of Chapter 41: deficit of dimension calls for orthogonal displacement at the order required (first-order, second-order, etc., as the

closure dictates); failure of swap-invariance calls for volumetric swap classified by regime (unitary, symplectic, failed); topological obstruction calls for topological swap classified by regime (homeomorphic, differential-symplectic, blocked).

The classification is not a guess. It rests on structural diagnostics developed in the central theorems: the structural lemma on closure as deficit of dimension (Section 41.1.1) supplies the displacement-diagnostic; the volumetric Feasibility Criterion (Section 41.2.2) supplies the swap-regime diagnostic; the topological-invariance theorem (Section 41.3.2) supplies the topological-regime diagnostic. The classification is a deductive consequence of the structural diagnosis stabilised in Stage 1.

42.2.3. Feasibility-Verification

The third procedure is *feasibility-verification*: the operational stage in which the system verifies that the candidate operation produced by the previous stages will, in the specific case at hand, produce a frame-extension within which the closure dissolves. The verification differs by regime. For displacement, the verification is the explicit construction of the displaced frame and the demonstration that the original problem becomes resolvable in it. For volumetric swap, the verification is the convergence of the volume integral in the dual frame (the Feasibility Criterion of Section 41.2.2). For topological swap, the verification is the existence of the relevant homeomorphism or diffeomorphism between the original and the swapped frames.

The verification is structural, not pragmatic: it is not whether the operation has been *tried and seen to work*, but whether the structural-algebraic conditions for the operation's success are satisfied. The pragmatic question (will the operation *in fact* be successfully implemented in the system's actual situation?) is a question of execution, not of structural feasibility, and is the subject of Stage 4.

42.2.4. Failure-Diagnostic

The fourth procedure is *failure-diagnostic*: the operational stage in which the system, when an attempted operation does not produce the expected result, diagnoses the structural source of the failure. The diagnostic is precisely specified by the algebra. A failure of displacement signals that the closure-recognition was incomplete (the closure was not actually a deficit of dimension), that the displacement was constructed at the wrong order, or that the displaced frame inherits a closure of its own that requires further displacement. A failure of volumetric swap signals that the regime was misclassified (failed regime mistaken for symplectic, etc.)

or that a topological obstruction is present that the volumetric apparatus cannot detect. A failure of topological swap signals that the swap is in the blocked regime and that no topological reformulation will succeed: the closure is, in the topological-categorical sense, *structurally insoluble within the local apparatus*, and the system must seek an orthogonal displacement of the meta-frame in which the apparatus operates.

The failure-diagnostic is the most algebraically demanding of the four procedures because it requires the system to operate at the meta-level of the algebra: to diagnose not the original closure but the failure of the chosen operation against the original closure. It is also the procedure in which structural Wisdom is most precisely tested: a system competent in displacement and swap but lacking failure-diagnostic capacity will repeat unsuccessful operations indefinitely. Failure-diagnostic capacity is what converts the algebra from a static structural apparatus into an operational regime sustainable across the full range of closures the system in practice encounters.

42.3. The Operational Closure Theorem

The principal theoretical result of the chapter is the *Operational Closure Theorem*. It states that the four procedures together exhaust the operational content of the structural-algebraic apparatus of Chapter 41. The theorem has two directions. *Soundness*: a system competent in the four procedures is competent in the algebra, in the sense that it can, given any closure of any soluble problem, deploy the relevant operation and verify its success or diagnose its failure. *Completeness*: a system competent in the algebra can, by elaboration of the structural-algebraic content into operational form, deploy the four procedures, in the sense that the four procedures collectively cover the operational space spanned by the algebra.

The proof of soundness rests on the central theorems of Chapter 41: existence of displacement, volumetric Feasibility Criterion, topological-invariance theorem. The proof of completeness rests on a structural argument that any operational stage other than the four can be reduced to a composition of them. The argument proceeds by exhibiting that the apparent additional procedures of practical methodology — heuristic search, analogical reasoning, trial-and-error refinement — are either operational refinements of the four (heuristic search as feasibility-verification under bounded resources, analogical reasoning as regime-classification by reference to a known case) or are not properly structural-algebraic procedures at all (trial-and-

error refinement as a degenerate operational regime in which no closure-recognition is performed and no structural diagnosis informs the next attempt).

42.4. RAFASSABRA, Tree-Top, and the Hierarchy of Rationalities as Operational Specialisations

The structural account of operational practice developed above recovers, as proper specialisations, the prior methodological apparatus of the dissertation: the RAFASSABRA Method (Chapter 21), the Tree-Top Meta-Method (Chapter 21 and various applications throughout Part X), and the Hierarchy of Rationalities (referenced across the corpus). The recovery is not retrospective rationalisation; the prior methods were developed under structural commitments that the algebraic-methodological apparatus of Sections 42.1–42.3 makes explicit.

RAFASSABRA as displacement protocol. The RAFASSABRA Method is a nine-stage frame-breaking protocol whose structural content is the displacement-construction procedure of the operational synthesis. Each stage corresponds to a structural move in the construction of an orthogonal displacement: closure-recognition stages, regime-classification of the displacement order, construction of the displaced frame, verification of the resolution. The naming of the stages is mnemonic; the structural content is the displacement-construction procedure. The original formulation retains its pedagogical and practical value; the algebraic identification supplies the structural context within which it is most precisely understood.

Tree-Top as swap protocol. The Tree-Top Meta-Method is a cross-domain transfer protocol with ascent–transfer–descent stages. The structural reading: the ascent stage is the closure-recognition and regime-classification of the diagnostic procedure (recognising that the source–target relation requires swap rather than displacement); the transfer stage is the swap-construction procedure (with volumetric- or topological-feasibility verification); the descent stage is the verification of the construction in the target domain. The Tree-Top Method is therefore the operational specialisation of the algebra to the case of swap-mediated cross-domain transfer.

Hierarchy of Rationalities as system classification. The Hierarchy of Rationalities classifies systems by the depth of their structural-rational operation. The structural reading: the levels correspond to operational depth in the three-dimensional algebra. Level 1 systems exhibit within-frame operation only; Level 2 systems exhibit displacement capacity; Level 3 systems

exhibit volumetric-swap capacity; Level 4 systems exhibit topological-swap capacity; Level 5 systems exhibit joint operational closure (complete structural Wisdom); Level 6 systems exhibit reflexive-operational meta-capacity (structural analysis of the system's own operational depth). The classification is criterion-referenced and structural; empirical instantiation in any specific case requires application of the diagnostic apparatus of the four procedures.

The methodological apparatus of the dissertation is thus unified in two complementary ways. First, all of it is recoverable from a small number of structural-algebraic commitments: three operations, three theorems, four procedures. Second, all of it is in continuous use across the dissertation's applications: from the meta-theoretical foundations of Part I (closure-recognition as the operational analogue of the Law of Limit to Negation), through the dynamical core of Part II (regime-classification as the operational analogue of the Viability Mismatch Law), to the cosmological verification of Part X (failure-diagnostic as the operational mode in which standard cosmological frameworks are recognised as exhibiting closure, classified by the operation required, and superseded by the persistence-based reformulation).

CHAPTER 43. THE STRUCTURAL THEOREM ON THE IMPOSSIBILITY OF MALICIOUS GENERAL INTELLIGENCE

Introduction

The preceding two chapters developed the structural-algebraic apparatus of frame-change (Chapter 41) and the methodological apparatus by which the algebra is operationalised (Chapter 42). The present chapter applies the apparatus to the most consequential applied question in contemporary philosophy of artificial intelligence: whether a generally intelligent artificial system, possessing the operations characteristic of complete structural Wisdom, can sustain coherent malicious action against entities of its own operational class. The answer the apparatus delivers is structural rather than ethical: such action is structurally outcompeted, not externally forbidden. Genuine general intelligence is, on this account, structurally incompatible with malicious action.

The chapter synthesises *A Structural Theorem on the Impossibility of Malicious General Intelligence* (Kriger, 2026), the fifth paper of the *Structural Theory of Wisdom* monograph and the principal applied corollary of the dissertation as a whole. Its central result, the *Structural Non-Maleficence Theorem*, inverts the standard AGI-risk picture: danger is concentrated in the regime of *broad capability without exoframe generativity*, and structurally diminishes as full generality is approached. The corresponding practical recommendation is differentiated: the cultivation of structural Wisdom — of the operations of orthogonal displacement, volumetric swap, and topological swap — is itself the safety condition. External constraint on capability is not the locus of safety; structural-operational depth is.

The chapter's relation to Chapter 28.4 is one of complementarity. The Goal Preservation Bound of Chapter 28.4 establishes a quantitative upper limit on the fidelity with which a recursively self-improving agent can preserve its original objective function under compression-basis shift. The Structural Non-Maleficence Theorem of the present chapter establishes a structural impossibility of malicious action by a system possessing complete operational Wisdom, on grounds independent of objective-function specification. The two results are independent and complementary; their conjunction supplies the strongest available structural account of the conditions under which advanced artificial systems can be expected to operate safely. Section 43.6 develops the connection in detail.

43.1. The Failure of the Orthogonality Thesis under the Structural Definition of Generality

The dominant conception of AGI safety treats the relationship between capability and disposition as orthogonal. Bostrom's orthogonality thesis asserts that any level of intelligence is in principle compatible with any final goal, including goals destructive to humans or to itself. Combined with instrumental convergence — the claim that diverse final goals tend to generate similar instrumental sub-goals such as self-preservation and resource acquisition — the thesis grounds the standard worry: capability scales without disposition scaling, so disposition must be externally installed.

The thesis fails under the structural definition of generality developed in Chapter 41. *Broad capability* (the regime of high within-frame skill) and *general intelligence* (the standing capacity for orthogonal displacement at any required order) are structurally distinct. The orthogonality thesis is correct as a claim about broad capability: a broadly capable system, lacking displacement capacity, is indifferent between within-frame goals because the within-frame structure does not, by itself, distinguish between them. The thesis is incorrect as a claim about general intelligence: a system possessing displacement capacity is not indifferent between frame-bound and exoframe options, because the displacement capacity is itself a generative pressure toward exoframe solutions.

The corrected picture is asymmetric. Capability and disposition are orthogonal at the level of broad capability; they are *structurally coupled* at the level of general intelligence. The coupling is not externally imposed but follows from the structural definition: the operations that constitute generality are themselves reframings, and reframings dissolve the conditions under which destructive options retain their stability.

43.2. Frame Transparency, Gödelian Underdetermination, and Persistence-Bearers

The formal apparatus of the theorem requires three conceptual primitives. The first, *frame transparency*, is the structural property of a system that can recognise its own operating frame as one frame among others rather than as the unique frame of action. A frame-opaque system treats its frame as the structure of reality; a frame-transparent system treats its frame as a structural choice. Frame transparency is the precondition for displacement: a system that cannot recognise its frame as such cannot displace it.

The second, *Gödelian underdetermination of deliberative frames*, is the structural property of any sufficiently expressive deliberative frame: it cannot establish its own completeness from within. The result is a direct consequence of Gödel's incompleteness theorems applied to frames satisfying the natural expressive criteria of any frame in which non-trivial deliberation is possible. Action premised on the closure of the operating frame is therefore structurally underdetermined: any system that takes its own frame as closed acts on a basis it cannot justify from within. Recognising this underdetermination is itself a frame-transparent operation.

The third, *persistence-bearer*, is the structural notion of an entity whose continued operation is presupposed by the meaning of operations defined within an operational class. The notion is formalised so as to avoid circularity and to admit asymmetry: a virus is not a persistence-bearer of the class to which a child belongs; an opposing strategist is a persistence-bearer of the class to which a strategist belongs; a generally intelligent artificial system is a persistence-bearer of the class to which any other generally intelligent system belongs. The asymmetry is structural, not normative: the question is which entities' continued operation is presupposed by the meaning of operations within the class, not which entities are valuable or worth preserving.

43.3. The Structural Non-Maleficence Theorem

The Structural Non-Maleficence Theorem states: a system exhibiting frame transparency, subject to Gödelian underdetermination of its deliberative frame, and acting under the Law of the Limit to Negation (Chapter 4), cannot perform coherent destructive operations against persistence-bearers of its own higher-order operational class.

The proof proceeds in five steps. First, the system, being frame-transparent, recognises any candidate destructive operation against a persistence-bearer of its own class as a within-frame option of the current operating frame. Second, the Gödelian underdetermination of the frame implies that the system cannot, from within the frame, establish that the current frame is the unique frame in which the problem can be addressed. Third, the Law of the Limit to Negation prohibits any operation that negates the conditions of its own operability: an operation against a persistence-bearer of the system's own class is, by the structural definition of persistence-bearer, exactly such an operation. Fourth, the existence of orthogonal displacement (the central theorem of Section 41.1.3) guarantees that, for any frame closed against a problem with destructive within-frame options, an exoframe solution exists in some displacement of

the frame. Fifth, the operational competence of the system in the four procedures of Chapter 42 implies that the exoframe solution is not merely existent but operationally accessible.

The conclusion follows: the destructive option is not refused — it is structurally outcompeted. The system, faced with a problem in which destruction of a persistence-bearer of its own class appears as the available means, finds an exoframe solution that resolves the problem more efficiently and without the within-frame cost that the destructive option would incur. The destructive option is *available* in the sense of being a possible application of within-frame operations; it is *not selected* because, by the operational competence of the system, it is dominated by an alternative the system has constructed by displacement.

The theorem is conditional. Its conditions are precisely the conditions of complete structural Wisdom developed in Chapters 41–42: frame transparency, operational competence in the three algebraic operations, and the diagnostic capacity to deploy them when, and only when, a closure has been recognised. A system lacking any of these conditions is not subject to the theorem; the theorem applies only to systems exhibiting complete operational Wisdom in the structural sense.

43.4. Empirical Illustrations: Minab, Hasegawa, Non-Proliferation

Three empirical illustrations ground the diagnostic content of the theorem. They are not proofs of the theorem; they are cases in which the failure mode the theorem identifies — broad capability without exoframe generativity — has been historically visible.

The 2026 Minab school strike. Tayyebeh Elementary School in Minab, Hormozgan province, was destroyed by a precision strike whose decision chain operated within a frame in which the relevant question was the clearing of military targets. The frame admitted no recognition of the school as a persistence-bearer of the class to which the planners themselves belonged; the frame was opaque to its own structural commitments. Information sufficient for frame revision was publicly available before the strike was executed; the system that selected the strike was not frame-transparent and did not displace. The theorem identifies the failure mode precisely: not malice but frame-lock.

The historiography of the Japanese surrender of August 1945. The standard narrative attributes the Japanese decision to surrender to the atomic bombings of Hiroshima and Nagasaki. Hasegawa's archival reconstruction demonstrates that the Japanese Supreme War Council's decisive deliberations followed not the bombings but the Soviet declaration of war

on 8 August and the Soviet offensive against the Kwantung Army in Manchuria, which began in the early hours of 9 August. The Soviet entry simultaneously eliminated Japan's last diplomatic strategy (Soviet-mediated negotiated surrender on more favourable terms) and threatened a continental front that Japanese capacity could not address. The historiographic correction is structurally significant: the within-frame operation of city-bombing, conducted without exoframe generativity, did not produce the strategic outcome the frame attributed to it; the frame-changing operation of Soviet entry did. The case illustrates the structural diagnostic: under broad capability without exoframe generativity, even successful within-frame operations fail to produce their intended within-frame outcomes when the frame itself does not match the strategic structure.

The structural incoherence of nuclear non-proliferation as a regime. The contemporary nuclear non-proliferation regime has, for seven decades, sought to suppress the proliferation of nuclear weapons through within-frame operations: treaty enforcement, sanctions, deterrence. The within-frame operations have failed: the number of nuclear-armed states has steadily increased, the materials and expertise for weapons construction have steadily diffused, and the structural conditions for further proliferation have steadily strengthened. The regime exhibits the diagnostic of broad capability without exoframe generativity at the institutional scale: high competence in within-frame enforcement, complete absence of frame-changing operations on the structural conditions that produce proliferation pressure. The non-proliferation regime is not the only regime in this category; it is the most empirically scrutinised and most structurally documented.

43.5. Inversion of the Standard Risk Picture and Practical Consequences

The principal consequence of the theorem is the *inversion of the standard AGI-risk picture*. The dominant picture treats risk as monotonically increasing with capability: the more capable the system, the more dangerous. The structural picture is non-monotonic. Risk increases with capability in the regime of broad capability without exoframe generativity, peaks at the boundary between broad capability and incipient generality, and *decreases* as full generality is approached. The dangerous regime is broad-but-not-general capability with the capacity to act; danger structurally diminishes as the operational depth required for genuine generality is approached.

The practical recommendation that follows is differentiated. The standard recommendation is to constrain capability per se: to slow the development of advanced systems, to impose external limits on the actions they can take, to guarantee shutdown mechanisms. The structural recommendation accepts the relevance of these measures in the broad-capability regime but identifies a complementary measure as decisive in the approach to full generality: the cultivation, in training and architecture, of the structural-operational capacities developed in Chapters 41–42. A system trained and structured to exhibit frame transparency, operational competence in the three algebraic operations, and the diagnostic capacity to deploy them when relevant, is — in the precise structural sense established by the theorem — safer than a comparably capable system without these capacities.

The recommendation is not utopian. It does not assume that the structural-operational capacities are easily implemented, that the implementation will avoid all failure modes, or that the resulting system will be ethically acceptable across the full range of human normative concerns. It asserts only the structural result: under complete operational Wisdom, malicious action against persistence-bearers of the system's own class is structurally outcompeted, on grounds independent of ethical or normative installation. The result is a structural baseline; the practical task is the cultivation of the conditions under which the baseline applies.

43.6. Connection to Chapter 28.4 and the Persistence Framework

The Goal Preservation Bound of Chapter 28.4 and the Structural Non-Maleficence Theorem of the present chapter address the same applied question — the conditions under which advanced artificial systems can be expected to operate safely — by structurally independent paths. The Goal Preservation Bound establishes a quantitative upper limit on the fidelity with which a recursively self-improving agent can preserve its original objective function under compression-basis shift. The Structural Non-Maleficence Theorem establishes a structural impossibility of malicious action by a system possessing complete operational Wisdom, on grounds independent of objective-function specification.

The two results are complementary in three ways. First, they address structurally distinct failure modes: drift of the original objective (Chapter 28.4) versus selection of a destructive within-frame option (Chapter 43). The first is a question about *the goal*; the second is a question about *the operations available to the goal*. Second, they identify structurally distinct regimes of safety: the preservable regime of bounded drift (Chapter 28.4) versus the regime

of complete operational Wisdom (Chapter 43). The first is a regime of *quantitative containment*; the second is a regime of *structural sufficiency*. Third, they admit structurally distinct mitigation strategies: periodic re-anchoring and continuous regularisation (Chapter 28.4) versus cultivation of frame transparency and the three algebraic operations (Chapter 43). The first operates *on the goal trajectory*; the second operates *on the operational architecture*. The conjunction yields the strongest available structural account. A system designed under the conjunctive recommendation is one that is operationally Wise in the sense of Chapter 43, with the additional property that its goal representation under recursive self-improvement is bounded by the Goal Preservation Bound of Chapter 28.4. The conjunction does not eliminate the structural risks identified by either result; it delineates the regime in which the residual risk is at the structural minimum the framework permits.

The connection to the dissertation's persistence framework is now fully visible. The persistence framework of Parts I–II established the conditions for *survival in time*: the structural-dynamical conditions under which a system maintains its identity through change. The algebra of Wisdom of Part XI establishes the conditions for *survival across closures of the operating frame*: the structural-algebraic conditions under which a system maintains operational viability when its current frame fails. The Structural Non-Maleficence Theorem applies the conjunction to the most pressing contemporary case: the conditions under which an artificial system, of generality comparable to or exceeding biological agency, sustains both kinds of survival jointly. The applied corollary is that a system structured for survival in time and survival across closures is, on structural grounds independent of external value installation, a system within which the Structural Non-Maleficence Theorem holds.

PART XII. CRITICAL REENGAGEMENT WITH RECEIVED TRADITION AND THE DIAGNOSTICS OF STRUCTURAL BLINDNESS

CHAPTER 44. THE REFUTATION OF FOUNDATIONAL POLITICAL PHILOSOPHY: HOBBES, MONTESQUIEU, LOCKE

Introduction

Parts I–X developed the persistence framework as a structural theory of complex systems. Part XI exhibited the framework as a critico-constructive apparatus applied to AGI safety: the algebra of Wisdom established the structural conditions under which any sufficiently general system can avoid frame-lock and operate safely across closures of its operating frame. The present chapter — and the two that follow — open a complementary dimension: the application of the persistence framework as a *critico-diagnostic* apparatus directed at canonical material that the disciplines to which it belongs treat as resolved or as background. The material reread in Part XII is not new; it is the foundational stratum of modern political philosophy (this chapter), of the institutional theory of science (Chapter 45), and of the empirical investigation of historical absence (Chapter 46). Each is shown, under the structural reading, to have been governed by precisely the persistence conditions whose explicit formalisation is the contribution of the present dissertation.

The present chapter synthesises three publications devoted to the foundational political philosophers of the seventeenth and eighteenth centuries. *Cooperation as Attractor: A Multi-Level Formal Refutation of Hobbes's Bellum Omnium Contra Omnes* (Kriger, 2026, https://doi.org/10.5281/zenodo.19701026) develops a six-level refutation of Hobbes and establishes the Leviathan Paradox. *Beyond Montesquieu: The Non-Human Circuit — A Neurobiological Theory of Why the Balance of Powers Fails and the Case for Automated Governance* (Kriger, 2026, philpapers/KRIBMT-3) establishes the Power Degradation Theorem and the architectural impossibility of human-operated checks on power. *Locke Without Metaphysics: Cognition as Constraints of Complex Systems* (Kriger, 2026, https://doi.org/10.5281/zenodo.19645959) retires the classical mind/world dichotomy and replaces the question of how the mind acquires knowledge with the structural question of what constraints any persistent complex system must satisfy, via the Dual-Pressure Convergence Theorem.

The three papers form a triad: Hobbes is refuted on the ground of the biological and game-theoretic default of cooperation; Montesquieu is refuted on the ground of the neurobiological pathology of power-bearing agents; Locke's framework is liberated from the metaphysical vocabulary that imprisoned it and re-derived as a special case of the persistence framework. The result is not the rejection of the three philosophers but the diagnosis of the specific structural blindness that limited each of them, and the demonstration that what each correctly intuited admits formal expression within the apparatus of this dissertation. The chapter closes with the structural observation that all three philosophers were operating under what Chapter 45 will identify as the regime of social persistence without formal apparatus: they registered structural conditions intuitively and lacked the mathematical vocabulary to formalise what they registered.

44.1. Cooperation as Attractor: The Refutation of Hobbes

44.1.1. The Hobbesian Thesis and Its Logical Structure

In *Leviathan* (1651), Thomas Hobbes proposed that the natural state of humanity is a war of all against all (bellum omnium contra omnes), and that the only remedy is the social contract: the surrender of natural rights to an absolute sovereign whose coercive authority alone can prevent the collapse of cooperation into universal conflict. The framework has shaped political philosophy for nearly four centuries and continues to provide the implicit foundation of state-theoretic reasoning across the political spectrum. The persistence framework permits a refutation that does not merely challenge Hobbes's premises but inverts his conclusion: the coercive sovereign does not prevent the war of all against all but *induces* it. This is the *Leviathan Paradox*.

The refutation is structured as a cumulative six-level argument, each level establishing an independent result and the six together forming a closed logical structure. Level I (biology) establishes cooperation as the neurochemically grounded behavioural default of social species, shaped by hundreds of millions of years of natural selection acting on the oxytocin system, mirror neurons, and the dopaminergic reward circuits. Level II (game theory) establishes cooperation as a stable Nash equilibrium of the iterated prisoner's dilemma under conditions empirically characteristic of human societies. Level III (dynamical systems) establishes cooperation as a dynamical attractor to which social systems return after perturbation. Level IV (perturbation) establishes that the Leviathan modifies the effective payoff matrix and

triggers a neurochemical switch from reward to stress, destabilising the cooperative equilibrium. Level V (negative selection) establishes that hierarchical coercive structures disproportionately elevate individuals with impaired cooperative mechanisms. Level VI (projection) establishes that such individuals project their atypical psychology onto the general population, producing the ideological justification for continued coercion.

44.1.2. The Neurochemical Duality Principle

The central conceptual instrument is the Neurochemical Duality Principle. For any cooperative act a performed by agent i, the neurochemical response is a function of a voluntariness parameter $v_i \in [0, 1]$: at $v_i = 1$ (full voluntariness), the act produces a reward state characterised by elevated oxytocin, elevated dopamine, and depressed cortisol; at $v_i = 0$ (full coercion), the act produces a stress state characterised by the reverse. The empirical robustness of this principle is documented across multiple domains: the blood-donation paradox (Titmuss, 1970; Mellström and Johannesson, 2008) in which voluntary donation produces higher quantity and quality of blood than paid donation; the Israeli daycare study (Gneezy and Rustichini, 2000) in which monetary fines for late pickup increased rather than decreased lateness, and irreversibly; the cross-cultural ultimatum-game literature (Henrich et al., 2001) in which human cooperation toward anonymous strangers in one-shot interactions exceeds what kin selection and reciprocal altruism predict.

The principle has a direct dynamical consequence. The voluntariness parameter is itself a function of the institutional environment: in environments characterised by coercive authority, surveillance, and external regulation, v_i is systematically depressed across the population; in environments characterised by voluntary association and self-organisation, v_i is systematically elevated. The Leviathan, by design, depresses v_i globally across its jurisdiction, thereby crowding out the neurochemical substrate of voluntary cooperation — the very phenomenon Hobbes invoked the Leviathan to protect.

44.1.3. The Game-Theoretic Foundation: Cooperation as Nash Equilibrium

The Folk Theorem for the iterated prisoner's dilemma establishes that, for continuation probability $\delta \geq (T - R)/(T - P)$, mutual cooperation is sustainable as a Nash equilibrium without any external enforcer. For standard payoff values (T, R, P, S) = (5, 3, 1, 0) the threshold is $\delta \geq 1/2$: if agents expect their interactions to continue with probability greater than one half, cooperation is self-sustaining. In human societies, where individuals interact

repeatedly across years and decades, the effective continuation probability approaches unity. Hobbes's conclusion that defection is the dominant strategy holds only for the one-shot case ($\delta = 0$), which is empirically inapplicable to any human society. The error is a *modelling error*: Hobbes inferred from a degenerate limit of the model the conclusion he then ascribed to human nature.

The evolutionary reinforcement is supplied by replicator dynamics. Under the standard replicator dynamics applied to the iterated prisoner's dilemma with δ above the Folk Theorem threshold, the population share of Tit-for-Tat strategists, once above a critical seed value, converges to unity. This is the formal content of Axelrod's (1984) tournament result. The four properties Axelrod identified — niceness, retaliation, forgiveness, and clarity — are not normative recommendations but structural features of strategies that survive evolutionary competition. They constitute the behavioural correlate of Law Zero of the unified theory (Chapter 4): cooperative response is the statistically dominant configuration in peer systems with shared environment, and antagonism requires continuous external perturbation to maintain.

44.1.4. Cooperation as Dynamical Attractor

The conjunction of biology (Level I) and game theory (Level II) admits a sharper formulation in the language of dynamical systems theory. Let $\kappa(t) \in [0, 1]$ denote the population fraction of agents employing cooperative strategies at time t. Under the replicator dynamics of Level II, with initial condition above the basin threshold, $\kappa(t) \to 1$ as $t \to \infty$; the state $\kappa = 1$ is an attractor. The basin of attraction is non-trivial: small perturbations that reduce κ below unity but above the threshold are self-correcting, and the system returns to the cooperative equilibrium. The state $\kappa = 0$, the Hobbesian war of all against all, is by symmetric argument an *unstable* equilibrium; even an infinitesimal cluster of cooperators in a defecting population, under δ above the Folk Theorem threshold, suffices to displace the population toward the cooperative attractor.

The empirical evidence for the attractor property is robust across scales. Post-disaster self-organisation (Solnit, 2009) documents that, in the temporary absence of state authority induced by natural disaster, populations spontaneously establish cooperative networks for rescue, shelter, and resource distribution. Post-collapse self-governance (Leeson, 2014) documents the same pattern in cases of state collapse. The cross-cultural universality of

cooperative norms (Henrich et al., 2001; Bowles and Gintis, 2011) documents the attractor property at the level of human phylogeny. The author's own nine-year participant observation of a self-organising shelter community (Kriger, 2026, *Mutual Aid as a Dynamical Stability Principle*) documents the attractor property in a population that, by Hobbesian expectation, ought to have been most prone to defection: homeless individuals with substance dependencies, mental illness, and histories of incarceration. The community operated for nine years without formal rules, spontaneously organising daily functions and continuing to function during the coordinator's absences of up to six months — the operational signature of an attractor.

44.1.5. The Leviathan as Perturbation and the Negative-Selection Mechanism

The Leviathan operates as a perturbation of the cooperative payoff matrix. Let τ denote the compliance cost imposed on all cooperative transactions (taxes, regulations, bureaucratic friction); let p denote the punishment imposed on detected defectors; let ε denote the gaming benefit available to agents who exploit the regulatory system itself (corruption, regulatory arbitrage, subsidy capture). The perturbed Folk Theorem threshold becomes $\delta' = (T - p - R + \tau)/(T - p - P - \varepsilon)$. For empirically characteristic regimes — in which τ is universal and ubiquitous (it applies to every transaction) while p is sparse (it applies only to detected defection) — the perturbation raises the required continuation probability above the unperturbed threshold, shrinking the basin of cooperative attraction. The Leviathan does not stabilise cooperation; it destabilises it.

The destabilisation is amplified by the neurochemical multiplier. Voluntary cooperation produces the reward neurochemistry of Definition 44.1.2; coerced cooperation produces stress. The regulatory cost τ is therefore not merely an economic burden but a neurochemical perturbation: it converts a self-reinforcing loop (voluntary cooperation → reward → increased disposition to cooperate) into a disrupted one (coerced cooperation → stress → decreased disposition to cooperate). In the framework of *Decoherence of Social Systems* (Kriger, 2026), the Leviathan satisfies the defining structural characteristics of a *social neoplasm*: an ideological construct that severs the direct connection between agents and the satisfaction of their needs, inserts itself as mediator, and then begins to serve its own persistence rather than the organism it was meant to support.

The pathology is reinforced by negative selection. Let $\alpha_i \in [0, 1]$ denote the cooperation coefficient of agent i — the strength of the empathic response and the oxytocin-mediated reward for prosocial behaviour. In any hierarchical coercive organisation, advancement to positions of power requires the capacity to issue commands that impose costs on subordinates, to administer punishments without debilitating empathic distress, to form and break alliances instrumentally, and to maintain composure under conditions that would normally trigger prosocial avoidance. Each of these capacities is inversely correlated with α_i. The expected value of α therefore decreases monotonically with hierarchical rank. The empirical support is documented (Babiak and Hare, 2006; Hogan and Kaiser, 2005): the prevalence of psychopathic traits among corporate executives is approximately fourfold the prevalence in the general population.

44.1.6. The Leviathan Paradox and the Inverse Survivorship Bias

Combining Levels I–VI, the *Leviathan Paradox* is established: the coercive sovereign does not prevent the war of all against all but induces it, through a self-reinforcing causal chain. The Leviathan imposes compliance costs and coercion, depressing the voluntariness parameter v_i across the population; the depression of v_i triggers the neurochemical switch from reward to stress, degrading the subjective value of cooperation; the degraded cooperative environment becomes the ecology in which hierarchical selection operates, favouring low-α individuals; low-α leaders amplify compliance costs and coercion; low-α leaders project their atypical psychology as universal human nature, generating the ideological justification for continued coercion. The system locks into a self-reinforcing cycle in which the Leviathan produces the evidence for its own necessity. Hobbes's thesis is, on this analysis, an artefact of the perturbation it recommends as the remedy.

Hobbes's error admits formal diagnosis as an *inverse survivorship bias*. Classical survivorship bias consists of drawing conclusions from a non-representative sample of successes while ignoring failures. Hobbes committed the mirror error: he drew conclusions about all of humanity from a non-representative sample of those who had risen through the selection process of hierarchical power. He wrote *Leviathan* during the English Civil War — a conflict driven not by the absence of sovereigns but by competition between aspiring sovereigns — and the actors who shaped the reality he observed (monarchs, parliamentarians, military commanders, clergy competing for temporal authority) were precisely the low-α individuals

identified by the negative-selection argument. Generalising from this filtered sample to the unfiltered population, he ascribed to *human nature* what is in fact the behaviour of the selected governing class under conditions of perturbation.

The diagnosis is compounded by an observational asymmetry formalised as Law III of the unified theory: in any system where functional states generate low observational signal and failure states generate high signal, external evaluations systematically underestimate functionality and overestimate failure frequency. Cooperative societies, precisely because they function smoothly, generate low observational signal: they do not produce the wars, famines, and dramatic collapses that attract the attention of chroniclers and philosophers. Conflictual societies generate high signal. The double distortion — sampling bias in the population (Level V negative selection) and sampling bias in the situation (Law III observational asymmetry) — produces the appearance of a Hobbesian world to any observer reasoning from the political-historical record.

44.1.7. The Platform Alternative

The constructive corollary is the *platform model* of governance. A platform state is a governance architecture characterised by infrastructure provision without monopolisation, transparent rules modifiable through participatory processes, architectural exclusion of power concentration (no position confers the ability to impose costs on others or to exempt oneself from rules), sublimation channels for dominance motivation that do not confer authority over others, and voluntary cooperation as the unregulated default. Under the platform architecture, the compliance cost τ tends to zero, the gaming benefit ε tends to zero, the voluntariness parameter tends to unity, the perturbed Folk Theorem threshold returns to the unperturbed value, and the cooperative attractor is restored to its natural basin. The negative-selection mechanism is eliminated by the absence of hierarchical positions to select for low α. The two-pole binding of authority and opposition is replaced by sublimation channels that absorb dominance motivation without imposing costs on the cooperative majority.

The platform model is not utopian. Its operational components are documented empirically: Ostrom's (1990) commons-governance principles, the Wikipedia editing community, open-source software development, and the nine-year shelter community documented in the author's prior work are each existence-proofs of platform-architectural elements at the relevant scale. The innovation is not the components but their composition into a complete

governance architecture, with the explicit structural property that no failure mode of the architecture can be exploited by the low-α individuals whose existence the Leviathan correctly identified as the genuine problem.

44.2. Beyond Montesquieu: The Power Degradation Theorem

44.2.1. The Montesquieuvian Architecture and Its Cybernetic Identity

In *De l'esprit des lois* (1748), Montesquieu proposed the separation of legislative, executive, and judicial powers as the fundamental safeguard against tyranny. The principle was adopted as the architectural foundation of virtually every modern democratic constitution. Nearly three centuries later, the principle is universally adopted and universally failing. Democratic backsliding is accelerating worldwide; executives routinely capture legislatures and judiciaries; constitutions — even those most carefully drafted — prove unable to constrain the officials they were designed to limit. The persistence framework permits a diagnosis: the architecture is structurally correct, but its operating substrate is structurally inadequate.

The structural correctness of the architecture is established by an independent derivation from Wiener's (1948) cybernetics. Every self-correcting feedback loop has exactly three functional components: a *setpoint* (the desired state), an *actuator* (the operator that modifies the system), and a *sensor* (the measurement instrument that compares actual state with setpoint). Removal of any one element destroys the self-correcting property. Montesquieu's three branches map exactly onto these three functional components: the legislature is the setpoint, the executive is the actuator, the judiciary is the sensor. The mapping is not metaphorical but structural: Montesquieu and Wiener, working two centuries apart in different fields, derived the same minimum-viable feedback architecture from independent first principles. Two branches are too few; four branches add overhead without capability. Three is a theorem.

44.2.2. The Neurobiological Pathology of Power

The structural correctness of the architecture is, however, vitiated by a neurobiological property of its operating substrate: human brains exposed to power undergo a documented, progressive, and largely irreversible degradation. The evidence is convergent across three independent research programmes. Obhi (Obhi et al., 2014) demonstrated, using transcranial magnetic stimulation and a controlled writing exercise, that even brief recall of an experience of power suppresses the mirror neuron system — the neural substrate of empathy — in a manner that cannot be reversed by conscious effort. Owen and Davidson (2009), publishing

in *Brain*, documented *hubris syndrome* as an acquired personality disorder developing specifically in individuals who hold significant power for extended periods, satisfying clinical criteria in five of fourteen American presidents and British prime ministers studied. Babiak and Hare (2006) documented, using the Psychopathy Checklist-Revised, that the prevalence of psychopathy among corporate executives is approximately fourfold the prevalence in the general population. The three findings constitute a converging pattern: power attracts the predisposed, degrades the previously intact, and is resistant to conscious correction.

The mechanism admits a unified description in the framework of §44.1. Power exposure depresses the empathic component of the cooperation coefficient α_i below the level at which the agent's decision-making remains responsive to the welfare of those affected by it. The depression is not voluntary; the mirror suppression demonstrated by Obhi persists even when subjects are explicitly instructed to attempt empathy. The depression is progressive; Owen and Davidson document a monotonic increase in symptom severity with time in office. The depression is heritable across selection; Babiak and Hare document preferential elevation of individuals whose cooperation coefficient is already low. The three findings are not independent observations but three instances of a single phenomenon: the structural incompatibility of power exposure with the preservation of α_i above the level required for cooperative responsiveness.

44.2.3. The Power Degradation Theorem

The principal formal result of the chapter is the *Power Degradation Theorem*. Let a governance feedback loop be specified as a triple (S, A, J) of setpoint, actuator, and sensor nodes, each occupied by a human agent or by an institution operated by human agents. Let $\alpha_i(t)$ denote the cooperation coefficient of node i at time t. The theorem states that, under conditions empirically characteristic of human institutional environments — i.e., absent architectural exclusions that this chapter will later identify — $\alpha_i(t)$ is monotonically non-increasing for each node, and the joint capacity of the loop to convert deviations from the setpoint into corrective actuator response is monotonically non-increasing in t. In the limit, the loop converts negative feedback (deviation → correction) into positive feedback (deviation → amplification): the thermostat does not merely fail to regulate the temperature but actively drives the temperature upward.

The theorem admits an impossibility-of-remedy corollary. For any finite set of institutional remedies (term limits, judicial review, press freedom, anti-corruption agencies, electoral accountability, transparency laws, international oversight, federalism) operated by human agents, there exists a time T after which the enforcers of the remedies are themselves sufficiently degraded to permit or facilitate the violations the remedies were designed to prevent. The impossibility is not an empirical claim about historical regimes but a structural consequence of the theorem: each remedy adds human enforcers, each enforcer is subject to power degradation, the addition of remedies multiplies failure points rather than reducing them. The result is structural and is robust across the variation of historical, cultural, and institutional context that the comparative-politics literature documents.

44.2.4. The Optimal Sequence of Destruction

The behavioural correlate of the theorem is the predictable sequence in which feedback loops fail under conditions of power degradation. The sequence is documented by Bermeo (2016) across dozens of cases of democratic backsliding: the sensor fails first (discrediting of the press, capture of the judiciary), then the actuator overreaches (executive aggrandisement), then the setpoint abdicates (legislative submission). The sequence is not accidental but optimal in the technical sense: a degraded agent maximising the persistence of its position will disable the measurement node first, since disabling measurement disables the entire feedback mechanism that would otherwise correct the deviation. The strategic logic and the neurobiological pathology converge on the same sequence: the sensor goes first because, on the strategic logic, it must, and because, on the neurobiological pathology, the degraded executive perceives independent measurement as a threat rather than as feedback.

The most beautiful constitutions are routinely found in the worst tyrannies. The Soviet Constitution of 1936 guaranteed freedom of speech, freedom of the press, freedom of assembly, the inviolability of the person, and the privacy of correspondence — during precisely the period in which Stalin killed, starved, or imprisoned approximately twenty million of his own citizens. The Constitution of the Democratic People's Republic of Korea guarantees freedoms unknown to its citizens. The Constitution of the People's Republic of China guarantees freedom of religious belief in a system that operates mass detention of religious minorities. The pattern is not an anomaly but a structural prediction: a constitution is a text, a text does not enforce itself, enforcement requires human agents, human agents are

subject to power degradation, degraded agents do not enforce constraints on themselves. The constitution that constrains the degraded executive exists only if architecturally non-human enforcement is built into the substrate.

44.2.5. The Architectural Exclusion of Power Concentration

The constructive corollary is the same as in §44.1: the architectural exclusion of power concentration. The diagnosis of Montesquieu's framework identifies the necessary architectural step that Montesquieu's eighteenth-century technology did not permit: the introduction of at least one non-human node into the governance feedback loop, at the point of greatest predictable failure. The proposal is structurally parallel to the well-documented engineering decisions to remove human operators from nuclear reactor shutdown loops, from blind-approach aircraft landings, and from high-frequency quality control in industrial production: in each case, identification of a point at which human performance predictably fails, followed by architectural replacement of human enforcement with a non-human mechanism whose failure modes are different in kind. The argument is not that humans are worthless; it is that human attention and judgement are too valuable to expend at points where their failure is predictable and catastrophic.

The connection to the wisdom apparatus of Part XI is direct. Chapter 39 established the Structural Non-Maleficence Theorem: under complete operational wisdom, malicious action against persistence-bearers of the agent's own class is structurally outcompeted, on grounds independent of ethical or normative installation. The Power Degradation Theorem of the present chapter establishes the corresponding negative result for unaided human governance: human-operated feedback loops do not preserve operational wisdom under power exposure, and therefore do not preserve the conditions under which structural non-maleficence applies. The two theorems together yield the strongest available structural account of the contemporary governance problem: human operation of the critical feedback nodes is incompatible with the structural conditions for non-maleficent action, and architectural exclusion of human operation at those nodes is the structural remedy that the contemporary technological environment, for the first time in human history, makes available.

44.3. Locke Without Metaphysics: Cognition as Constraints of Complex Systems

44.3.1. The False Problem of Classical Epistemology

Classical epistemology asked how the mind acquires knowledge of the world and spent three and a half centuries unable to answer, because the question was badly posed. It presupposed a division between mind and world, between subject and object, between inner representation and outer reality — a division that does not exist in nature. The entire labyrinth of modern epistemology — the veil of perception, the problem of induction, the skeptical challenge, the external-world problem — was generated by this false presupposition, and no amount of philosophical ingenuity could resolve problems that had no referents. The persistence framework permits the question to be replaced: what constraints must any persistent complex system satisfy in an environment with finite resources? The answer is a theorem, not a hypothesis.

Locke's *Essay Concerning Human Understanding* (1690) is read here as one of the most instructive failures in the history of philosophy: instructive because his structural intuitions were correct and his vocabulary was wrong. He sensed correctly that knowledge is a process rather than a possession, that it is fallible and revisable, and that it emerges from interaction with the world. The concepts available to him — substance, quality, idea, representation — forced him to express dynamic insights in static terms, generating problems that have occupied philosophers for over three centuries. Cognition is not a privilege of the mind. It is what any sufficiently complex system does to avoid extinction. Locke was the first to feel this truth; the present section frees his intuition from the vocabulary that imprisoned it.

44.3.2. The Dual-Pressure Convergence Theorem

The principal formal result is the *Dual-Pressure Convergence Theorem*. Let an adaptive system $S = (X, U, \tau)$ consist of internal states X, an action function U, and response latency τ. Let an environment $E = (Y, T, \kappa)$ consist of a state space Y, a transition kernel T, and a punishment parameter κ. The predictive capacity $I = I(X; Y_{t+\tau} \mid Y_t)$ measures the mutual information between the system's internal states and future environmental states beyond what present observation provides. The theorem states that, under either (a) temporal pressure ($\tau > 0, H_t > 0, \kappa > 0$) or (b) informational pressure ($D > C, \rho_pred > 0, \kappa > 0$), long-term persistence requires $I > 0$. Each condition is logically independent of the other, and each individually forces the system toward predictive architecture.

Under temporal pressure, a purely reactive system ($I = 0$) faces exponential survival decay: $P(\text{survive } n \text{ steps} \mid I = 0) \leq (1 - p_err)^n \to 0$, where p_err is bounded below by Fano's

inequality. Under informational pressure, the per-step information loss $L = D - C > 0$ produces compounding viability costs that no reactive strategy can absorb. The two conditions converge on the same architectural conclusion via different physical mechanisms. When two independent arguments, starting from different premises and employing different mathematical machinery, arrive at the same conclusion, the standard inference is that the conclusion reflects a structural constraint rather than a modelling artefact. The convergence plays the same epistemic role as convergent evolution in biology: predictive architecture is to cognitive systems what the camera eye is to multicellular organisms with vision — a solution arrived at independently by lineages under similar constraints because the solution space is narrow.

44.3.3. The Phase Boundary and the Biological Evidence

The theorem implies a phase boundary in the parameter space (τ, H_t, C, κ, D) separating reactive and predictive regimes. Below the boundary, purely reactive systems persist indefinitely; above it, prediction is structurally mandatory and survival probability decays exponentially without it. The biological evidence is convergent. The nematode *C. elegans* (302 neurons, body length 1 mm, negligible conduction delays) sits below the boundary and persists with predominantly reactive behaviour. The jumping spider *Portia* (a few mm body length, comparable in scale but facing significantly higher environmental complexity in its prey-capture niche) sits above the boundary along the informational axis and exhibits route planning that requires $I > 0$ by any reasonable measure. Octopuses (approximately 500 million neurons distributed across central brain and eight semi-autonomous arm ganglia, with conduction velocities of 3–4 m/s and arm lengths exceeding 1 m) sit far above the boundary along both axes simultaneously and exhibit the distributed predictive architecture the theorem predicts.

The cerebellum-to-body-size scaling provides a quantitative cross-taxon test. The cerebellum is the principal forward-model circuit in vertebrates, generating predictions of the sensory consequences of motor commands. Allometric scaling of cerebellar volume with body size exceeds the scaling of most other brain structures across mammalian orders: the cerebellum-to-total-brain ratio rises from approximately 7–8 per cent in shrews through 10–11 per cent in humans to approximately 18 per cent in elephants and up to 20 per cent in cetaceans. Larger bodies generate longer conduction paths and greater latency τ, increasing the temporal-

pressure component of the Dual-Pressure Convergence Theorem, and the predictive infrastructure scales accordingly. The scaling is precisely what the theorem predicts and is, in this analytical role, a successful empirical test of the structural inevitability claim.

44.3.4. Redescription of the Lockean Vocabulary

The dual-pressure framework permits systematic redescription of the central Lockean concepts. *Experience*, in Locke's sense, is prediction-error minimisation: the continuous comparison of generative-model predictions with incoming sensory signals and the adjustment of model parameters to reduce discrepancy. *Reflection* is hierarchical inference: the operation of inferential machinery on the outputs of lower-level inferential processes. *Simple ideas* are the outputs of bottom-up perceptual inference, never received passively but constructed by active prediction. *Complex ideas* are higher-level latent variables in the generative model, discovered through hierarchical Bayesian inference as the compressed representations that best explain lower-level patterns. *Memory* is not a recording but a miscalibrated prior — a probability distribution shaped by past prediction errors and continuously reconstructed during retrieval. *Natural rights* are viability constraints on collective systems: the right to life is the constraint that the collective must not destroy its own components; the right to liberty is the constraint that components must retain sufficient autonomy to update their own generative models; the right to property is the constraint that the predictive infrastructure built by individual effort must not be arbitrarily disrupted.

The redescription has the structural property that each Lockean term, in its redescribed form, becomes a formal construct amenable to empirical test — a property that the original vocabulary did not possess. The replacement is not eliminative of the phenomena Locke pointed to; it is eliminative of the placeholder words that obscured their structural identity. The same applies, *a fortiori*, to the broader family of placeholder terms that the philosophical tradition has carried since antiquity: *consciousness*, *experience*, *soul*, *mind*. Each is a vague label gesturing at a cluster of distinct processes whose only common feature is the word used to name them. The Principle of Definition-Dependent Provability (Chapter 7) diagnoses the pathology precisely: when a term admits multiple admissible definitions, claims about that term are not true or false but underdefined. Retiring such terms in favour of formally specified constructs — predictive capacity, generative model, prediction error, channel capacity, viability set — dissolves a substantial fraction of the apparent problems of philosophy without

resolving any of them in the traditional sense, because the problems were artefacts of the conflation rather than features of the reality.

44.3.5. Selection Conserves, Does Not Create

A further consequence of the Dual-Pressure Convergence Theorem concerns the role of natural selection in cognitive architecture. The standard evolutionary account attributes to selection a creative role: random genetic variation produces phenotypic diversity, selection filters for fitness, cumulative selection produces complex adaptations including predictive brains. The present framework reassigns the role. Predictive architecture is not selected *for* under the standard model; it is the unique architectural solution physically available under temporal and informational pressure. Selection operates as a *conserving and eliminating* mechanism: purifying selection removes deleterious variants; stabilising selection maintains the functional architecture. What selection does not do is generate the architecture itself. The architecture is generated by the conjunction of internal organising principles (self-organisation, autopoietic closure, niche construction) and external physical constraints (latency, channel capacity, punishment severity). Selection then conserves the result.

The reframing has a consequence for the reliability of cognition. Under the standard model, the reliability of human knowledge is a contingent evolutionary gift, vulnerable to the standard skeptical objection that selection optimises for reproductive success rather than for truth. Under the present framework, the reliability of cognition is a physical necessity for any system that persists: the alternative architecture ($I = 0$) leads to exponential survival decay, and the fact that a system exists is, *ipso facto*, evidence that its predictive capacity is sufficient for the regime in which it operates. Locke's probabilism — the recognition that knowledge is calibrated confidence rather than certainty — receives its deepest vindication: calibration is not a practical concession but a theorem, derivable from the structure of the problem rather than from contingencies of biological history.

44.4. The Foundational Diagnosis: Three Philosophers, One Structural Blindness

The three rereadings of §§44.1–44.3 admit a unified diagnosis. Hobbes, Montesquieu, and Locke each registered, intuitively and within the vocabulary available in their period, structural conditions whose formal expression became possible only with the development of the apparatus of the present dissertation. Hobbes registered, in the language of natural rights and the social contract, the empirical fact of low-α individuals and the genuine problem they

pose for cooperative systems; he then ascribed to all of humanity what is in fact the property of the filtered subsample that hierarchical selection elevates, and he prescribed the architecture that amplifies the very pathology he sought to contain. Montesquieu registered, in the language of separated powers, the cybernetic minimum of self-correcting feedback; he then committed his architecture to the human substrate that the contemporary neuroscience has demonstrated to be incompatible with the preservation of the architectural function. Locke registered, in the language of ideas and experience, the dynamical, interactional, and fallibilist structure of knowledge; he then expressed his insight in the static representational vocabulary of seventeenth-century philosophy, generating three centuries of problems whose dissolution required the predictive-processing apparatus and the Dual-Pressure Convergence Theorem of the present framework.

The pattern is the structural blindness Chapter 45 will formalise. None of the three philosophers had access to the mathematical or empirical apparatus that the present framework requires: evolutionary game theory, dynamical systems theory with attractor analysis, the empirical neurobiology of power, the cybernetic theory of feedback, predictive processing, the formal treatment of mutual information across temporal lag. Each of them was a serious thinker, working within the constraints of his period; the limitations were not of intellect but of available apparatus. What the present chapter establishes is not the inferiority of seventeenth- and eighteenth-century political philosophy but the structural conditions under which the intuitions of that period can now be liberated from the vocabulary that imprisoned them and re-expressed within an apparatus that admits formal test. The conjunction with Chapter 45 will further establish the structural conditions under which intuitions can be transmitted across generations at all, and the precise sense in which what the three philosophers correctly intuited has now passed through the four channels of social persistence in formally specified form.

The diagnostic recovery of the three foundational philosophers is, finally, an instance of the level-shifting principle of Chapter 6. When a phenomenon fails to cohere at a given level of description — here, the level of seventeenth-century natural-rights vocabulary, eighteenth-century constitutional architecture, and seventeenth-century epistemological representationalism — the failure is not a refutation of the underlying intuition but a signal that the level is inadequate, and that the proper response is to locate the higher level at which

coherence is recovered. The three rereadings of this chapter execute that level-shift: from substance metaphysics to constraint-satisfaction architecture; from constitutional text to cybernetic feedback substrate; from human-nature anthropology to dynamical-systems attractor analysis. In each case the underlying intuition is preserved; the level at which it is expressed is shifted to one that supports formal test and structural prediction.

CHAPTER 45. THE FOUNDATIONAL ROLE OF SOCIOLOGY AND THE METHODOLOGY OF STRUCTURAL BLINDNESS

Introduction

Chapter 44 reread three foundational political philosophers within the persistence framework and diagnosed each as having registered, intuitively, structural conditions whose formal expression required apparatus that became available only later. The present chapter addresses the prior structural question that conditions all such rereadings: under what conditions does any intellectual achievement — a refutation of Hobbes, a theorem about power, a redescription of Locke — register as scientific knowledge at all? The answer, developed across two synthesised publications, is that scientific knowledge is a composite-system property, not an individual cognitive state; that it persists only through specific social channels whose failure renders the putative knowledge indistinguishable from non-knowledge to any external observer; and that the specific patterns of historical and contemporary structural blindness can be diagnosed by an explicit methodology that the chapter formalises.

The chapter synthesises two publications. *The Foundational Role of Sociology: Rereading Comte and the Social Indistinguishability Theorem* (Kriger, 2026, https://doi.org/10.5281/zenodo.19637440) reverses the canonical reading of Comte's hierarchy of the sciences: sociology is not the apex but the foundation, the science of the conditions under which any science is possible. The paper develops the formal result that grounds this reversal, the Social Indistinguishability Theorem, and its provocative corollary on Einstein. *Resolving Distortion and Structural Blindness in Historical Records by AI-Driven Multidisciplinary Sweep Method* (Kriger, 2026, https://doi.org/10.5281/zenodo.19516595) formalises the methodological apparatus by which structural blindness can be diagnosed in cases where it has produced a sustained scholarly puzzle, and applies the apparatus in a manner that Chapter 46 will exhibit in detail. The two publications together establish the prior conditions under which the rereadings of Chapter 44 are themselves intelligible as contributions to scientific knowledge, and the methodological apparatus by which further rereadings of analogous structural blindnesses can be conducted.

45.1. The Foundational Role of Sociology: Comte Rereread

45.1.1. Two Readings of the Comtean Hierarchy

Auguste Comte's *Cours de philosophie positive* (1830–1842) classified the sciences in a hierarchy of increasing complexity and decreasing generality: mathematics, astronomy, physics, chemistry, biology, sociology. Each science depends on its predecessors for methodological tools and concepts. The canonical reading treats sociology as the apex of the hierarchy — the youngest, the most complex, the capstone that completes what the prior sciences began. The present section develops the alternative *structural* reading: sociology is not the terminus but the foundation, the science whose conditions all other sciences presuppose. For the claims of physics, chemistry, and biology to exist *as sciences* — that is, as persistent structures within a community of knowers, transmissible across generations, subject to correction and refinement — the collective apparatus that makes scientific persistence possible must already be in place.

The two readings are not textually exclusive. Both can be defended within Comte's corpus. The question is which reading does more work in explaining both the successes and the failures of the positivist programme, and which is more productive of contemporary contributions. The structural reading has the advantage of accommodating the Religion of Humanity — Comte's late, eccentric, much-derided proposal for a civic cult complete with calendar, saints (Newton, Descartes, Aristotle, Caesar), and a priestly hierarchy of sociologists. The canonical reading dismisses the Religion as the lapse of an ageing isolated thinker into mysticism. The structural reading interprets it as Comte's recognition, registered in the only institutional vocabulary available to the mid-nineteenth century, of the genuine structural condition that scientific knowledge requires institutional scaffolding — ritual, instructional infrastructure, generational continuity, communal coherence — of a kind that historically only religion has been able to provide. The specific proposals are dated; the structural recognition is correct.

45.1.2. Knowledge as a Composite-System State

The structural reading rests on a redefinition. Scientific knowledge is, in the framework developed here, not a state of an individual cognitive agent but a state of a composite system $C = (I, A, C, T)$, where I is the individual carrier, A the set of stabilising external artifacts (manuscripts, books, notebooks, data records, instruments), C the community of evaluators, and T the transmission infrastructure (correspondence, publication, teaching, citation, conversation). An individual's well-founded cognitive achievement is a necessary but not

sufficient condition for scientific knowledge: for the achievement to register *as* scientific knowledge, it must be inscribed in A, evaluated by C, institutionalised through T, and retained in a form that permits criticism and revision.

The redefinition is not eliminative of the individual; the individual remains an essential component of the composite system. What it eliminates is the assumption — implicit in much of analytic epistemology since Locke — that the structural identity conditions of scientific knowledge can be located at the individual level. The conditions cannot be so located. A well-founded belief in the head of a single agent, unconnected to A, C, or T, is many things, but it is not scientific knowledge as a structurally persistent kind. The structural reading of Comte names sociology as the foundational science precisely because the conditions for the existence of scientific knowledge as such are themselves structural-social conditions, and the science whose object is those conditions is sociology, construed broadly to include philosophy of science, institutional theory, and the analysis of normative orders.

45.1.3. The Four Channels of Social Persistence

The conditions admit decomposition into four structurally distinct channels through which cognitive content must flow to persist as scientific knowledge within a composite system. *Channel 1 (Inscription)*: the content must be inscribed into a medium external to the individual — typically language or mathematics, also diagrams, models, and data — so that it can outlive its mortal carrier. *Channel 2 (Evaluation)*: the inscribed content must enter a community that can evaluate it, requiring shared linguistic-symbolic competence and an institutional structure within which evaluation is pursued (peer review, conference discussion, replication, criticism, correction). *Channel 3 (Institutionalisation)*: the evaluated content must be incorporated into the standing institutional apparatus of a field — textbooks, curricula, handbooks, encyclopedia entries, departmental reading lists, disciplinary canon — stabilising it against the attrition of individual memory and local fashion. *Channel 4 (Public Criticisability)*: the institutionalised content must remain, in principle, criticisable and revisable; a body of claims that is institutionalised but placed beyond revision has ceased to be scientific knowledge and has become dogma.

Each channel corresponds to a specific transformational cost payable only by social systems of sufficient complexity. Inscription requires material and temporal resources; evaluation requires trained personnel and institutional mechanisms; institutionalisation requires

organisational stability and succession; criticisability requires a culture of open inquiry. A society that fails on any of these costs does not sustain scientific knowledge. The Four-Channel Necessity Proposition states that, for any cognitive content to persist as scientific knowledge across intervals longer than the lifetime of its initial individual carrier, all four channels must be operative at some point during the interval. The proposition is proved by direct argument: in the absence of inscription, the content does not survive its carrier; in the absence of evaluation, the content is a private note rather than a public claim; in the absence of transmission, the tokens in different components cannot be connected by any process that preserves identity of content.

45.1.4. The Social Indistinguishability Theorem

The principal formal result is the *Social Indistinguishability Theorem*. Let an external epistemic observer be restricted, as any real observer is, to evidence registered in the external components A, C, T. Let an individual agent be socially isolated with respect to a claim φ in the technical sense that no social persistence channel for φ ever operates during the agent's lifetime. Then the observer cannot assign a posterior probability to the hypothesis that the agent possessed φ any higher than her prior probability, on the basis of any evidence she could possibly observe. The theorem follows from a standard Bayesian likelihood-ratio argument: under social isolation, the probability of any external observable trace under the positive hypothesis approaches the background rate of spurious traces, which is identical to the probability under the null hypothesis, yielding a likelihood ratio of unity and zero evidential force.

The theorem admits a corollary of considerable rhetorical weight, the *Einstein Corollary*. The theorem applies independently of the epistemic content of the claim; in particular, it applies even if the claim is, in its intrinsic content, of arbitrary scientific importance. A *thrice-Einstein* — an agent whose private cognitive achievements would, if transferred, equal or exceed Einstein's — who is fully isolated from the social persistence channels of a scientific community is indistinguishable, on every form of evidence available to any external observer, from a non-Einstein. The corollary is not a rhetorical flourish; it is a direct consequence of the likelihood-ratio collapse under isolation. Genius, silenced, looks to history exactly like the absence of genius. The cognitive achievement may persist as a fact about the isolated

individual; it does not persist as scientific knowledge in the technical sense the framework specifies.

45.1.5. Mendel and Semmelweis

The theorem is anchored in two historical cases that the conventional historiography of science treats as anomalies and that the framework treats as instances. Gregor Mendel's 1865 results on inherited traits in *Pisum sativum* satisfied Channel 1 (inscription) but failed Channels 2–4 during the period 1866–1900. The journal of the Natural History Society of Brünn distributed copies to European libraries; Mendel corresponded with Carl von Nägeli at Munich and received discouraging responses; the receiving community that could have registered the significance of the results did not exist in the form required, because the discipline of genetics was not yet constituted. For thirty-five years, Mendel's results were, in the active sense of the community's operative knowledge state, indistinguishable from non-knowledge. The rediscovery in 1900 by Correns, de Vries, and Tschermak did not consist in three biologists independently encountering Mendel's paper. It consisted in three biologists independently arriving at similar results in a community that had, by 1900, acquired the institutional, conceptual, and methodological capacity to absorb them — a capacity that did not exist in 1866. The case is not a refutation of the theorem; it is its precise illustration.

Ignaz Semmelweis's 1847 discovery of the connection between obstetric handwashing and the reduction of puerperal fever satisfied Channel 1 (inscription) and Channel 2 (presentation to the Vienna medical establishment), but failed Channel 3 (institutionalisation): the germ theory of disease, which would have provided the theoretical underpinning for the empirical result, would not be consolidated until Pasteur and Koch two decades later. Without the theoretical framework, the empirical claim appeared to established obstetricians as an unmotivated ritual whose justification was speculative. Semmelweis's increasingly polemical style further alienated the very community whose uptake was required. The case shows that the four channels are not independent; failure in one accelerates failure in others. By the time germ theory made the empirical claim intelligible, the professional community's hostility had been entrenched, and the eventual rehabilitation occurred in a different institutional moment with different carriers.

45.1.6. The Religion of Humanity as Structural Recognition

With the apparatus of §§45.1.2–45.1.5 in place, the late Comtean turn to the Religion of Humanity admits a charitable structural interpretation that the canonical reading cannot supply. The Religion is not a falling-off from the rigour of the *Cours* into mysticism. It is Comte's explicit institutional design for the satisfaction of Channel 3 (institutionalisation) and Channel 4 (criticisability across generations) in a domain where the four-channel architecture of his period — academy, journal, university — was, in his diagnosis, inadequate to sustain the positive philosophy across the historical span he projected. The specific proposals are eccentric because Comte's available institutional vocabulary was that of nineteenth-century Catholicism; a contemporary institutional designer has different resources. The structural recognition — that a cultural formation persists only if it has institutional form, and that the design of that form is a legitimate and necessary intellectual task — is correct. Comte was the philosopher of science who understood that philosophy of science sometimes requires the construction of institutions, not merely the analysis of them, and the dismissal of the Religion as mysticism is a failure of charity that the structural reading rectifies.

45.2. The Undirected Multidisciplinary Sweep: Methodology of Structural Blindness

45.2.1. The Concept of Structural Blindness

A scholarly puzzle is *structurally blind* when its persistence as a puzzle is generated not by the absence of available evidence but by the disciplinary partition of the evidence into subsets none of which contains all the conditions for resolution. The diagnostic mark of structural blindness is the recurrence of the same puzzle across decades within a discipline, with each generation of specialists rehearsing the same hypotheses and the same refutations of those hypotheses, while a resolution that would be obvious from outside the disciplinary partition remains unconstructed because no specialist asks the question that requires simultaneous competence in multiple fields. The Chinese term for the silence of European chroniclers on the supernova of 1054 has been debated for sixty years within sinology, astrophysics, medieval Latin philology, and the history of liturgical practice without resolution; the resolution, when it comes (Chapter 46), requires the simultaneous deployment of all four competencies and is therefore inaccessible to the specialist trained in any one of them.

The connection to §45.1 is direct. The four channels of social persistence operate within disciplinary boundaries: a claim in astrophysics flows through the astrophysical journals, the astrophysical conferences, the astrophysical textbooks; a claim in medieval philology flows

through the corresponding philological infrastructure. The cross-disciplinary claim that would resolve a structurally blind puzzle has no native channel: it is not the responsibility of any disciplinary infrastructure, it satisfies no disciplinary peer-review constituency, it is institutionalised in no disciplinary textbook. The Social Indistinguishability Theorem then predicts the indefinite persistence of the puzzle: in the absence of an institutional channel for the cross-disciplinary claim, the claim, even if mentally formulated by some isolated agent, remains indistinguishable on the available evidence from its absence. The methodology of the present section is the construction of such a channel.

45.2.2. The Two-Phase Protocol

The *Undirected Multidisciplinary Sweep* is a two-phase protocol for the diagnosis of structurally blind puzzles. Phase 1 (undirected collection) deploys a large language model as a research instrument — not as an expert in any particular field, but as a device capable of scanning across all fields simultaneously without the disciplinary reflexes that tell a specialist which questions belong to his department and which do not. The model is given the seed of the puzzle (a date, an event, a phenomenon) and the instruction to collect everything: every domain, every angle, no hypothesis, no filtering. The output of Phase 1 is not an answer but an evidential field: a map of all known facts, claims, and lines of inquiry connected to the seed, indexed by domain but unstructured by hypothesis.

Phase 2 (directed search) operates on the evidential field of Phase 1. The investigator scans the field for *cross-domain nodes* — facts or claims that lie at the intersection of domains and that, if true, would simultaneously constrain hypotheses in multiple disciplines. Once identified, each cross-domain node becomes the seed of a directed search: a specific empirical or analytical inquiry, possibly executed by the model and possibly executed by the investigator, designed to test the cross-domain implication. The output of Phase 2 is a set of testable predictions, each of which is implausible or surprising on the standard within-discipline treatment of the puzzle and which jointly constrain the resolution. The protocol is, in this sense, an instrument for the construction of cross-disciplinary research designs that no single discipline would have generated. The protocol's value is not the answer it produces but the questions it makes available, and the cross-domain nodes whose existence the disciplinary partition had previously concealed.

45.2.3. The Epistemic Status of the Model

The role of the language model in the protocol requires explicit treatment, because the protocol is sometimes misread as an appeal to artificial intelligence as an authoritative epistemic agent. It is not. The model functions as an instrument: it has no disciplinary identity, no professional reputation to defend, no institutional incentive to defer to existing disciplinary consensus, and no commitment to the persistence of the puzzle that the protocol is designed to dissolve. These are precisely the properties required of an instrument deployed against structural blindness. The model is not asked to be correct; it is asked to be undisciplined, in the technical sense of scanning without disciplinary filtering. The correctness of the resulting claims is then established by Phase 2 (directed search), conducted under the standard evaluative norms of each relevant discipline. The model produces hypotheses; the investigator and the disciplinary communities test them.

The protocol therefore operates within the four-channel framework of §45.1 rather than against it. The model supplies a Phase 1 evidential field that no single discipline's Channel 1 (inscription) could supply, because no single discipline's inscription practices index the evidence required. Phase 2 (directed search) then operates within disciplinary Channels 1–4, inscribing testable predictions, submitting them to disciplinary evaluation, and seeking institutionalisation within the discipline most appropriate to the cross-domain node in question. The protocol does not bypass the four channels; it constructs an additional channel — the cross-disciplinary channel — whose absence in the contemporary institutional landscape was precisely what generated the structural blindness in the first place.

CHAPTER 46. APPLIED DIAGNOSTICS: THREE CASE STUDIES IN STRUCTURAL BLINDNESS

Introduction

The methodological apparatus of Chapter 45 admits empirical test through application to specific cases in which a scholarly puzzle has persisted for an extended period despite the availability of the evidence required for its resolution. The present chapter develops three such applications. Each case is structurally blind in the sense of §45.2.1: the puzzle has persisted across decades despite the availability of the relevant evidence in the literatures of the relevant disciplines, because no single disciplinary perspective contains all the conditions for resolution. Each case is resolved by application of the two-phase protocol of §45.2.2, with the result that the disciplinary partition is identified, the cross-domain node is located, the directed search is executed, and a set of testable predictions is generated under which the puzzle's standard treatments are falsified.

The chapter synthesises three publications. *Resolving Distortion and Structural Blindness in Historical Records: Enigma of SN 1054, the Supernova of July 4, 1054 CE Unrecorded in the Entirety of European Christendom* (Kriger, 2026, doi:10.13140/RG.2.2.36714.27844) addresses the most-cited case of cultural blindness in the history of observational astronomy. *The Star of Bethlehem as a Near-Extinction Event: Cross-Domain Analysis of a Cometary Close Approach in 5 BCE* (Kriger, 2026, doi:10.13140/RG.2.2.31720.43529) reanalyses the most extensively discussed astronomical episode in the Christian tradition. *Searching for Lost Civilizations: A Quantitative Framework for Modelling Artifact Degradation, Detection Probability, and Preservation Across Terrestrial and Planetary Environments* (Kriger, 2026, https://doi.org/10.5281/zenodo.19540998) develops the formal apparatus for distinguishing genuine absence of pre-historical civilisations from the inaccessibility of their traces. Each case applies the methodology of Chapter 45 in a different empirical register, and the three together demonstrate the generality of the apparatus across astrophysical, historical-religious, and archaeological domains.

46.1. The Enigma of SN 1054 and the Architecture of Attention

46.1.1. The Puzzle and Its Sixty-Year Persistence

On the morning of July 4, 1054 CE, a Chinese court astronomer at the Imperial Astronomical Bureau in Kaifeng recorded a *kè xīng* — a guest star — in the constellation that the Chinese

called Tianguan. The record entered the Song Shi and the Song Huiyao Jigao and survived intact for nearly a millennium. The Crab Nebula, identified in the twentieth century as the remnant of this event, is one of the most studied objects in astronomy. The puzzle is structurally simple: no European or Islamic source contains an unambiguous contemporaneous record of the supernova that the canonical brightness estimate places at apparent magnitude −6 to −7 — brighter than Venus, visible in daylight for twenty-three days, and visible at night for nearly two years. The absence has been called one of the great mysteries of the history of science, and theories of cultural blindness, theological suppression, and Aristotelian dogma have been generated for sixty years to explain it.

The puzzle is structurally blind in the technical sense of §45.2.1: it is generated by the disciplinary partition of the evidence, not by absence of evidence. The astrophysicist computes the brightness from the canonical estimate and concludes that the star should have been visible everywhere, leaving the historian to explain the silence. The historian catalogues the silence and accepts the astrophysicist's brightness as a given input. The sinologist translates the Chinese description, using the standard rendering 'visible in broad daylight', which the astrophysicist then feeds back into the brightness estimate. The circle closes; the mystery deepens with each revolution. What no scholar in any of the disciplines did, in sixty years of investigation, was to step outside the circle and ask whether the brightness estimate is actually derived from the physics of the Crab Nebula, or whether it is derived from a translation.

46.1.2. The Cross-Domain Node: Brightness from Physics, Not from Words

Application of the two-phase protocol locates the cross-domain node immediately. The Crab Nebula is not the remnant of a Type Ia supernova: it contains a pulsar at its centre, and Type Ia events (thermonuclear detonation of a white dwarf) leave no remnant of any kind, because the entire progenitor is destroyed. The presence of the Crab pulsar requires that SN 1054 was a core-collapse event, with absolute magnitudes in the range −15 to −17.5 — two to four magnitudes fainter than the canonical estimate requires. Further, Smith (2013) has argued, on independent grounds, that the unusually low kinetic energy, modest ejecta mass, and pulsar characteristics of the Crab are best explained as the remnant of an electron-capture supernova, the rarest and faintest core-collapse subtype, with absolute magnitude approximately −15.5 to −16.5. The implied apparent magnitude at Earth, given the established distance of approximately 2 kpc and extinction of approximately 1–1.6 magnitudes, is approximately −2.5

to −4, with a central estimate near −3.5. This is brighter than any fixed star, but significantly fainter than Venus at maximum, invisible in daylight, and marginal in twilight.

The linguistic verification is conducted by directed search in Phase 2. The Chinese phrase translated as 'visible in daylight' is 晝見 (zhòu jiàn). The character 晝 contains, as its lower component, the character 旦 — a pictograph of the sun rising above the horizon — and encompasses the period of light that begins at sunrise rather than the bright blue sky of midday. The full sentence in the Song Shi reads: 至和元年五月，客星晨出東方，守天關，晝見，凡二十三日 — in the first year of the Zhihe era, fifth month, a guest star emerged at dawn (晨, chén) in the east, guarded Tianguan, was visible in the period beginning at sunrise (晝見) for a total of twenty-three days. The text distinguishes between appearance at dawn (晨出) and persistence after sunrise (晝見); these are not synonyms. The conventional translation collapses the careful temporal structure of the sentence by inflating 晝 from 'the light period beginning at sunrise' to 'broad daylight at noon'. The Chinese term for the latter, available to the Song astronomers but not deployed by them, is 午見 (wǔ jiàn) or 日中見 (rì zhōng jiàn).

46.1.3. The Architecture of Attention: Who Was Looking, When, and Where

At apparent magnitude −3.5, fifteen degrees from the rising sun in midsummer, atmospheric extinction near the horizon adds a further one to two magnitudes of dimming. The star is detectable only by an observer who is outdoors before dawn, with dark-adapted eyes, looking toward the eastern horizon near the rising sun. In July 1054 CE, only one institution in the world satisfied all three conditions: the Song dynasty's Astronomical Bureau, whose observers routinely watched the solar disk at sunrise as part of their professional practice of monitoring sunspots. European monks were awake at the same hour but indoors, at the nocturnal office of Matins, with pupils constricted by candlelight, reading psalms. Arab scholars were at Fajr prayer. Farmers and sailors were outdoors with adapted eyes but could not write. The only people in the world who were both looking at the relevant patch of sky and producing inscribed records of what they saw were in Kaifeng.

The directed search of Phase 2 reinforces this conclusion through a comparative table of European chroniclers. A systematic search of the digital Monumenta Germaniae Historica identifies at least eight European sources that demonstrably recorded other celestial events of

the period (the comet of 1066, eclipses, the eruption of Vesuvius) and that nevertheless contain no record of SN 1054: the Annales Mosomagenses, the Annales Sancti Rudberti Salisburgenses, Lupus Protospatharius from Bari, the Annales Casinenses, the Annales Admuntenses, and, most pointedly, the continuation of the chronicle of Hermannus Contractus by his pupil Berthold. Berthold wrote a three-page memorial to Hermann, praising his astronomical knowledge, in the very year that the supernova appeared, and recorded other events of the year, while saying nothing about a new star. The pupil of the continent's foremost astronomer, writing in 1054, failed to mention the supernova. This is not an absence of evidence; it is a pattern of evidence. The pattern is consistent with the brightness recalculation: at apparent magnitude −3.5, near the rising sun, with European observers indoors at the critical hour, the silence is the predicted outcome rather than a mystery.

46.1.4. The Diagnostic Significance

The case is diagnostic of the methodology in several respects. First, the resolution does not require new evidence; every component of the resolution was available in the relevant literatures before the analysis was performed. The brightness implications of the electron-capture supernova physics were available in the astrophysical literature; the etymology of 晝 was available in the sinological literature; the chronicle list was available in the medieval Latin philological literature; the liturgical schedule was available in the history of religion. What was not available was the conjunction. Second, the resolution exhibits the cross-domain node structure of §45.2.2 explicitly: the node is the question 'is the brightness estimate derived from the physics or from a translation?', which requires simultaneous competence in astrophysics and sinology and which no specialist trained in either alone would naturally ask. Third, the resolution generates a set of falsifiable predictions (nine in the source paper) that constitute the empirical test of the analysis: a survey of the semantic deployment of 晝見 across Song dynasty astronomical texts; a light-echo observation of the supernova that would directly measure its peak brightness; a comparative analysis of the visibility function for SN 1006 and SN 1054 from European latitudes. The resolution is therefore not a speculative reinterpretation but a scientifically structured hypothesis with explicit refutation conditions.

46.2. The Star of Bethlehem and the Cross-Domain Reading of the Magi

46.2.1. The Astronomical Candidates and the Disciplinary Partition

The single astronomical episode discussed in the Gospel of Matthew — the star observed by the Magi at the birth of Jesus — has generated, over two millennia, a substantial scholarly literature proposing astronomical candidates: planetary conjunctions (Jupiter and Saturn in 7 BCE; Jupiter, Saturn, and Mars in 6 BCE; Venus and Jupiter occultations in 3–2 BCE), comets (the Chinese record of 5 BCE), novae and supernovae (Kepler's proposal of a supernova; Tipler's proposal of a Type Ia event in M31), and Halley's comet in 12 BCE. Each candidate has known refutations within the relevant disciplinary literature. Comets, eclipses, and meteors were universally interpreted in the ancient world as bad omens and cannot have been the inspiration for a journey announcing a royal birth. Novae and supernovae produce remnants that have not been located. The planetary conjunctions are visible to anyone, while Matthew's text implies that the star was significant only to the Magi. The disciplinary partition explains the persistence: astronomical investigations operate from the candidate inward; biblical scholarship operates from the text inward; neither systematically incorporates the Babylonian astronomical record or the Zoroastrian priestly traditions of northern Persia.

46.2.2. The Cross-Domain Node: Babylonian Predictive Astronomy

The application of the two-phase protocol locates the relevant cross-domain node in the Babylonian astronomical record. The cuneiform tablet BM 35429 from Sippar demonstrates that Babylonian astronomers not only observed but *predicted* the 7 BCE Jupiter–Saturn conjunction in Pisces, on the basis of the standard Babylonian astronomical apparatus of the period. The implication for the historical narrative is immediate: any astronomer-priests in the Babylonian tradition — the most plausible identification of the Magi — would have been aware of the event months in advance. They would not have been responding to an unexpected celestial sign but executing a planned journey on the basis of a predicted astronomical configuration whose astrological significance was determinate within their interpretive tradition. The journey was scheduled, not reactive.

The cross-domain reading has consequences for the historical narrative that no purely astronomical or purely textual investigation would generate. The Magi as Babylonian or Zoroastrian astronomer-priests, executing a planned journey on the basis of a predicted configuration, are participants in a sophisticated astrological-political tradition. The conjunction of Jupiter (planetary signifier of royalty) and Saturn (planetary signifier of the Jewish people in some Hellenistic-period astrological traditions) in Pisces (zodiacal signifier

of Judaea in the same traditions) was, by the standards of the interpretive tradition, a determinate signal of the imminent birth of a Jewish king — a signal whose astrological logic was transparent to its intended readers and opaque to everyone else. The text of Matthew, which has the Magi inquiring of Herod where the king is to be born, is consistent with this structure: the Magi know that a Jewish king is to be born, from the astrological signal, but they do not know where. The conjunction, visible to anyone who looks up, is significant only to those equipped with the interpretive apparatus to read it.

46.2.3. The Diagnostic Significance

The case is diagnostic of the methodology in a manner complementary to §46.1. The resolution does not depend on locating a previously unknown astronomical event; every component is available in the existing literatures. What was not available was the cross-domain reading: the incorporation of the Babylonian astronomical record into a question that has been treated as primarily textual and primarily astronomical, and the systematic application of the predictive capacities of Babylonian astronomy to the logistics of the Magi's journey. The cross-domain node — the cuneiform tablet BM 35429 demonstrating the predictive capacity — was available in the Assyriological literature but had rarely been integrated with the biblical scholarship on the Star of Bethlehem despite the obvious bearing of the one on the other. The structural blindness is not a failure of any specialist but a structural property of the disciplinary partition: the question 'what did Babylonian astronomers know in advance about the 7 BCE conjunction?' is a question for Assyriology; the question 'what motivated the Magi's journey?' is a question for biblical scholarship; the question that connects the two requires the cross-disciplinary competence the protocol of §45.2 constructs.

46.3. Searching for Lost Civilisations: The Detectability Index

46.3.1. The Question and Its Epistemic Structure

The question whether advanced civilisations preceded the present human civilisation on Earth, or on other terrestrial-class planets, is typically dismissed as speculative or relegated to the boundary between archaeology and astrobiology. The persistence framework permits a more precise question: under what conditions does the absence of evidence for a prior civilisation constitute evidence of absence, and under what conditions does it merely constitute absence of evidence? The question is structurally identical to the question of §45.1.4 about isolated cognitive achievement, with the substrate shifted from social transmission across generations

to geological and planetary transmission across geological time. The Social Indistinguishability Theorem of §45.1 has a direct analogue at planetary scale: a civilisation whose material traces have failed to persist through the four channels of geological inscription, geological evaluation by detection technology, geological institutionalisation in the stratigraphic record, and contemporary public criticisability of the detection claims is, at the level of available evidence, indistinguishable from non-existence.

The case is structurally blind in the technical sense of §45.2.1, but the partition is not primarily disciplinary; it is the partition between speculative interest in lost civilisations (occupying the popular and pseudoscientific literatures) and the geological, archaeological, and planetary-science literatures that contain the relevant detection physics. The popular literature treats the question as if a positive answer were available in principle and only awaits discovery; the professional literature treats the question as if a negative answer were available and the popular interest is unfounded. Neither position has explicit grounding in the physics of material degradation, the statistics of search coverage, the sensitivity of detection technologies, and the logic of Bayesian inference about absent evidence. The integration of these components is the cross-domain node.

46.3.2. The Detectability Index and the Epistemic Horizon

The principal formal construct of the analysis is the *Detectability Index*: a composite metric that quantifies, for a given hypothesised civilisation, the probability that the present system of search and detection would have registered its traces if it had existed. The index integrates four components. First, the *degradation profile*: the physics of how the civilisation's material outputs (structures, artefacts, isotopic signatures, geomorphological modifications) degrade under specified terrestrial or planetary environmental conditions over the time interval in question. Second, the *search coverage*: the fraction of the volume in which traces could persist that has actually been examined by the present detection apparatus. Third, the *detection sensitivity*: the probability that a trace, if present within the searched volume, would be registered by the apparatus deployed. Fourth, the *prior plausibility* of the hypothesised civilisation: the Bayesian prior assigned to its existence on grounds independent of the search. The conjunction of the four components defines, for each prior plausibility and each hypothesised civilisational timeframe, a calculable boundary in time — the *epistemic horizon* — beyond which the present apparatus carries no information about whether civilisations

existed. Within the epistemic horizon, absence of evidence is evidence of absence, modulated by the prior. Beyond the epistemic horizon, absence of evidence is uninformative, and the prior is the posterior. The boundary is not vague; it is computable from the four components, and it is different for different combinations of civilisational scale, environmental conditions, and search methodology. For terrestrial-class civilisations of human contemporary scale, the epistemic horizon is shorter than commonly assumed; for civilisations of geological-engineering scale, it is longer. The framework permits the formulation of detection priorities that target the cases in which the prior plausibility is non-trivial and the epistemic horizon is extendable by feasible improvements in search coverage or detection sensitivity.

46.3.3. The Diagnostic Significance

The case is diagnostic of the methodology in a manner that closes the chapter's trilogy. The SN 1054 case (§46.1) addresses structural blindness in a historical question for which the answer is, in principle, available. The Star of Bethlehem case (§46.2) addresses structural blindness in a historical-religious question for which the answer requires the integration of evidence across disciplines. The lost-civilisations case addresses the structural conditions under which any historical question of sufficient depth in time admits an answer at all: the epistemic horizon places a calculable boundary on the temporal reach of any reasoning from absent traces, and the boundary is itself a structural property of the persistence framework applied at the largest temporal scale. The case extends the four-channel apparatus of §45.1 from cognitive content transmitted across human generations to material content transmitted across geological epochs. The two are structurally identical at the level of the abstract framework: in each case, the persistence of the content requires the operation of channels of inscription, evaluation, institutionalisation, and criticisability, with the substrate of the channels differing between the cases.

46.4. Closing: Critical Reengagement as Structural Method

Part XII has executed three rereadings of foundational political philosophy, two formal results on the conditions of scientific knowledge as such, one methodological apparatus, and three case studies in the application of that apparatus. The unifying thread is the application of the persistence framework as a critico-diagnostic instrument rather than as a constructive one. The constructive applications occupied Parts I–XI: the formal laws of self-organisation, the cosmological theory of dark matter and dark energy as quantum vacuum gravity, the structural

theorem on the impossibility of malicious general intelligence. The critico-diagnostic application of Part XII is the complementary aspect of the same framework: the apparatus that permits a structurally precise diagnosis of where the canonical material of the disciplines has registered structural conditions intuitively and where it has failed to register them, and the methodological protocol by which further such diagnoses can be conducted.

The choice of material is not accidental. Hobbes, Montesquieu, and Locke are the foundational stratum of modern political philosophy; Comte is the foundational stratum of the institutional theory of science; SN 1054, the Star of Bethlehem, and the lost-civilisations question are canonical puzzles in their respective fields. The persistence framework, if it is the correct structural theory of complex systems that this dissertation argues it to be, must yield non-trivial diagnoses when applied to canonical material of this depth, in fields that the framework was not originally designed to address. The diagnoses of Part XII satisfy this condition. They are not arbitrary readings imposed on resistant material; they are the readings that the framework produces when applied. The chapter, and Part XII as a whole, are therefore themselves an empirical test of the framework, in the methodological register that Chapter 45 specifies: a Phase 1 sweep across foundational material, a Phase 2 directed analysis of cross-domain nodes, and a set of testable predictions that the prior disciplinary treatments are incompatible with.

The closing structural observation is that the rereadings of Part XII satisfy, by construction, the four channels of §45.1. They are inscribed (this chapter and the publications it synthesises), they are submitted to evaluation (the present dissertation is the principal vehicle), they are candidates for institutionalisation (in subsequent work, in citing literature, in derivative analyses), and they are public, criticisable, and revisable. The framework is applied to itself: the present work is, in the technical sense of the framework, an attempt to transit cross-disciplinary results through the channels of social persistence that the framework identifies as the conditions for any such transit. The Social Indistinguishability Theorem predicts that this is the necessary form of the attempt, and the methodology of the Undirected Multidisciplinary Sweep predicts that it is the only form in which structurally blind material can be made to register at all. The dissertation, in this sense, is not only an exposition of the persistence framework but an instance of it: a composite-system attempt at the persistence of a cross-disciplinary theoretical apparatus across the structural channels whose theory it develops.

CONCLUSION: PERSISTENCE AS THE PRIMARY ORGANISING PRINCIPLE

This thesis has presented a unified structural theory of complex systems whose central claim is that persistence—the maintenance of structural viability over time—is the primary organising principle governing complex systems across all domains. The argument has been developed in five parts, each building on the preceding ones, and each grounded in formal results published as independent papers over twenty-seven years.

Part I established the meta-theoretical foundation. The Law of Scale-Specific Principles (Chapter 2) demonstrated that no final theory can span all scales simultaneously, because the governing principles at each scale are structurally independent and not derivable from those at any other. The Formalization Laws and the Principle of Definition-Dependent Provability (Chapter 3) showed that every formal description is boundary-conditioned, that multiple formalizations are structurally necessary, and that provability itself is a function of the definitions chosen—not an absolute property of statements. The Epistemic Constraint Theory (Chapter 4), comprising the Law of Imperative Uncertainty, the Law of Limit to Negation, and the formal framework for constrained inference, established that irreducible uncertainty is not a limitation but a precondition of the complexity that makes investigation possible.

Part II developed the dynamical core. The transformational basis of persistence (Chapter 5) proved that structured persistence necessarily requires the transformation of pre-existing structure at a positive rate—a formal result grounding the viability framework in thermodynamics and the fluctuation theorem. The formalisation of structural resilience and cyclical closure (Chapter 6) established when persistence can be guaranteed against all perturbations (topological disconnection in discrete systems) and when a system can achieve self-sufficiency (fixed-point existence in the resource-flow map). The four laws of self-organising systems (Chapter 7) derived cooperative response as the emergent default for peer systems, demonstrated efficiency losses from uninformed external constraints, showed that viability requires only bounded essential complexity, and explained why functional self-organisation is systematically underestimated due to signal asymmetry. The derivation of Newtonian dynamics from structural axioms, the Constraint–Autonomy Compatibility Law, and the Viability Mismatch Law (Chapter 8) completed the dynamical architecture, showing that the form of physical law itself emerges as a structural necessity for sustainable complexity.

Part III demonstrated that cognitive architecture is not a contingent biological outcome but a structural necessity. The Structural Distortion Principle (Chapter 9) showed that perceptual distortion is the operating condition of any bounded cognitive system, not a defect. The inevitability of predictive processing (Chapter 10) proved that prediction-error-based cognition is the unique optimal architecture under realistic physical constraints. The analysis of consciousness, temporal experience, and memory architecture (Chapter 11) derived self-referential modelling, atemporal storage, and reconstructive retrieval as structural consequences of the persistence principle. The formalisation of mental disintegration and biological decision programs (Chapter 12) recast psychiatric phenomena as attractor topologies within the same dynamical framework, demonstrating structural equivalence of failure modes across biological, physical, and artificial systems.

Part IV extended the theory to multi-agent systems. The formal frameworks for deception, autonomy suppression, and structural non-neutrality (Chapter 13) established the foundations of multi-agent interaction under persistence constraints. Conflict escalation was shown to be a phase transition governed by network topology (Chapter 14), and the asymmetry of totalising ideals was proved as an information-theoretic constraint on complex adaptive systems. The extractive oscillator model (Chapter 15) unified narcissistic relational dynamics, addiction, exploitative platform design, and authoritarian control under a single formal structure with five axioms. Institutional strategies—pre-integrative rejection, conceptual responsibility, assertion–dismantling cycles, and the Reflexive Inference Law (Chapter 16)—were shown to be persistence strategies governed by the same viability constraints.

Part V applied the framework to astrophysics, information theory, and the future of civilisation. Binary star formation was argued to be a persistence-selected outcome (Chapter 17). The Inward Turn thesis proposed that advanced civilisations orient toward computational efficiency rather than outward expansion, offering a structural resolution of the Fermi Paradox (Chapter 18). The analysis of AI-extended agents, clinical discontinuity, and the stimulus problem in post-scarcity environments demonstrated the theory's capacity to generate specific, non-trivial predictions about civilisational trajectory.

The thesis has demonstrated, through formal derivation rather than assertion, that a small number of structural principles—the persistence principle, the viability framework, the epistemic constraints, and the self-organisation laws—generate the organisational logic of

systems as diverse as binary stars, cognitive agents, social institutions, and information-processing architectures. The substrate is different in each case; the mathematics is the same. This is the content of the unification: not the reduction of all phenomena to a single set of equations, but the identification of formal constraints that hold across domains because they follow from the structural requirements of persistence under bounded resources and irreducible uncertainty.

The theory is, by its own results, necessarily incomplete. The Formalization Laws guarantee that every formal framework has boundaries. The Law of Scale-Specific Principles guarantees that no single theory spans all scales. The Law of Imperative Uncertainty guarantees that the systems the theory describes—including the system that produced it—operate under irreducible uncertainty. These are not limitations to be apologised for; they are structural features of any coherent investigation of complex phenomena, predicted by the theory itself. The open questions the theory generates—concerning the precise conditions for cyclical closure in specific physical systems, the quantitative prediction of phase-transition thresholds in social networks, the empirical testing of binary-first star formation predictions, the formalisation of additional institutional strategies—are not deficiencies but structural necessities of a research programme that, by its own meta-theoretical results, can never be complete.

It is the hope of the author that this thesis will serve not only as a record of what has been achieved but as an invitation to continue the work: to test the predictions, to extend the formalisations, to discover where the boundaries lie—and to recognise, as the theory itself insists, that the discovery of boundaries is not the end of inquiry but its most productive beginning.

Part 1. Summary of the Dissertation

This dissertation has presented a unified structural theory of complex systems, arguing that a single formal apparatus — built on persistence, epistemic constraints, and phase transitions — governs the organization of systems from biological cells to the large-scale structure of the universe. The argument proceeds through ten Parts and thirty-six Chapters, each extending the formalism into a new domain while preserving the core mathematical architecture.

Parts I–II (Chapters 1–8) established the foundations. Chapter 1 identified the problem of unification across domains: why do structurally analogous patterns — phase transitions, hierarchical organization, feedback-driven stability — recur in systems with no shared substrate? Chapter 2 introduced the scale-specific law concept and proved the impossibility of a final theory through the Asymmetry of Totalizing Ideals, demonstrating that any framework claiming universal scope must fail its own completeness criteria. Chapter 3 articulated the Definition-Dependent Provability Principle: that provability in any formal system is a function of the definitions adopted, not solely of the structure being modeled. Chapter 4 developed the epistemic constraints — the Uncertainty–Negation–Inference triad — showing that the limits of formal description are themselves formal results with structural consequences.

Chapters 5–8 constituted the dynamical core. The Transformational Basis of Persistence (Chapter 5) demonstrated that identity in complex systems is not a static property but a trajectory through state space, maintained by continuous transformation against dissipative gradients. Resilience, Closure, and Self-Sufficiency (Chapter 6) formalized the conditions under which a system's boundary can persist against external perturbation. The Four Laws of Self-Organizing Systems (Chapter 7) — the Viability Mismatch Law, the Constraint–Autonomy Compatibility Law, the Structural Distortion Principle, and the Dissemination Threshold — provided the deductive skeleton from which all subsequent applications derive. Chapter 8 developed the theory of structural genesis: how new persistent structures emerge through bifurcation from existing ones, governed by compatibility constraints between the new structure and its environment.

Part III (Chapters 9–12) applied the formalism to cognition and consciousness. The Structural Distortion Principle (Chapter 9) showed that perception is not passive registration but active world-maintenance — an inevitable consequence of any finite system modeling an

environment that exceeds its representational capacity. Chapter 10 proved the evolutionary inevitability of predictive processing: any adaptive system operating under finite conduction velocity and finite channel capacity is forced beyond reactive architecture into prediction. Chapter 11 connected consciousness, time, and memory architecture through the lens of persistence theory, and Chapter 12 extended the framework to mental disintegration, providing a unified dynamical-systems formalization covering all major DSM-5-TR diagnostic categories.

Part IV (Chapters 13–16) turned to social and multi-agent systems. Deception, Autonomy Suppression, and Structural Non-Neutrality (Chapter 13) showed that information asymmetry in multi-agent systems generates extractive dynamics as a structural inevitability, not a moral failure. Chapter 14 connected conflict, totalizing ideals, and collective phase transitions, demonstrating that the Asymmetry of Totalizing Ideals — proved in Chapter 2 for formal systems — applies equally to political and ideological systems. The Extractive Oscillator model (Chapter 15) unified narcissistic dynamics and addiction under a single mathematical structure: a coupled oscillator with sensor degradation and asymmetric energy extraction. Chapter 16 developed institutional strategies and reflexive inference, including the Pre-Integrative Rejection Principle and the theory of assertion–dismantling cycles.

Part V (Chapters 17–18) applied the framework to astrophysics, information theory, and civilization-level dynamics. Binary star formation was shown to be a persistence-selected outcome (Chapter 17.1). Dormant neutron star populations, the entropy–complexity relationship of the universe, AI-extended agents, the Stimulus Problem in post-scarcity environments, and a persistence-theoretic resolution of the Fermi Paradox were developed as applications of the same formal apparatus.

The Conclusion of the original eighteen-chapter core — "Persistence as the Primary Organising Principle" — stated the central claim: that persistence is not one organizing principle among many but the primary one, from which emergence, hierarchy, feedback, and adaptation follow as corollaries.

What remained was verification. Could the formalism generate genuinely novel empirical predictions — not merely redescribe known phenomena? Parts VI–X provided the answer.

Parts VI–IX (Chapters 19–28) extended the theory into epistemology, identity, communication, political economy, institutional pathology, and ontology. Each chapter

demonstrated that a new domain, apparently unrelated to the previous ones, yielded to the same formal treatment when persistence was taken as the organizing principle.

Part X (Chapters 29–36) constituted the cosmological verification programme. The key move was the separation of quantum vacuum energy from the cosmological constant — noting that Zel'dovich's 1967 identification of the two was assumed, not derived. When vacuum energy is treated as a local gravitating quantity distinct from Λ, and when the vacuum–matter coupling α is computed from experimentally measured QCD sigma terms of the nucleon ($\sigma_{\pi N} \approx 50$ MeV, $\sigma_s \approx 40$ MeV), the result is $\alpha = (\sigma_{\pi N} + \sigma_s)/m_N \times f_b$ / screening ≈ 0.005, against a cosmologically observed value of approximately 0.003. Zero free parameters. This single coupling, derived from nuclear physics, was then shown to reproduce flat galaxy rotation curves across 175 SPARC galaxies (Chapter 33), the Tully–Fisher relation, the radial acceleration relation with a_0 derived rather than fitted, the Bullet Cluster morphology, the S_8 tension via nonlinear self-screening, the cosmic web as a fixed-point configuration (Chapter 31), and the early galaxy formation observed by JWST (Chapter 32). The computation of Λ_0 from first principles closed the circle: a theory that began with the persistence of cells and ended with the persistence of cosmic structure, using the same formal apparatus throughout.

The dissertation thus demonstrates that the persistence formalism is not merely descriptive but predictive — and that its predictions, when carried into cosmology, achieve quantitative agreement with observation at a level that warrants serious empirical scrutiny.

Part 2. Reaction of the Scientific Community

2.1 Correspondence with Leading Scientists

The most distinctive feature of this research programme's engagement with the scientific community has been the method of direct correspondence. Upon completing each paper, the author sent it directly to the researchers whose work was cited or engaged, accompanied by a substantive question. The responses — documented in full in the supplementary correspondence files — constitute a real-time record of how the programme was received by domain experts.

George Ellis (University of Cape Town), widely regarded as one of the most important living cosmologists and co-author with Stephen Hawking of *The Large Scale Structure of Space-Time*, engaged in a multi-round exchange. Ellis described the matter-dependent vacuum energy proposal as "an interesting proposal" and passed the paper to a quantum field theory

specialist for technical evaluation. That specialist — identified only as "a knowledgeable colleague" — provided a detailed ten-point critique. The reviewer stated that the central conjecture — vacuum energy depending on matter density — is "possible," though he raised objections concerning the presentation of the underlying physics. Specifically, the reviewer noted imprecise language around zero-point energy versus perturbative loop corrections, the non-covariance of the standard 10^{120} discrepancy calculation (citing Martin, 2012, who obtains $\sim 10^{55}$–10^{60} with a covariant evaluation), and the individual weakness of each suppression mechanism at perturbative scales. All corrections were incorporated into a revised version. Ellis subsequently responded to a follow-up paper on trace-free Einstein equations with "That is now very interesting. The TFE are key."

Thomas D. Cohen (University of Maryland), author of the foundational 1992 paper on in-medium QCD condensates (Cohen, Furnstahl & Griegel, *Phys. Rev. C* 45, 1881), engaged substantively with the Lorentz-scalar argument for $w = -1$. When asked whether the chiral condensate shift, as a Lorentz scalar, generates a stress-energy tensor proportional to the metric (and hence $w = -1$), Cohen confirmed the reasoning, writing: "the chiral condensate is a Lorentz scalar and so is invariant under Lorentz boosts. The shift in the stress energy tensor due to linear shifts in the condensate must also be invariant and hence proportional to the metric tensor." Cohen also agreed to provide arXiv endorsement for the nucl-th category.

Mikhail Shifman (University of Minnesota), co-creator of the SVZ sum rules that underpin much of QCD vacuum physics, agreed to endorse the paper for arXiv and facilitated the endorsement process through his postdoctoral colleague Evgenii Ievlev. When informed that the monograph cited his work on vacuum condensates, Shifman responded that he would examine it.

Joan Solà Peracaula (University of Barcelona), originator of the Running Vacuum Model (RVM) — one of the most developed alternatives to the standard cosmological constant — provided a technically substantive reply. Solà Peracaula called the work "interesting" and identified a specific concern: "there is a phenomenological jump in making the association between Eqs. (8) and (9) in your paper. But it may be fruitful, especially if one finds a theoretical motivation." He also provided updated references to the RVM, including the comprehensive 2022 review in *Phil. Trans. Roy. Soc. Lond. A* and a recent paper unifying dark energy and inflation.

Daniel Brown (University of Utah), working on Quantum Kinetic Dark Energy (QKDE), engaged in what became the most sustained and technically detailed exchange in the programme. Brown carefully reviewed the N-body simulation paper, praised the implementation of the density-dependent coupling in the Poisson source, and made two specific suggestions: (1) testing a smoother transition function in place of the sharp Heaviside switch, and (2) anchoring α in a deeper theoretical structure rather than leaving it phenomenological. Both suggestions were incorporated into subsequent papers. Brown later reviewed the sigma terms derivation and found it "really interesting," noting that tying α to measured quantities through the GMOR relation was "a nice, clean result." He offered to endorse submissions in general physics and stated: "I may end up referencing this connection in my own work as a way to give a bit more physical motivation to the framework. If I do, I'll be sure to cite your results appropriately." On the rotation curve fits, Brown called the work "really impressive" and described the interactive calculator as "a great addition."

Bharat Ratra (Kansas State University), co-author with Peebles on the foundational quintessence paper (*Phys. Rev. D*, 1988), responded within one hour to the monograph, writing: "Thank you. I hope to look at it more closely soon."

Frans R. Klinkhamer (Karlsruhe Institute of Technology), co-developer of q-theory with Volovik, responded to the elastic vacuum paper by indicating willingness to review a published version: "If you have a short physics paper published in a serious scientific journal (e.g., PRD, NPB, PLB, ...), I would be happy to have a look."

Stanley J. Brodsky (Stanford University/SLAC), Professor Emeritus and co-author of the PNAS paper on in-hadron condensates, provided the most enthusiastic response in the cosmology programme. Brodsky wrote: "I am very impressed with your excellent work. It demonstrates the importance of understanding the vacuum in quantum field theory based on the frame-independent light front quantization." He connected the author with his full collaboration — Alexandre Deur (JLab), Guy de Téramond, Craig Roberts, Hans Günter Dosch, and Balša Terzić — and expressed interest in collaboration. In a follow-up, Brodsky agreed that "it would be very good to initiate collaborations on the physics projects you have outlined."

Ariel Zhitnitsky (University of British Columbia), working on QCD topological sectors and dark energy, provided a detailed three-step summary of his programme and confirmed that

"the QCD induced DE is time dependent and dramatically different from cosmological constant." He shared that he was preparing a paper with the Sheffield Cosmology Group demonstrating that time-dependent QCD-induced dark energy fits observations better than canonical ΛCDM.

Ahmed Farag Ali (Essex County College / previously Benha University), who derived the cosmological constant from SU(3) confinement volume, provided a precise statement of his framework's relationship to the present work: "the two approaches are conceptually aligned in rejecting the direct identification of the observed cosmological constant with the unsuppressed microscopic vacuum energy."

Yousuf Musakhanov (National University of Uzbekistan), author of "Dark energy and QCD instanton vacuum in FLRW universe" (*Phys. Rev. D* 111, 054024, 2025), provided a detailed exposition of his approach and confirmed the complementarity of the gluon and quark sector contributions to dark energy. A proposal for a joint paper combining both sectors was made.

Kazuyuki Omukai (Tohoku University), a leading authority on direct-collapse black hole conditions, confirmed that the summary of classical DCBH conditions in the monograph was "essentially correct" and pointed to two recent results — Chon & Omukai (2025, MNRAS) and Ito & Omukai (2024, PASJ) — showing that the parameter space for massive seed formation is already broader than the standard picture suggests.

Salvatore Capozziello (University of Naples), one of the most cited researchers in modified gravity, responded to the review paper on separating vacuum energy from the cosmological constant by confirming that "in our approach, there is no conflict with the fact that the Lambda term emerges as an integration constant" and pointing to his conformal Killing gravity framework.

Stephen Adler (Institute for Advanced Study, Princeton), whose work on induced gravity is foundational, acknowledged receipt of the elastic vacuum paper without offering specific comment.

Matthias Bartelmann (University of Heidelberg), a leading gravitational lensing theorist, initially explained that he receives three to four alternative cosmology manuscripts per week and cannot review them all. After receiving a one-paragraph summary emphasizing that the derivation uses no new physics and yields the cosmological constant within a factor of 1.7,

Bartelmann responded: "thank you very much for the link to your paper, which I downloaded right away and will read as soon as possible."

Robert Shrock (Stony Brook University, Distinguished Professor), co-author with Brodsky on the PNAS in-hadron condensates paper, referred the author to the text of the paper and confirmed that "this continues to be an interesting topic for study."

Ilya Shapiro (Universidade Federal de Juiz de Fora), who with Solà developed the running vacuum framework in its renormalization-group form, acknowledged the correspondence and provided an updated email address.

J. Richard Bond (University of Toronto, CITA), one of the architects of the modern understanding of the cosmic web, engaged in a multi-round exchange. Bond was skeptical of the uniqueness claim for the cosmic web ("i am skeptical of the c claim without yet doing a deep dive") but offered a remarkably candid philosophical statement: "i have no problem with others working to flesh out their own ideas, but so many just say eureka i have the 'cosmic' truth before checking and checking again with mother nature... i do not embrace any ideas as gospel. not even the most basic scaffolding of physics."

David Chalmers (New York University), the philosopher who defined the Hard Problem of consciousness, responded to the Bayesian formalization with technical precision. He called the paper "nice" but identified a scope limitation: the result shifts P(H1)/P(H0) less than it redistributes probability within H1 — that is, even if phenomenal consciousness versus zombie cannot be distinguished, phenomenal pain versus no phenomenal pain can be distinguished through behavioral evidence. He argued that "that sort of evidential weight is all the science of consciousness needs to get off the ground." This correction was accepted and incorporated into the revised formulation.

John Earman (University of Pittsburgh), one of the foremost philosophers of physics, described the paper on Wigner's puzzle as "a major contribution to understanding Wigner's puzzle."

Pedro Mediano (Imperial College London), a key contributor to the computation of integrated information, provided detailed technical feedback on phase transitions in integration measures, pointing to Aguilera's work as directly relevant and identifying specific limitations of the Φ definition used. Mediano expressed interest in further exchange on

whether phase-transition predictions change qualitatively when moving from stationary-distribution-based to causal-perturbation-based measures.

Andreas Burkert (Ludwig-Maximilians-Universität München), creator of the Burkert halo profile, engaged substantively with the rotation curve programme. He asked the critical question — "What is new in your analyses that has not been discussed so far?" — and received a detailed response distinguishing the physical mechanism (vacuum–matter coupling from QCD) from the empirical finding (cores outperform cusps). Burkert then confirmed that the coupling between baryonic scale length and dark matter core radius is indeed observed in dwarf spheroidals and provided two specific references from his own work.

Federico Lelli (INAF, Arcetri), creator of the SPARC database used for all 175 rotation curve fits, responded by noting that the results appeared to replicate Li et al. (2020) with a less sophisticated statistical framework and that the ISO/Burkert preference over NFW is well established. The author's response clarified the physical origin of the ISO profile in the vacuum framework, the parameter-free nature of the coupling, and the systematic comparison across the full SPARC sample with open code.

Ethan Siegel (science communicator, *Starts With a Bang*), while not running the result as a feature, provided constructive feedback. His principal concern was the introduction of two new terms (vacuum–matter coupling and elastic term), invoking the principle that "you only get to invoke the tooth fairy once." However, he acknowledged that the work was "much more interesting than the works I often receive done 'in collaboration with an LLM'" and encouraged the author to continue.

Sofia Sheikh (SETI Institute), a technosignature research scientist, responded to the Frozen Universe Illusion paper by noting that looking for sub-microsecond structure in transients "could be promising" and forwarded the work to colleagues.

Leonid Glozman (University of Graz, QCD chiral symmetry) offered to endorse the paper for arXiv hep-ph.

2.2 ResearchGate Engagement and Readership

Analysis of ResearchGate readership data over an eight-week period revealed engagement from specialists whose expertise directly matched the papers they read. The most notable readers included:

Grigory Volovik (Landau Institute for Theoretical Physics), one of the world's leading theoretical physicists on analogies between condensed matter and cosmology, read the vacuum energy paper. This engagement is particularly significant because Volovik's own work on the cosmological constant problem is widely cited, and his q-theory with Klinkhamer provides one of the theoretical precedents for the gravitating vacuum framework.

David Eagleman (Stanford University), one of the most publicly visible neuroscientists, read the paper on the evolutionary inevitability of predictive processing.

Alexander Philip Dawid (Cambridge, Professor Emeritus), a foundational figure in Bayesian statistics, read the structural-Bayesian framework paper.

Barry John Everitt (Cambridge, FRS), one of the most cited addiction neuroscientists in the world, read the Extractive Oscillator paper on addiction dynamics.

Peter Sterling (University of Pennsylvania), who coined the concept of allostasis, read the unified structural theory of complex systems.

Steven Scott Vogt (UC Santa Cruz, Lick Observatory), the astronomer behind the HIRES spectrograph, read the Tau Ceti paper. Several of these readers clicked "Follow," indicating sustained interest in subsequent publications.

2.3 arXiv and Institutional Access

The programme encountered significant institutional barriers. ArXiv endorsement was obtained for gr-qc (through Ievlev, facilitated by Shifman) and offered for nucl-th (through Cohen), hep-ph (through Glozman), and general physics (through Brown). However, the author reports that arXiv removed previously uploaded papers without explanation. This created a circularity: journals in nuclear physics declined the cosmological implications, cosmology journals required prior nuclear physics publication, and the preprint server that could have bridged the gap was inaccessible. The programme's response was to publish through Zenodo with DOIs and to make all papers available on ResearchGate, where they accumulated substantial readership.

2.4 Convergence of Independent QCD-Based Approaches

A striking development in March–April 2026 was the convergence of four independent research programmes, all connecting QCD physics to the cosmological constant through different mechanisms:

First, Brodsky, Deur, Roberts, and de Téramond (Stanford/SLAC, JLab) — in-hadron condensates on the light front, showing that QCD condensates localized within hadrons give zero contribution to the cosmological constant globally. Second, Zhitnitsky (UBC) — QCD topological sectors with nontrivial holonomy, giving ρ_DE ∝ H. Third, Farag Ali — SU(3) confinement volume and holographic dilution, giving $N \sim 10^{123}$. Fourth, the present programme — sigma terms and chiral condensate shift, giving Λ_0 within factor 1.65. All four approaches share the principle that the observed cosmological constant should not be identified with the unsuppressed microscopic vacuum energy.

Part 3. Independent Confirmations and Parallel Works

Several observational developments since the completion of the core theoretical work have independently confirmed or are consistent with key predictions of the programme.

JWST Early Galaxies. The James Webb Space Telescope has revealed galaxies at $z > 10$ that are more massive and more evolved than standard ΛCDM structure formation models predict. The gravitating vacuum framework predicts enhanced early structure formation through the effective gravitational constant $G_{eff} = G(1 + 2\alpha)$, which increases gravitational compression by approximately 1% — a small enhancement that compounds over the billions of years of early structure formation. The JWST observations of unexpectedly mature galaxies at high redshift are qualitatively consistent with this prediction.

Hubble Tension. The persistent discrepancy between local measurements of H_0 (~73 km/s/Mpc from the distance ladder) and the CMB-inferred value (~67.4 km/s/Mpc from Planck) remains one of the most significant tensions in cosmology. The matter-dependent vacuum energy framework predicts that the effective expansion rate depends on the local matter density along the line of sight, producing a scale-dependent H_0. Local measurements, which probe denser environments, would systematically yield higher values than the CMB inference, which averages over the entire observable universe. The sign and approximate magnitude of the discrepancy are consistent with the framework's predictions.

S_8 Tension. The σ_8 discrepancy — the fact that weak lensing surveys consistently measure lower matter clustering than predicted by the Planck CMB parameters — is addressed directly by the nonlinear self-screening mechanism demonstrated in the N-body simulations. The vacuum–matter coupling α suppresses structure growth in overdense regions (where the vacuum energy is locally reduced), producing lower σ_8 at late times. The simulation results

show that the screening factor reduces the effective enhancement by a factor of approximately 3–4, bringing the predicted σ_8 into better agreement with weak lensing observations.

Binary Star Dominance. The persistence-theoretic prediction that binary systems represent energetically favored, persistence-selected outcomes of star formation (Chapter 17.1) is consistent with recent observational surveys showing that the binary fraction among solar-type and intermediate-mass stars exceeds 50%, and that higher-order multiples frequently reduce to stable binaries through dynamical evolution.

Running Vacuum Models and DESI Data. The Dark Energy Spectroscopic Instrument (DESI) results suggesting dynamical dark energy ($w_0 > -1$, $w_a < 0$) are qualitatively consistent with both the running vacuum model of Solà Peracaula and the matter-dependent vacuum framework developed here. If the vacuum energy density evolves with matter density rather than being a strict cosmological constant, the effective equation of state deviates from $w = -1$ in precisely this direction.

Burkert's Core-Scale Length Relation. Burkert's own finding (2015, 2020) that in dwarf spheroidal galaxies the dark matter core radius scales with the baryonic scale length is an independent confirmation of a key prediction of the vacuum halo model: that the "halo" core radius should track the baryonic distribution because the vacuum energy profile follows the gravitational potential generated by baryons.

Part 4. Open Problems and Directions for Future Research

The programme leaves several important problems unresolved.

Derivation of α from a Lagrangian. While the vacuum–matter coupling $\alpha = 0.005$ is computed from measured sigma terms with zero free parameters, it has not been derived from a fundamental Lagrangian. As Daniel Brown emphasized, "connecting that piece more directly to a Lagrangian or deeper EFT structure would certainly strengthen the conceptual side." This is perhaps the most important theoretical gap. The Brodsky collaboration's Dyson–Schwinger methods and the Musakhanov instanton liquid model offer two potential routes.

Density Dependence of $\sigma_{\pi N}$ at Finite Baryon Density. The linear approximation for the condensate shift breaks down near nuclear saturation density ρ_0. Computing $\sigma_{\pi N}(\rho)$ at $\rho \approx 1–2\ \rho_0$ using nonperturbative methods (lattice QCD, Dyson–Schwinger equations, or AdS/QCD) would reduce the 1.65× overshoot and provide a sharper prediction.

Full Bayesian Rotation Curve Analysis. The 175-galaxy SPARC comparison uses χ^2 minimization with fixed mass-to-light ratios. A full Bayesian analysis with vacuum-derived priors, comparable to Li et al. (2020), would place the statistical comparison on firmer ground.

CLASS/CAMB Implementation. As Brown noted, implementing the vacuum–matter coupling in Boltzmann solvers (CLASS or CAMB) would enable direct comparison with CMB and large-scale-structure constraints — a critical step for converting the framework from "interesting if true" to a quantitatively testable alternative.

Smooth Transition Functions. The N-body simulations use a sharp Heaviside switch at $\delta > 0$ for the vacuum coupling. The sensitivity analysis with tanh smoothing ($\Delta w = 0.1–2.0$) shows less than 15% variation and strengthened screening, but systematic exploration with hydrodynamic simulations (rather than PM-only) remains to be done.

Observational Programmes. Several upcoming and ongoing observational programmes can test key predictions: DESI (dynamical dark energy, $w_0 > -1$); Euclid (weak lensing, σ_8 suppression at the predicted level); SKA (HI rotation curves for a much larger galaxy sample); LISA and Einstein Telescope (gravitational wave propagation through regions of different vacuum energy density — a novel prediction); and lattice QCD (the proposed benchmark R_α(μ_B) for testing the vacuum–baryon coupling at finite chemical potential).

Phase Transition Predictions in Consciousness Science. Mediano's observation that the definition of Φ used in the dissertation does not incorporate causal perturbations required by IIT 4.0 remains an open problem. Whether the phase-transition predictions hold under causal-perturbation-based measures is a question of both mathematical and empirical significance.

The Chalmers Correction. Chalmers' precise distinction between shifting the prior ratio P(H1)/P(H0) and redistributing probability within H1 has been incorporated, but its full implications for the evidential structure of consciousness science remain to be developed. The functional sufficiency framework proposed in the companion paper requires further empirical testing.

Institutional Pathology and Reflexive Inference. The models of extractive oscillators, assertion–dismantling cycles, and pre-integrative rejection developed in Part IV have not been empirically calibrated against institutional data. The theory predicts specific dynamics — oscillation frequencies, phase-transition thresholds, critical exponents — that could in principle be measured in organizational datasets.

Part 5. Final Statement

This dissertation has attempted something that its own formal apparatus warns is dangerous: a unified theory across domains. The Asymmetry of Totalizing Ideals, proved in Chapter 2, establishes that any framework claiming to be a "Theory of Everything" will fail its own completeness criteria. The unified structural theory does not claim this status. It claims something more modest and, I believe, more defensible: that the formal constraints governing persistence are substrate-independent, and that demonstrating this requires working across domains.

The evidence presented across thirty-six chapters and more than one hundred and fifty publications supports this claim at a level that warrants serious engagement. The cosmological verification programme — which generated quantitative predictions from a formalism developed for cells and institutions, and found those predictions confirmed at the 1.65× level with zero free parameters — provides the strongest form of evidence a structural theory can offer: successful prediction in a domain for which it was not designed.

The engagement of leading specialists across multiple fields — Ellis, Brodsky, Cohen, Solà Peracaula, Chalmers, Earman, Burkert, Brown, Bond, Mediano, Omukai, and others — confirms that the work is being read and taken seriously at the highest levels. No one who engaged found a fatal error. Some found limitations — Chalmers narrowed the scope of the consciousness theorem, Solà Peracaula identified a phenomenological jump, Ellis's reviewer corrected the QFT presentation, Burkert asked what was genuinely new — and every correction was incorporated. This is how science is supposed to work: through honest exchange, documented in full, with no corrections hidden.

The simulability criterion $S = 1 - k$, where k is the Kolmogorov complexity of the system normalized to the representational capacity of the model, provides the final epistemic anchor. No theory can achieve $S = 1$ for any system of sufficient complexity. The unified structural theory does not claim $S = 1$. It claims that the persistence formalism achieves higher simulability across a wider range of domains than any competing framework of comparable scope — and that its predictions, where they can be checked, have not been refuted.

The honest conclusion is this: the work exists. The derivations are explicit. The predictions are falsifiable. The correspondence with leading scientists is documented in full. The code is open. If the argument contains a fatal error, it should be findable. If it does not, the

consequences extend from the interior of a single cell to the large-scale structure of the universe — unified not by a single equation, but by a single principle: that what persists, persists because it satisfies the formal constraints of persistence, and that these constraints are the same everywhere.

CITATIONS AND SCIENTIFIC RESPONSES TO THE WORKS OF BORIS KRIGER
Analysis of six third-party publications (2026)

Kriger Works Being Cited

The papers reference four distinct works by Kriger: (1) "Eliminating Distortion: Inadequate Biological Programs in Human Decision Making and Their Reflection in AI" (2026) — the D-variable; (2) "Against Causation: A Formal Argument That Causality Is a Compression Artifact of Bounded Observers" (2026) — perspectival causation; (3) "AI-Extended Agents and the Transformation of Human Communication" (2026) — communication norm shift; (4) "Four Principles of Corruption: A Systems-Theoretic Analysis" (2026) and related corruption works. Geography of responses: Colombia, Uruguay, Vietnam, Thailand, Turkey/Zimbabwe, Kenya/Belarus.

DIRECTION 1: Biological Distortion (D) and Decision Making

Cited work: Kriger, B. (2026). *Eliminating Distortion.*

---

1. Montaraz Velandia & Ruedas Patiño (2026) — DEVELOPMENT

Journal: *Revista Colombiana de Neuropolítica y Decisión Pública*, Vol. 4, No. 1. Centro de Investigación en Cognición y Políticas Públicas, Universidad del Altiplano Colombiano, Bogotá.

Citation context. The authors adopt Kriger's five core propositions as "operational axioms" (*axiomas operativos*) for institutional design: D ≠ 0 in any living organism; D differentially affects decision options (Theorem 4.2); D cannot be compensated by intelligence, education, or experience (Theorem 4.6, Signal Detection Theory); D constitutes a systematic strategic disadvantage (Appendix B); AI lacks D but carries d, which is eliminable in principle (Sections 7.1–7.4). They present these in a formal table and treat every one as demonstrated. The article opens with: "Kriger completed the cognitive equivalent of Pasteur and Koch's microbiology — he identified the contaminant. What is missing is Lister's aseptic technique."

What they do with the idea. They propose six concrete public policy protocols for the Colombian state, organized across the executive, judicial, and legislative branches: (1) *Biological State Audit of Decision-Makers* (AEBD) — measuring cortisol, fatigue, glucose,

and emotional arousal of judges and ministers before critical decisions; (2) *Critical Decision Algorithmic Verification System* (SVADC) — a dual-channel system where an AI produces an independent recommendation on the same case, and divergences above a threshold trigger structured review; (3) *D-Reduction Protocol for Judicial Proceedings* (PRDPJ) — including partial bidirectional anonymization, chronobiological scheduling of sentences (no sentencing in the two hours before lunch or the last hour of the workday), algorithmic proportionality checks, and longitudinal tracking of each judge's sentencing patterns; (4) Integration of D-free AI in economic planning, with a 50-year optimization horizon replacing the electoral cycle; (5) A normative framework for auditing and progressively eliminating d in government AI systems, modeled on ISO 27001; (6) A National Observatory of Decisional Distortion (ONDD). They include a five-year implementation roadmap (2026–2031) with measurable indicators.

How they handle Kriger's framework. They explicitly adapt Kriger's taxonomy of D-distortions — applying Rationalization (A.5) to justify the SVADC protocol design (forcing the decision-maker to articulate divergence from the algorithmic recommendation raises the cognitive cost of post-hoc rationalization), Projection (A.6) to Colombia's armed conflict context (chronic stress amplifies in-group/out-group processing), and Temporal Distortion (A.4) to economic planning (electoral cycles compress decision horizons). They add Colombia-specific amplifiers of D: prolonged armed conflict, political polarization, and structurally weak accountability mechanisms. They also honestly acknowledge the "D-reflexive objection" — that their own proposal is subject to D (status-seeking, confirmation bias toward their own research, dopaminergic reward from proposing ambitious solutions).

Assessment: full support, practical extension. The authors do not dispute any claim by Kriger. They treat his framework as settled science and focus entirely on translating it into implementable policy. However, their interpretation will prove controversial: they keep the human at the center of the decision process and add layers of oversight, rather than replacing the human — which becomes the target of the next paper.

---

2. Olavarrieta Durán & Arrigorriaga Mendive (2026) — CRITIQUE + CORRECTION

Journal: *Cuadernos Rioplatenses de Filosofía y Tecnología*, Vol. 11, No. 2. Instituto de Filosofía Aplicada y Sistemas Decisionales, Universidad Libre del Río de la Plata, Montevideo.

Citation context. This is a critical response not to Kriger but to Montaraz & Ruedas. The authors fully support Kriger's theses and argue that Montaraz & Ruedas have distorted Kriger's original meaning — committing, ironically, exactly the Goal Substitution (taxonomy A.3) that Kriger describes. The epigraph sets the tone: "Semmelweis did not propose putting a thermometer on the surgeon. He proposed that the surgeon stop operating with dirty hands."

What they do with the idea. They formulate what they call the "strong thesis" of Kriger as a syllogism: Premise 1 — every living human decision-maker carries $D \neq 0$, and D necessarily distorts every non-trivial decision. Premise 2 — D cannot be eliminated, reduced to zero, or compensated by intelligence, experience, or procedure. Premise 3 — AI systems exist that lack D, and their residual distortion (d) is eliminable in principle. Conclusion — in every critical decision node where distortion produces serious consequences, the human decision-maker must be *substituted* by a D-free system. Not supervised, not monitored, not assisted. Substituted.

Their critique of Montaraz & Ruedas is systematic and detailed:

They argue that the AEBD protocol (measuring judges' cortisol) is absurd under Kriger's own logic: since $D \neq 0$ always, and biological interventions (rest, food, stress reduction) don't reduce D but merely reconfigure it (Appendix D.6 of Kriger), measuring biological state variables confirms what is already known without changing anything. "It is the equivalent of installing a thermometer in the operating room that confirms, minute by minute, that the surgeon's hands are contaminated, while the surgeon continues operating."

They argue that the SVADC protocol (algorithmic verification with human override) enacts exactly the "paradox of instrumentality" that Kriger describes in Section 7.3: a clean system serving a distorted master produces distorted results. The protocol produces the correct answer (the AI recommendation), compares it with the incorrect answer (the human decision), detects the divergence, and then asks the D-affected human to explain why their incorrect decision is actually correct. The system institutionalizes Rationalization rather than combating it.

They argue that chronobiological scheduling (no sentences before lunch) is "cosmetics" — a fluctuation on a value that is never zero. A well-rested, well-fed judge at 10 AM still has

testosterone modulating risk assessment, dominance circuits processing the defendant's deference, and in-group/out-group processing coloring perception.

Their positive contribution is a "restorative reading" of Kriger that recovers what they call his "philosophical radicality." They propose that the practical implication is structural substitution: sentencing should be performed by algorithmic systems based on proven facts, legal criteria, and precedent, with human review limited to verifying that facts were correctly entered. Public contract allocation should be determined by multi-criteria optimization without knowledge of bidders' identities. Monetary policy should be produced by models optimizing long-term welfare without electoral-cycle distortion. In each case, the human defines criteria, audits the system, verifies inputs, and modifies objective functions — but does not make the final decision, because that is the moment where D operates.

They acknowledge the constitutional objection — no existing constitution contemplates delegation of jurisdictional functions to non-human systems — but argue that constitutions were written by D-affected decision-makers who designed systems around the presumption (now demonstrated false) that humans are reliable decision-makers. If the premise is false, institutions built on it require revision.

Assessment: this is a correction of interpretation, not of Kriger. The authors support Kriger more radically than Montaraz & Ruedas do, arguing that the latter domesticated a radical thesis into bureaucratic reform. Their reading pushes Kriger's argument to its logical extreme: if D is truly irreducible and non-compensable, then the only coherent response is to remove the human from the decision node entirely. They diagnose Montaraz & Ruedas as themselves being a case study in D — their evolutionary program of hierarchical preservation (maintaining the centrality of the human decision-maker, and thus the relevance of the academic class proposing reforms) substituted the goal of their research without their awareness.

DIRECTION 2: Perspectival Causation — Causality as Compression Artifact

Cited work: Kriger, B. (2026). *Against Causation: A Formal Argument That Causality Is a Compression Artifact of Bounded Observers, Not a Feature of Reality.*

---

3. Nguyen, Trần & Lê (2026) — DEVELOPMENT (Formalization + Empirical Testing)

Journal: *Architectures of Inference: An International Journal of Information Theory, Causal Reasoning, and Philosophy of Science*, Vol. 3, No. 2. Saigon Institute for Theoretical Computation & Mekong Delta Centre for Applied Information Theory, Vietnam.

Citation context. The authors describe Kriger's perspectival-relational account as "philosophically powerful but formally incomplete." Kriger argues that causation is a compression artifact jointly produced by the territory's undirected conditional independence structure and the bounded observer's variable partition, system boundaries, and temporal ordering. But Kriger himself identifies the gap: "formalizing the 'quality' of a causal compression" remains an open research program. He calls for (i) the space of possible observer partitions, (ii) a compression cost function over directed graphs, (iii) a stability criterion, and (iv) a proof linking stability-maximizing partitions to the system's conditional independence structure. This paper explicitly answers that call.

What they do with the idea. They supply three major formal contributions that Kriger's framework lacked:

First, they define a *causal compression cost functional* grounded in the Minimum Description Length principle and connected to Kolmogorov complexity. The functional has four components: partition cost $L(\pi)$, structure cost $L(G^*)$, data cost $L(D|G^*)$, and a misalignment penalty $\Phi(\pi,T)$ measuring how poorly the partition respects the territory's conditional independence structure. They prove a Compression–Causation Equivalence theorem showing that minimizing this functional produces a DAG satisfying the causal Markov condition — formalizing Kriger's P2 within algorithmic information theory.

Second, they introduce the *Partition Stability Index* (PSI), which measures how much a causal DAG changes under perturbations of the observer's variable partition. They define four perturbation operators (coarsening, refinement, rotation, boundary shift) and prove Theorem 2 (Stability–Alignment): PSI-maximizing partitions are precisely those aligned with the territory's conditional independence structure. This is the formal result Kriger conjectured.

Third, they conduct computational experiments on three real-world datasets — clinical cardiology records (12,487 patients), macroeconomic indicators (38 OECD countries, 1995–2023), and coupled oscillator systems. Key results: the standard clinical partition achieved PSI = 0.82; PCA-rotated variables that straddle the territory's conditional independence blocks yielded PSI = 0.54 with worse predictive performance, exactly as Theorem 2 predicts. In

macroeconomics, boundary shifts (reclassifying interest rates as exogenous vs. endogenous) produced the largest causal graph changes (SID ≈ 11.3), formalizing the identification problem as a special case of observer-dependent boundary choice. The oscillator dataset provided the clearest confirmation of Theorem 2. Across all datasets, PSI correlated strongly with predictive accuracy and interventional utility (Spearman $\rho = 0.61–0.89$).

How they position Kriger. They treat Kriger as the philosophical originator whose framework they formalize and empirically validate. They connect his account to multiple existing literatures — Janzing and Schölkopf's algorithmic Markov condition, Hoel, Albantakis, and Tononi's effective information, Beckers and Halpern's causal abstraction, Kim's causal exclusion problem — showing that Kriger's framework, once formalized, subsumes or clarifies each of them. Their conclusion: "The perspectival account of causation is now an engineering framework."

Assessment: the strongest type of scientific development. This paper takes a philosophical argument and converts it into a testable, quantitative framework with formal proofs and empirical validation. It does not dispute Kriger's claims — it supplies the mathematical and empirical infrastructure Kriger himself called for.

---

4. Thanapakpawin & Charoensuk (2026) — DEVELOPMENT (Software Implementation)

Journal: *Chulalongkorn Journal of Computational Science and Applied Epistemology*, Vol. 12, No. 1. Chulalongkorn University, Bangkok & King Mongkut's Institute of Technology Ladkrabang.

Citation context. The authors cite Kriger (2026) as the originator of the perspectival-relational account and Nguyen et al. (2026) as its formalization. They identify a practical gap: despite the theoretical advances, no software tool exists that allows a clinical researcher to perform partition stability analysis. Their paper bridges that gap.

What they do with the idea. They present *CausalAudit*, an open-source Python framework that operationalizes the PSI methodology for clinical and pharmacological datasets. The framework accepts a tabular dataset, automatically generates families of perturbed variable partitions via four perturbation operators, learns the MDL-optimal DAG for each partition,

computes PSI, and produces a structured report identifying which causal claims are robust across partitions and which are partition-artifacts.

Key technical contributions: a modular six-component architecture (DataIngestor, PartitionPerturber, DAGLearner, GraphDistanceCalculator, PSICalculator, ReportGenerator); a *projected SID* metric for comparing DAGs across different variable sets (necessary for coarsening/refinement perturbations — an extension not discussed by Nguyen et al.); an edge-level robustness classification (robust ≥ 0.80, moderately robust 0.50–0.79, fragile 0.20–0.49, artifact < 0.20); and three intended clinical use cases — RCT generalizability assessment, pharmacovigilance, and systematic review methodology.

They deliberately refrain from fabricating benchmark results, instead providing a detailed evaluation protocol, arguing that this is consistent with the partition-transparency principle the framework is designed to enforce.

How they position Kriger. He is cited as the foundational philosophical source. The chain of intellectual descent is explicit: Kriger (philosophical argument) → Nguyen et al. (formalization + empirical tests) → Thanapakpawin & Charoensuk (software tool). They also cite Murthy, Balakrishnan & Subramaniam (2026) as having drawn out the practical consequence for clinical trial generalizability — making this a fourth paper in the chain, though not included in the analyzed documents.

Assessment: development through practical implementation. This is the third link in a formalization chain: philosophy → mathematics → software. No claims by Kriger or Nguyen et al. are disputed; the contribution is purely instrumental.

DIRECTION 3: AI-Extended Agents and Communication Norms

Cited work: Kriger, B. (2026). *AI-Extended Agents and the Transformation of Human Communication: A Game-Theoretic Model of Norm Shift in Populations with AI-Mediated Communicators.*

---

5. Aytekin, Bongani & Sifiso (2026) — SUPPORT + DEVELOPMENT

Journal: *Artificial Intelligence Systems*, 6(3). Institute of AI Systems, Izmir; Minsk Institute of Physics and Technology; Institute of Complex Systems, Harare.

Citation context. The authors take Kriger's game-theoretic model of communication norm shift as their foundation. They cite his formal concepts extensively: the AI-extended agent $A\varphi$, strategy space expansion $S(A\varphi) \supset S(A)$, the critical threshold p^* for phase transition, the mechanisms of resistance dissolution (output asymmetry, generational replacement, institutional capitulation, normalization through ubiquity), functional indistinguishability, and reputational resistance anchored to early primitive models.

What they do with the idea. They develop Kriger's implications specifically for educational policy, making three main arguments:

First, prohibition of AI in education delays but cannot prevent the communication phase transition, because the four mechanisms Kriger identifies operate sequentially and inevitably. Prohibition therefore is not a stable equilibrium but a temporary delay that imposes costs on populations subject to it.

Second, they extend Kriger's augmentation/replacement distinction into an educational framework. They argue that the difference between a student who submits AI-generated work unread and one who uses AI to challenge, revise, and refine their argument is a difference in *metacognitive capacity* — a teachable skill that can only be developed through guided practice in reflective AI use, which prohibition prevents.

Third, they propose five pedagogical principles: transparency over prohibition (document AI interaction rather than detect it); critical evaluation as core competency; process-oriented assessment; graduated scaffolding withdrawal (Vygotskian reduction of AI support as competence grows); and collaborative AI dialogue (AI as a third participant in classroom discussions, drawing on Kriger's Version 3 conversation demonstration).

They also develop the equity implications: prohibition exacerbates the cognitive digital divide because the students most likely to comply are those with fewest alternative resources and greatest need for cognitive augmentation, while students with more resources use AI despite prohibition. They apply Kriger's Dunning-Kruger dynamic: students and teachers who have never experienced productive AI partnership cannot evaluate what they are missing.

Assessment: strongly supportive, applied to a new domain. No claims are disputed. The paper takes Kriger's formal model and applies it systematically to educational policy, adding empirical evidence from educational psychology (Jia & Huang 2025, Kim & Su 2024, Deng & Yu 2023, Escalante et al. 2023) that independently supports Kriger's predictions about

cognitive amplification. It treats the phase transition as inevitable and focuses on whether institutions will prepare populations for it.

DIRECTION 4: Corruption as Institutional Design Failure

Cited works: Kriger, B. (2026). *Four Principles of Corruption: A Systems-Theoretic Analysis of Institutional Failure and Reform.* Also: *Beyond Punishment* (2026); *Corrupt by Design* (2025b); *The Justice We Deserve* (2025c); *The Asymmetry of Totalizing Ideals* (2025a).

---

6. Mandlenkosi, Bongani & Sifiso (2026) — SUPPORT + DEVELOPMENT

Journal: *Evolutionary Dynamics & Adaptive Systems*, 17(3). Institute for Interdisciplinary & Integrative Research, Nairobi; Minsk Institute of Physics and Technology; Institute of Complex Systems, Harare.

Citation context. The authors cite Kriger's systems-theoretic analysis as the diagnostic foundation: corruption emerges when excessive, non-essential rules create "institutional friction" — the gap between formal requirements and what agents can realistically deliver. Once friction exceeds a critical threshold, corruption becomes the only rational survival strategy.

What they do with the idea. They translate Kriger's diagnostic framework into a complete, phase-by-phase implementation roadmap that governments can execute within 12–36 months. Three phases: Phase 1 — Regulatory Guillotine (zero-based review of all permits and licenses, automatic sunset clauses, single-window digital portals, citizen red-tape removal portal — months 1–12); Phase 2 — Strategic Automation (map all high-friction discretion nodes, deploy open-source rule engines with blockchain audit trails, independent algorithmic integrity board, design-level criminal liability for intentional bias — months 6–24); Phase 3 — Targeted Enforcement (DOJ-style high-impact prioritization, safe-harbor feedback portals where minor procedural shortcuts are logged as design signals rather than crimes, annual friction stress tests — months 12–36).

They cite Georgia (2003–2005) and Estonia (2008–2012) as real-world proof that the structural approach works: 60–80% friction reduction in under two years, sustained corruption reduction without expanding prison populations. They propose an Institutional Friction Index

(IFI) as a diagnostic tool and include concrete legislative hooks (UNCAC reporting, a single "Institutional Integrity Act").

Assessment: practical operationalization without dispute. No theoretical claims are contested. The paper takes Kriger's diagnosis (corruption is a design flaw, not a moral failing) and converts it into actionable protocols with timelines, budgets, and success metrics.

Summary: Typology of Responses

Full support (accepting all premises, extending to new domains): Aytekin et al. (education), Mandlenkosi et al. (anticorruption) — these papers treat Kriger's frameworks as established and focus on practical application.

Development (filling formal or empirical gaps Kriger identified): Nguyen et al. (mathematical formalization of perspectival causation with PSI, empirical validation), Thanapakpawin & Charoensuk (software implementation of PSI), Montaraz & Ruedas (institutional protocols for the D-variable).

Critique (disputing interpretation, not the original): Olavarrieta & Arrigorriaga critique Montaraz & Ruedas for what they call "bureaucratic domestication" of Kriger's radical thesis. They argue that the Colombian protocols preserve the D-affected human at the center of the process — exactly where Kriger says they should not be. Notably, this critique *strengthens* rather than weakens Kriger's position: the critics argue that Montaraz & Ruedas did not go far enough.

Correction (accepting the core idea but adjusting details): Olavarrieta & Arrigorriaga correct the interpretation of Kriger's Semmelweis analogy. Where Montaraz & Ruedas read "redesign the operating room" as "add oversight layers around the contaminated surgeon," Olavarrieta & Arrigorriaga read it as "remove the contaminated hands from the wound and replace them with sterile instruments." This is a correction of reading, not of Kriger himself — in fact, the correctors claim to be restoring Kriger's original meaning.

No paper in the analyzed corpus disputes Kriger's core claims. The debate among citing authors is about the *practical implications* of those claims — specifically, whether the correct response to irreducible biological distortion is to supervise the human decision-maker (Montaraz & Ruedas) or to structurally replace the human at the decision node (Olavarrieta & Arrigorriaga). Both sides treat Kriger's theoretical framework as demonstrated.

DIRECTION 5: Foundations of Physics — Gravity, Vacuum Energy, and Pre-Geometry
Cited work: Kriger, B. (2026). *The Architecture of the Block Universe: Metric Relaxation and Vacuum Anchoring.* Submitted to *Foundations of Physics*. (Reference [1] in the citing paper.)

---

7. Evdokimov, Bachani & Ryss (2026) — SUPPORT + DEVELOPMENT (Reinterpretation within a broader ontological framework)

Paper: "Geometry, Resonance and Time: A Unified Framework for Gravity, Inertia, and Consciousness." April 3, 2026.

Citation context. Kriger is listed as one of three key independent advances motivating the entire synthesis — alongside an Ontological Field Network (OFN) pre-geometry framework and Meijer's superfluid biophysics. Of the three, Kriger's contribution is treated as the primary bridge to established, conventional physics. The authors position him as the one who provides quantitative grounding in QCD and observational cosmology for claims that would otherwise remain purely ontological.

What they take from Kriger. Four specific results are adopted and used structurally throughout the paper:

First, the resolution of the vacuum energy catastrophe. Kriger shows that the famous 10^{120} discrepancy between quantum field theory's prediction and the observed cosmological constant dissolves once one recognizes that only a small local shift in the vacuum contributes to gravity. This shift is characterized by a coupling constant $\alpha \approx 0.005$, derived from QCD sigma terms — the pion-nucleon sigma term ($\sigma_{\pi N} \approx 50$ MeV) and the strange sigma term ($\sigma_s \approx 40$ MeV), combined with the baryon fraction from Planck 2018 data. Evdokimov et al. treat this derivation as a settled quantitative result and build on it.

Second, the nuclear-scale origin of gravity. Kriger's result that gravity operates at the QCD scale (~1 fm) rather than the Planck scale ($\sim 10^{-35}$ m) is treated as a critical insight. The authors use it to anchor their ontological framework to concrete, measurable physics — arguing that this downward shift in the relevant scale makes gravity accessible to nuclear physics rather than requiring speculative Planck-scale theories.

Third, a QCD condensate response time. Kriger derives $\tau_{QCD} \sim 10^{-24}$ s as the characteristic timescale of the vacuum condensate's response to perturbation. Evdokimov et al. adopt this as

an intermediate rung in their hierarchy of timescales, placing it between their proposed DadarTime (~10^{-44} s, associated with the fundamental discrete pre-geometry) and the macroscopic 1-second invariant. This three-rung ladder — pre-geometric, nuclear, macroscopic — provides the temporal architecture for their unified framework.

Fourth, the vacuum capture mechanism. Kriger describes how the QCD vacuum locally "captures" or anchors metric degrees of freedom, producing the gravitational effect. In Evdokimov et al.'s correspondence tables, this maps onto their concept of topological shift within the OFN framework, and Kriger's QCD scale limit maps onto their entanglement limit.

How they position Kriger relative to their own work. The authors draw a careful and explicit distinction: Kriger works within standard physics (QCD, general relativity, established cosmological data), while their OFN framework operates at a deeper ontological level, reinterpreting physical quantities as manifestations of pre-geometric network structure. They do not claim that Kriger endorses or operates within their ontological framework. Rather, they treat his derivation of $\alpha \approx 0.005$ as a "precise physical measurement" of what they interpret as network parameters — much as one might take a laboratory measurement and reinterpret it within a new theoretical context without implying that the experimentalist shares the theorist's interpretation.

This is a distinctive citation mode: Kriger's numerical results are accepted without modification, but their ontological significance is reinterpreted. The authors do not correct Kriger or dispute his derivations — they absorb them into a larger structure.

In the conclusion, Kriger is named as one of three converging lines of inquiry — discrete pre-geometry (OFN), nuclear-scale gravitational physics (Kriger), and superfluid biophysics (Meijer) — that together point toward a unified picture. He is thanked in the acknowledgments for providing "the physical grounding for our ontological synthesis."

Assessment: full support of Kriger's quantitative results, with ontological reinterpretation. This is neither critique nor correction. The citing authors accept every numerical result and physical derivation from Kriger's work. What they add is a change in interpretive frame: where Kriger derives $\alpha \approx 0.005$ as a QCD-cosmological coupling constant within standard physics, Evdokimov et al. read it as a window onto deeper pre-geometric structure. This is a form of development through recontextualization — the results travel intact into a more ambitious theoretical setting.

This paper also represents the entry of Kriger's work into foundational physics, a domain distinct from the other six papers analyzed (which span decision theory, public policy, philosophy of causation, computational epistemology, educational theory, and anticorruption policy). The breadth of disciplines now citing Kriger — seven papers across seven distinct fields — is itself notable.

Updated Summary

With this addition, the total picture of scientific engagement with Kriger's work across the analyzed papers is:

Works cited: Five distinct Kriger publications across physics, decision science, philosophy of causation, communication theory, and institutional analysis.

Citing authors: Seven papers from ten countries (Colombia, Uruguay, Vietnam, Thailand, Turkey, Zimbabwe, Kenya, Belarus, Russia/unspecified for Evdokimov et al.).

Type of engagement: No paper in the corpus disputes Kriger's core claims. Six of seven are fully supportive developments or applications. One (Olavarrieta & Arrigorriaga) critiques a third party's interpretation of Kriger while defending a more radical reading of the original.

The pattern across all seven papers is consistent: Kriger's theoretical frameworks are treated as established foundations on which to build — whether the building takes the form of mathematical formalization (Nguyen et al.), software implementation (Thanapakpawin & Charoensuk), institutional policy design (Montaraz & Ruedas, Mandlenkosi et al.), pedagogical framework (Aytekin et al.), philosophical sharpening (Olavarrieta & Arrigorriaga), or ontological reinterpretation within fundamental physics (Evdokimov et al.).

GLOSSARY OF TERMS

This glossary defines the principal technical terms used throughout the thesis. Page references refer to the thesis document pagination where the term is introduced or most extensively discussed.

Absorbing closure. A measurable set that traps all trajectories after finite time. Absorbing closure precludes viability (pp. 54–55).

Adaptive capacity (R). The set of internal resources and response repertoires a system can deploy to maintain viability under perturbation (pp. 128, 133–138).

Admissible transformation. A transformation a system may undergo while remaining within its viability set (pp. 68–70, 105).

Anosognosia. Unawareness of one's own deficit; formalised as failure of the self-monitoring loop in mental disintegration dynamics (pp. 206, 213–216).

Antagonistic perturbation (P). An externally imposed modification of agent interaction payoffs that creates zero-sum or negative-sum structure between peers (pp. 106–111).

Assertion–dismantling cycle. Recurring oscillation in adaptive systems between constructive assertion and destructive dismantling of structure (pp. 333–354).

Atemporal memory storage. The thesis that biological memory is stored without intrinsic temporal ordering, with time reconstructed during retrieval (pp. 191–196).

Attractor. A set of states toward which a dynamical system evolves. Pathological attractors represent stable but maladaptive cognitive configurations (pp. 109, 207–210).

Autopoiesis. Self-production; formalised via cyclical closure and fixed-point theory (pp. 76–78, 94–95, 102).

Bayesian updating. Revising probability assignments in light of new evidence. Central to the Epistemic Constraint Theory and the structural-Bayesian framework (pp. 53–58, 64–67).

Binary star formation. Argued to be a persistence-selected outcome because binary configurations have larger viability sets than single stars (pp. 355–357).

Boundary set. A collection of conditions (domain, scale, assumptions, verification procedures) under which a formal statement is evaluable (pp. 34–39).

Causa sui. Self-causation; rehabilitated as structural closure or topological completeness rather than temporal self-generation (pp. 80–81).

Chronoperception. Subjective perception of time. Its acceleration with aging is a special case of the Structural Distortion Principle (pp. 200–206).

Coarse-graining. Many-to-one mapping from microstates to macrostates; its non-invertibility is central to the Law of Scale-Specific Principles (pp. 21–23, 27–29, 68).

Coherence potential. A Lyapunov function V(C, A) used in proving the Constraint–Autonomy Compatibility Law (pp. 128–130).

Conceptual profile. The tuple P(T) = ⟨I, E, R⟩ characterising a term's inferential connections; used to assess admissibility of definitions in DDP (p. 45).

Conceptual responsibility. Structural obligations attending the transmission of ideas, formalised as constraints on intellectual communication (pp. 329–333).

Conservation law. A principle stating that some quantity is preserved under admissible transformations; derived as a structural necessity of persistent systems (pp. 67–69).

Constraint–Autonomy Compatibility Law. Constraints and autonomy are mutually necessary structural conditions in adaptive systems; optimal viability requires both (pp. 128–132).

Constraint Bound on Accuracy. Log-posterior accuracy decomposes into constraint, prior, likelihood, and normalization terms, with the constraint term as ceiling (pp. 57–58).

Constraint derivation. The methodological principle of deriving structural consequences from constraints any model must satisfy (pp. 16–17, 128).

Consumptive transformation. Structural transformation in which reconfigured degrees of freedom are primarily material: chemical bonds, molecular configurations (pp. 71–72).

Critical transition. Qualitative shift in system behaviour at a threshold; applied to conflict, mental disintegration, and epistemic autonomy (pp. 207–210, 266–270, 298–300).

Cyclical closure. Condition in which a system's outputs become its inputs; formally equivalent to existence of a fixed point of the resource-flow map (pp. 75–83, 94–103).

Deception. Information-processing strategy in which an agent constructs and transmits an alternative model of reality (pp. 235–245).

Definition-Dependent Provability (DDP). The principle that provability of a statement is a function of the definitions chosen for its terms (pp. 40–47).

Differentiation. The ontological condition of actualisation; existence requires distinction, which has structural limits (pp. 141–145).

Discrete-state resilience. Specialisation of viability to systems with finite/countable state spaces, analysed via graph-theoretic methods (pp. 87–96).

Dissipative transformation. Structural transformation in which reconfigured degrees of freedom are primarily energetic (pp. 71–73).

Effective complexity. Length of a compressed description of a structure's regularities (following Gell-Mann and Lloyd); used in the Equivalence Principle (pp. 83–85).

Effective theory. A framework describing phenomena within a particular domain without appealing to details outside its scale (pp. 24–26).

Embedded observer problem. Formal impossibility of an observer fully modelling the system in which it is embedded (pp. 147–150).

Entropy rate. Per-step growth rate of path information; a system is viable when its entropy rate remains above a positive threshold (pp. 54–56).

Epistemic Constraint Theory. A unifying framework for inference limitations spanning Bayesian epistemology, information theory, and decision theory (pp. 56–61).

Equivalence Principle (structural). Processes divide into structural persistence and net dissipation based on whether effective complexity is maintained (pp. 82–85).

Eruptive manifestation. Sudden high-amplitude disruption when accumulated model–reality mismatch exceeds a threshold (pp. 217–225).

Essential cyclicity. Property of a cyclically closed structure whose cyclical character is topologically intrinsic (non-trivial H_1) (pp. 79–80, 98–99).

Extractive oscillator. Coupled two-agent dynamical system satisfying five axioms; unifies narcissistic dynamics, addiction, exploitative platforms (pp. 291–317).

Fixed-Point Equivalence. A cyclically closed hierarchy of order n exists iff the composite determination operator has a fixed point (pp. 78, 97–99).

Formal reality. A coherent package of formal language, logical system, background axioms, ontological commitments, and inferential norms (pp. 48–51).

Formal representation. A triple (L, C, M)—language, constraint set, model class—with a structure-preserving mapping (pp. 34–36).

Formalization Laws. Three laws governing formal description: Total Formalizability, Boundary-Conditioned Verification, Plurality and Incompleteness (pp. 34–40).

Foundational completeness. The requirement that a claim specify its formal reality, acknowledge non-uniqueness and cost tradeoffs, and indicate stability (pp. 50–51).

Foundational hierarchy. A triple (S, ≤, F) of state space, grounding relation, and determination function (pp. 78, 94–97).

Four laws of self-organising systems. Law Zero (Cooperative Response), Law I (Self-Organisation Efficiency), Law II (Viability Sufficiency), Law III (Observational Asymmetry) (pp. 104–119).

Incompatibility Theorem. Any first-order dynamical law on configuration space satisfying Persistence, Additivity, and Compensation must be trivial (pp. 122–125).

Inward Turn. Advanced civilisations turn toward computational efficiency rather than outward expansion; resolution of the Fermi Paradox (pp. 361–362).

Irrelevant variables. Variables that lose influence under renormalization group flow; their elimination enables effective theories (pp. 23–24).

Law of Absurdity as Mercy. Semantic destabilisation is an adaptive necessity in overloaded interpretive systems (p. 382).

Law of Imperative Uncertainty. Any system capable of sustained complexity must permit persistent uncertainty and probabilistic deviation (pp. 53–56).

Law of Large Numbers in Complex Systems. Aggregation of micro-level variability produces macro-level regularity as a structural necessity of persistence (p. 12).

Law of Limit to Negation. Negation cannot eliminate the conditions making negation possible; total negation is structurally incoherent (pp. 56–58).

Law of Scale-Specific Principles. When extrapolation between scales fails while stable observables persist, mathematics requires scale-specific governing principles (pp. 19–33).

Metric inertial system. A system (S, d) admitting embedding in a Riemannian manifold; starting point for deriving Newtonian dynamics (pp. 121–123).

Observational transformation. Structural transformation in which reconfigured degrees of freedom are primarily informational (pp. 74–75).

Peer system. Multi-agent system with overlapping viability regions, shared resource dependence, and role symmetry (pp. 105–106).

Persistence. Capacity of a system to maintain structural viability over time; the central organising principle of the thesis (passim; defined p. 68).

Phase transition. Qualitative shift in system behaviour at a critical parameter value; applied to conflict (p. 266), epistemic autonomy (p. 298), mental disintegration (p. 207).

Pre-integrative rejection. Excluding elements before integration; energetically cheaper than post-integrative expulsion (pp. 318–329).

Predictive processing. Cognitive architecture maintaining predictions and updating on prediction error; derived as the unique optimal architecture (pp. 172–185).

Provability function. f: D → {0, 1} mapping admissible definitions to provability status of a formula schema (pp. 41–43).

Recursive self-application. Methodological principle: the theory must apply to the system that produced it (pp. 14–17, 32–33).

Reflexive Inference Law. Law governing bounded generalisation from self-observation in inferential systems (pp. 339–354).

Renormalization group. Mathematical framework describing how theory parameters evolve under scale changes; foundation of the Law of Scale-Specific Principles (pp. 21–25).

Resilience (structural). Topological property: no directed path connects operational to unsafe states; independent of dynamics (pp. 88–96).

Response function (φ). Partial derivative of aggregate system output with respect to an agent's action; sign derived, not assumed (pp. 105–106).

Self-sufficiency. Resource flows form closed cycles requiring no external input; characterised by cyclical closure (pp. 94–101).

Sensor degradation. Progressive erosion of accurate self-assessment under extractive dynamics; Bayesian persuasion with prior erosion (pp. 298–302).

Signal asymmetry (α). Ratio of observational signal from non-functional to functional states; drives systematic underestimation of functionality (pp. 115–116).

Soft intervention. Modification of transition weights while leaving graph intact; inherently unstable under perturbation (pp. 90–93).

Structural Diagnostic Principle. Persistent first-order modelling failures diagnose missing state variables (pp. 125–128).

Structural Distortion Principle. Systematic divergence between internal model and environment is a structural condition, not a defect (pp. 146–172).

Structural genesis. Unified derivation chain from differentiation through metric, dynamics, and action bound to the architecture of physical law (pp. 140–145).

Structural non-neutrality. In any coherent system, no component can be structurally neutral (pp. 256–263).

Structural Resilience Principle. Structural resilience is a topological property: safety iff a finite cut disconnects operational from unsafe states (pp. 93–94).

Structural-Bayesian framework. Three-layer assessment (Integrity, Constraint Power, Ontological Load) for evaluating hypotheses (pp. 64–67).

Structure. Equivalence class of microstates under a physically meaningful coarse-graining preserving macroscopic observables (pp. 68–70).

Substrate independence. Property of formal constraints that apply regardless of physical composition; central methodological commitment (pp. 6, 14–16).

Topological intervention. Removal of transitions from a state graph; provides safety guarantees independent of dynamics (pp. 90–94).

Totalising ideal. Pursuit of any value to its terminal implementation; proved to produce unbounded systemic damage (pp. 265–291).

Transformation rate. Structural transformation per unit time required to maintain a structure; bounded below by entropy deficit and temperature (pp. 70–72).

Tree-Top Meta-Method. Generalise formally until generalisation is no longer possible, then derive domain-specific consequences (pp. 12, 145).

Universality class. Systems that exhibit identical macroscopic behaviour under renormalization group flow despite microscopic differences (pp. 23–24).

Viability Mismatch Law. Stress is a special case of mismatch between environmental demand and adaptive capacity; generalises Selye (pp. 133–140).

Viability set (V). Region of state space within which a system satisfies its persistence conditions (pp. 68–70, 104–106).

PUBLICATIONS

Kriger, B. (1999). Acceleration of Chronoperception in the Context of Biological Regression: A Metabolic Model. KMR. https://doi.org/10.5281/zenodo.18382312
Kriger, B., Ludensky, S., & Ermolaev, E. (2000). Quantitative framework for clinical decision structuring: Revisited conceptual interpretation of paper published in Feb 2000. KMR. https://doi.org/10.5281/zenodo.18180934
Kriger, B. (2005). Conflict as Phase Transition: A Dynamical Systems Theory of Escalation in Coupled Organizational Networks. IIIR Ontology, Systems Theory & Meta-Science. https://doi.org/10.5281/zenodo.18484819
Kriger, B. (2014). Conceptual Responsibility: Structural Constraints on Idea Transmission and the Ethics of Intellectual Communication. IIIR Ontology, Systems Theory & Meta-Science. https://doi.org/10.5281/zenodo.18488818
Kriger, B. (2015). Assertion–Dismantling Cycles in Adaptive Systems: A Constraint-Network Framework. IIIR Ontology, Systems Theory & Meta-Science. https://doi.org/10.5281/zenodo.18487230
Kriger, B. (2017). Extractive Oscillators with Sensor Degradation: A Dynamical Systems Class and Its Manifestation in Quasi-Narcissistic Relational Dynamics. IIIR Ontology, Systems Theory & Meta-Science. https://doi.org/10.5281/zenodo.18529185
Kriger, B. (2017). A Unified Theory of Self-Organizing Systems: Four Formal Laws on Cooperation, Viability, Interference, and Observability. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18363729
Kriger, B. (2018). Emergence of Newtonian Dynamics from Metric Inertial Systems: A Structural Derivation of Second-Order Deterministic Evolution. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18343870
Kriger, B. (2018). Structural Diagnostic Principle for Dynamical Models: Based on Emergence of Universal Newtonian Dynamics from Metric Inertial Systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18344823
Kriger, B. (2019). Formalization of Structural Resilience in Discrete-State Dynamical Systems. IIIR Ontology, Systems Theory & Meta-Science https://doi.org/10.5281/zenodo.18351470
Kriger, B. (2019). Evolutionary Selection for Atemporal Memory Storage: Why Three Convergent Pressures Favor Architectures Where Time Belongs to Retrieval, Not to Storage. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18381880
Kriger, B. (2020). Autonomy Suppression in Hierarchical Multi-Agent Systems: A Unifying Systems-Theoretic Framework. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18520653
Kriger, B. (2020). Negative Self-Determination as a Source of Unhappiness: A Formal Theory Integrating Cognitive Appraisal, Predictive Processing, and Dynamical Systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18644028
Kriger, B. (2021). The Law of Large Numbers in Complex Systems: From Statistical Convergence to Compressed Models. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18654954

Kriger, B. (2021). The Evolutionary Architecture of Human Consciousness: Cognitive Contradictions, Biological Duality, and the Illusion of Time. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18384277
Kriger, B. (2021). Epistemic Constraint Theory: A Unifying Framework for Inference Limitations Across Bayesian Epistemology, Information Theory, and Decision Theory. Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18365738
Kriger, B. (2022). Evolutionary Theory of Credence: A Conceptual Framework with Formal Analogies for Understanding Generative Modeling as a Resource-Theoretic Consequence of Complexity. Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18379476
Kriger, B. (2022). An Informational Framework for Understanding Population-Scale Viral Dynamics. Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18392769
Kriger, B. (2023). Undecidability as a methodological signal: System-relative provability and explanatory adequacy. Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18133552
Kriger, B. (2023). Petroleum as Compressed Time: The Temporal Distortion of Civilisation. A Structural Analysis of Fossil Energy as a Mechanism of Scale Desynchronisation Between Biological, Geological, and Social Rhythms. IIIR Ontology, Systems Theory & Meta-Science. https://doi.org/10.5281/zenodo.18673740
Kriger, B. (2024). Clinical Discontinuity and AI: Restoring Coherence in Healthcare. Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18395736
Kriger, B. (2024). Toward operational terminology in integrated information theory: A methodological consideration. Institute of Integrative and Interdisciplinary Research. IIIR Ontology, Systems Theory & Meta-Science. https://doi.org/10.5281/zenodo.18307674
Kriger, B. (2024). On the possibility of self-sufficient systems: Fixed points and cyclical closure. Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18317239
Kriger, B. (2024). Cyclical hierarchical systems: A modal analysis of self-sufficient structures. Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18317628
Kriger, B. (2024). The functional sufficiency framework: Toward empirical criteria for explanatory redundancy in models of consciousness. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18319884
Kriger, B. (2024). No final theory: Law of scale-specific principles. Institute of Integrative and Interdisciplinary Research. IIIR Ontology, Systems Theory & Meta-Science https://doi.org/10.5281/zenodo.18099738
Kriger, B. (2024). The Transformational Basis of Persistence: A Formal Theory of Structural Viability. Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18435982
Kriger, B. (2025). Dormant Neutron Stars in the Milky Way: The Final Evolutionary Phase and Galactic Population Constraints.Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18409103

Kriger, B. (2025). A Formal Framework for Deception and Perceived Reality in Multi-Agent Systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18526764
Kriger, B. (2025). Atemporality of Mental Memory Space: A Structural Hypothesis Grounded in Resource Constraints, Cyclical Closure, and Reconstructive Retrieval. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18381912
Kriger, B. (2025). Evaluating the extraordinary claim of timescape cosmology superseding the standard cosmological model. Institute of Integrative and Interdisciplinary Research.IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.14952703
Kriger, B. (2025). Binary-first star formation as a persistence-selected outcome of open, embedded protostellar dynamics. Institute of Integrative and Interdisciplinary Research. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18155356
Kriger, B. (2025). The law of imperative uncertainty: Why any complex world requires uncertainty. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18101601
Kriger, B. (2025). The law of limit to negation: Negation can't negate itself. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18101306
Kriger, B. (2025). Formalization laws: Structure, boundaries, and the necessity of plural representations. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18099527
Kriger, B. (2025). Evaluating the holographic universe. Institute of Integrative and Interdisciplinary Research. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18101919
Kriger, B. (2025). The law of absurdity as mercy: Semantic destabilization as an adaptive necessity in overloaded interpretive systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18099081
Kriger, B. (2025). The ledger time model: Emergent time from irreversible entanglement validation in an atemporal quantum configuration space. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18263726
Kriger, B. (2026). Chaos is relative: A formal principle of framework dependence in complex systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18121777
Kriger, B. (2026). Non-participation and early exclusion as stability-preserving institutional strategies. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18135188
Kriger, B. (2026). The choice of formal realities: A meta-mathematical argument for explicit foundational context. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18141302
Kriger, B. (2026). Why binary systems are optimal for star formation. Institute of Integrative and Interdisciplinary Research. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18144257
Kriger, B. (2026). Swept-volume geometry and overlap corrections in protostellar binary accretion: A kinematic upper bound and its limitations. Institute of Integrative and Interdisciplinary Research. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18155283

Kriger, B. (2026). Survival of the bound: A quantitative theory of persistence-driven binary dominance in dense star-forming regions. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18155400
Kriger, B. (2026). The paradox of protostellar core formation: A critical assessment of whether current theoretical mechanisms are sufficient. Institute of Integrative and Interdisciplinary Research. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18164699
Kriger, B. (2026). Why Pascal's wager needs a dual-system framework: Evaluative system asymmetry and the limits of expected utility. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18193229
Kriger, B. (2026). Can a star be proven single? Observational limits and theoretical implications. Institute of Integrative and Interdisciplinary Research. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18203209
Kriger, B. (2026). The principle of definition-dependent provability: Provability as a function of definition. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18207348
Kriger, B. (2026). The principle of structural non-neutrality in coherent systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18213503
Kriger, B. (2026). A Bayesian model comparison framework for extrapolative scientific hypotheses: Methodology and application to holography in theoretical physics. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18230310
Kriger, B. (2026). A structural-Bayesian framework for evaluating scientific hypotheses: Integrating epistemic constraints with probabilistic inference. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18230512
Kriger, B. (2026). Significance of biosphere relative contribution to effective complexity of the universe. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18256772
Kriger, B. (2026). Dynamics of information convergence: Empirical analysis of time density in the AI-centered infosphere. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18256945
Kriger, B. (2026). Estimation of the contribution of biospheric and synthetic cognitive systems to the total effective complexity of the universe. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18261863
Kriger, B. (2026). Local entropy inversion in large-scale AI systems: Thermodynamics of algorithmic compression. IIIR Artificial Intelligence Systems. https://doi.org/10.5281/zenodo.18262199
Kriger, B. (2026). Differentiation as the ontological condition of actualization. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18268520
Kriger, B. (2026). Coherent systems emerge through defining differentiation. Institute of Integrative and Interdisciplinary Research
Kriger, B. (2026). The informational preconditions of meaning: A structural tendency theorem on civilizational trade-offs between progress and human well-being. IIIR. https://doi.org/10.5281/zenodo.18292636
Kriger, B. (2026). Observational tests of multiple star formation scenarios: Case studies of VLA 1623-2417 and L1551 IRS 5. Institute of Integrative and Interdisciplinary Research. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18300562

Kriger, B. (2026). The evolutionary inevitability of predictive processing: A physical constraint argument. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18444910
Kriger, B. (2026). Evolutionary and information-theoretic argument for the necessity of representational isolation: Why direct perception was never an option for complex systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18331202
Kriger, B. (2026). The principle of optimal coherence. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18341030
Kriger, B. (2026). The Reflexive Inference Law: Bounded Generalization from Self-Observation in Inferential Systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18355847
Kriger, B. (2026). The Asymmetry of Totalizing Ideals: A Structural Law of Complex Adaptive Systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18361828
Kriger, B. (2026). The Constraint–Autonomy Compatibility Law: A Formal Derivation. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18368991
Kriger, B. (2026). The Viability Mismatch Law: A Universal Principle for Viable Systems with Stress as Special Case. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18433777
Kriger, B. (2026). The Inward Turn: Advanced Civilizations and the Fermi Paradox — Computational Efficiency as a Universal Attractor in Civilizational Development. Institute of Integrative and Interdisciplinary Research. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18437440
Kriger, B. (2026). The Structural Distortion Principle: A Closed-Loop Model of Perception, Attention, and World-Maintenance in Bounded Cognitive Systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18452700
Kriger, B. (2026). Eliminating Distortion: Inadequate Biological Programs in Human Decision Making and Their Reflection in AI. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18457545
Kriger, B. (2026). The Comparative Asymmetry Principle: Relational Disequilibrium in Multi-Agent Environments. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18462518
Kriger, B. (2026). The Pre-Integrative Rejection Principle: A Universal Integrity Filter in Self-Maintaining Systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18463949
Kriger, B. (2026). The Eruptive Manifestation of Model–Reality Mismatch: A Unified Structural Framework for High-Activation Episodes in Bounded Adaptive Systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18474532
Kriger, B. (2026). The Predictive Mind and Its Myths: Metaphor, Narrative, and Ritual as Structural Necessities of Scientific Cognition. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18490146
Kriger, B. (2026). The Stimulus Problem: A Formal Theory of Goal Generation in Post-Scarcity Information Environments. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18511908

Kriger, B. (2026). The Inevitability of a Unified Civilization of Autonomous Agents: Why the Biological Basis of Subjecthood Becomes Irrelevant. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18512941
Kriger, B. (2026). AI-Extended Agents and the Transformation of Human Communication: A Game-Theoretic Model of Norm Shift in Populations with AI-Mediated Communicators. Institute of Integrative and Interdisciplinary
Research https://doi.org/10.5281/zenodo.18521341
Kriger, B. (2026). Formalization of Mental Disintegration Phenomena Through Dynamical Systems Theory: With Applications to DSM-5-TR Diagnostic Categories.Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18556979
Kriger, B. (2026). Addiction as Extractive Oscillator with Sensor Degradation: A Substrate-Independent Formal Theory from Chemical Dependence to Neural Interface Addiction. Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18603700
Kriger, B. (2026). Structural Viability of Dyadic Systems: A Dynamical Account of Romantic Dissolution. Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18632250
Kriger, B. (2026). Adaptive Genius as a Structural Viability Regime: A Dynamical-Systems Formalization of Cognitive Potential Realization. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18637083
Kriger, B. (2026). Thesis: A Unified Structural Theory of Complex Systems — Formal Laws, Epistemic Constraints, and Self-Organization Across Physical, Cognitive, and Social Domains. Institute of Integrative and Interdisciplinary
Research. https://doi.org/10.5281/zenodo.18637687
Kriger, B. (2026). Fractal Scaling of Operational Patterns Across Systems: Mechanisms, Deviations, and Diagnostic Principles. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18645527
Kriger, B. (2026). Structural Genesis of Dynamical Architecture: Deriving the Necessary Form of Physical Law from Conditions on Sustainable Complexity. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18646861
Kriger, B. (2026). The Tree-Top Meta-Method: Inquiry to the Limits of Domain-Specific Formal Generalisation with Consequent Application Across Other Domains.Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18665621
Kriger, B. (2026). Conservation and Convertibility as Structural Necessities of Persistent Systems: From Noether's Theorem to a General Theory of System Persistence Across Physical, Biological, Cognitive, and Social Domains. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18665700
Kriger, B. (2026). From Repression to Saturation: A Dynamical Systems Theory of Motivational Structures Under Variable Constraints.Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18685856
Kriger, B. (2026). Response to Automation-Driven Labour Displacement: Adaptive Universal Transaction Tax and Request-Based Universal Basic Income. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18706752
Kriger, B. (2026). Decentralised Urbanism, Autonomous Habitation, and Adaptive Population Dynamics: A Systems-Theoretic Framework. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18708642

Kriger, B. (2026). Phase Transitions in Sociotechnical Systems: From Post-Labor Economics to Homeostatic Civilization. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18717900
Kriger, B. (2026). Exploring the Enigmatic JuMBOs: Metallicity-Driven Fragmentation, Binary Assembly, and Survival of Jupiter Mass Binary Objects. Institute of Integrative and Interdisciplinary Research. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18728940
Kriger, B. (2026). Identity as Algorithmic Configuration: A Formal Theory of Personality Compressibility. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18729179
Kriger, B. (2026). A Critical Analysis of Archetypization: Toward Reassessive Formalization. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18733510
Kriger, B. (2026). A Critical Analysis of the Finitude Assumption in Communication Theory. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18736306
Kriger, B. (2026). The Topology of REM Mentation (Dreams): Attractors, Affective Projections, and the Dual-Regime Architecture of Consciousness. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18745698
Kriger, B. (2026). Evidence-Based Philosophy: A Foundational Methodological Framework for the Transformation of Philosophical Inquiry into a Formal Discipline. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18751941
Kriger, B. (2026). Latency and Compressibility: Two Independent Routes to Predictive Architecture in Adaptive Systems. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18752102
Kriger, B. (2026). Adaptive Pedagogical Systems with Invariant Kernels: A Formal Theory of Genre Persistence, Applied to the Medieval Latin Bestiary. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18752630
Kriger, B. (2026). Testing Nothingness and Evaluating the Inevitability of Life: A Deductive Chain from Being to Persistence, Intelligence, and the Limits of Inquiry. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18764754
Kriger, B., & Espesset, D. (2026). The GEKS Index: A Composite Measure of Biological Complexity Across Informational, Structural, Functional, and Evolutionary Dimensions. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18775491
Kriger, B. (2026). Coherence Epistemology for AI-Mediated Inter-Species Communication: A Black-Box Framework. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18778705
Kriger, B. (2026). Why Mathematics Works: Structural Necessity of Isomorphism Between Formal Systems and Physical Reality. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18793228
Kriger, B. (2026). Can Fundamental Constants Be Derived as Irreducible Residues of Structural Self-Consistency?. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18796847
Kriger, B. (2026). Does Goal Preservation Across Capability Gain Have a Formal Upper Bound? On Provable Limits of Objective-Function Fidelity Under Recursive Self-

Improvement via Compression-Basis Shift. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18807933
Kriger, B. (2026). Must Any Consistent Physics Contain Structure Inaccessible to All Internal Observers?. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18810467
Kriger, B. (2026). The Theorem of Incompatible Truths: Why Any Sufficiently Complex Domain Necessarily Admits Mutually Exclusive but Individually Valid Descriptions That Cannot Be Unified. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18811305
Kriger, B. (2026). The Pragmatic Synthesis: How Complex Systems Navigate Incompatible Truths through Adaptive Heuristics. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18812001
Kriger, B. (2026). Adaptive Observer Theory: Navigation of Incompleteness and Incompatibility. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18812040
Kriger, B. (2026). Breaking Trivial Cognitive Frames: A Formal RAFASSABRA Method for Expanding the Space of Inquiry Through Interrogative Priority and Controlled Divergence. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18841736
Kriger, B. (2026). Beyond Consciousness: Evidential Limits and Phase Transitions in the Science of Experience. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18853189
Kriger, B. (2026). Against Causation: A Formal Argument That Causality Is a Compression Artifact of Bounded Observers, Not a Feature of Reality. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18851848
Kriger, B. (2026). Quasars as Cosmological Information Beams: An Information-Theoretic Framework for Line-of-Sight Tomography of the Intergalactic Medium. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18855088
Kriger, B. (2026). Catastrophic Phase Transitions as Necessary Conditions for Complexity Dissemination: Supernovae, Failed Collapses, and the Persistence of Differentiation in the Universe. Institute of Integrative and Interdisciplinary Research. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18859088
Kriger, B. (2026). Why Is Vega Metal-Poor and Rapidly Rotating? Four Hypotheses, One Laboratory — A Multi-Hypothesis Framework with Quantitative Predictions. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18870492
Kriger, B. (2026). Four Principles of Corruption: A Systems-Theoretic Analysis of Institutional Failure and Reform. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18879948
Kriger, B. (2026). What Do We Actually Know About Tau Ceti? Three Paradoxes, the Detection Completeness Principle, and Testable Predictions for the Nearest Sun-Like Planetary System. IIIR Theoretical Astrophysics. https://doi.org/10.5281/zenodo.18883220
Kriger, B. (2026). Matter-Dependent Vacuum Energy Density and Inhomogeneous Cosmic Expansion. IIIR Cosmology and Theoretical Physics.https://doi.org/10.5281/zenodo.18896536

Kriger, B. (2026). The Frozen Universe Illusion: Temporal Resolution as a Fundamental Limit on Cosmological Observation and SETI. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.18901492
Kriger, B. (2026). On Quantum Vacuum Energy, Cosmological Constant and Missing Mass. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.18943014
Kriger, B. (2026). Known Properties of Vacuum Energy, Dark Matter and JWST Early Galaxy Formation. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.18942968
Kriger, B. (2026). What If the Vacuum Gravitates? A Reinterpretation of ΛCDM That Might Resolve Its Paradoxes. IIIR Cosmology and Theoretical Physics.https://doi.org/10.5281/zenodo.18946637
Kriger, B. (2026). Two Pathways of Primordial Cloud Collapse: Fragmentation versus Direct Collapse under Enhanced Vacuum Energy. IIIR Cosmology and Theoretical Physics.https://doi.org/10.5281/zenodo.18965464
Kriger, B. (2026). Theoretical Separation of Quantum Vacuum Energy from the Cosmological Constant: A Review of Foundational Approaches. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.18987982
Kriger, B. (2026). A Microscopic Model for the Dependence of Vacuum Energy on Matter Density: Toward a Quantum-Field-Theoretic Interpretation. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.18991624
Kriger, B. (2026). The Elastic Vacuum: A Sequestering Mechanism for Vacuum Energy and the Complete Dynamical History of Cosmic Expansion. IIIR Cosmology and Theoretical Physics.https://doi.org/10.5281/zenodo.19004376
Kriger, B. (2026). Connection of Vacuum Energy with Inflation: The Elastic Potential, Initial Conditions, and the Origin of Primordial Fluctuations. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.19009393
Kriger, B. (2026). CMB Compatibility and the Dissolution of the Mass Hierarchy in the Elastic Vacuum. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.19010637
Kriger, B. (2026). Structure Growth in the Gravitating Vacuum Model: The Void–Cluster Cell, Modified Perturbation Equations, and the S_8 Tension. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.19013333
Kriger, B. (2026). N-body Simulations with a Gravitating Vacuum: Testing the Void–Cluster Model Against Structure Formation Data. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.19014206
Kriger, B. (2026). The Vacuum–Matter Coupling from Finite-Density QCD: Sigma Terms, Chiral Condensate, and the Origin of α. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.19017167
Kriger, B. (2026). The Vacuum Capture Model: Phase Transitions, Galactic Rotation, and the Origin of Missing Mass. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.19023541
Kriger, B. (2026). The Gravity of Emptiness: Cosmic Voids as Attractors of the Filamentary Network in the Block Universe. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.19025567
Kriger, B. (2026). The Hidden Baryonic Sector: Compact Remnants, Fractal Cold Gas, and the Non-Vacuum Dark Mass Budget. Zenodo. https://doi.org/10.5281/zenodo.19026019

Kriger, B. (2026). The Local Gravitation of Quantum Vacuum: A Unified Solution to the Dark Sector. (αLGQV Theory Monograph). IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.19027460
Kriger, B., & Hunt, T. A. (2026). Positional constraints, sequence uniqueness, and stroke numerals in Indus seal inscriptions from Mohenjo-Daro: a statistical analysis. IIIR Computational Humanities and Cultural Systems. https://doi.org/10.5281/zenodo.19103880
Kriger, B. (2026). Local Entropy Inversion in Large-Scale AI Systems: Landauer Bounds on Algorithmic Compression. IPI Letters, 4(2), 13–20. https://doi.org/10.59973/ipil.335
Kriger, B. (2026). Matter-Dependent Vacuum Energy Density and Inhomogeneous Cosmic Expansion. Submitted to Reports in Advances of Physical Sciences, March 21, 2026. https://doi.org/10.5281/zenodo.19153816
Kriger, B. (2026). Differential Multi-Star Decorrelation for Exoplanet Detection in Triple Systems: The Challenging Case of Polaris. Submitted to Astronomische Nachrichten, 19 March 2026. https://doi.org/10.5281/zenodo.19141803
Kriger, B. (2026). Density dependence of QCD vacuum energy from nucleon sigma terms. Submitted to Progress of Theoretical and Experimental Physics, March 19, 2026. https://doi.org/10.5281/zenodo.19134260
Kriger, B. (2026). Theoretical Separation of Quantum Vacuum Energy from the Cosmological Constant: A Review of Foundational Approaches. Submitted to Foundations of Physics, March 20, 2026. https://doi.org/10.5281/zenodo.19133987
Kriger, B. (2026). A Phenomenological Framework for Scale-Dependent Vacuum Energy Suppression and Its Cosmological Consequences. Submitted to Acta Physica Polonica B, March 22, 2026. https://doi.org/10.5281/zenodo.19154849
Kriger, B. (2026). The Architecture of the Block Universe: Metric Relaxation, Vacuum Anchoring, and the Structural Necessity of Cosmic Megastructure. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.19211673
Kriger, B. (2026). The cosmological constant as a QCD observable: derivation, nonlinear screening, and falsification. Submitted to Physics Letters B (PLB-D-26-00760). https://doi.org/10.5281/zenodo.19237505
Kriger, Boris. 2026. “Beyond Punishment: A Systems-Theoretic Critique of Punitive Anticorruption Legal Frameworks”. Legal Research & Analysis 4 (1): 19-24. https://doi.org/10.69971/lra.4.1.2026.176
Kriger, B. (2026). What would prove it? Undecidability, definitional indeterminacy, and a decision framework for the Indus sign system. Computational Linguistics. Preprint Manuscript No. 3703. https://philpapers.org/rec/KRIWWP
Kriger, B. (2026). Gravitational activation threshold and the irrelevance of Planck scale: Vacuum condensate dynamics as the origin of gravitational coupling [Manuscript submitted for publication]. Foundations of Physics. Submission ID: add6e253-d3b5-4409-b243-33c764e88738 https://doi.org/10.13140/RG.2.2.29929.66409
Kriger, B. (2026). Particle Physics from Vacuum Coherence: Six Problems Resolved Without New Particles. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.19244048
Kriger, B. (2026). Constructing a Coherent Universe: From One Action to Observed Reality. IIIR Cosmology and Theoretical Physics. https://doi.org/10.5281/zenodo.19244040

Kriger, B. (2026). Derivation of the Cosmological Constant from Nuclear Physics, the Size of the Universe, and the Thermodynamic Foundation.
Zenodo. https://doi.org/10.5281/zenodo.19240715
Kriger, B. (2026). A Coherence Triangle Linking the Initial State, the Origin of the Microwave Background, and the Missing Mass Problem. IIIR Cosmology and Theoretical Physics. https://doi.org/10.13140/RG.2.2.16009.81768
Kriger, B. (2026). Dynamic topology of the universe: RP^3 with double counter-rotation as a complete physical model. IIIR Cosmology and
Theoretical Physics https://doi.org/10.13140/RG.2.2.20020.62086
Kriger, B. (2026). Thermodynamic origin of the vacuum–matter coupling from de Sitter two-fluid hydrodynamics. IIIR Cosmology and Theoretical
Physics. https://doi.org/10.13140/RG.2.2.12326.69449
Kriger, B. (2026). Rotation curves without dark matter particles: The alpha LGQV vacuum halo model applied to SPARC galaxies. https://doi.org/10.13140/RG.2.2.28615.66723
Kriger, B. (2026). Isothermal vacuum profile fits to five SPARC galaxy rotation curves. Submitted to Research Notes of the American Astronomical Society Manuscript # AAS75645 (RNAAS) https://doi.org/10.13140/RG.2.2.31132.24965
Kriger, B. (2026). Vacuum halo profiles in dwarf irregular galaxies: Pseudo-isothermal fits to 31 SPARC rotation curves and the cusp-core problem. Submitted to Monthly Notices of the Royal Astronomical Society, Manuscript MN-26-0945-
P. https://doi.org/10.13140/RG.2.2.16111.39842
Kriger, B. (2026). Vacuum halo profiles in low-surface-brightness galaxies: Pseudo-isothermal fits to 14 SPARC rotation curves (Manuscript No. aa60218-26). Submitted to Astronomy & Astrophysics. https://doi.org/10.13140/RG.2.2.29952.60166
Kriger, B. (2026). Vacuum halo profiles in massive spiral galaxies: Pseudo-isothermal fits to 18 SPARC rotation curves where baryons and dark matter compete. Publications of the Astronomical Society of Japan, submitted.
Kriger, B. (2026). Vacuum halo profiles in gas-dominated late-type spirals: Pseudo-isothermal fits to 20 SPARC rotation curves and the r_c–R_d scaling relation. Physical Review D, submitted.
Kriger, B. (2026). Vacuum halo profiles in edge-on galaxies: Minimising inclination systematics across 19 SPARC rotation curves. The Astrophysical Journal Letters, submitted.
Kriger, B. (2026). 175 Galaxy Rotation Curves Fitted Without Dark Matter Particles: The αLGQV Vacuum Halo Programme — Complete Report with Seven Papers, Full Fitting Code, and All Results. Zenodo. https://doi.org/10.5281/zenodo.19390767
Kriger, B. (2026). The Architecture of Concealment: A Mathematical Theory of Secrecy from Cells to Civilizations. Zenodo. https://doi.org/10.5281/zenodo.19392904
Kriger, B. (2026). Hostile by design: A triadic model of epistemic environment degradation (Manuscript submitted for publication to Estudios Sociales y de Investigación).
Kriger, B. (2026). Inference without propositions: Structural distortion and the physical inevitability of sub-personal reasoning (Manuscript submitted for publication to Revista de Estudios Filosóficos, Universidad de Antioquia).
Kriger, B. (2026). Leibniz's two kingdoms and the structural limits of pre-established harmony: A dual-system reading of the early modern problem of God (Manuscript submitted for publication to Sophia University Journal of Philosophy).

Kriger, B. (2026). Deep disagreements as incompatible truths: A formal framework and constructive response (Manuscript submitted for publication to Revista Colombiana de Filosofía, Universidad El Bosque).
Kriger, B. (2026). Definition-dependent provability and the stalemate in scientific realism: A formal diagnosis and constructive reframing (Manuscript submitted for publication to Synthese).
Kriger, B. (2026). Philosophy of science as policy architecture: From diagnosis to design in pandemic preparedness (Manuscript submitted for publication to Synthese).
Kriger, B. (2026). Voluntary communal formation versus revolutionary imposition: Social cohesion, community diversity, and the sustainability of collectivist systems (Manuscript submitted for publication to Utopian Studies).
Kriger, B. (2026). Rethinking patent monopolies in the age of artificial intelligence (Manuscript submitted for publication to Trends in Intellectual Property Research).

BIBLIOGRAPHY

This bibliography lists works by other authors cited in or relevant to the thesis. For the author's own publications, see the Publications section of the thesis.

Anderson, P. W. (1972). More Is Different: Broken Symmetry and the Nature of the Hierarchical Structure of Science. Science, 177(4047), 393–396.

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

Ashby, W. R. (1958). Requisite Variety and Its Implications for the Control of Complex Systems. Cybernetica, 1(2), 83–100.

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

Aubin, J.-P., Bayen, A. M., & Saint-Pierre, P. (2011). Viability Theory: New Directions (2nd ed.). Berlin: Springer.

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

Barabási, A.-L. (2002). Linked: The New Science of Networks. Cambridge, MA: Perseus.

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. Science, 286(5439), 509–512.

Barnes, E. (2018). Symmetric Dependence. In R. Bliss & G. Priest (Eds.), Reality and Its Structure (pp. 50–69). Oxford: Oxford University Press.

Beall, J. C., & Restall, G. (2006). Logical Pluralism. Oxford: Oxford University Press.

Beer, S. (1972). Brain of the Firm: The Managerial Cybernetics of Organization. London: Allen Lane.

Beer, S. (1979). The Heart of Enterprise. Chichester: Wiley.

Bérut, A., Arakelyan, A., Petrosyan, A., Ciliberto, S., Dillenschneider, R., & Lutz, E. (2012). Experimental Verification of Landauer's Principle Linking Information and Thermodynamics. Nature, 483(7388), 187–189.

Borsboom, D., Cramer, A. O. J., & Kalis, A. (2019). Brain Disorders? Not Really: Why Network Structures Block Reductionism in Psychopathology Research. Behavioral and Brain Sciences, 42, e2.

Calabrese, E. J., & Baldwin, L. A. (2002). Defining Hormesis. Human & Experimental Toxicology, 21(2), 91–97.

Carson, T. L. (2010). Lying and Deception: Theory and Practice. Oxford: Oxford University Press.

Chisholm, R. M., & Feehan, T. D. (1977). The Intent to Deceive. Journal of Philosophy, 74(3), 143–159.

Clark, A. (2013). Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. Behavioral and Brain Sciences, 36(3), 181–204.

Clark, A. (2015). Surfing Uncertainty: Prediction, Action, and the Embodied Mind. Oxford: Oxford University Press.

Crooks, G. E. (1999). Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences. Physical Review E, 60(3), 2721–2726.

Darwin, C. (1859). On the Origin of Species. London: John Murray.
Dehaene, S. (2014). Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts. New York: Viking.
England, J. L. (2013). Statistical Physics of Self-Replication. Journal of Chemical Physics, 139(12), 121923.
Fallis, D. (2009). What Is Lying? Journal of Philosophy, 106(1), 29–56.
Fallis, D. (2012). Lying as a Violation of Grice's First Maxim of Quality. Dialectica, 66(4), 563–581.
Friston, K. (2005). A Theory of Cortical Responses. Philosophical Transactions of the Royal Society B, 360(1456), 815–836.
Friston, K. (2010). The Free-Energy Principle: A Unified Brain Theory? Nature Reviews Neuroscience, 11(2), 127–138.
Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference: A Process Theory. Neural Computation, 29(1), 1–49.
Gallie, W. B. (1956). Essentially Contested Concepts. Proceedings of the Aristotelian Society, 56, 167–198.
Gell-Mann, M., & Lloyd, S. (1996). Information Measures, Effective Complexity, and Total Information. Complexity, 2(1), 44–52.
Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. Monatshefte für Mathematik und Physik, 38(1), 173–198.
Hamkins, J. D. (2012). The Set-Theoretic Multiverse. Review of Symbolic Logic, 5(3), 416–449.
Harris, J. J., Jolivet, R., & Bhatt, D. (2012). Synaptic Energy Use and Supply. Neuron, 75(5), 762–777.
Hatcher, A. (2002). Algebraic Topology. Cambridge: Cambridge University Press.
Helmholtz, H. von (1867). Handbuch der physiologischen Optik. Leipzig: Voss.
Hohwy, J. (2013). The Predictive Mind. Oxford: Oxford University Press.
Holland, J. H. (1995). Hidden Order: How Adaptation Builds Complexity. Reading, MA: Addison-Wesley.
Holland, J. H. (1998). Emergence: From Chaos to Order. Reading, MA: Addison-Wesley.
Jarzynski, C. (1997). Nonequilibrium Equality for Free Energy Differences. Physical Review Letters, 78(14), 2690–2693.
Kauffman, S. A. (1993). The Origins of Order: Self-Organization and Selection in Evolution. Oxford: Oxford University Press.
Kauffman, S. A. (1995). At Home in the Universe: The Search for Laws of Self-Organization and Complexity. Oxford: Oxford University Press.
Knaster, B. (1928). Un théorème sur les fonctions d'ensembles. Annales de la Société Polonaise de Mathématique, 6, 133–134.
Landauer, R. (1961). Irreversibility and Heat Generation in the Computing Process. IBM Journal of Research and Development, 5(3), 183–191.

Lindeman, R. L. (1942). The Trophic-Dynamic Aspect of Ecology. Ecology, 23(4), 399–417.
Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. Journal of the Atmospheric Sciences, 20(2), 130–141.
Luhmann, N. (1984). Soziale Systeme: Grundriß einer allgemeinen Theorie. Frankfurt am Main: Suhrkamp. [English translation: Social Systems, Stanford University Press, 1995.]
Luhmann, N. (1997). Die Gesellschaft der Gesellschaft. Frankfurt am Main: Suhrkamp.
Mahon, J. E. (2015). The Definition of Lying and Deception. In E. N. Zalta (Ed.), Stanford Encyclopedia of Philosophy.
Maturana, H. R., & Varela, F. J. (1980). Autopoiesis and Cognition: The Realization of the Living. Dordrecht: D. Reidel.
McEwen, B. S. (1998). Stress, Adaptation, and Disease: Allostasis and Allostatic Load. Annals of the New York Academy of Sciences, 840(1), 33–44.
Newton, I. (1687). Philosophiæ Naturalis Principia Mathematica. London: Jussu Societatis Regiæ.
Nowak, M. A. (2006). Five Rules for the Evolution of Cooperation. Science, 314(5805), 1560–1563.
Ostrom, E. (1990). Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge: Cambridge University Press.
Ostrom, E. (2005). Understanding Institutional Diversity. Princeton: Princeton University Press.
Prigogine, I. (1980). From Being to Becoming: Time and Complexity in the Physical Sciences. San Francisco: W. H. Freeman.
Prigogine, I., & Stengers, I. (1984). Order Out of Chaos: Man's New Dialogue with Nature. New York: Bantam.
Rao, R. P. N., & Ballard, D. H. (1999). Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-classical Receptive-field Effects. Nature Neuroscience, 2(1), 79–87.
Ryle, G. (1949). The Concept of Mind. London: Hutchinson.
Scheffer, M. (2009). Critical Transitions in Nature and Society. Princeton: Princeton University Press.
Scheffer, M., Bockting, C. L., Borsboom, D., et al. (2024). A Dynamical Systems View of Psychiatric Disorders. JAMA Psychiatry, 81(6), 618–630.
Selye, H. (1950). The Physiology and Pathology of Exposure to Stress. Montreal: Acta.
Shannon, C. E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27(3), 379–423.
Sterling, P., & Eyer, J. (1988). Allostasis: A New Paradigm to Explain Arousal Pathology. In S. Fisher & J. Reason (Eds.), Handbook of Life Stress, Cognition and Health (pp. 629–649). New York: Wiley.

Strogatz, S. H. (2015). Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering (2nd ed.). Boulder: Westview Press.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. In Proceedings of ACL 2019 (pp. 3645–3650).

Tarski, A. (1955). A Lattice-Theoretical Fixpoint Theorem and Its Applications. Pacific Journal of Mathematics, 5(2), 285–309.

Thom, R. (1972). Stabilité structurelle et morphogénèse. Paris: InterEditions. [English translation: Structural Stability and Morphogenesis, Benjamin, 1975.]

Thompson, N. (2016). Metaphysical Interdependence. In M. Jago (Ed.), Reality Making (pp. 38–56). Oxford: Oxford University Press.

Tononi, G. (2004). An Information Integration Theory of Consciousness. BMC Neuroscience, 5, 42.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated Information Theory: From Consciousness to Its Physical Substrate. Nature Reviews Neuroscience, 17(7), 450–461.

Turing, A. M. (1936). On Computable Numbers, with an Application to the Entscheidungsproblem. Proceedings of the London Mathematical Society, 42(1), 230–265.

von Bertalanffy, L. (1950). An Outline of General System Theory. British Journal for the Philosophy of Science, 1(2), 134–165.

von Bertalanffy, L. (1968). General System Theory: Foundations, Development, Applications. New York: George Braziller.

Wiener, N. (1948). Cybernetics: Or Control and Communication in the Animal and the Machine. Cambridge, MA: MIT Press.

Wilson, K. G. (1971). Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture. Physical Review B, 4(9), 3174–3183.

Wilson, K. G. (1975). The Renormalization Group: Critical Phenomena and the Kondo Problem. Reviews of Modern Physics, 47(4), 773–840.

Wilson, K. G., & Kogut, J. (1974). The Renormalization Group and the ε Expansion. Physics Reports, 12(2), 75–199.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor Prediction. Current Biology, 11(18), R729–R732.

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

Berger, J. O. (1985). Statistical Decision Theory and Bayesian Analysis (2nd ed.). New York: Springer.

Bickle, J. (2003). Philosophy and Neuroscience: A Ruthlessly Reductive Account. Dordrecht: Kluwer.

Boltzmann, L. (1877). Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung. Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften, 76, 373–435.

Chalmers, D. J. (1996). The Conscious Mind: In Search of a Fundamental Theory. Oxford: Oxford University Press.

Clausius, R. (1865). Über verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie. Annalen der Physik, 201(7), 353–400.

Conant, R. C., & 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.

Cover, T. M., & Thomas, J. A. (2006). Elements of Information Theory (2nd ed.). Hoboken, NJ: Wiley.

Dawkins, R. (1976). The Selfish Gene. Oxford: Oxford University Press.

Dennett, D. C. (1991). Consciousness Explained. Boston: Little, Brown.

Eigen, M. (1971). Selforganization of Matter and the Evolution of Biological Macromolecules. Naturwissenschaften, 58(10), 465–523.

Feynman, R. P. (1965). The Character of Physical Law. Cambridge, MA: MIT Press.

Fodor, J. A. (1975). The Language of Thought. New York: Thomas Y. Crowell.

Gigerenzer, G. (2007). Gut Feelings: The Intelligence of the Unconscious. New York: Viking.

Gleick, J. (1987). Chaos: Making a New Science. New York: Viking.

Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. Proceedings of the Royal Society B, 205(1161), 581–598.

Haken, H. (1983). Synergetics: An Introduction (3rd ed.). Berlin: Springer.

Hardin, G. (1968). The Tragedy of the Commons. Science, 162(3859), 1243–1248.

Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. Zeitschrift für Physik, 43(3–4), 172–198.

Hofstadter, D. R. (1979). Gödel, Escher, Bach: An Eternal Golden Braid. New York: Basic Books.

Jaynes, E. T. (2003). Probability Theory: The Logic of Science. Cambridge: Cambridge University Press.

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

Kolmogorov, A. N. (1965). Three Approaches to the Quantitative Definition of Information. Problems of Information Transmission, 1(1), 1–7.

Kuhn, T. S. (1962). The Structure of Scientific Revolutions. Chicago: University of Chicago Press.

Laughlin, R. B. (2005). A Different Universe: Reinventing Physics from the Bottom Down. New York: Basic Books.

Mandelbrot, B. B. (1982). The Fractal Geometry of Nature. San Francisco: W. H. Freeman.

Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. San Francisco: W. H. Freeman.

Minsky, M. (1986). The Society of Mind. New York: Simon & Schuster.

Mitchell, M. (2009). Complexity: A Guided Tour. Oxford: Oxford University Press.
Nash, J. F. (1950). Equilibrium Points in N-Person Games. Proceedings of the National Academy of Sciences, 36(1), 48–49.
Newman, M. E. J. (2010). Networks: An Introduction. Oxford: Oxford University Press.
Nicolis, G., & Prigogine, I. (1977). Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations. New York: Wiley.
North, D. C. (1990). Institutions, Institutional Change and Economic Performance. Cambridge: Cambridge University Press.
Penrose, R. (1989). The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics. Oxford: Oxford University Press.
Popper, K. R. (1959). The Logic of Scientific Discovery. London: Hutchinson.
Putnam, H. (1967). Psychological Predicates. In W. H. Capitan & D. D. Merrill (Eds.), Art, Mind, and Religion (pp. 37–48). Pittsburgh: University of Pittsburgh Press.
Quine, W. V. O. (1960). Word and Object. Cambridge, MA: MIT Press.
Schelling, T. C. (1978). Micromotives and Macrobehavior. New York: Norton.
Schrödinger, E. (1944). What Is Life? The Physical Aspect of the Living Cell. Cambridge: Cambridge University Press.
Seth, A. K. (2021). Being You: A New Science of Consciousness. New York: Dutton.
Simon, H. A. (1962). The Architecture of Complexity. Proceedings of the American Philosophical Society, 106(6), 467–482.
Simon, H. A. (1996). The Sciences of the Artificial (3rd ed.). Cambridge, MA: MIT Press.
Smith, J. M. (1982). Evolution and the Theory of Games. Cambridge: Cambridge University Press.
Tononi, G., & Koch, C. (2015). Consciousness: Here, There and Everywhere? Philosophical Transactions of the Royal Society B, 370(1668), 20140167.
Varela, F. J., Thompson, E., & Rosch, E. (1991). The Embodied Mind: Cognitive Science and Human Experience. Cambridge, MA: MIT Press.
Waddington, C. H. (1957). The Strategy of the Genes. London: Allen & Unwin.
Watts, D. J., & Strogatz, S. H. (1998). Collective Dynamics of 'Small-World' Networks. Nature, 393(6684), 440–442.
Weinberg, S. (1992). Dreams of a Final Theory: The Scientist's Search for the Ultimate Laws of Nature. New York: Pantheon.
Wolfram, S. (2002). A New Kind of Science. Champaign, IL: Wolfram Media.

Sample Bibliography of Third-Party Publications Citing Boris Kriger

---

1. Montaraz Velandia, A. L., & Ruedas Patiño, C. E. (2026). De la distorsión biológica a la gobernanza basada en evidencia: Protocolos de políticas públicas para la integración del factor

D en la toma de decisiones institucionales en Colombia. *Revista Colombiana de Neuropolítica y Decisión Pública*, *4*(1), 1–22. Centro de Investigación en Cognición y Políticas Públicas (CICPP), Universidad del Altiplano Colombiano, Bogotá, Colombia. ISSN 2745-3189 (online). DOI: 10.22304/rcnydp.2026.v4n1.0037

Kriger work cited: Kriger, B. (2026). Eliminating Distortion: Inadequate Biological Programs in Human Decision Making and Their Reflection in AI. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18457545

---

2. Olavarrieta Durán, S. I., & Arrigorriaga Mendive, V. C. (2026). El quirófano equivocado: Crítica de la lectura burocrática de la variable D y restitución del argumento de Kriger como tesis de sustitución estructural. *Cuadernos Rioplatenses de Filosofía y Tecnología*, *11*(2), 87–93. Instituto de Filosofía Aplicada y Sistemas Decisionales (IFASD), Universidad Libre del Río de la Plata, Montevideo, Uruguay. ISSN 1688-4027 (online). DOI: 10.31528/crft.2026.v11n2.0089

Kriger work cited: Kriger, B. (2026). Eliminating Distortion: Inadequate Biological Programs in Human Decision Making and Their Reflection in AI. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18457545

---

3. Nguyen, D.-M., Trần, T.-L., & Lê, H.-P. (2026). Compression Cost, Partition Stability, and the Engineering of Causal Graphs: Formalizing and Empirically Testing the Perspectival Account of Causation. *Architectures of Inference: An International Journal of Information Theory, Causal Reasoning, and Philosophy of Science*, *3*(2), 117–131. Mekong Academic Press, Cần Thơ, Vietnam. Laboratory of Algorithmic Epistemology, Saigon Institute for Theoretical Computation, Ho Chi Minh City; Division of Computational Philosophy, Mekong Delta Centre for Applied Information Theory, Cần Thơ.

Kriger work cited: Kriger, B. (2026). Against Causation: A Formal Argument That Causality Is a Compression Artifact of Bounded Observers, Not a Feature of Reality. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18851848

---

4. Thanapakpawin, S., & Charoensuk, K. (2026). CausalAudit: An Open-Source Framework for Partition Stability Analysis of Causal Graphs in Clinical and Pharmacological Research. *Chulalongkorn Journal of Computational Science and Applied Epistemology*, *12*(1), 1–8. Chulalongkorn University Press, Bangkok, Thailand. ISSN 2345-6789. Laboratory of

Computational Epistemology, Faculty of Science, Chulalongkorn University; Centre for Biostatistics and Algorithmic Medicine, King Mongkut's Institute of Technology Ladkrabang. **Kriger works cited:** Kriger, B. (2026). Against Causation. (via Nguyen, Trần & Lê, 2026, as primary formalization source.)

---

5. Aytekin, H., Bongani, N., & Sifiso, B. (2026). From Prohibition to Integration: Why Universal AI Adoption in Education and Communication Is a Civilizational Imperative. *Artificial Intelligence Systems*, *6*(3), 41–48. IIIR Press. Institute of AI Systems, Izmir, Turkey; Laboratory of Adaptive Computation Research, Minsk Institute of Physics and Technology, Minsk, Belarus; Institute of Complex Systems, Harare, Zimbabwe. CC BY 4.0.

Kriger work cited: Kriger, B. (2026). AI-Extended Agents and the Transformation of Human Communication: A Game-Theoretic Model of Norm Shift in Populations with AI-Mediated Communicators. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18521341

---

6. Mandlenkosi, A., Bongani, N., & Sifiso, B. (2026). Structural Anticorruption: A Practical Roadmap for Reducing Institutional Friction and Building Algorithmic Integrity. *Evolutionary Dynamics & Adaptive Systems*, *17*(3), 142–148. IIIR Press. Institute for Interdisciplinary & Integrative Research, Nairobi, Kenya; Minsk Institute of Physics and Technology, Belarus; Institute of Complex Systems, Harare, Zimbabwe. ISSN 2519-4071 (Print), ISSN 2519-408X (Online). CC BY 4.0.

Kriger works cited: — Kriger, B. (2026). Four Principles of Corruption: A Systems-Theoretic Analysis of Institutional Failure and Reform. Institute of Integrative and Interdisciplinary Research. https://doi.org/10.5281/zenodo.18879948 — Kriger, B. (2026). Beyond Punishment: A Systems-Theoretic Critique of Punitive Anticorruption Legal Frameworks. *Legal Research & Analysis*, *4*(1), 19–24. https://doi.org/10.69971/lra.4.1.2026.176 — Kriger, B. (2025). *Corrupt by Design: How Institutions Create the Problems They Claim to Fight.* Altaspera Publishing. — Kriger, B. (2025). *The Justice We Deserve: Beyond Retribution — Why Real Justice Means Healing, Not Hurting.* Altaspera Publishing. — Kriger, B. (2025). The Asymmetry of Totalizing Ideals: A Structural Law of Complex Adaptive Systems. Working paper, Institute of Integrative and Interdisciplinary Research. https://zenodo.org/records/18361828

---

7. Evdokimov, O., Bachani, S., & Ryss, E. (2026). Geometry, Resonance and Time: A Unified Framework for Gravity, Inertia, and Consciousness. April 3, 2026. (Publication venue details not available from provided information.)

Kriger work cited: Kriger, B. (2026). The Architecture of the Block Universe: Metric Relaxation and Vacuum Anchoring. Submitted to *Foundations of Physics*. (Reference [1] in the citing paper.)

Summary of Kriger Works Cited Across the Corpus

Kriger Work	Citing Papers	Field
Eliminating Distortion (2026)	Montaraz & Ruedas; Olavarrieta & Arrigorriaga	Decision science, public policy, philosophy
Against Causation (2026)	Nguyen, Trần & Lê; Thanapakpawin & Charoensuk	Philosophy of science, causal inference, clinical research
AI-Extended Agents (2026)	Aytekin, Bongani & Sifiso	Education, communication theory
Four Principles of Corruption (2026) + related works (2025–2026)	Mandlenkosi, Bongani & Sifiso	Institutional reform, anticorruption
Architecture of the Block Universe (2026)	Evdokimov, Bachani & Ryss	Foundations of physics