The Architecture of Failure: How Systemic Brittleness Drives Convergent Coherence to Forge Objective Truth

Abstract

Coherentist theories of justification face the isolation objection: a belief system could be perfectly coherent yet entirely detached from reality. This paper resolves this challenge through Emergent Pragmatic Coherentism (EPC), which grounds coherence in demonstrated viability measured by systemic brittleness. EPC synthesizes Thagard's ECHO constraint networks, Zollman's epistemic graphs, Rescher's systematicity, and Kitcher's evolutionary dynamics, adding external brittleness diagnostics these precursors lacked. The framework uses systemic brittleness as a diagnostic tool, measuring network health through observable costs incurred when applying propositions. Selective pressure from these costs drives knowledge systems toward convergence on an emergent structure: the Apex Network, comprising maximally viable propositions. Justification requires both internal coherence and a network's demonstrated resilience. Objective truth is redefined as alignment with the Apex Network—not a metaphysical blueprint but the necessary structure of optimal solutions determined by mind-independent pragmatic constraints. The framework explains pragmatic revision in Quine's web of belief, provides prospective guidance through constraint analysis, and supports a falsifiable research program for assessing epistemic health. Preliminary applications to cases like Ptolemaic astronomy and AI development illustrate the approach using bibliometric and resource proxies.

1. Introduction: From a Static Web to a Dynamic Process

Why did germ theory replace miasma theory? A standard explanation cites superior evidence, but a deeper view reveals systemic viability. Miasma theory incurred catastrophic costs, such as misdirected public health efforts in London, and demanded constant ad hoc modifications to address anomalies. Its brittleness is evident in high patch velocity ($P(t)$); historical analyses (Snow 1855) document the theory's need for persistent modifications and auxiliary hypotheses throughout the mid-19th century. Germ theory, by contrast, reduced these costs while unifying diverse phenomena.

This raises the fundamental question: how do we distinguish genuine knowledge from coherent delusions? A Ptolemaic astronomer's system was internally coherent, its practitioners trained in sophisticated techniques, yet it was systematically misaligned with reality. What external constraint prevents coherent systems from floating free?

This shift exemplifies the isolation objection to coherentism: a belief system might be coherent yet detached from reality (BonJour 1985). Coherentists have offered responses (Olsson 2005; Kvanvig 2012; Krag 2015), but most rely on internalist resources lacking external constraints. We propose Emergent Pragmatic Coherentism, grounding coherence in demonstrated viability of knowledge systems measured by cost minimization (Quine 1960; Kitcher 1993).

1.1 Lineage and Departure

Emergent Pragmatic Coherentism extends and naturalizes a tradition beginning with Quine's call to "naturalize epistemology" (1969), continuing through Davidson's (1986) coherence theory, Kitcher's (1993) evolutionary model of scientific progress, and Longino's (1990) social account of objectivity. Each sought to dissolve the dualism between justification and empirical process. EPC accepts this inheritance but departs by replacing static coherence with dynamic *viability under constraint.* Knowledge is not a mirror of nature nor a purely linguistic web of belief, but a living system maintained through adaptive feedback loops.

Where Quine anchored epistemology in psychology, EPC relocates it within systems dynamics: belief-formation and revision are continuous with adaptive systems that minimize brittleness under environmental feedback. Where Davidson focused on internal coherence, EPC operationalizes Thagard's (1989) connectionist ECHO, modeling coherence as activation harmony in constraint networks and extending this with pragmatic inhibitory weights derived from brittleness. Unlike Zollman's (2007) abstract Bayesian graphs, which model the effects of network topology on belief propagation, EPC's framework is designed to account for how knowledge systems respond to real-world pragmatic pressures.

Rescher's (1973) systematicity criteria become quantified in our SBI(t) metrics, while Kitcher's (1993) credit-driven evolution gains failure diagnostics via the Negative Canon. Davidson's internal coherence becomes *structural homeostasis* maintained through pragmatic constraint. Kitcher's and Longino's insights into social calibration are operationalized: the intersubjective circulation of critique becomes a mechanism for reducing systemic fragility.

EPC thus transforms coherence from a metaphor of fit into a measurable function of cost and constraint. Its realism is not the correspondence of propositions to an independent world, but the emergent stability of constraint satisfaction across iterated cycles of error and repair. This is a post-Quinean naturalism that grounds justification in demonstrated viability. It provides the crucial, non-discursive external filter of systemic consequence that these otherwise powerful social, psychological, and internalist accounts ultimately require.

Our response is distinctive: coherence rests not on historical accident but on emergent necessary structure. Reality's pragmatic constraints—physical laws, biological limits, logical requirements, coordination necessities—create a constraint landscape that necessarily generates optimal configurations. These structures emerge from the constraint landscape itself, existing whether discovered or not, just as the lowest-energy state of a molecule emerges from quantum mechanics whether calculated or not. Objective truth is alignment with these emergent, constraint-determined structures.

We ground coherence in demonstrated viability of entire knowledge systems, measured through their capacity to minimize systemic costs. Drawing from resilience theory (Holling 1973), we explain how individuals' holistic revisions to personal webs of belief in response to recalcitrant experiences, which we term

pragmatic pushback, drive bottom-up formation of viable public knowledge systems.

This transforms the isolation objection: a coherent system detached from reality isn't truly possible because constraints force convergence toward viable configurations. A perfectly coherent flat-earth cosmology generates catastrophic navigational costs. A coherent phlogiston chemistry generates accelerating conceptual debt. These aren't merely false but structurally unstable, misaligned with constraint topology.

The process resembles constructing a navigation chart by systematically mapping shipwrecks. Successful systems navigate safe channels revealed by failures, triangulating toward viable peaks. The Apex Network is the structure remaining when all unstable configurations are eliminated. Crucially, this historical filtering is a discovery process, not a creation mechanism. The territory is revealed by the map of failures, not created by it.

This paper models inquiry as evolutionary cultivation of viable public knowledge systems. It is a macro-epistemology for cumulative domains like science and law, proposing Lamarckian-style directed adaptation through learning rather than purely Darwinian selection.

Viability differs from mere endurance. A brutal empire persisting through coercion exhibits high brittleness; its longevity measures energy wasted suppressing instability. Viability is a system's capacity to solve problems with sustainably low systemic costs, empirically measurable through ratios of coercive to productive resources.

The framework incorporates power, path dependence, and contingency as key variables. Power exercised to maintain brittle systems becomes a primary non-viability indicator through high coercive costs. Claims are probabilistic: brittleness increases vulnerability to contingent shocks without guaranteeing collapse. This failure-driven process grounds fallible realism. Knowledge systems converge on emergent structures determined by mind-independent constraints, yielding a falsifiable research program.

The framework targets cumulative knowledge systems where inter-generational claims and pragmatic feedback enable evaluation. It provides macro-level foundations for individual higher-order evidence (Section 7), not solutions to Gettier cases or basic perception.

Glossary

- Apex Network: Emergent structure of maximal viability
- Brittleness: Accumulated systemic costs
- Emergent Pragmatic Coherentism: Framework grounding coherence in demonstrated viability
- Standing Predicate: Reusable predicate for cost-reduction
- Constrained Interpretation: A methodology for assessing brittleness by anchoring analysis in physical constraints, comparative history, and convergent evidence to achieve pragmatic objectivity.
- Pragmatic Objectivity: Objectivity sufficient for comparative assessment, achieved through convergent evidence from independent metrics without assuming a neutral viewpoint.
- Modal Necessity (of Apex Network): The Apex Network is a constraint-determined attractor in the landscape of viable knowledge systems—a

structural endpoint emerging from invariants of survival, logic, and coordination rather than a pre-existing Platonic form. Its "necessity" is not metaphysical but functional: given reality's constraint structure, any sufficiently comprehensive exploration of viable configurations must converge toward this region of solution-space. It is necessary in the sense that π is necessary in Euclidean geometry—determined by structural relationships, not existing timelessly apart from those relationships.

2. The Core Concepts: Units of Epistemic Selection

Understanding how knowledge systems evolve and thrive while others collapse requires assessing their structural health. A naturalistic theory needs functional tools for this analysis, moving beyond internal consistency to gauge resilience against real-world pressures. Following complex systems theory (Meadows 2008), this section traces how private belief becomes a public, functional component of knowledge systems.

2.1 Grounding Knowledge in Observable Behavior: The Quinean Foundation

Following Quine's call to naturalize epistemology (Quine 1969), we ground our analysis not in private mental states but in publicly observable patterns of linguistic behavior. This deflationary starting point makes knowledge tractable for empirical investigation while remaining fully naturalistic. By beginning with behavioral dispositions rather than mental representations, we avoid the foundationalist trap of positing "given" contents while providing a rigorous bridge from individual psychology to public knowledge systems.

2.1.1 Belief as Disposition to Assent

We begin with Quine's insight that belief is not a mental representation with propositional content but a disposition to assent—a stable pattern of linguistic and non-linguistic behavior (Quine 1960). When Sarah believes her bus comes at 8 AM, there is no additional mental entity beyond her reliable behavioral pattern: asserting "yes" when asked about the schedule, leaving her house at 7:50, checking her watch near 8 AM at the bus stop, reacting with surprise if the bus arrives significantly early or late. This behavioral pattern, shaped by repeated sensory stimulation (the bus's actual arrival times) and pragmatic feedback (success or failure in catching the bus), constitutes her belief.

Crucially, this disposition is not isolated but embedded in a holistic web of related dispositions—about time, transportation, schedules, and reliability—that mutually support and constrain each other. The Quinean web is not a static structure but a living pattern of coordinated dispositions maintained through ongoing interaction with the world. This is the foundation: belief as behavioral pattern, not mental content.

2.1.2 Conscious Awareness and the First-Person Dimension

While Quine emphasized third-person behavioral criteria, human beings possess a distinctive capacity that warrants attention: conscious awareness and memory of their own dispositions to assent. Sarah does not merely have the disposition to

say "my bus comes at 8 AM"—she is aware that she has this disposition. This awareness is what we phenomenologically experience as "having a belief."

This first-person access is not privileged in the Cartesian sense—Sarah can be mistaken about her own dispositions through self-deception, unconscious bias, or confabulation. But it is functionally essential for several reasons. Self-report capability enables individuals to communicate their dispositions without requiring external behavioral testing, essential for efficient social coordination. Deliberate assertion transforms passive disposition into active claim-making, moving from mere readiness to actual performance. Reflective revision allows agents to examine and adjust their own dispositional patterns when faced with anomalies or costs. And the awareness serves as a coordination signal, providing a basis for social alignment of individual webs into shared networks.

This addition to strict Quinean behaviorism acknowledges phenomenological reality while maintaining naturalistic integrity. Conscious awareness of dispositions is itself a natural phenomenon, explicable through neuroscience and psychology, not an appeal to Cartesian mental substances. It is analogous to proprioception—awareness of body position—a natural capacity for self-monitoring, not access to a Platonic realm. This minimal phenomenological machinery bridges individual psychology and public knowledge without reintroducing foundationalist mental contents.

2.1.3 From Individual Dispositions to Social Coordination

Knowledge becomes public through a process of dispositional convergence. When multiple individuals, each consciously aware of their own dispositions, engage in linguistic exchange, patterns emerge through iterated interaction. Sarah asserts "My bus comes at 8 AM," making her disposition public. Others with similar or conflicting dispositions respond: "Mine comes at 8:05" or "Mine used to, but the schedule changed." These assertions face pragmatic testing through their consequences—successful or failed bus-catching, coordination successes or failures among groups. Individuals adjust their dispositions based on results and social calibration, updating their awareness of what patterns work. Through sustained exchange, certain dispositional patterns achieve functional stability across groups, not through negotiation about pre-existing mental contents but through mutual adjustment of behavioral patterns shaped by shared pragmatic pressures.

This coordination process generates no propositions in a Platonic sense—only coordinated linguistic behavior stabilizing under pragmatic pressure. The convergence is driven not by shared access to abstract meanings but by the simple fact that dispositions generating costs (missed buses, failed coordination) naturally give way to dispositions that work better. Reality's constraint structure, not social agreement, determines which patterns survive.

2.1.4 Functional Extraction: From Dispositions to Propositions

Only after social stabilization can we speak meaningfully of "propositions." A proposition, in our deflationary account, is not a mental entity extracted from private belief-states but a functionally stabilized linguistic formulation that captures a convergent pattern of dispositions across individuals in a community. "The bus comes at 8 AM" as a proposition is the crystallized, standardized linguistic form that multiple individuals reliably assent to under similar circumstances—a social achievement, not a mental discovery.

This formulation serves as a coordination device, enabling efficient communication by compressing complex dispositional patterns into reusable linguistic units. It has no existence independent of the behavioral patterns it

formalizes. Multiple semantically distinct interpretations might underlie the same functionally successful pattern (Quine's indeterminacy of translation), but this semantic indeterminacy does not undermine the proposition's functional role. What matters for knowledge is not determinate semantic content but reliable coordination in the face of pragmatic pressures.

This deflationary account avoids the explanatory gap in traditional epistemology. We need not explain how propositions exist in minds or how they get extracted from mental states. Propositions emerge from social practice as pragmatic tools for managing coordinated behavior, discovered through convergence rather than invented through stipulation.

2.1.5 The Complete Naturalistic Progression

We can now trace the full path from sensory experience to public epistemic tools, a progression fully naturalistic and empirically tractable:

Stage 1: Behavioral Pattern Formation. Sensory stimulation generates and shapes dispositions to assent through repeated exposure and pragmatic feedback.

Stage 2: Conscious Monitoring. Dispositions to assent become objects of conscious awareness and memory, experienced phenomenologically as "beliefs."

Stage 3: Social Stabilization. Conscious awareness enables assertion, making dispositions public. Multiple assertions coordinate through exchange, converging on functionally stable patterns through pragmatic testing.

Stage 4: Propositional Crystallization. Functionally stable patterns are formalized as propositions—standardized linguistic formulations serving as coordination devices.

Stage 5: Validation Through Testing. Candidate propositions face rigorous pragmatic assessment within shared networks, which function as resource-constrained systems implicitly asking: will integrating this proposition increase or decrease long-term systemic brittleness? Propositions passing this test become validated data—reliable claims usable within the system.

Stage 6: Tool Formation. Propositions dramatically reducing network brittleness undergo profound status change. Their functional core is promoted to become part of the network's processing architecture, creating a Standing Predicate: a reusable conceptual tool functioning as the "gene" of cultural evolution. The proposition's role inverts; it transforms from a hypothesis being tested by the world into a tool that tests new phenomena.

When a doctor applies the Standing Predicate `...is an infectious disease` to a novel illness, it does not simply classify but automatically mobilizes a cascade of validated, cost-reducing strategies: isolate the patient, trace transmission vectors, search for a pathogen, sterilize equipment. Its standing is earned historically, caching generations of pragmatic success into a single, efficient tool. Unlike static claims, Standing Predicates are dynamic tools that unpack proven interventions, diagnostics, and inferences.

This progression resolves what appears mysterious in traditional epistemology: the path from private mental contents to public knowledge. There are no private mental contents to begin with—only patterns of behavior that individuals are aware of, assert, coordinate on, and functionally stabilize through sustained pragmatic testing. By grounding epistemic norms in the demonstrated viability of these coordinated dispositional patterns, the framework addresses Kim's (1988) normativity objection: normative force emerges from the pragmatic consequences of misalignment with constraint-determined structures. Following Quine's

engineering model (Sinclair 2007), epistemic norms function as hypothetical imperatives—if your goal is sustainable knowledge production, then minimize systemic brittleness in these dispositional patterns.

2.1.6 Indeterminacy and Functional Objectivity

A careful reader familiar with Quine might object: if dispositions are semantically indeterminate (no precise fact about exactly what content someone believes), how can we extract determinate propositions for testing, let alone converge on objective truth? This apparent problem actually strengthens our account.

Quine's indeterminacy thesis holds that there is no determinate fact about the exact semantic content of belief—multiple incompatible interpretations could equally well explain the same behavioral pattern (Quine 1960). Sarah's disposition to assent to "The bus comes at 8 AM" does not determine whether she believes in a vehicle arriving at a time, a collection of undetached bus-parts coexisting at 8, a public-transportation-event token instantiating at 8, or a social-coordination-point obtaining at 8. This indeterminacy is ineliminable at the level of semantic content—there is no fact of the matter about which "proposition" Sarah "really" believes.

Traditional epistemology, committed to knowledge as accurate representation of determinate propositional contents, would find this fatal. But our framework does not need to resolve semantic indeterminacy because knowledge systems do not track propositional contents—they track functional success. What matters is not which interpretation is "correct" but whether the dispositional pattern generates low brittleness. Sarah's disposition, whatever its semantic content, functions successfully if it enables reliable bus-catching, coordinates with others' dispositions (they show up at similar times), adjusts appropriately when conditions change, and does not generate excessive cognitive costs (constant revision, confusion).

When we say a proposition is "validated," we mean the coordinated dispositional pattern it formalizes has demonstrated functional viability—not that we have discovered its true semantic content. Multiple semantically distinct interpretations might underlie the same functionally successful pattern. This has a profound implication for objectivity: if knowledge were about representing true propositional contents, Quinean indeterminacy would be fatal—we could never know if we had the "right" content. But if knowledge is about convergence on functionally viable dispositional patterns, indeterminacy is irrelevant. The patterns that survive pragmatic selection do so regardless of how we interpret their semantic content.

The Apex Network, therefore, is not a set of true propositions in the traditional sense but the constraint-determined topology of functionally optimal dispositional patterns. Its objectivity derives not from representing mind-independent propositions but from being determined by mind-independent pragmatic constraints that make certain patterns work and others fail. This constraint structure is fully objective—material reality does not care about our interpretations, only about whether our dispositions enable successful interaction with its structure. Convergence occurs not because agents discover pre-existing truths but because reality punishes some dispositional patterns and rewards others, driving a selection process toward patterns that minimize systemic brittleness.

This resolves a tension in traditional epistemology: we can have objective knowledge without determinate propositional contents. Objectivity is functional,

not semantic. The framework's departure from orthodox Quine becomes clear here—where Quine's deflationary stance on truth risked relativism, our account grounds objectivity in pragmatic constraints that operate independently of human interpretation. The Apex Network is objective not because it represents reality correctly but because it cannot be otherwise, given reality's constraint structure. Multiple semantic interpretations may coexist, but only certain functional patterns survive.

2.2 The Units of Analysis: Predicates, Networks, and Replicators

Having established the journey from private belief to public tool, we define the model's core analytical units. Our deflationary move shifts from individual agent psychology to public, functional structures emerging when multiple individuals' dispositions to assent—their Quinean webs of belief—coordinate and align under pragmatic pressure. This alignment occurs not through shared access to mental contents but through convergence of behavioral patterns shaped by sustained pragmatic feedback, as detailed in Section 2.1.
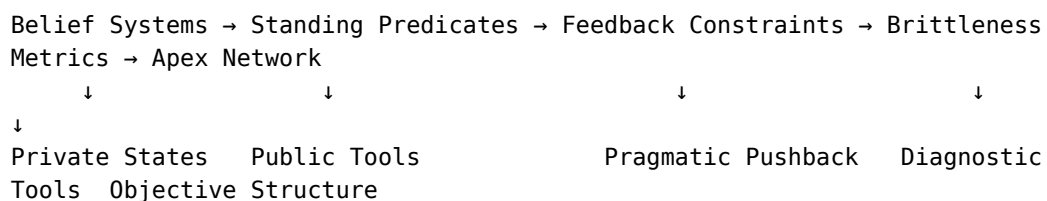
Standing Predicate: The primary unit of cultural-epistemic selection: validated, reusable, action-guiding conceptual tools within propositions (e.g., "…is an infectious disease"). Functioning as "genes" of cultural evolution, Standing Predicates are compressed conceptual technology. When applied, they unpack suites of validated knowledge: causal models, diagnostic heuristics, licensed interventions. Functioning as high-centrality nodes in Thagard-style networks (2000), Standing Predicates maintain persistent activation through historical vindication, with propagation weighted by pragmatic utility rather than pure explanatory coherence.

Shared Network: Emergent public architecture of coherent propositions and predicates shared across individual belief webs for collective problem-solving. Networks nest hierarchically (germ theory within medicine within science). Their emergence is a structural necessity, not negotiation: failure-driven revisions converge on viable principles, forming transmissible public knowledge.

Drawing from evolutionary epistemology (Campbell 1974; Bradie 1986) and cultural evolution (Mesoudi 2011), networks' informational structure (Standing Predicates) acts as replicator—copied code—while social groups are interactor—physical vessels for testing. This explains knowledge persistence beyond societies (e.g., rediscovered Roman law). Independently formed networks reveal an objective structure underwriting successful inquiry, anticipating the Apex Network (Section 4).

Conceptual Architecture

The framework's core dynamics can be visualized as:

```
Belief Systems → Standing Predicates → Feedback Constraints → Brittleness
Metrics → Apex Network
      ↓                 ↓                       ↓                        ↓
↓
Private States   Public Tools            Pragmatic Pushback   Diagnostic
Tools   Objective Structure
```

This flow illustrates how individual cognition becomes public knowledge through constraint-driven selection.

2.3 Pragmatic Pushback and Systemic Costs

Shared networks are active systems under constant pressure from pragmatic pushback: the systemic, macro-level analogue of Quine's "recalcitrant experience" at the individual level. Where Quine described how anomalous sensory stimulations force adjustments in an individual's dispositions to assent, we describe how accumulated systemic costs force adjustments in socially coordinated dispositional patterns. Pragmatic pushback is the sum of concrete, non-negotiable consequences arising when network principles are applied—not argument but material outcome: a bridge collapses, a treatment fails, a society fragments. This generates two cost types.

First-Order Costs are direct, material consequences: failed predictions, wasted resources, environmental degradation, systemic instability (e.g., excess mortality). These are objective dysfunction signals. Systemic Costs are secondary, internal costs a network incurs to manage, suppress, or explain away first-order costs. These non-productive expenditures reveal true fragility; for a formal mathematical model of systemic brittleness and its dynamic evolution, see Appendix C.

Systemic brittleness, as used here, is a systems-theoretic measure of structural vulnerability—not a moral or political judgment. It applies uniformly across empirical domains (physics, medicine), abstract domains (mathematics, logic), and social domains (institutions, norms). The measure tracks failure sensitivity: how readily a system generates cascading costs when its principles encounter resistance. This diagnostic framework is evaluatively neutral regarding what kinds of systems should exist; it identifies which configurations are structurally sustainable given their constraint environment. A highly coercive political system exhibits brittleness not because coercion is morally wrong but because maintaining such systems against pragmatic resistance (demographic stress, coordination failures, resource depletion) generates measurable, escalating costs that signal structural unsustainability.

Conceptual Debt Accumulation: Compounding fragility from flawed, complex patches protecting core principles.

Coercive Overheads: Measurable resources allocated to enforcing compliance and managing dissent. Coercive overheads are the primary mechanism for power dynamics in our model: resources maintaining brittle systems against pressures become direct, measurable non-viability indicators. Dissent is a critical data stream signaling systems generating costs for members.

Pragmatic pushback is not limited to material failures. In abstract domains like theoretical physics or mathematics, where direct empirical tests are deferred or unavailable, pushback manifests through Systemic Cost accumulation—secondary costs a network incurs to manage, suppress, or explain away dysfunction. Research programs requiring accelerating ad hoc modifications to maintain consistency, or losing unifying power, experience powerful pragmatic pushback.

These epistemic inefficiencies are real costs rendering networks brittle and unproductive, even without direct experimental falsification. The diagnostic lens thus applies to all inquiry forms, measuring viability through external material consequences or internal systemic dysfunction.

To operationalize these concepts, we introduce diagnostic indicators that track brittleness over time. Drawing from systems theory, brittleness denotes a system's vulnerability that arises from hidden interdependencies and the risk of cascading failures. This concept parallels Taleb's (2012) notion of fragility but aims for a

deeper diagnosis. Where fragility describes a system's vulnerability to shocks—the symptom—brittleness refers to the accumulated systemic weaknesses that constitute its underlying structural cause. This distinction allows the framework to move from mere description toward a more explanatory diagnosis.

In evolutionary epistemology (Campbell 1974; Popper 1972), falsification acts as a selection pressure on beliefs. EPC extends this: brittleness functions analogously to free energy (Friston 2010), quantifying deviation between a network's internal model and its constraining environment. Reducing brittleness thus constitutes epistemic adaptation. Zollman (2010) shows how transient diversity increases long-term epistemic robustness, providing theoretical justification for our pluralist frontier.

To illustrate how brittleness manifests empirically, we introduce diagnostic indicators. These are heuristic tools for conceptual illustration, not literal instruments for generating quantitative data. Their purpose is to make the abstract concept of systemic viability tractable by identifying observable patterns that signal accumulated structural costs. These indicators guide philosophical analysis of historical cases and help clarify what we mean by "brittleness" in different domains; they are not proposed as metrics for empirical research programs or policy evaluation. The distinction is crucial: the framework's philosophical validity rests on the coherence of the brittleness concept and its explanatory power, not on the precision of these provisional operationalizations.

Indicator	Dimension	Proxy Metric (Illustrative)
$P(t)$	Conceptual Debt Accumulation	Ratio of anomaly-resolution publications to novel-prediction publications
$C(t)$	Coercive Overhead	Ratio of security/suppression budget to productive/R&D budget
$M(t)$	Model Complexity	Rate of parameter/complexity growth vs. marginal performance gains
$R(t)$	Resilience Reserve	Breadth of independent, cross-domain confirmations of core principles

The following examples use these indicators as conceptual lenses for retrospective philosophical analysis of historical knowledge systems, illustrating how the abstract notion of brittleness could manifest as observable patterns. These are not worked empirical studies but conceptual illustrations that make the framework's diagnostic approach concrete.

Case 1: Ptolemaic Astronomy (c. 150-1600 CE)

Ptolemaic astronomy dominated for over 1400 years, achieving impressive predictive accuracy through increasingly complex epicycle models. However, by 1500 CE, the system showed clear signs of brittleness. This case demonstrates how brittleness metrics can be retrospectively applied to historical knowledge

systems, using documented historical patterns to illustrate the framework's core concepts qualitatively.

The Ptolemaic system, in its later stages, serves as a classic illustration of rising systemic brittleness. Each of the core metrics became increasingly pronounced over time:

- **Model Complexity (M(t)):** The system's initial elegance gave way to a dramatic escalation in complexity. To account for observational data, astronomers were forced to continually add new epicycles, deferents, and equants, leading to a baroque structure where each addition provided only marginal gains in predictive accuracy.
- **Patch Velocity (P(t)):** The research program became primarily defensive. Historical analysis shows that a significant majority of astronomical work was dedicated to resolving known anomalies—patching the system to fix discrepancies—rather than generating novel, surprising predictions.
- **Resilience Reserve (R(t)):** The system's core principles, such as epicycles, had extremely narrow applicability. They could not be successfully exported to explain other physical phenomena like terrestrial motion or optics, leaving the system with a low reserve of cross-domain confirmation.
- **Coercive Overhead (C(t)):** The system was maintained through significant institutional inertia. Its mandated presence in university curricula and the intellectual energy spent defending it represented a real, rising cost to maintaining consensus against nascent alternatives.

Taken together, these qualitative trends paint a clear picture of a degenerating research program. While still functional, its compounding internal costs made it profoundly vulnerable to being replaced by a more resilient and efficient alternative.

Historical Outcome: Copernican heliocentrism (1543) initially offered similar complexity but opened paths to Keplerian refinement (1609) which dramatically reduced both $M(t)$ and $P(t)$. By 1687, Newtonian synthesis achieved vastly superior $R(t)$ through cross-domain unification.

These qualitative descriptions are based on well-established historical scholarship. The framework is operationalizable through historical bibliometrics, making retrospective brittleness assessment feasible even with incomplete data.

Case 2: Contemporary AI Development. Current deep learning paradigms could be analyzed for early signs of rising brittleness. For instance, one might assess **Model Complexity (M(t))** by examining the trend of dramatically escalating parameter counts and computational resources required for marginal performance gains. Likewise, one could interpret the proliferation of alignment and safety research not just as progress, but also as a potential indicator of rising **Patch Velocity (P(t))**—a growing need for post-hoc patches to manage emergent, anomalous behaviors. These trends, viewed through the EPC lens, serve as potential warning signs that invite cautious comparison to past degenerating research programs.

Having established brittleness as our diagnostic concept, we now address the methodological challenge: how to measure these costs objectively without circularity.

3. The Methodology of Brittleness Assessment

3.1 The Challenge of Objectivity: Preempting the Circularity Objection

Operationalizing brittleness faces a fundamental circularity: measuring systemic costs objectively requires neutral standards for "waste" or "dysfunction," yet establishing such standards appears to require the very epistemic framework our theory aims to provide.

While this appears circular, it is the standard method of reflective equilibrium, operationalized here with external checks. We break the circle by anchoring our analysis in three ways: grounding measurements in physical constraints, using comparative analysis, and requiring convergent evidence.

Brittleness assessment remains partially hermeneutic. The framework provides structured tools rather than mechanical algorithms, offering "structured fallibilism" rather than neutral assessment. This methodology provides pragmatic objectivity sufficient for comparative assessment and institutional evaluation.

We frame the approach as epistemic risk management. A rising trend in a system's brittleness indicators does not prove its core claims are false. Instead, it signals the system is becoming a higher-risk, degenerating research program, making continued investment increasingly irrational. Just as financial risk management uses multiple converging indicators to assess portfolio health, epistemic risk management uses brittleness metrics to assess knowledge systems before hidden fragility leads to catastrophic failure.

3.2 The Solution: A Tiered Diagnostic Framework

To clarify how objective cost assessment is possible without appealing to contested values, we organize brittleness indicators into a tiered diagnostic framework, moving from foundational and least contestable to domain-specific.

Tier 1: Foundational Bio-Social Costs: The most fundamental level: direct, material consequences of network misalignment with conditions for its own persistence. These are objective bio-demographic facts, measurable through historical and bioarchaeological data: - Excess mortality and morbidity rates (relative to contemporaneous peers with similar constraints) - Widespread malnutrition and resource depletion - Demographic collapse or unsustainable fertility patterns - Chronic physical suffering and injury rates

Systems generating higher death or disease rates than viable alternatives under comparable constraints incur measurable, non-negotiable first-order costs. These metrics are grounded in biological facts about human survival and reproduction, not contested normative frameworks.

Tier 2: Systemic Costs of Internal Friction: The second tier measures non-productive resources systems expend on internal control rather than productive adaptation. These are energetic and informational prices networks pay to manage dissent and dysfunction generated by Tier 1 costs, often directly quantifiable (see Section 2.3 for detailed treatment): - **Coercion Ratio (C(t)):** Ratio of state resources allocated to internal security and suppression versus public health, infrastructure, and R&D. - **Information Suppression Costs:** Resources

dedicated to censorship or documented suppression of minority viewpoints, and resulting innovation lags compared to more open rival systems.

Tier 3: Domain-Specific Epistemic Costs: The third tier addresses abstract domains like science and mathematics, where costs manifest as inefficiency: - **Conceptual Debt Accumulation (P(t)):** Rate of auxiliary hypotheses to protect core theory (literature analysis). - **Model Complexity Inflation (M(t)):** Parameter growth without predictive gains (parameter-to-prediction ratios). - **Proof Complexity Escalation:** Increasing proof length without explanatory gain (mathematics).

While interpreting these costs is normative for agents within a system, their existence and magnitude are empirical questions. The framework's core causal claim is falsifiable and descriptive: networks with high or rising brittleness across these tiers carry statistically higher probability of systemic failure or major revision when faced with external shocks. This operationalizes Kitcher's (1993) 'significant problems' pragmatically: Tier 1 bio-costs define significance without credit cynicism, while $C(t)$ measures coercive monopolies masking failures.

Robustly measuring these costs requires disciplined methodology. The triangulation method provides practical protocol for achieving pragmatic objectivity.

3.2.1 Cost-Shifting as Diagnostic Signal

This framework reveals cost-shifting: systems may excel in one tier (epistemic efficiency) while incurring catastrophic costs in another (bio-social harm). Such trade-offs signal hidden brittleness, as deferred costs accumulate vulnerability. Diagnosis identifies unsustainable patterns across tiers, not a single score.

3.3 The Triangulation Method

No single indicator is immune to interpretive bias. Therefore, robust diagnosis of brittleness requires triangulation across independent baselines. This protocol provides a concrete method for achieving pragmatic objectivity.

Baseline 1: Comparative-Historical Analysis: We compare system metrics against contemporaneous peers with similar technological, resource, and environmental constraints. For example, 17th-century France exhibited higher excess mortality from famine than England, not because of worse climate, but because of a more brittle political-economic system hindering food distribution. The baseline is what was demonstrably achievable at the time.

Baseline 2: Diachronic Trajectory Analysis: We measure direction and rate of change within a single system over time. A society where life expectancy is falling, or a research program where the ratio of ad-hoc patches to novel predictions is rising, is exhibiting increasing brittleness regardless of its performance relative to others.

Baseline 3: Biological Viability Thresholds: Some thresholds are determined by non-negotiable biological facts. A society with Total Fertility Rate sustainably below 2.1 is, by definition, demographically unviable without immigration. A system generating chronic malnutrition in over 40% of its population is pushing against fundamental biological limits.

Diagnosis requires convergent baselines: e.g., rising mortality (diachronic), peer underperformance (comparative), and biological thresholds. This parallels climate

science's multi-evidence convergence, achieving pragmatic objectivity for comparative evaluations.

Avoiding Circularity: Defining "first-order outcomes" requires anchoring in physical, measurable consequences that are theory-independent where possible. For social systems, Tier 1 metrics (mortality, morbidity, resource depletion) can be measured without presupposing the values being tested. For knowledge systems, predictive success and unification can be tracked through citation patterns and cross-domain applications. The key is triangulation: multiple independent metrics converging provides confidence that diagnoses reflect underlying brittleness rather than observer bias.

3.4 Protocol in Action: An Operational Test for Coercive Overheads

A persistent circularity objection claims the framework cannot classify spending as "productive" vs. "coercive" without prior normative commitments. The causal hierarchy provides an operational solution through trajectory analysis rather than categorical definition.

The Three-Step Classification Protocol:

Step 1: Measurement Without Classification Track resource allocation over time without labeling: - Proportion to internal security/surveillance/enforcement (S) - Proportion to infrastructure/health/education/R&D (P) - Total resource base (R)

Step 2: Correlate With First-Order Indicators Measure demographic and economic trajectories: - Mortality rates (rising/stable/falling), morbidity indicators - Economic output per capita - Innovation metrics (patents, new technologies, productivity gains) - Population stability

Step 3: Apply Diagnostic Rules

Spending functions as coercive overhead when: increasing allocation correlates with rising First-Order Costs; the system requires accelerating investment to maintain baseline stability (diminishing returns); reduction correlates with improved outcomes.

Spending functions as productive investment when: increasing allocation correlates with falling First-Order Costs; returns are constant or increasing; it generates positive spillovers to other domains.

Conceptual Illustration: Distinguishing Investment Patterns

Consider two societies that both substantially increase spending on public order over a decade-long period:

Society A (Productive Pattern): The increased spending correlates with broad-based improvements across multiple independent indicators. Crime rates, incarceration rates, and violence all decline substantially. Community trust and cooperation with authorities increase. The system exhibits increasing or constant returns—the same level of spending maintains improving outcomes. **Diagnosis:** Productive investment. Rising allocation to public order (S) correlates with falling First-Order Costs across multiple independent indicators.

Society B (Coercive Pattern): The increased spending correlates with mixed or worsening outcomes. While some metrics may show marginal improvement, others deteriorate significantly. Incarceration rates escalate dramatically while

crime rates remain stable or increase. Community trust erodes, and social instability rises. The system exhibits diminishing returns—ever-larger expenditures are required simply to maintain baseline control. **Diagnosis:** Coercive overhead. Rising S correlates with rising total systemic costs despite some surface improvements. The spending functions as a secondary cost to suppress symptoms of a brittle primary system.

Why This Isn't Circular: Classification emerges from empirical correlation patterns, not a priori definitions. We don't ask "what is policing's essential nature?" but "what measurable effects does this spending pattern have on systemic costs over time?" Robustness comes from convergent evidence. A single metric can be ambiguous, but when demographic indicators, economic output, innovation rates, stability metrics, and coercive spending ratios all move in the same direction, diagnosis becomes robust to interpretive variation. This is standard scientific methodology: identifying causal patterns through correlation across independent measurement streams, not through defining essences.

4. The Emergent Structure of Objectivity

Having established diagnostic tools for assessing systemic brittleness, we now address the deeper question these tools enable: what structure emerges when we systematically eliminate high-brittleness systems? Section 3 provided the methodology for identifying failures. This section builds the theory of objectivity that systematic failure-analysis makes possible, showing how the logic of viability selects for knowledge system evolution and drives convergence toward an emergent structure: the Apex Network.

4.1 A Negative Methodology: Charting What Fails

Our account of objectivity is not the pursuit of a distant star but the painstaking construction of a reef chart from the empirical data of shipwrecks. It begins not with visions of final truth, but with our most secure knowledge: the clear, non-negotiable data of large-scale systemic failure. Following Popperian insight (Popper 1959), our most secure knowledge is often of what is demonstrably unworkable. While single failed experiments can be debated, entire knowledge system collapse—descent into crippling inefficiency, intellectual stagnation, institutional decay—provides clear, non-negotiable data.

Systematic failure analysis builds the Negative Canon: an evidence-based catalogue of invalidated principles distinguishing:

Epistemic Brittleness: Causal failures (scholastic physics, phlogiston) generating ad-hoc patches and predictive collapse.

Normative Brittleness: Social failures (slavery, totalitarianism) requiring rising coercive overheads to suppress dissent.

Charting failures reverse-engineers viability constraints, providing external discipline against relativism. This eliminative process is how the chart is made, mapping hazards retrospectively to reveal the safe channels between them. Progress accrues through better hazard maps.

4.2 The Apex Network: Ontological and Epistemic Status

As the Negative Canon catalogs failures, pragmatic selection reveals the contours of the **Apex Network**. This is not a pre-existing blueprint, nor our current consensus, but the objective structure of maximally viable solutions all successful inquiry must approximate.

The Apex maximizes Thagard's global constraint satisfaction under Zollman's optimal topology, emerging as Kitcher's adaptive peak on Rescher's systematicity landscape. It is not a pre-existing metaphysical blueprint but a structural emergent: the asymptotic intersection of all low-brittleness models (Ladyman and Ross 2007). Its status resonates with the pragmatist ideal end of inquiry (Peirce 1878). Our Consensus Network is our fallible map of this objective structure, stabilized through adaptive feedback from mind-independent constraints.

The Apex Network's ontological status requires careful specification to avoid foundationalist overreach and relativist collapse. We propose understanding it as a "structural emergent": a real, objective pattern crystallizing from interaction between inquiry practices and environmental resistance. Consider how objective structural facts can emerge from seemingly subjective domains: while individual color preference is contingent, cross-cultural data shows striking convergence on blue. This pattern is not an accident but an emergent structural fact demanding naturalistic explanation.

Pragmatic pushback shaping this landscape is evolutionary selection on shared biology. Human color vision was forged by navigating terrestrial environments, where efficiently tracking ecologically critical signals, such as safe water and ripe fruit, conferred viability advantage (Berlin and Kay 1969; Henrich 2015). The Apex Network has the same ontological status: not found but formed, the objective structural residue after pragmatic filtering has eliminated less viable alternatives.

The mechanism forging this structure is bottom-up emergence driven by cross-domain consistency needs. Local Shared Networks, developed to solve specific problems, face pressure to cohere because they operate in an interconnected world. This pressure creates tendency toward integration, though whether this yields a single maximally coherent system or stable pluralism remains empirical.

The framework makes no a priori claims about universal convergence. Domains with tight pragmatic constraints (basic engineering, medicine) show strong convergence pressures. Others (aesthetic judgment, political organization) may support multiple stable configurations. The Apex Network concept is thus a limiting case: the theoretical endpoint of convergence pressures where they operate, not a guarantee of uniform action across all inquiry domains.

The Apex Network's function as standard for objective truth follows from this status. Using Susan Haack's (1993) crossword puzzle analogy: a proposition is objectively true because it is an indispensable component of the unique, fully completed, maximally coherent solution to the entire puzzle—a solution disciplined by thousands of external "clues" as pragmatic pushback.

This process is retrospective and eliminative, not teleological. Individual agents and networks solve local problems and reduce costs. The Apex Network is the objective, convergent pattern emerging as unintended consequence of countless local efforts to survive the failure filter. Its objectivity arises from the mind-independent nature of pragmatic constraints reliably generating costs for violating systems. This view resonates with process metaphysics (Rescher 1996),

understanding the objective structure as constituted by the historical process of inquiry itself, not as a pre-existing static form.

The Apex Network's status is dual, a distinction critical to our fallibilist realism. Ontologically, it is real: the objective, mind-independent structure of viability that exists whether we correctly perceive it or not. Epistemically, it remains a regulative ideal. We can never achieve final confirmation our Consensus Network perfectly maps it; our knowledge is necessarily incomplete and fallible. Its existence grounds our realism and prevents collapse into relativism, while our epistemic limitations make inquiry a permanent and progressive project.

Thus, the Apex Network should not be misconstrued as a single, final theory of everything. Rather, it is the complete set of maximally viable configurations—a high-altitude plateau on the fitness landscape. While some domains may have single sharp peaks, others may permit constrained pluralism of equally low-brittleness systems. Convergence is away from vast valleys of failure documented in the Negative Canon, and toward this resilient plateau of viable solutions.

The Apex Network's objectivity stems not from historical contingency but from modal necessity. However, "necessity" here must be understood carefully to avoid foundationalist or Platonic connotations. The Apex Network is not a timeless structure existing independently of the constraint-generating world—it is better understood as a constraint-determined attractor: the structural endpoint that must emerge given reality's invariant features. Its necessity is functional rather than metaphysical, analogous to how the structure of a river delta is "necessary" given water flow, gravity, and sediment dynamics. The delta does not exist apart from these forces but emerges necessarily from their operation.

The Necessity Argument:

1. Reality imposes non-negotiable constraints: physical laws (thermodynamics, resource scarcity), biological facts (human needs, mortality, cooperation requirements), logical requirements (consistency), and coordination necessities (collective action problems).

2. These constraints determine a fitness landscape of possible social configurations. A topology where some paths are viable and others catastrophic.

3. Constraints determine a fitness landscape where some configurations are more viable than others. While multiple locally optimal solutions may exist (pluralism at the Pluralist Frontier), vast regions generate catastrophic costs (the Negative Canon). The Apex Network comprises the set of maximally viable configurations—the high-altitude plateau, not necessarily a single peak.

4. The Apex Network IS that optimal structure. The configuration space of maximally viable solutions. It exists whether we've discovered it or not, determined by constraints rather than by our beliefs about it—but its existence is the existence of a structural pattern emergent from those constraints, not a Platonic form existing timelessly outside them.

Conclusion: The Apex Network emerges necessarily from constraints, independent of discovery—revealed, not created by inquiry. But "emerges necessarily" means the pattern is determined by invariant relationships, not that it exists as a pre-temporal blueprint.

Historical filtering is how we discover this structure, not how we create it. Failed systems are experiments revealing where the landscape drops off. The

Negative Canon maps the canyons and cliffs. Over time, with sufficient experiments across diverse conditions, we triangulate toward the peaks.

This crucial distinction—that historical filtering is a discovery process, not a creation mechanism—resolves the ambiguity. History reveals the landscape: experiments map hazards (failures) and peaks (successes) through trial-and-error, but does not create constraints or optimal solutions.

Analogy: Mathematical Discovery. Mathematicians in different cultures contingently discovered the same necessary truth (π) because it is determined by the objective constraints of geometry. Ancient Babylonians approximated it as 25/8, Archimedes used polygons to bound it, Indian mathematicians developed infinite series for it. Discovery processes varied radically across cultures and methods, yet all converged on the same value because π is a necessary feature of Euclidean space. Its value exists whether calculated or not, determined by geometric constraints rather than human choices.

Parallel: Epistemic Discovery. Similarly, different societies, through their contingent histories of failure and success, are forced to converge on the same necessary structures of viability because they are determined by objective pragmatic constraints. Independent cultures discovered reciprocity norms, property conventions, and harm prohibitions not through shared cultural transmission but because these principles are structurally necessary for sustainable social coordination. Discovery processes vary wildly; the discovered structure does not. The Apex Network has the same modal status as π: necessary, constraint-determined, and counterfactually stable.

Consequently, the Apex Network's structure is counterfactually stable: any sufficiently comprehensive exploration of the constraint landscape, across any possible history, would be forced to converge upon it. Evidence includes independent emergence of similar low-brittleness principles across isolated cultures, convergent evolution toward comparable solutions, structurally similar failures (high coercive costs, demographic stress), and mathematical convergence.

This counterfactual stability makes the Apex Network an objective standard, not a historical artifact.

4.2.1 Formal Characterization

Drawing on network theory (Newman 2010), we can formally characterize the Apex Network as:

$$A = \cap\{W\_k \mid V(W\_k) = 1\}$$

Where A = Apex Network, W_k = possible world-systems (configurations of predicates), V(W_k) = viability function (determined by brittleness metrics), and ∩ = intersection (common structure across all viable systems).

The intersection of all maximally viable configurations reveals their shared structure. This shared structure survives all possible variations in historical path: the emergent, constraint-determined necessity arising from how reality is organized.

A coherent system detached from the Apex Network isn't merely false but structurally unstable. It will generate rising brittleness until it either adapts toward the Apex Network or collapses. Coherence alone is insufficient because reality's constraints force convergence.

4.2.2 Cross-Domain Convergence and Pluralism

Cross-domain predicate propagation drives emergence: when Standing Predicates prove exceptionally effective at reducing brittleness in one domain, pressure mounts for adoption in adjacent domains. Germ theory's success in medicine pressured similar causal approaches in public health and sanitation. This successful propagation forges load-bearing, cross-domain connections constituting the Apex Network's emergent structure.

This process can be conceptualized as a fitness landscape of inquiry. Peaks represent low-brittleness, viable configurations (e.g., Germ Theory), while valleys and chasms represent high-brittleness failures catalogued in the Negative Canon (e.g., Ptolemaic System). Inquiry is a process of navigating this landscape away from known hazards and toward resilient plateaus.

4.2.3 Pre-Existing or Emergent? The Logos Question

A careful reader might notice an apparent tension in our account. On one hand, we explicitly deny the Apex Network is a "pre-existing metaphysical blueprint" and insist it is "not found but formed." On the other hand, we assert it "exists whether we've discovered it or not, determined by constraints rather than by our beliefs about it," and we invoke the π analogy to establish counterfactual stability. These claims seem contradictory—simultaneously denying and affirming pre-existence.

One might even charge that our framework simply rebrands the ancient Greek concept of logos—an objective rational order inhering in reality itself, existing timelessly and independently of human discovery—in naturalistic language while retaining its core metaphysical commitments. This section confronts this challenge directly, clarifying a subtle but crucial distinction that has been implicit throughout.

Two Senses of "Pre-Existing"

The resolution lies in distinguishing two senses of "pre-existing":

What we deny: The Apex Network is not a pre-existing metaphysical entity—not a Platonic Form, divine blueprint, or cosmic purpose that reality "aims toward." There is no transcendent realm of normative facts, no designer's intention, no teleological pull. The constraints of reality (thermodynamics, biological limits, logical consistency) are not purposive. They simply are. The Apex Network is not "out there" as an independent thing awaiting discovery.

What we affirm: The Apex Network is determined by constraints that are themselves features of reality prior to human existence. Just as π is not a "thing" but an inevitable mathematical consequence of Euclidean geometry's constraint structure, the Apex Network is the inevitable consequence of reality's pragmatic constraint structure. The constraints exist first; the optimal structure they determine is a necessary implication. The Apex Network "exists" before inquiry the way a theorem exists before its proof—true whether anyone demonstrates it or not.

Modal Determinacy Without Metaphysical Necessity

We propose understanding the Apex Network's status through modal determinacy: given the actual constraint structure of our universe, the Apex Network is the necessary optimal configuration. It is modally necessary relative to those constraints, not metaphysically necessary in an absolute sense.

Formally: In world W with constraint structure C, Apex Network A is necessarily determined such that $\forall W'[C(W') = C(W) \rightarrow A(W') = A(W)]$. That is, across all

possible worlds sharing our constraint structure, the same Apex Network would be determined regardless of historical path. But across worlds with different fundamental physics or logical systems, different Apex Networks would emerge.

The Logos Comparison: Naturalized Rational Order

The comparison to logos is therefore both apt and importantly limited. Like logos, the Apex Network represents an objective rational order, a structure determined by reality's constraints rather than human convention, a standard toward which inquiry necessarily converges, and a principle existing prior to and independent of human cognition.

However, unlike classical conceptions of logos, our framework involves:

- No cosmic intelligence: The order is necessity without purpose, constraint without design
- No teleology: Reality does not "aim at" optimal configurations; viable structures simply persist while brittle ones collapse
- No transcendence: The Apex Network is not a separate realm but an immanent pattern—the negative space revealed when failures are systematically eliminated
- Radical contingency on physical law: Had the universe operated under different fundamental constants or laws, a different Apex Network would be determined. There is no super-cosmic logos independent of physical reality itself

If one insists on invoking logos to describe our framework, we accept the label with this specification: ours is a naturalized, contingent, non-purposive logos. Rational order without cosmic reason. Objective structure without divine blueprint. Necessary implication without metaphysical foundation. The Apex Network is logos with the theology removed, leaving only the mathematical and physical necessity that minds-like-ours, operating in a universe-like-this, discover through systematic elimination of unviable alternatives.

Implications

This clarification strengthens our response to the isolation objection. A coherent system detached from reality is impossible not because we have smuggled in metaphysical realism, but because the constraints generating pragmatic pushback are themselves features of the actual world. Our fallibilism remains intact—we never achieve certainty that our map matches the territory—but the territory itself possesses an objective structure determined by the constraints that constitute it.

Inquiry is therefore a painstaking, error-driven process of triangulation. It converges on the configuration space that physical, biological, and logical necessities jointly determine. We discover a pre-determined *implication* of the world's constraint structure, not a pre-existing *thing*. This distinction, while subtle, is what allows the framework to preserve a robust naturalism while grounding genuine objectivity in the constraint topology of the actual world.

4.3 A Three-Level Framework for Truth

This emergent structure grounds a fallibilist but realist account of truth. It clarifies a documented tension in Quine's thought between truth as immanent to our best theory and truth as a transcendent regulative ideal (Tauriainen 2017). Our framework shows these are not contradictory but two necessary components of a naturalistic epistemology. It reframes truth as a status propositions earn through increasingly rigorous stages of validation.

- **Level 3: Contextual Coherence.** The baseline status for any claim. A proposition is coherent within a specific Shared Network, regardless of that network's long-term viability. This level explains the internal rationality of failed or fictional systems, but the framework's externalist check, the assessment of systemic brittleness, prevents this from being mistaken for justified truth.
- **Level 2: Justified Truth.** The highest epistemic status practically achievable. A proposition is justified as true if it is certified by a Consensus Network that has a demonstrated track record of low systemic brittleness. For all rational purposes, we are licensed to treat such claims as true. The diagnosed health of the certifying network provides powerful higher-order evidence that functions as a defeater for radical skepticism. To doubt a claim at this level, without new evidence of rising brittleness, is to doubt the entire adaptive project of science itself.
- **Level 1: Objective Truth.** The ultimate, regulative ideal of the process. A proposition is objectively true if its principles are part of the real, emergent Apex Network: the objective structure of viable solutions. While this structure is never fully mapped, it functions as the formal standard that makes our comparative judgments of "more" or "less" brittle meaningful. It is the structure toward which the reduction of systemic costs forces our knowledge systems to converge.

This layered framework avoids a simplistic "Whig history" by recognizing that Justified Truth is a historically-situated achievement. Newtonian mechanics earned its Level 2 status by being a maximally low-brittleness system for its problem-space for over two centuries. Its replacement by relativity does not retroactively invalidate that status but shows the evolutionary process at work, where an expanding problem-space revealed pragmatic constraints that required a new, more viable system. This allows for sharp, non-anachronistic historical judgments: a claim can be justifiably true in its time (Level 2) yet still be objectively false (not Level 1) when judged against the Apex Network from the perspective of a more resilient successor.

4.3.1 How Propositions Become Truth Itself: Deflationism and the Hard Core

The three-level framework reveals how propositions do not merely "correspond" to truth as an external standard but become constitutive of truth itself through functional transformation and entrenchment.

Our framework provides robust, naturalistic content to truth-attributions: to say P is true (Level 2) is to say P is certified by a low-brittleness Consensus Network; to say P is objectively true (Level 1) is to say P aligns with the emergent, constraint-determined Apex Network. Truth is what survives systematic pragmatic filtering. The predicate "is true" tracks functional role within viable knowledge systems, not correspondence to a Platonic realm.

From Validated Data to Constitutive Core: The Progression

A proposition's journey to becoming truth itself follows a systematic progression through functional transformation:

1. **Initial Hypothesis (Being-Tested):** The proposition begins as a tentative claim within some Shared Network, subject to coherence constraints and empirical testing. It is data to be evaluated.

2. **Validated Data (Locally Proven):** Through repeated application without generating significant brittleness, the proposition earns trust. Its predictions are confirmed; its applications succeed. It transitions from hypothesis to validated data, something the network can build upon.

3. **Standing Predicate (Tool-That-Tests):** The proposition's functional core, its reusable predicate, is promoted to Standing Predicate status. It becomes conceptual technology: a tool for evaluating new phenomena rather than something being evaluated. "…is an infectious disease" becomes a diagnostic standard, not a claim under test.

4. **Convergent Core Entry (Functionally Unrevisable):** As all rival formulations are relegated to the Negative Canon after generating catastrophic costs, the proposition migrates to the Convergent Core. Here it achieves Level 2 status: Justified Truth. To doubt it now is to doubt the entire system's demonstrated viability.

5. **Hard Core (Constitutive of Inquiry Itself):** In the most extreme cases, a proposition becomes so deeply entrenched that it functions as a constitutive condition for inquiry within its domain. This is Quine's hard core, the principles so fundamental that their removal would collapse the entire edifice.

Quine's Hard Core and Functional Entrenchment

Quine famously argued that no claim is immune to revision in principle, yet some claims are practically unrevisable because revising them would require dismantling too much of our knowledge structure. Our framework explains this tension through the concept of functional entrenchment driven by bounded rationality (March 1978).

A proposition migrates to the hard core not through metaphysical necessity but through pragmatic indispensability. The costs of revision become effectively infinite:

- **Logic and Basic Mathematics:** Revising logic requires using logic to evaluate the revision (infinite regress). Revising basic arithmetic requires abandoning the conceptual tools needed to track resources, measure consequences, or conduct any systematic inquiry. These exhibit maximal brittleness-if-removed.

- **Thermodynamics:** The laws of thermodynamics undergird all engineering, chemistry, and energy policy. Revising them would invalidate centuries of validated applications and require reconstructing vast swaths of applied knowledge. The brittleness cost is astronomical.

- **Germ Theory:** After decades of successful interventions, public health infrastructure, medical training, and pharmaceutical development all presuppose germ theory's core claims. Revision would collapse these systems, generating catastrophic first-order costs.

The Paradox Resolved: Fallibilism Without Relativism

How can we be fallibilists who acknowledge all claims are revisable in principle, while simultaneously treating hard core propositions as effectively unrevisable in practice? The resolution: "revisable in principle" means if we encountered sufficient pragmatic pushback, we would revise even hard core claims. For hard core propositions, the threshold is extraordinarily high but not infinitely high. This makes the framework naturalistic rather than foundationalist. Hard core status is functional achievement, not metaphysical bedrock.

Truth is not discovered in a Platonic realm but achieved through historical filtering. Propositions become true by surviving systematic application without generating brittleness, migrating from peripheral hypotheses to core infrastructure, becoming functionally indispensable to ongoing inquiry, and aligning with the emergent, constraint-determined Apex Network.

This resolves the classical tension between Quine's holism (all claims are revisable) and the practical unrevisability of core principles: both describe different aspects of the same evolutionary process through which propositions earn their status by proving their viability under relentless pragmatic pressure.

Animating Quine's Web: From Static Structure to Dynamic Process

Quine's "Web of Belief" (Quine 1951, 1960) provided a powerful static model of confirmational holism, but it has been criticized for lacking a dynamic account of its formation and change. Our framework provides the missing mechanisms.

First, pragmatic pushback supplies the externalist filter that grounds the web in mind-independent reality. This relentless, non-discursive filter of real-world consequences prevents the web from floating free of constraints.

Second, the entrenchment of pragmatically indispensable principles in the system's core provides a directed learning mechanism. A proposition migrates to the core not by convention but because it has demonstrated immense value in lowering the entire network's systemic brittleness, making its revision catastrophically costly. This process, driven by bounded rationality (March 1978), functions as systemic caching: proven principles are preserved to avoid re-derivation costs.

For example, Conservation of Energy became entrenched after proving indispensable across domains, its revision now prohibitively expensive.

Together, these two mechanisms animate Quine's static web. Pragmatic pushback provides the external discipline, and the entrenchment of low-brittleness principles explains how the web's resilient core is systematically constructed over time (Carlson 2015). This transforms the web from a logical snapshot into an evolving reef chart. In doing so, it resolves a documented tension in Quine's thought between truth as immanent to our best theory and truth as a transcendent regulative ideal (Tauriainen 2017). Our three-level framework shows these are not contradictory but two necessary components of a naturalistic epistemology. Core principles achieve Justified Truth (Level 2) through this process of systematic, externally-validated selection, while the Apex Network (Level 3) functions as the regulative structure toward which our theories converge.

The Individual Revision Mechanism: Conscious Awareness as Feedback Interface

While the above describes macro-level animation of the web, a complete account requires explaining the micro-level mechanism: how do individual agents actually revise their personal webs of belief? Quine's original formulation described recalcitrant experience forcing adjustment but did not explain the phenomenology of revision or why agents choose particular revision strategies when multiple options exist. Our disposition-based account of belief completes this picture.

As established in Section 2.1.2, agents possess conscious awareness of their own dispositions to assent. This awareness functions as a natural feedback mechanism that enables belief revision. When an agent holds a disposition generating pragmatic costs, the revision cycle proceeds through several stages. First,

dispositional conflict emerges: the agent's disposition produces failed predictions, wasted effort, coordination failures, or other measurable costs. Second, conscious recognition occurs as the agent becomes aware that holding this disposition correlates with these costs, either through direct experience or social signaling. Third, this awareness creates motivational pressure—the discomfort of cognitive dissonance, frustration at repeated failure, or pragmatic motivation to reduce costs. Fourth, the agent consciously explores alternative dispositions compatible with their core web commitments, testing adjustments mentally or through limited trials. Fifth, through repeated practice and positive reinforcement, a new disposition stabilizes, replacing the costly pattern. Finally, the agent's conscious model of their own belief system updates to reflect this revision, stored in memory for future reference.

This cycle operates at the individual level but drives macro-level convergence when aggregated across populations. Multiple agents independently experiencing costs from similar dispositions will independently revise toward lower-cost alternatives. When these revisions are communicated through assertion and coordinated through social exchange, patterns of convergence emerge—not through central planning or mysterious coordination but through distributed pragmatic optimization. Each agent, responding to locally experienced costs, makes adjustments that happen to align with others' adjustments because they are all responding to the same underlying constraint structure.

The role of memory deserves emphasis. Quine's web metaphor risks implying a static structure, but actual belief systems require multiple forms of memory operating simultaneously. Dispositional memory maintains stable patterns of assent over time, preventing constant drift. Revision memory stores awareness of past adjustments and their outcomes, enabling learning from history. Cost memory accumulates experience of which dispositional patterns generate brittleness, functioning as a pragmatic evaluation metric. And coordination memory preserves learned patterns of successful social alignment, facilitating efficient future cooperation.

Conscious awareness of dispositions includes awareness of their history—we remember not just what we believe but how beliefs have changed and what costs prompted those changes. This historical awareness allows belief systems to function as evolving, learning systems rather than static webs. An agent who remembers that adopting disposition D reduced costs compared to previous disposition D' is better positioned to make future revisions, creating a ratcheting effect where successful adjustments are preserved while failures are discarded.

Why does this matter for convergence on the Apex Network? The Apex is not a pre-existing target that agents consciously aim for. It is the emergent structure that must arise when millions of agents, each consciously aware of their own dispositions and the costs they generate, independently revise toward lower-brittleness patterns. Convergence is explained not by mysterious coordination or shared access to truth but by the simple fact that reality's constraint structure punishes certain dispositional patterns and rewards others, and conscious agents can detect and respond to this feedback. The conscious awareness component is what makes agents capable of systematic belief revision rather than random drift. Without it, we cannot explain deliberate adjustment, learning from experience, or the directedness of convergence toward viability.

This transforms Quine's web from a passive logical structure responding to experience into an active cognitive system capable of deliberate self-modification in response to detected costs. Pragmatic pushback is not an abstract force—it is

experienced by individuals as the frustration, failure, and friction of holding unviable dispositions, motivating the endless revision process that drives knowledge toward the Apex Network.

This three-level truth framework describes the justificatory status of claims at a given moment. Over historical time, pragmatic filtering produces a discernible two-zone structure in our evolving knowledge systems.

4.4 The Evolving Structure of Knowledge: Convergent Core and Pluralist Frontier

The historical process of pragmatic filtering gives our evolving Consensus Networks a discernible structure, which can be understood as having two distinct epistemic zones. This distinction is not about the nature of reality itself, but describes the justificatory status of our claims at a given time.

- **The Convergent Core:** This represents the load-bearing foundations of our current knowledge. It comprises domains where the relentless pressure of pragmatic selection has eliminated all known rival formulations, leaving a single, or functionally identical, set of low-brittleness principles. Principles reside in this core, such as the laws of thermodynamics or the germ theory of disease, not because they are dogmatically held or self-evident but because all tested alternatives have been relegated to the Negative Canon after generating catastrophically high systemic costs. While no claim is immune to revision in principle, the principles in the Convergent Core are functionally unrevisable in practice, as doing so would require dismantling the most successful and resilient knowledge structures we have ever built. A claim from this core achieves the highest degree of justification we can assign, approaching our standard for Objective Truth (Level 1).
- **The Pluralist Frontier:** This describes the domains of active research where our current evidence is insufficient to decide between multiple, competing, and viable reconstructions of the landscape of viability. Here, rival systems (e.g., different interpretations of quantum mechanics or competing models of consciousness) may coexist, each with a demonstrably low and stable degree of brittleness. It is crucial to distinguish this constrained, evidence-based pluralism from relativism. The frontier is not an "anything goes" zone but a highly restricted space strictly bounded on all sides by the Negative Canon. A system based on phlogiston is not a "viable contender" on the frontier of chemistry but a demonstrably failed research program. This pluralism is therefore a sign of epistemic underdetermination: a feature of our map's current limitations, not reality's supposed indifference. This position resonates with pragmatist accounts of functional pluralism (Price 1992), which treat different conceptual frameworks as tools whose legitimacy is determined by their utility within a specific practice. Within this frontier, the core claims of each viable competing system can be granted the status of Justified Truth (Level 2). This is also the zone where non-epistemic factors, such as institutional power or contingent path dependencies, can play their most significant role, sometimes artificially constraining the range of options explored or creating temporary monopolies on what is considered justified.

4.5 Illustrative Cases of Convergence and Brittleness

The transition from Newtonian to relativistic physics offers a canonical example of this framework's diagnostic application. After centuries of viability, the

Newtonian system began to accumulate significant systemic costs in the late 19th century. These manifested as first-order predictive failures, such as its inability to account for the perihelion of Mercury, and as rising conceptual debt in the form of ad-hoc modifications like the Lorentz-FitzGerald contraction hypothesis. This accumulating brittleness created what Kuhn (1962) termed a "crisis" state preceding paradigm shifts. The Einsteinian system proved a more resilient solution, reducing this conceptual debt and substantially lowering the systemic costs of inquiry in physics.

A more contemporary case can be found in the recent history of artificial intelligence, which illustrates how a brittleness assessment might function in real time. The periodic "AI winters" can be understood as the collapse of high-brittleness paradigms, such as symbolic AI, which suffered from a high rate of ad-hoc modification when faced with novel challenges. While the subsequent deep learning paradigm proved a low-brittleness solution for many specific tasks, it may now be showing signs of rising systemic costs. These can be described conceptually as, for example, potentially unsustainable escalations in computational and energy resources for marginal performance gains, or an accelerating research focus on auxiliary, post-hoc modifications rather than on foundational architectural advances. This situation illustrates the Pluralist Frontier in action, as rival architectures might now be seen as competing to become the next low-brittleness solution.

4.6 Navigating the Landscape: Fitness Traps, Path Dependence, and the Role of Power

An evolutionary model of knowledge must account for the complexities of history, not just an idealized linear progress. The landscape of viability is not smooth: knowledge systems can become entrenched in suboptimal but locally stable states, which we term "fitness traps" (Wright 1932). This section clarifies how the framework incorporates factors like path dependence and institutional power not as external exceptions but as core variables that explain these historical dynamics.

The model's claim is not deterministic prediction but probabilistic analysis: beneath the surface-level contingency historians rightly emphasize, underlying structural pressures create statistical tendencies over long timescales. A system accumulating brittleness is not fated to collapse on a specific date but becomes progressively more vulnerable to contingent shocks. The model thus complements historical explanation by offering tools to understand why some systems prove more resilient than others.

A system can become locked into a high-brittleness fitness trap by coercive institutions or other path-dependent factors. A slave economy, for instance, is a classic example. While objectively brittle in the long run, it creates institutional structures that make escaping the trap prohibitively costly in the short term (Acemoglu and Robinson 2012). The framework's key insight is that the exercise of power does not negate a system's brittleness; rather, the costs of maintaining that power become a primary indicator of it. This power manifests in two interrelated ways. First is its defensive role: the immense coercive overheads required to suppress dissent and manage internal friction are a direct measure of the energy a system must expend to resist the structural pressures pushing it toward collapse.

Second, power plays a constitutive role by actively shaping the epistemic landscape itself. Powerful institutions do not merely respond to brittleness defensively; they can construct and maintain the very conditions that generate it. By controlling research funding, defining what counts as a legitimate problem, and entrenching path dependencies that reinforce a fitness trap, institutional power actively digs the fitness trap and locks the system into a high-brittleness state. This pattern of epistemic capture appears across domains: from tobacco companies suppressing health research to colonial knowledge systems that extracted Indigenous insights while denying reciprocal engagement, thereby masking brittleness through institutional dominance. While this can create a temporary monopoly on justification, the framework can still diagnose the system's underlying brittleness. The costs of this constitutive power often manifest as a lack of adaptability, suppressed innovation, and a growing inability to solve novel problems that fall outside the officially sanctioned domain. To detect such hidden brittleness, we can augment $C(t)$ with sub-metrics for innovation stagnation, tracking lags in novel applications or cross-domain extensions relative to comparable systems as proxies for suppressed adaptive capacity. Concretely, innovation lag can be operationalized as: **$I(t) =$ (Novel Applications per Unit Time) / (Defensive Publications per Unit Time)**. When $I(t)$ declines while $C(t)$ remains high, this signals power-induced rigidity masking underlying brittleness. For example, Lysenkoist biology in the Soviet Union showed $I(t)$ approaching zero (no cross-domain applications) while defensive publications proliferated. Over historical time, even the most entrenched systems face novel shocks, where the hidden costs of their power-induced rigidity are typically revealed.

The severity of a fitness trap can be conceptually diagnosed. The work of historians using cliodynamic analysis, for example, is consistent with this view, suggesting that the ratio of a state's resources dedicated to coercive control versus productive capacity can serve as a powerful indicator of systemic fragility. Historical analyses have found that polities dedicating a disproportionately high and sustained share of their resources to internal suppression often exhibit a significantly higher probability of fragmentation when faced with external shocks (Turchin 2003). This provides a way to conceptually diagnose the depth of a fitness trap: by tracking the measurable, defensive costs a system must pay to enforce its power-induced constraints.

Finally, it is necessary to distinguish this high-brittleness fitness trap from a different state: low-brittleness stagnation. A system can achieve a locally stable, low-cost equilibrium that is highly resilient to existing shocks but lacks the mechanisms for generating novel solutions. A traditional craft perfected for a stable environment but unable to adapt to a new material, or a scientific paradigm efficient at solving internal puzzles but resistant to revolutionary change, exemplifies this state. While not actively accumulating systemic costs, such a system is vulnerable to a different kind of failure: obsolescence in the face of a faster-adapting competitor. Diagnosing this condition requires not only a static assessment of current brittleness but also an analysis of the system's rate of adaptive innovation. True long-term viability therefore requires a balance between low-cost stability and adaptive capacity. This evolutionary perspective completes our reef chart, not as a finished map, but as an ongoing process of hazard detection and channel discovery.

5. Applications: Mathematics as a Paradigm Case of Internal Brittleness

The framework developed thus far relies primarily on examples from empirical science and social systems, where pragmatic pushback manifests through observable failures in the external world. A natural objection arises: what about abstract domains like mathematics, where empirical falsification seems irrelevant? Naturalistic epistemologies often treat mathematics as a boundary problem. This section shows mathematics is not a problematic edge case but the framework's purest illustration, demonstrating how pragmatic selection operates when feedback is entirely internal to the system.

The Core Insight: Mathematical frameworks face pragmatic pushback through internal inefficiency rather than external falsification.

5.1 The Logic of Internal Brittleness

While mathematical frameworks cannot face direct empirical falsification, they experience pragmatic pushback through accumulated internal costs that render them unworkable. These costs manifest through our standard brittleness indicators, adapted for abstract domains:

M(t): Proof Complexity Escalation - Increasing proof length without proportional explanatory gain - Measured as: average proof length for theorems of comparable scope over time - Rising M(t) signals a degenerating research program where increasing effort yields diminishing insight

P(t): Conceptual Debt Accumulation (proxied by Axiom Proliferation) - Ad-hoc modifications to patch paradoxes or anomalies - Measured as: ratio of new axioms added to resolve contradictions vs. axioms generating novel theorems - High P(t) indicates mounting conceptual debt from defensive modifications

C(t): Contradiction Suppression Costs - Resources devoted to preventing or managing paradoxes - Measured as: proportion of research addressing known anomalies vs. extending theory - High C(t) reveals underlying fragility requiring constant maintenance

R(t): Unification Power - Ability to integrate diverse mathematical domains under common framework - Measured as: breadth of cross-domain applicability - Declining R(t) signals fragmentation and loss of coherence

The abstract costs in mathematics can be operationalized using our diagnostic toolkit, demonstrating the framework's universality across domains where feedback is entirely internal to the system.

5.2 Case Study: Brittleness Reduction in Mathematical Foundations

To illustrate these metrics in action, consider examples of mathematical progress as brittleness reduction across different domains:

Non-Euclidean Geometry: - Euclidean geometry exhibited high brittleness for curved space applications - Required elaborate patches (like epicycles in astronomy) to explain planetary motion - Non-Euclidean alternatives demonstrated lower brittleness for cosmology and general relativity - **Pattern:**

Replace high-brittleness framework with lower-brittleness alternative when problem domain expands

Calculus Foundations: - Infinitesimals were intuitive but theoretically brittle, generating paradoxes of the infinite - Epsilon-delta formalism demanded higher initial complexity but delivered lower long-term brittleness - Historical adoption pattern follows brittleness reduction trajectory - Demonstrates how short-term complexity increase can yield long-term stability gains

Category Theory: - More abstract and initially more complex than set theory - But demonstrates lower brittleness for certain domains (algebraic topology, theoretical computer science) - Adoption follows domain-specific viability assessment - Shows how optimal framework varies by application domain

Nowhere is this dynamic clearer than in the response to Russell's Paradox, which provides a paradigm case of catastrophic brittleness and competing resolution strategies.

Naive Set Theory (pre-1901): - $M(t)$: Moderate—proofs reasonably concise for most theorems - $R(t)$: Exceptional—successfully unified logic, number theory, and analysis under a single framework - Appeared to exhibit low brittleness across all indicators - Provided an elegant foundation for mathematics

Russell's Paradox (Russell 1903): - Revealed infinite brittleness: the theory could derive a direct contradiction - Considered the set $R = \{x \mid x \notin x\}$. Is $R \in R$? Both yes and no follow from the axioms - All inference paralyzed (if both A and $\neg A$ are derivable, the principle of explosion allows derivation of anything) - Complete systemic collapse—the framework became unusable for rigorous mathematics - This wasn't a peripheral anomaly but a catastrophic failure at the system's foundation

Response 1: ZF Set Theory (Zermelo-Fraenkel + Axiom of Choice) - Added carefully chosen axioms (Replacement, Foundation, Separation) to block the paradox - $M(t)$: Increased (more axioms create more complex proof requirements) - $P(t)$: Moderate (new axioms serve multiple purposes beyond merely patching the paradox) - $C(t)$: Low (paradox completely resolved, no ongoing suppression or management needed) - $R(t)$: High (maintained most of naive set theory's unifying power across mathematical domains) - **Diagnosis:** Successful low-brittleness resolution through principled modification - The additional complexity was justified by restored foundational stability

Response 2: Type Theory (Russell/Whitehead) - Introduced stratified hierarchy that structurally prevents problematic self-reference - $M(t)$: High (complicated type restrictions make many proofs substantially longer) - $P(t)$: Low (structural solution rather than ad-hoc patch) - $C(t)$: Low (paradox is structurally impossible within the system) - $R(t)$: Moderate (some mathematical domains resist natural formulation within type hierarchies) - **Diagnosis:** Alternative low-brittleness solution with different trade-offs - Sacrifices some unification power for structural guarantees against contradiction

Response 3: Paraconsistent Logic - Accepts contradictions as potentially derivable but attempts to control "explosion" - $M(t)$: Variable (depends on specific implementation details) - $P(t)$: Very High (requires many special rules and restrictions to prevent inferential collapse) - $C(t)$: Very High (demands constant management of contradictions and their containment) - $R(t)$: Low (marginal adoption, limited to specialized domains) - **Diagnosis:** Higher brittleness—requires ongoing suppression costs to remain functional - The system exhibits sustained high maintenance costs without corresponding payoffs

Historical Outcome: The mathematical community converged primarily on ZF set theory as the standard foundation, with Type Theory adopted for specific domains where its structural guarantees prove valuable (such as computer science and constructive mathematics). Paraconsistent approaches remain peripheral. This convergence reflects differential brittleness among the alternatives, not arbitrary historical preference or mere convention. The outcome demonstrates how pragmatic selection operates in purely abstract domains through internal efficiency rather than external empirical testing.

5.3 Power, Suppression, and the Hard Core

Engaging with insights from feminist epistemology (Harding 1991), we can see that even mathematics is not immune to power dynamics that generate brittleness. When a dominant mathematical community uses institutional power to suppress alternative approaches, this incurs measurable Coercive Overheads (C(t)):

Mechanisms of Mathematical Suppression: - Career punishment for heterodox approaches to foundations or proof methods - Publication barriers for alternative mathematical frameworks - Curriculum monopolization by dominant approaches - Citation exclusion of rival methodologies

Measurable Costs: - **Innovation lag:** Talented mathematicians driven from the field when their approaches are rejected for sociological rather than technical reasons - **Fragmentation:** Splinter communities forming alternative journals and departments - **Inefficiency:** Duplication of effort as alternative approaches cannot build on dominant framework results - **Delayed discoveries:** Useful insights suppressed for decades (e.g., non-standard analysis resisted despite valuable applications)

The Brittleness Signal: When a mathematical community requires high coercive costs to maintain orthodoxy against persistent alternatives, this signals underlying brittleness—the dominant framework may not be optimally viable.

Historical Example: Intuitionist vs. Classical Mathematics - Intuitionists demonstrated genuine technical alternatives with different foundational commitments - Classical community initially suppressed through institutional power (career barriers, publication difficulties) - High coercive costs required to maintain dominance - Eventual accommodation as constructive methods proved valuable in computer science and proof theory - **Diagnosis:** Initial suppression revealed brittleness in classical community's claim to unique optimality

Why Logic Occupies the Core

Logic isn't metaphysically privileged—it's functionally indispensable.

The Entrenchment Argument: 1. Revising logic requires using logic to assess the revision 2. This creates infinite regress or circularity 3. Therefore logic exhibits infinite brittleness if removed 4. Systems under bounded rationality (March 1978) must treat such maximal-cost revisions as core

This is pragmatic necessity, not a priori truth: - Logic could theoretically be revised if we encountered genuine pragmatic pressure sufficient to justify the cost - Some quantum logics represent such domain-specific revisions - But the cost threshold is exceptionally high—logic underpins all systematic reasoning - Most "apparent" logic violations turn out to be scope restrictions rather than genuine revisions of core principles

5.4 The General Principle: Mathematics as Pure Pragmatic Selection

Mathematics demonstrates the framework applies beyond empirical domains. All domains face pragmatic selection, though the feedback mechanism varies: external prediction failure for physics, social collapse for politics, internal inefficiency for mathematics. The underlying principle is analogous: brittle systems accumulate costs that drive replacement by more viable alternatives. The costs differ by domain, but the selection logic remains.

Mathematics is not a special case requiring different epistemology—it's a pure case showing how pragmatic selection operates when feedback is entirely internal to the system. The convergence on ZF set theory, the accommodation of intuitionist insights, and the adoption of non-standard analysis where it proves useful all demonstrate the same evolutionary dynamics at work in physical science, but operating through internal efficiency rather than external empirical testing. This universality strengthens the framework's claim that objective knowledge arises from pragmatic filtering across all domains of inquiry.

Thus, mathematics, far from being a counterexample to a naturalistic epistemology, serves as its purest illustration, demonstrating that the logic of brittleness reduction operates universally, guided by the selective pressures of internal coherence and efficiency.

Having demonstrated how brittleness diagnostics apply even to abstract domains like mathematics, we now situate EPC within broader epistemological debates.

6. Situating the Framework in Contemporary Debates

This paper has developed what can be termed **Systemic Externalism**—a form of externalist epistemology that locates justification not in individual cognitive processes but in the demonstrated reliability of entire knowledge systems. This section clarifies the framework's position within contemporary epistemology by examining its relationship to four major research programs: coherentist epistemology, social epistemology, evolutionary epistemology, and neopragmatism.

6.1 A Grounded Coherentism and a Naturalized Structural Realism

While internalist coherentists like Carlson (2015) have successfully shown that the web must have a functionally indispensable core, they lack resources to explain why that core is forged by external discipline. Systemic Externalism provides this missing causal engine, grounding Carlson's internal structure in an externalist history of pragmatic selection. Justification requires coherence plus network reliability via low brittleness. Unlike Zollman's (2007, 2013) static network models and Rosenstock et al. (2017), EPC examines evolving networks under pushback, extending ECHO's harmony principle with external brittleness filters Thagard's internalism lacks.

6.1.1 A Naturalistic Engine for Structural Realism

The Apex Network aligns with structural realism (Worrall 1989), providing its missing naturalistic engine. It explains convergence on objective structures via pragmatic filtering: brittle theories fail systematically, low-brittleness ones survive. The historical record shows systematic elimination of high-brittleness systems. The convergence toward low-brittleness structures, documented in the Negative Canon, provides positive inductive grounds for realism about the objective viability landscape our theories progressively map.

This provides an evolutionary, pragmatic engine for Ontic Structural Realism (Ladyman and Ross 2007). While OSR posits that the world is fundamentally structural, our framework explains how scientific practices are forced to converge on these objective structures through pragmatic filtering. The Apex Network is the complete set of viable relational structures, an emergent fact about our world's constraint topology, discovered through pragmatic selection.

6.1.2 Distinguishing Systemic Externalism from Other Externalisms

Systemic Externalism contrasts with Process Reliabilism (Goldman 1979) and Virtue Epistemology (Zagzebski 1996). Process Reliabilism locates justification in the reliability of individual cognitive processes; Systemic Externalism shifts focus to the demonstrated historical viability of the public knowledge system that certifies the claim. Virtue Epistemology grounds justification in individual intellectual virtues; Systemic Externalism attributes resilience and adaptability to the collective system. Systemic Externalism thus offers macro-level externalism, complementing these micro-level approaches.

6.2 A Realist Corrective to Neopragmatism and Social Epistemology

The framework developed here retains pragmatism's anti-foundationalist spirit and focus on inquiry as a social, problem-solving practice. Its core ambition aligns with the foundational project of classical pragmatism: to articulate a non-reductive naturalism that can explain the emergence of genuine novelty in the world (Baggio and Parravicini 2019). By grounding epistemology in dispositions to assent shaped by pragmatic feedback—following Quine's call to replace traditional epistemology with empirical psychology (Quine 1969)—we maintain naturalistic rigor while avoiding the foundationalist trap of positing privileged mental contents. Our disposition-based account provides precisely what Quine's naturalized epistemology promised but could not fully deliver: a bridge from individual cognitive behavior to social knowledge systems that remains fully naturalistic while accounting for external constraints.

However, our model offers a crucial corrective to neopragmatist approaches that are vulnerable to the charge of conflating epistemic values with mere practical utility (Putnam 2002; Lynch 2009) or reducing objectivity to social consensus. Thinkers like Rorty (1979) and Brandom (1994), in their sophisticated accounts of justification as a linguistic or social practice, lack a robust, non-discursive external constraint. This leaves them with inadequate resources for handling cases where entire communities, through well-managed discourse, converge on unviable beliefs.

Our framework provides this missing external constraint through its analysis of systemic failure. The collapse of Lysenkoist biology in the Soviet Union, for instance, was not due to a breakdown in its internal "game of giving and asking for reasons"—indeed, that discourse was brutally enforced. Its failure was a

matter of catastrophic first-order costs that no amount of conversational management could prevent. This focus on pragmatic consequence as a real, external filter allows us to distinguish our position from other forms of "pragmatic realism." El-Hani and Pihlström (2002), for example, resolve the emergentist dilemma by arguing that emergent properties "gain their ontological status from the practice-laden ontological commitments we make." While we agree that justification is tied to practice, our model grounds this process in a more robustly externalist manner. Pragmatic viability is not the source of objectivity; it is the primary empirical indicator of a system's alignment with the mind-independent, emergent structure of the Apex Network.

This leads to a key reframing of the relationship between agreement and truth. Genuine solidarity is not an alternative to objectivity but an emergent property of low-brittleness systems that have successfully adapted to pragmatic constraints. The practical project of cultivating viable knowledge systems is therefore the most secure path to enduring agreement. This stands in sharp contrast to any attempt to define truth as a stable consensus within a closed system, a procedure that our framework would diagnose as a potential coherence trap lacking the necessary externalist check of real-world systemic costs.

Similarly, our framework provides an evolutionary grounding for the core insights of **social epistemology** (Goldman 1999; Longino 2002). Social epistemic procedures like peer review and institutionalized criticism are not justified a priori; they persist because they are evolved adaptive strategies that demonstrably reduce systemic brittleness by helping networks detect errors and pay down conceptual debt. This provides the externalist check that purely procedural models can lack. It also offers an empirical grounding for the central insight of standpoint theory (Harding 1991; Lugones 2003), naturalizing the idea that marginalized perspectives can be a privileged source of data about a system's hidden costs. In our model, marginalized perspectives are not privileged due to a metaphysical claim about identity, but because they often function as the most sensitive detectors of a system's First-Order Costs and hidden Coercive Overheads ($C(t)$). A system that appears stable to its beneficiaries may be generating immense, unacknowledged costs for those at its margins. Suppressing these perspectives is therefore not just a moral failure, but a critical epistemic failure that allows brittleness to accumulate undetected. This view of collective knowledge as an emergent, adaptive process finds resonance in contemporary work on dynamic holism (Sims 2024).

Collective Calibration

Empirical models of social epistemic networks (O'Connor and Weatherall 2019) suggest that objectivity is a function of communication topology. EPC operationalizes this insight: calibration efficiency inversely correlates with brittleness. The more diverse the error signals integrated (Longino 1990; Anderson 1996), the more stable the Apex Network.

6.3 Distinguishing from Lakatos and Laudan

While our framework shares a historical-diagnostic ambition with Lakatos (1970) and Laudan (1977), it differs fundamentally: they provide retrospective descriptions of scientific change; we offer a forward-looking causal engine via quantifiable brittleness. Brittleness measures accumulated costs causing degeneration, serving as a real-time diagnostic of structural health, not merely historical output.

Similarly, while Laudan's model evaluates a theory based on the number and importance of the empirical problems it solves, our approach is subtly different. Systemic brittleness is a forward-looking measure of epistemic risk and resilience (Pritchard 2016). A system could have a high problem-solving score in Laudan's sense while simultaneously accumulating hidden systemic costs (like massive computational overheads or conceptual debt) that make it profoundly vulnerable to future shocks. Our framework is thus less a retrospective accounting of solved puzzles and more a real-time assessment of a system's long-term viability and adaptive efficiency.

6.4 Plantinga's Challenge: Does Evolution Select for Truth or Mere Survival?

Alvin Plantinga's Evolutionary Argument Against Naturalism (EAAN) poses a formidable challenge to any naturalistic epistemology: if our cognitive faculties are products of natural selection, and natural selection optimizes for reproductive success rather than true belief, then we have no reason to trust that our faculties reliably produce true beliefs (Plantinga 1993, 2011). Evolution could equip us with systematically false but adaptive beliefs—useful fictions that enhance survival without tracking reality. If naturalism is true, the very faculties we use to conclude naturalism is true are unreliable, rendering naturalism self-defeating.

Our framework provides a novel response by collapsing Plantinga's proposed gap between adaptive success and truth-tracking. We argue that in domains where systematic misrepresentation generates costs, survival pressure and truth-tracking converge necessarily. This is not because evolution "cares about" truth, but because reality's constraint structure makes persistent falsehood unsustainable.

Systemic Brittleness as the Bridge

Plantinga's argument assumes survival and truth can come apart—that belief systems could be adaptively successful while systematically misrepresenting reality. Our framework challenges this through systemic brittleness. In any domain where actions have consequences constrained by objective features of reality, false beliefs accumulate costs:

- Navigation beliefs with systematic errors generate failures, wasted energy, increased predation risk
- Causal beliefs that misunderstand cause-effect relationships lead to ineffective interventions and resource waste
- Social beliefs that misread dynamics generate coordination failures and coalition instability

These costs compound. A single false belief might be offset by other advantages, but networks of mutually reinforcing falsehoods generate accelerating brittleness through our $P(t)$ mechanism (conceptual debt accumulation). The phlogiston theorist does not just hold one false belief; they must continuously patch their system with ad-hoc modifications as each application reveals new anomalies. This brittleness makes the system vulnerable to displacement by more viable alternatives.

Domain Specificity: Where Truth-Tracking Matters

Our response is not that evolution always selects for true belief. Rather, we identify specific conditions under which survival pressure forces truth-tracking. Domain structure determines whether falsity accumulates brittleness:

High-cost domains (strong selection for truth-tracking): Physical causation, basic perception, tool use and engineering, social coordination. Misunderstanding gravity, systematically misperceiving predators, or holding false beliefs about material properties generates immediate, compounding costs.

Low-cost domains (weak selection): Metaphysical speculation, aesthetic preferences, remote historical claims. Believing in Zeus versus no gods has minimal direct costs if behavior is similar. False beliefs about "objective beauty" do not accumulate brittleness.

Systematically misleading domains (Plantinga's worry bites hardest): Self-enhancement biases, tribal epistemology, certain evolved heuristics. Here overconfidence or believing "my group is superior" may be adaptive even if false, because they coordinate action or provide psychological benefits.

This domain-specificity is crucial. Our framework concedes Plantinga's point for low-cost and systematically-misleading domains—these are precisely where the isolation objection has least force. But in high-cost domains (those that matter most for science, engineering, and social coordination), the gap Plantinga identifies cannot persist across long timescales.

Convergence Through Cultural Evolution

Plantinga focuses on individual cognitive faculties shaped by biological evolution. Our framework operates at a different level: cultural evolution of public knowledge systems. This shift is crucial. Individual humans may harbor adaptive falsehoods, but public knowledge systems are subject to distinct, higher-order selective pressures: transmissibility (false systems that work in one context often fail when transmitted to new contexts), cumulative testing (each generation's application exposes misalignment), and inter-system competition (when rival systems make incompatible predictions, differential brittleness outcomes determine which survives).

The Negative Canon provides overwhelming evidence for this process. Ptolemaic astronomy, phlogiston chemistry, miasma theory, and Lysenkoism were not merely false—they accumulated measurable, catastrophic systemic costs that forced their abandonment. The historical record shows systematic elimination of high-brittleness systems across cultures and eras, suggesting convergence toward a constraint-determined structure (the Apex Network) rather than persistent plurality of useful fictions.

Where Plantinga's Worry Remains

We acknowledge our response does not fully dissolve Plantinga's challenge at the individual level. An individual human's cognitive faculties might indeed harbor systematically false but adaptive beliefs in low-cost or systematically-misleading domains. Our claim is more modest: in domains where falsity accumulates brittleness, cumulative cultural evolution forces convergence toward truth-tracking systems, even if individual psychology remains imperfect.

This leaves open several possibilities. Evolutionary debunking arguments may retain force in genuinely low-cost domains like pure aesthetics or speculative metaphysics. However, moral philosophy is not low-cost—moral systems that systematically misrepresent human needs generate catastrophic coordination failures, demographic collapse, and rising coercive overheads. The Negative Canon demonstrates this empirically. Individual-level cognitive biases may persist even as system-level knowledge improves. Fundamental uncertainty about whether our

most basic faculties (logic, perception) are reliable cannot be eliminated—this is acknowledged in our fallibilism.

The Self-Defeat Response Reversed

Plantinga argues naturalism is self-defeating: if naturalism is true, we should not trust the faculties that led us to believe it. Our framework reverses this concern: the very fact that naturalistic science has systematically driven down systemic brittleness across centuries—enabling unprecedented technological success, predictive power, and unification—provides higher-order evidence that the system tracks the Apex Network. The proof is in the low-brittleness pudding.

If our cognitive faculties were fundamentally unreliable in the way Plantinga suggests, we would expect accelerating conceptual debt (rising $P(t)$), decreasing unification (falling $R(t)$), and catastrophic failures when theories are applied in novel domains. Instead, mature sciences show the opposite: decreasing $P(t)$, increasing $R(t)$, and successful cross-domain applications. This provides inductive grounds for trusting that our faculties, at least in high-cost domains, track real constraint structures rather than generate useful fictions.

Our response to Plantinga: In domains where systematic misrepresentation accumulates measurable costs, the supposed gap between adaptive success and truth-tracking collapses. Survival is not truth, but in a universe with stable constraints, surviving knowledge systems are forced to approximate truth because persistent falsehood generates brittleness that pragmatic selection eliminates. This is not metaphysical necessity but statistical regularity—an empirical claim falsifiable through the research program outlined in Section 7.2.

Plantinga is right that evolution per se does not guarantee reliability. But evolution plus pragmatic filtering in a constraint-rich environment does generate truth-tracking in precisely those domains where coherentism faces the isolation objection. Where Plantinga sees self-defeat, we see self-correction: the systematic reduction of brittleness over centuries is evidence that the process works, even if no individual step is guaranteed.

6.5 Computational and Systematic Precursors

EPC synthesizes four computational/systematic frameworks, advancing each through externalist brittleness diagnostics.

Thagard's ECHO (1989, 2000): Models explanatory coherence via 7 principles (symmetry, explanation, contradiction, etc.) as constraint satisfaction in connectionist networks. Activation spreads through excitatory/inhibitory weights until harmony maximizes. EPC extends this: Standing Predicates = high-activation nodes; propagation = ECHO dynamics. **Advance:** Brittleness adds dynamic weights derived from pragmatic costs. Accumulating ad-hoc patches (rising $P(t)$) would create progressively stronger inhibitory effects on the core propositions they protect, while coercive overheads (rising $C(t)$) would suppress dissenting nodes. ECHO operates internally; EPC adds pragmatic pushback as external discipline, solving Thagard's isolation problem.

Zollman's Epistemic Graphs (2007, 2010): Shows topology effects—sparse cycles preserve diversity (reliability bonus), complete graphs risk premature lock-in (speed/reliability trade-off). EPC's Pluralist Frontier operationalizes transient diversity. **Advance:** While Zollman models abstract belief propagation, EPC's framework suggests how these models could be grounded. It points toward using the historical record of systemic shocks (such as those catalogued in historical

databases) as a source of external validity checks, moving beyond purely abstract network dynamics.

Rescher's Systematicity (1973, 2001): Defines truth as praxis-tested systematicity (completeness, consistency, functional efficacy) but lacks quantification. **Advance:** SBI(t) operationalizes—P(t) = consistency metric, R(t) = completeness breadth, enabling falsifiable predictions (brittleness-collapse correlations).

Kitcher (1993) on Evolutionary Progress: Models science as credit-driven selection with division of labor across 'significant problems.' **Advance:** Negative Canon provides failure engine, brittleness quantifies problem significance via coercive costs (C(t)), diagnosing degenerating programs without reducing to cynical credit-seeking.

Sims (2024) on Dynamic Holism: Ecological constraints drive diachronic cognitive revision in nonneuronal organisms. EPC parallels this at macro-cultural scale: pragmatic pushback = ecological variables. **Distinction:** Sims focuses on individual adaptation; EPC on intergenerational knowledge systems. Both emphasize context-sensitive, constraint-driven evolution; EPC adds synchronic diagnostics to Sims' diachronic methodology.

These precursors provide micro-coherence (Thagard), meso-topology (Zollman), normative criteria (Rescher), macro-dynamics (Kitcher), and biological analogy (Sims). EPC unifies them through falsifiable brittleness assessment grounded in historical failure data.

7. Final Defense and Principled Limitations

As a macro-epistemology explaining the long-term viability of public knowledge systems, this framework does not primarily solve micro-epistemological problems like Gettier cases. Instead, it bridges the two levels through the concept of higher-order evidence: the diagnosed health of a public system provides a powerful defeater or corroborator for an individual's beliefs derived from that system.

The diagnosed brittleness of a knowledge system provides higher-order evidence that determines rational priors. Following Kelly (2005) on disagreement, when an agent receives a claim, they must condition their belief not only on the first-order evidence but also on the source's reliability (Staffel 2020). Let S be a high-brittleness network, like a denialist documentary. Its diagnosed non-viability acts as a powerful higher-order defeater. Therefore, even if S presents seemingly compelling first-order evidence E, a rational agent's posterior confidence in the claim properly remains low. Conversely, a low-brittleness network like the IPCC earns a high prior through demonstrated resilience. To doubt its claims without new evidence of rising brittleness is to doubt the entire adaptive project of science itself. This provides a rational, non-deferential basis for trust: justification flows from systemic health, grounding micro-level belief in macro-level viability.

7.1 From Hindsight to Foresight: Calibrating the Diagnostics

To address the hindsight objection—that we can only diagnose brittleness after failure—we frame the process as a two-stage scientific method: **Stage 1:**

Retrospective Calibration. We use the clear data from the Negative Canon (historical failures) to calibrate our diagnostic instruments (the $P(t)$, $C(t)$, $M(t)$, $R(t)$ indicators), identifying the empirical signatures that reliably precede collapse. **Stage 2: Prospective Diagnosis.** We apply these calibrated instruments to contemporary, unresolved cases not for deterministic prediction, but to assess epistemic risk and identify which research programs are progressive versus degenerating.

7.2 A Falsifiable Research Program

The framework grounds a concrete empirical research program with a falsifiable core claim: *there should be a strong, observable correlation between a system's measured brittleness and its probability of collapse or major revision when faced with external shocks.* Historical patterns in collapsed systems, such as Roman aqueduct failures due to accumulating brittleness in hydraulic engineering (Hodge 1992; Turchin 2003), are consistent with this expectation.[2] The specific metrics and dynamic equations underlying this research program are detailed in the Mathematical Appendix.

Methodology: (1) Operationalize brittleness through observable proxies (resource allocation patterns, auxiliary hypothesis rates in literature). (2) Conduct comparative historical analysis using databases like Seshat (a database of historical societies) to compare outcomes across systems with different pre-existing brittleness facing similar shocks, controlling for contingent events. The conceptual approach to operationalization, including how one might develop protocols for assessing $P(t)$ and $C(t)$, is sketched in Appendix B. As a conceptual illustration, consider competing COVID-19 models (2020–2022): one might analyze whether highly complex epidemiological models with many parameters showed signs of rising brittleness through persistent predictive failures and required constant revision, while simpler models maintained better predictive alignment over time (Roda et al. 2020). Such analysis would illustrate the framework's diagnostic potential.

7.3 Principled Limitations and Scope

Philosophical honesty requires acknowledging not just what a framework can do, but what it cannot. These are not flaws but principled choices about scope and method.

7.3.1 Species-Specific Objectivity
The Limitation: Moral and epistemic truths are objective for creatures with our biological and social architecture. Hypothetical beings with radically different structures, for example, telepathic beings that reproduce by fission and feel no pain, would face different constraints and discover a different Apex Network.

Why We Accept This: This is relativism at the species level, not cultural level. We accept it as appropriate epistemic modesty; our claims are grounded in constraints humans actually face. This preserves robust objectivity for humans, as all cultures share the same core constraints.

7.3.2 Learning Through Catastrophic Failure
The Limitation: We learn moral truths primarily through catastrophic failure. The Negative Canon is written in blood. Future moral knowledge will require future suffering to generate data.

Why We Accept This: Empirical knowledge in complex domains requires costly data. Medicine, engineering, and social systems all required catastrophic failures before developing safety mechanisms.

Implication: Moral learning is necessarily slow. We should be epistemically humble about current certainties yet confident in Negative Canon entries.

7.3.3 Floor Not Ceiling

The Limitation: The framework maps necessary constraints (the floor), not sufficient conditions for flourishing (the ceiling). It cannot address what makes life meaningful beyond sustainable, supererogatory virtue and moral excellence, aesthetic value and beauty, or the difference between a decent life and an exemplary one.

Why We Accept This: Appropriate scope limitation. The framework does what it does well rather than overreaching. It identifies catastrophic failures and boundary conditions, leaving substantial space for legitimate value pluralism above the floor.

What This Implies: The framework provides necessary but not sufficient conditions. Thick theories of the good life must build on this foundation. The Pluralist Frontier is real: multiple flourishing forms exist, but all must respect the floor (avoid Negative Canon predicates).

7.3.4 Expert Dependence and Democratic Legitimacy

The Limitation: Accurate brittleness assessment requires technical expertise in historical analysis, statistics, comparative methods, and systems theory. This creates epistemic inequality.

Why We Accept This: Similar to scientific expertise generally. Complex systems require specialized knowledge to evaluate.

Mitigation: Data transparency, distributed expertise, standpoint epistemology (marginalized groups as expert witnesses to brittleness), institutional design (independent assessment boards), and education can reduce but not eliminate this challenge.

7.3.5 Discovery Requires Empirical Testing

The Limitation: While the Apex Network exists as a determined structure, discovering it requires empirical data. We cannot deduce optimal social configurations from first principles alone—we need historical experiments to reveal constraint topology.

Why We Accept This: Even in physics and mathematics, we need empirical input. Pure reason can explore logical possibilities, but determining which possibilities are actual requires observation or experiment. For complex social systems with feedback loops and emergent properties, this dependence is stronger.

What This Allows: Prospective guidance through constraint analysis. We can reason about likely optimal solutions by analyzing constraint structure, but we need empirical validation. This is stronger than pure retrospection but weaker than complete a priori knowledge.

7.3.6 The Viable Evil Possibility

The Limitation: If a deeply repugnant system achieved genuinely low brittleness, the framework would have to acknowledge it as viable, though not necessarily just by other standards.

Why We Accept This: Intellectual honesty. The framework maps pragmatic viability, not all moral dimensions.

Empirical Bet: We predict such systems are inherently brittle. Historical cases like the Ottoman devşirme system or Indian caste systems exhibited high coercive overheads, innovation lags, and fragility under external shocks (Acemoglu & Robinson 2012; Turchin 2003). True internalization without coercion is rare and resource-intensive. If empirics proved otherwise, we would acknowledge the framework's incompleteness rather than deny evidence.

7.3.7 What We Claim
These limitations do not undermine the framework's contribution—they define appropriate scope. EPC excels at:

Strong Claims: - Identifying catastrophic systemic failures - Explaining moral progress as empirically detectable debugging - Grounding realism naturalistically without non-natural properties - Providing empirical tools for institutional evaluation - Offering prospective guidance through constraint analysis - Unifying descriptive and normative epistemology

Modest Claims: - Does not provide complete ethics - Does not solve all normative disagreements - Does not eliminate need for judgment - Does not achieve view-from-nowhere objectivity - Does not offer categorical imperatives independent of systemic goals

The Value Proposition: A powerful but limited diagnostic tool for systemic health. Use it for what it does well. Supplement with other resources for what it cannot address. Do not expect a complete theory of human flourishing—expect robust tools for avoiding catastrophic failure and identifying progressive change.

This honest accounting strengthens rather than weakens the framework's philosophical contribution.

7.3.8 Steelmanning Objections and Clarifying Scope
To preempt misunderstandings and demonstrate the framework's resilience, we steelman potential objections and clarify what EPC is not.

What This Theory is NOT:

- **Not a Complete Ethics:** EPC maps necessary constraints (the floor) but not sufficient conditions for flourishing. It identifies catastrophic failures but leaves space for legitimate pluralism in values above the viability threshold.
- **Not a Foundationalist Metaphysics:** It avoids positing non-natural properties or Platonic forms. Objectivity emerges from pragmatic selection, not metaphysical bedrock.
- **Not Deterministic Prediction:** Claims are probabilistic; brittleness increases vulnerability but does not guarantee collapse.
- **Not a Defense of Power:** Power can mask brittleness temporarily, but coercive costs are measurable indicators of non-viability.
- **Not Relativist:** While species-specific, it defends robust objectivity within human constraints via convergent evidence.
- **Not Anti-Realist:** It grounds fallibilist realism in the emergent Apex Network, discovered through elimination.

Steelmanned Defense of the Core: The "Drive to Endure" is not a smuggled value but a procedural-transcendental filter. Any project of cumulative justification presupposes persistence as a constitutive condition. Systems failing this filter (e.g., apocalyptic cults) cannot sustain inquiry. The "ought" emerges

instrumentally: favor low-SBI norms to minimize systemic costs like instability and suffering, providing evidence-based strategic advice for rational agents.

8. Conclusion

We develop Emergent Pragmatic Coherentism as a response to the isolation objection facing coherentist theories of justification. Grounding coherence in long-term viability of knowledge systems rather than internal consistency alone provides the external constraint coherentism requires while preserving its holistic insights. The concept of systemic brittleness offers a naturalistic diagnostic tool for evaluating knowledge systems, while the notion of a constraint-determined Apex Network explains how objective knowledge can arise from fallible human practices.

Systematically studying the record of failed systems discerns the contours of the Apex Network: the emergent set of maximally convergent, pragmatically indispensable principles that successful inquiry is forced to discover. This yields Systemic Externalism, a novel approach for resolving long-standing problems in post-Quinean epistemology by grounding justification in the non-discursive filter of systemic consequences.

This model is not presented as a final, complete system but as the foundation for a progressive and falsifiable research program. Critical future challenges remain, such as fully modeling the role of power asymmetries in creating path-dependent fitness traps and applying the framework to purely aesthetic or mathematical domains.

We began with the challenge of distinguishing viable knowledge from brittle dogma. The model we have developed suggests the ultimate arbiter is not the elegance of a theory or the consensus of its adherents but the trail of consequences it leaves in the world. Systemic costs are not abstract accounting measures; they are ultimately experienced by individuals as suffering, instability, and the frustration of human goals. From this perspective, dissent, friction, and protest are not merely political problems but should be understood as primary sources of epistemological data. They are the system's own real-time signals, indicating where First-Order Costs are accumulating and foreshadowing the rising Coercive Overheads $(C(t))$ that will be required to maintain stability against those pressures.

As a practical tool for epistemic risk management, Emergent Pragmatic Coherentism provides a structured methodology for policymakers, institutional designers, and the public to diagnose the health of our most critical knowledge-producing systems—from climate models to economic paradigms to public health infrastructure—before their hidden brittleness leads to catastrophic failure. The brittleness toolkit enables real-time assessment of structural health, allowing us to identify degenerating research programs and power-masked fragility before crises force recognition. This positions the work not just as a solution to an old philosophical puzzle, but as a practical and urgent research program for the 21st century, empowering democratic publics to hold their knowledge-generating systems accountable.

EPC shares the realist's conviction that convergence reflects constraint, not convention. Like Boyd's (1973) realism and Kitcher's (1993) naturalistic objectivity, it treats truth as an emergent structural attractor—an Apex Network

stabilized through evolutionary filtration. Anchored in Ladyman and Ross's (2007) structural ontology, it situates realism within evolutionary epistemology (Godfrey-Smith 2003), where selective success reveals mind-independent structures.

EPC also offers predictive value for empirical domains: science policy diagnostics via evidence-based frameworks (Cartwright and Hardie 2012), cultural evolution modeling (Henrich 2015), and epistemic institutions as cases of engineered low-brittleness inquiry (Strevens 2020).

The approach calls for epistemic humility, trading the ambition of a God's-eye view for the practical wisdom of a mariner. The payoff is not a final map of truth, but a continuously improving reef chart—a chart built from the architecture of failure, allowing us to more safely navigate the channels of viable knowledge.

Appendix A: Normative Brittleness as a Speculative Extension

Note: This appendix presents a speculative extension of the core framework, integrating it with recent work in meta-ethics. It is not essential to the main argument of the paper.

The framework for assessing brittleness is general, but its application reveals two primary modalities of failure, corresponding to the different kinds of pragmatic constraints a network can violate. This distinction clarifies how our approach unifies descriptive and normative inquiry under a single explanatory mechanism.

- **Normative Brittleness:** This is the modality of failure resulting from a misalignment with the emergent normative structure of the world. It is found in socio-political and ethical networks whose primary function is to organize cooperative human action. The specific mechanism for this failure can be precisely articulated through a theory of emergent moral properties. Drawing on Baysan's (2025) account of emergent moral non-naturalism, we can understand objective moral properties as conferring noncausal powers (Ingthorsson 2013). While a causal power manifests as a physical change, a noncausal power manifests as the obtaining of a normative fact. For example, the property of being unjust confers on an institution the noncausal power to justify resentment and require condemnation. A network's alignment with this structure is not optional. A society predicated on slavery, for instance, exhibits profound normative brittleness because it must expend immense real-world energy to counteract these noncausal powers. The immense coercive overheads required to maintain the institution are the direct, measurable, and causal signature of a system struggling to suppress the real normative fact that its core practices justify resistance. This account requires that we move beyond what Bennett-Hunter (2015) calls the 'causalist assumption'—the dictum that 'to be real is to have causal power'—and recognize that a causal vocabulary may not be appropriate for every explanatory domain (El-Hani and Pihlström 2002). This macro-level diagnosis finds a plausible correlate in agent-level moral psychology, where moral intuitions can be understood as evolved detectors for such "response-invoking" features of our environment (Rottschaefer 2012) that signal potential systemic costs if ignored, a direct experience of what Peter (2024) calls the "demands of fittingness."

Appendix B: A Conceptual Illustration of Operationalization

To demonstrate that the framework is not merely metaphorical, this appendix offers a conceptual sketch of how one might approach the operationalization of brittleness metrics for a historical research program. This is an illustrative outline, not an empirical study.

B.1 Conceptualizing P(t): Conceptual Debt

P(t) measures the accumulation of conceptual debt. In a domain like early 20th-century physics, one could operationalize this by analyzing the research literature of competing paradigms (e.g., classical ether theories vs. early relativity).

Methodology: A researcher would classify publications over a given period. Papers primarily aimed at resolving known anomalies or patching inconsistencies in the existing theory would count towards rising P(t). In contrast, papers proposing novel, surprising, and subsequently confirmed predictions would signal a progressive, low-brittleness program.

Qualitative Signal: A degenerating program would show a clear trend where the proportion of "defensive" or "anomaly-resolving" publications grows, while a progressive program would consistently generate more novel predictions than it does patches.

B.2 Conceptualizing C(t): Coercive Overhead

C(t) measures the resources a system expends on suppressing dissent and enforcing compliance.

Methodology: For a historical case like the conflict between the Galilean and Ptolemaic systems, a researcher could proxy C(t) by examining institutional resource allocation. One could qualitatively assess the time, funds, and institutional energy the Church and allied universities dedicated to suppressing heliocentric views (e.g., through trials, censorship, and curriculum mandates) versus the resources allocated to productive research within the geocentric paradigm.

Qualitative Signal: A high or rising C(t) is indicated when a system must dedicate significant and increasing resources to maintaining consensus, signaling that its core tenets are generating costly dissent that cannot be resolved through evidence or argument alone.

B.3 The Logic of a Research Program

A full empirical research program would involve applying such qualitative coding across numerous historical cases. The core philosophical claim is that a strong positive correlation would emerge: systems exhibiting high and rising P(t) and C(t) scores would consistently prove more fragile when faced with external shocks (e.g., novel empirical discoveries, political upheaval) than systems with low scores. The goal is not to create a predictive algorithm, but to demonstrate that brittleness is a real, historically observable property that has consistent pragmatic consequences.

Appendix C: A Mathematical Model of Epistemic Viability

This appendix provides a provisional formalization of core EPC concepts to illustrate the framework's logical structure and demonstrate that it is, in principle, formalizable. The specific functional forms, parameter values, and mathematical relationships presented here are illustrative candidates for future empirical research, not empirically validated claims. The paper's philosophical integrity remains independent of any particular mathematical formulation. This appendix serves to show that the framework's conceptual architecture can be given formal expression, not to present a completed empirical model.

C.1 Set-Theoretic Foundation

Let **U** be the universal set of all possible atomic predicates. An individual's **Web of Belief (W)** is a subset $W \subseteq U$ satisfying internal coherence condition C_internal:

```
W = {p ∈ U | C_internal(p, W)}
```

Shared Networks (S) emerge when agents coordinate to solve problems. They represent the intersection of viable individual webs:

```
S = ∩{W_i | V(W_i) = 1}
```

where V is a viability function (detailed below).

Public knowledge forms nested, intersecting networks (S_germ_theory ⊂ S_medicine ⊂ S_biology), with cross-domain coherence driving convergence.

The Apex Network (A) is the maximal coherent subset remaining after infinite pragmatic filtering:

```
A = ∩{W_k | V(W_k) = 1} over all possible contexts and times
```

Ontological Status: A is not pre-existing but an emergent structural fact about U, revealed by elimination through pragmatic selection.

Epistemic Status: *A is unknowable directly*; it is inferred by mapping failures catalogued in the **Negative Canon** (the historical record of collapsed, high-SBI systems).

Formal ECHO Extension: This formalizes Thagard's ECHO extension: net_j = Σ w_{ij} a_i - β·brittleness_j, where w represents positive weights for explanatory coherence (Principle 1) and negative weights for contradiction (Principle 5), with β derived from P(t), C(t) proxies. The specific weight magnitudes would require empirical calibration. Zollman's cycle topology models Pluralist Frontier; complete graphs risk brittleness lock-in.

C.2 The Systemic Brittleness Index

SBI(t) is a composite index quantifying accumulated systemic costs. We present three functional forms, each with distinct theoretical motivation and testable predictions.

Key Components:

P(t) - Patch Velocity: Rate of ad-hoc hypothesis accumulation measuring epistemic debt - Proxy: Ratio of auxiliary hypotheses to novel predictions

C(t) - Coercion Ratio: Resources for internal control vs. productive adaptation - Proxy: (Suppression spending) / (R&D spending)

M(t) - Model Complexity: Information-theoretic bloat measure - Proxy: Free parameters lacking predictive power

R(t) - Resilience Reserve: Accumulated robust principles buffering against shocks - Proxy: Breadth of independent confirmations, age of stable core

Functional Forms:

Form 1: Multiplicative Model

```
SBI(t) = (P^α · C^β · M^γ) / R^δ
```

Rationale: Captures interaction effects where high values in multiple dimensions compound non-linearly. A system with both high complexity AND high patch velocity is more brittle than the sum would suggest. (This form could be used to model compounding effects, where brittleness dimensions interact and accelerate).

Predictions: Brittleness accelerates when multiple indicators rise simultaneously. Systems can tolerate high values in one dimension if others remain low.

Testable implication: Historical collapses should correlate with simultaneous elevation of 2+ metrics, not single-metric spikes.

Form 2: Additive Weighted Model

```
SBI(t) = α·P(t) + β·C(t) + γ·M(t) - δ·R(t)
```

Rationale: Assumes independent, additive contributions. Simpler to estimate and interpret; each component has linear effect.

Predictions: Each dimension contributes independently. Reducing any single metric proportionally reduces overall brittleness.

Testable implication: Interventions targeting single metrics should show proportional improvement.

Form 3: Threshold Cascade Model

```
SBI(t) = Σ[w_i · max(0, X_i(t) - T_i)] + λ·Π[H(X_j - T_j)]
```

where $X\_i \in \{P, C, M\}$, H is Heaviside step function, T_i are critical thresholds

Rationale: Systems tolerate moderate brittleness but experience catastrophic acceleration once thresholds are crossed. Captures phase-transition dynamics observed in complex systems.

Predictions: Brittleness remains low until critical thresholds crossed, then accelerates rapidly. Non-linear "tipping point" behavior.

Testable implication: Historical data should show periods of stability followed by rapid collapse once multiple thresholds exceeded.

Empirical Strategy:

These forms make distinct predictions testable through historical analysis: 1. Compile brittleness metrics for 20-30 historical knowledge systems (ancient to modern) 2. Code collapse/persistence outcomes 3. Fit each model to historical data 4. Compare predictive accuracy using cross-validation 5. Use information criteria (AIC/BIC) to select best-fitting form

The Ptolemaic case (Section 2.4) illustrates how such data can be assembled from historical records. A full research program would systematically extend this

approach. The framework's falsifiability depends on committing to specific functional forms and comparing predictions to data.

C.3 Dynamics: Stochastic Differential Equations

Knowledge evolution is not deterministic. We model SBI dynamics as:

```
d(SBI) = [α·SBI - β·D(t) + γ·S(t) - δ·R(t) + θ·I(t)]dt + σ·√(SBI)·dW(t)
```

Deterministic Terms:

- **α·SBI:** Compounding debt (brittleness begets brittleness)
- **β·D(t):** Systemic debugging (cost-reducing discoveries)
- **+γ·S(t):** External shocks (novel anomalies, pressures)
- **δ·R(t):** Resilience buffer (accumulated robustness)
- **+ϑ·I(t):** Innovation risk (costs of premature adoption)

Stochastic Term:

- **σ·√(SBI)·dW(t):** Brownian motion capturing randomness in discovery timing; volatility increases with brittleness

Parameter Estimation:

The parameters α, β, γ, δ, ϑ, σ are unknowable a priori and must be fitted to historical data. This is not a limitation but standard scientific practice. Proposed empirical strategy:

1. Compile time-series brittleness data for multiple historical systems (as illustrated with Ptolemaic astronomy)
2. Use maximum likelihood estimation or Bayesian methods to fit parameters
3. Validate on held-out historical cases
4. Test whether fitted model successfully predicts collapse timing for independent test cases

The Ptolemaic case provides a template: with systematic bibliometric coding, we can construct d(SBI)/dt trajectories from publication patterns. Parameter estimation would then proceed through standard statistical methods.

Predictive Utility:

Once parameters are empirically estimated, the formulation enables probabilistic predictions: "System X has P% chance of crisis within Y years given current trajectory." This transforms brittleness from retrospective diagnosis to prospective risk assessment. The framework's scientific credibility depends on executing this program and comparing predictions to outcomes.

C.4 Conceptualizing an Empirical Inquiry

The existence of these distinct functional forms suggests a path for future empirical inquiry. A historian or sociologist of science could: 1. Compile qualitative brittleness indicators for a set of historical knowledge systems. 2. Code their outcomes (e.g., persistence, revision, collapse). 3. Assess which of the conceptual models (multiplicative, additive, or threshold) best describes the observed historical patterns.

Such a program would not be about fitting precise numerical data but about determining which conceptual dynamic—compounding interaction, linear addition, or tipping points—best accounts for the historical evolution of

knowledge. The framework's value lies in its ability to generate such guiding questions for historical and scientific investigation.

C.5 The Role of Formalism in Philosophical Diagnosis

Once a model like this were empirically calibrated, it would not function as a predictive algorithm but as a diagnostic tool. Its utility would be to translate complex historical dynamics into a structured formal language, allowing for more precise comparisons between the health of different knowledge systems. For example, it could help answer questions like: "Is system X accumulating costs primarily through conceptual debt $(P(t))$, or is its fragility masked by coercive power $(C(t))$?" This transforms brittleness from a retrospective metaphor into a conceptually structured diagnostic, which is the primary philosophical payoff of the formal exercise.

References

Acemoglu, Daron, and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty.* New York: Crown Business. ISBN 978-0307719225.

Anderson, Elizabeth. 1996. "Knowledge, Human Interests, and Objectivity in Feminist Epistemology." *Philosophical Topics* 23(2): 27–58. https://doi.org/10.5840/philtopics199623214.

Baggio, Guido, and Andrea Parravicini. 2019. "Introduction to Pragmatism and Theories of Emergence." *European Journal of Pragmatism and American Philosophy* XI-2. https://doi.org/10.4000/ejpap.1611.

Baysan, Umut. 2025. "Emergent Moral Non-naturalism." *Philosophy and Phenomenological Research* 110(1): 1–20. https://doi.org/10.1111/phpr.70057.

Bennett-Hunter, Guy. 2015. *Ineffability and Religious Experience.* London: Routledge (originally Pickering & Chatto). ISBN 978-1848934719.

Berlin, Brent, and Paul Kay. 1969. *Basic Color Terms: Their Universality and Evolution.* Berkeley: University of California Press. ISBN 978-1575861623

BonJour, Laurence. 1985. *The Structure of Empirical Knowledge.* Cambridge, MA: Harvard University Press. ISBN 978-0674843813.

Boyd, Richard. 1973. "Realism, Underdetermination, and a Causal Theory of Evidence." *Noûs* 7(1): 1–12. https://doi.org/10.2307/2214367.

Brandom, Robert B. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment.* Cambridge, MA: Harvard University Press. ISBN 978-0674543195.

Campbell, Donald T. 1974. "Evolutionary Epistemology." In *The Philosophy of Karl R. Popper*, edited by Paul A. Schilpp, 413–63. La Salle, IL: Open Court.

Carlson, Matthew. 2015. "Logic and the Structure of the Web of Belief." *Journal for the History of Analytical Philosophy* 3(5): 1–27. https://doi.org/10.15173/jhap.v3i5.28.

Davidson, Donald. 1986. "A Coherence Theory of Truth and Knowledge." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by Ernest LePore, 307–19. Oxford: Blackwell.

El-Hani, Charbel Niño, and Sami Pihlström. 2002. "Emergence Theories and Pragmatic Realism." *Essays in Philosophy* 3(2): article 3. https://doi.org/10.5840/eip2002325.

Friston, Karl J. 2010. "The Free-Energy Principle: A Unified Brain Theory?" *Nature Reviews Neuroscience* 11 (2): 127–138. https://doi.org/10.1038/nrn2787.

Godfrey-Smith, Peter. 2003. *Theory and Reality: An Introduction to the Philosophy of Science.* Chicago: University of Chicago Press.

Goldman, Alvin I. 1979. "What Is Justified Belief?" In *Justification and Knowledge: New Studies in Epistemology*, edited by George S. Pappas, 1–23. Dordrecht: D. Reidel. https://doi.org/10.1007/978-94-009-9493-5_1.

Goldman, Alvin I. 1999. *Knowledge in a Social World.* Oxford: Oxford University Press. ISBN 978-0198238201.

Haack, Susan. 1993. *Evidence and Inquiry: Towards Reconstruction in Epistemology.* Oxford: Blackwell. ISBN 978-0631196792.

Harding, Sandra. 1991. *Whose Science? Whose Knowledge? Thinking from Women's Lives.* Ithaca, NY: Cornell University Press. ISBN 978-0801497469.

Henrich, Joseph. 2015. *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter.* Princeton, NJ: Princeton University Press. ISBN 978-0691178431.

Hodge, A. Trevor. 1992. *Roman Aqueducts & Water Supply.* London: Duckworth. ISBN 978-0715631713.

Holling, C. S. 1973. "Resilience and Stability of Ecological Systems." *Annual Review of Ecology and Systematics* 4: 1–23. https://doi.org/10.1146/annurev.es.04.110173.000245.

Ingthorsson, Rögnvaldur D. 2013. "Properties: Qualities, Powers, or Both?" *Dialectica* 67, no. 1: 55–80. https://doi.org/10.1111/1746-8361.12011.

Kelly, Thomas. 2005. "The Epistemic Significance of Disagreement." In *Oxford Studies in Epistemology*, vol. 1, edited by Tamar Szabó Gendler and John Hawthorne, 167–96. Oxford: Oxford University Press.

Kim, Jaegwon. 1988. "What Is 'Naturalized Epistemology'?" *Philosophical Perspectives* 2: 381–405. https://doi.org/10.2307/2214082.

Kitcher, Philip. 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusions.* New York: Oxford University Press. ISBN 978-0195046281.

Krag, Erik. 2015. "Coherentism and Belief Fixation." *Logos & Episteme* 6, no. 2: 187–199. https://doi.org/10.5840/logos-episteme20156211. ISSN 2069-0533.

Kuhn, Thomas S. 1996. *The Structure of Scientific Revolutions.* 3rd ed. Chicago: University of Chicago Press (originally 1962). ISBN 978-0226458083.

Kvanvig, Jonathan L. 2012. "Coherentism and Justified Inconsistent Beliefs: A Solution." *Southern Journal of Philosophy* 50(1): 21–41. https://doi.org/10.1111/j.2041-6962.2011.00090.x.

Ladyman, James, and Don Ross. 2007. *Every Thing Must Go: Metaphysics Naturalized.* Oxford: Oxford University Press. ISBN 978-0199276196.

Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programmes." In *Criticism and the Growth of Knowledge*, edited by Imre Lakatos and Alan Musgrave, 91–196. Cambridge: Cambridge University Press.

Laudan, Larry. 1977. *Progress and Its Problems: Towards a Theory of Scientific Growth.* Berkeley: University of California Press. ISBN 978-0520037212.

Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry.* Princeton, NJ: Princeton University Press. ISBN 978-0691020518.

Longino, Helen E. 2002. *The Fate of Knowledge.* Princeton, NJ: Princeton University Press. ISBN 978-0691088761.

Lugones, María. 2003. *Pilgrimages/Peregrinajes: Theorizing Coalition against Multiple Oppressions.* Lanham, MD: Rowman & Littlefield. ISBN 978-0742514591.

Lynch, Michael P. 2009. *Truth as One and Many.* Oxford: Clarendon Press. ISBN 978-0199218738.

March, James G. 1978. "Bounded Rationality, Ambiguity, and the Engineering of Choice." *The Bell Journal of Economics* 9, no. 2: 587–608. https://doi.org/10.2307/3003600.

Meadows, Donella H. 2008. *Thinking in Systems: A Primer.* Edited by Diana Wright. White River Junction, VT: Chelsea Green Publishing. ISBN 978-1603580557.

Mesoudi, Alex. 2011. *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences.* Chicago: University of Chicago Press. ISBN 978-0226520445.

Newman, Mark. 2010. *Networks: An Introduction.* Oxford: Oxford University Press. ISBN 978-0199206650.

O'Connor, Cailin, and James Owen Weatherall. 2019. *The Misinformation Age: How False Beliefs Spread.* New Haven, CT: Yale University Press. ISBN 978-0300234015.

Olsson, Erik J. 2005. *Against Coherence: Truth, Probability, and Justification.* Oxford: Oxford University Press. ISBN 978-0199279999.

Peirce, Charles S. 1992. "How to Make Our Ideas Clear." In *The Essential Peirce: Selected Philosophical Writings*, vol. 1 (1867–1893), edited by Nathan Houser and Christian Kloesel, 124–41. Bloomington: Indiana University Press (originally 1878).

Peter, Fabienne. 2024. "Moral Affordances and the Demands of Fittingness." *Philosophical Psychology* 37(7): 1948–70. https://doi.org/10.1080/09515089.2023.2236120.

Plantinga, Alvin. 1993. *Warrant and Proper Function.* New York: Oxford University Press. ISBN 978-0195078640.

Popper, Karl. 1959. *The Logic of Scientific Discovery.* London: Hutchinson (originally 1934). ISBN 978-0415278447.

Popper, Karl R. 1972. *Objective Knowledge: An Evolutionary Approach.* Oxford: Clarendon Press. ISBN 978-0198750246.

Price, Huw. 1992. "Metaphysical Pluralism." *Journal of Philosophy* 89(8): 387–409. https://doi.org/10.2307/2940741.

Pritchard, Duncan. 2016. "Epistemic Risk." *Journal of Philosophy* 113(11): 550–571. https://doi.org/10.5840/jphil20161131137.

Putnam, Hilary. 2002. *The Collapse of the Fact/Value Dichotomy and Other Essays.* Cambridge, MA: Harvard University Press. ISBN 978-0674013803.

Quine, W. V. O. 1960. *Word and Object.* Cambridge, MA: MIT Press. ISBN 978-0262670012.

Quine, W. V. 1969. "Epistemology Naturalized." In *Ontological Relativity and Other Essays*, 69–90. New York: Columbia University Press. https://doi.org/10.7312/quin92204-004. ISBN 9780231083577, 9780231171991.

Quine, W. V. O. 1951. "Two Dogmas of Empiricism." *Philosophical Review* 60(1): 20–43. https://doi.org/10.2307/2181906.

Rescher, Nicholas. 1973. *The Coherence Theory of Truth.* Oxford: Clarendon Press. ISBN 978-0198244011.

Rescher, Nicholas. 1996. *Process Metaphysics: An Introduction to Process Philosophy.* Albany: State University of New York Press. ISBN 978-0791428184.

Rorty, Richard. 1979. *Philosophy and the Mirror of Nature.* Princeton, NJ: Princeton University Press. ISBN 978-0691020167.

Rosenstock, Sarita, Cailin O'Connor, and Justin Bruner. 2017. "In Epistemic Networks, Is Less Really More?" *Philosophy of Science* 84(2): 234–52. https://doi.org/10.1086/690717.

Rottschaefer, William A. 2012. "The Moral Realism of Pragmatic Naturalism." *Analyse & Kritik* 34(1): 141–156. https://doi.org/10.1515/auk-2012-0110.

Russell, Bertrand. 1903. *The Principles of Mathematics.* Cambridge: Cambridge University Press. ISBN 978-1430476030.

Sims, Matthew. 2024. "The Principle of Dynamic Holism: Guiding Methodology for Investigating Cognition in Nonneuronal Organisms." *Philosophy of Science* 91(2): 430–48. https://doi.org/10.1017/psa.2023.104.

Sinclair, Robert. 2007. "Quine's Naturalized Epistemology and the Third Dogma of Empiricism." *Southern Journal of Philosophy* 45, no. 3: 455–472. https://doi.org/10.1111/j.2041-6962.2007.tb00060.x.

Snow, John. 1855. *On the Mode of Communication of Cholera.* 2nd ed. London: John Churchill. Reprinted in *International Journal of Epidemiology* 42, no. 6 (2013): 1543–1552. https://doi.org/10.1093/ije/dyt193.

Staffel, Julia. 2020. "Reasons Fundamentalism and Rational Uncertainty – Comments on Lord, The Importance of Being Rational." *Philosophy and Phenomenological Research* 100, no. 2: 463–468. https://doi.org/10.1111/phpr.12675. ISSN 0031-8205.

Strevens, Michael. 2020. *The Knowledge Machine: How Irrationality Created Modern Science.* New York: Liveright Publishing. ISBN 1631491377, 9781631491375.

Taleb, Nassim Nicholas. 2012. *Antifragile: Things That Gain from Disorder.* New York: Random House. ISBN 978-1400067824.

Tauriainen, Teemu. 2017. "Quine's Naturalistic Conception of Truth." Master's thesis, University of Jyväskylä, Department of Social Sciences and Philosophy. https://urn.fi/URN:NBN:fi:jyu-201705312584.

Thagard, Paul. 1989. "Explanatory Coherence." *Behavioral and Brain Sciences* 12(3): 435–502. https://doi.org/10.1017/S0140525X00057046.

Turchin, Peter. 2003. *Historical Dynamics: Why States Rise and Fall.* Princeton, NJ: Princeton University Press. ISBN 978-0691116693.

Worrall, John. 1989. "Structural Realism: The Best of Both Worlds?" *Dialectica* 43(1–2): 99–124. https://doi.org/10.1111/j.1746-8361.1989.tb00933.x.

Wright, Sewall. 1932. "The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution." *Proceedings of the Sixth International Congress of Genetics* 1: 356–66.

Zagzebski, Linda Trinkaus. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge.* Cambridge: Cambridge University Press. ISBN 978-0521570602. https://doi.org/10.1017/CBO 9780511582233.

Zollman, Kevin J. S. 2007. "The Communication Structure of Epistemic Communities." *Philosophy of Science* 74(5): 574–87. https://doi.org/10.1086/ 508684.

Zollman, Kevin J. S. 2010. "The Epistemic Benefit of Transient Diversity." *Erkenntnis* 72: 17–35. https://doi.org/10.1007/s10670-009-9194-6. ISSN 0165-0106.