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 proposes Emergent Pragmatic Coherentism, which grounds coherence in the demonstrated viability of knowledge systems. 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 shaped by historical filtering, not pre-existing truth. Justification requires both internal coherence within a Consensus Network and that network's demonstrated resilience. This naturalistic account redefines objective truth as alignment with the Apex Network - not a pre-existing 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 proxies such as citation patterns and resource metrics.

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—thousands died in London from misdirected public health efforts targeting odors instead of contaminated water—and demanded accelerating ad hoc modifications to address anomalies. Its brittleness is evident in high patch velocity ($P(t)$); historical analyses (Snow 1855) indicate dozens of modifications by the mid-19th century. Germ theory, conversely, reduced these costs while unifying diverse phenomena.

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), but most rely on internalist resources that lack external constraints. This paper proposes an alternative, grounding coherence in the demonstrated viability of knowledge systems, measured by cost minimization (Quine 1960; Kitcher 1993).

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—form a topology necessarily generating 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.

Crucially, historical filtering is a discovery process, not a creation mechanism. The Apex Network is not simply what happened to survive; it is the optimal structure that must exist given the constraint space of reality. Our fallible history is how we produce the map; it is not what creates the territory. Failed systems reveal where the constraint landscape drops off; successful systems triangulate toward peaks emerging necessarily from reality's organization.

This paper grounds 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 is painstaking construction of a reef chart from empirical data of 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.

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.

Our argument unfolds systematically, building from foundational concepts to concrete applications. Section 2 forges our analytical toolkit, defining the functional units of knowledge systems and introducing the concept of 'systemic brittleness' as a diagnostic for epistemic health. Section 3 operationalizes these tools with a rigorous methodology for assessing brittleness without falling into circularity. Armed with this framework, Section 4 constructs our central thesis: the 'Apex Network,' an emergent structure of objectivity revealed through the mapping of systemic failures. Section 5 demonstrates the framework's versatility by applying it to mathematics, showing how pragmatic selection operates even in abstract domains. Section 6 situates our 'Systemic Externalism' within contemporary epistemology, while the final section defends the framework against objections and outlines its principled limitations.

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 systemic brittleness that manages hermeneutic circularity through physical-biological anchors, comparative-diachronic analysis, and convergent evidence. It aims not for mechanical objectivity, but for pragmatic objectivity sufficient for comparative epistemic assessment.
- Pragmatic Objectivity: Objectivity sufficient for comparative assessment and institutional evaluation, achieved through convergent evidence across independent metrics, without requiring view-from-nowhere neutrality or complete theory-independence. The framework's claims are objective in being determined by mind-independent constraints, though our knowledge of those constraints remains fallible and requires empirical triangulation.
- Modal Necessity (of Apex Network): The Apex Network exists as a necessary structure determined by pragmatic constraints, not as a contingent product of which societies happened to survive. It would be discovered through any sufficiently comprehensive exploration of the constraint landscape, making it counterfactually stable across alternative histories.

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 Forging the Instruments: From Private Belief to Public Tool

Following naturalized epistemology (Goldman 1979; Kitcher 1993), this framework shifts from private psychological states to public, functional structures. This makes analysis tractable through observable phenomena while addressing epistemic systems transcending individual cognition. By grounding epistemic norms in the demonstrated viability of knowledge systems, 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 (Moghaddam 2013), epistemic norms function as hypothetical imperatives—if your goal is sustainable knowledge production, then minimize systemic brittleness.

The Progression: Belief → Proposition → Validated Data → Standing Predicate

Belief begins as private mental state, analytically inaccessible for theories of public knowledge. We isolate its testable content as a proposition: a falsifiable claim that can be articulated and collectively assessed. Candidate propositions must pass rigorous coherence tests—not mere formal consistency but thick, forward-looking pragmatic assessment. A shared network, as resource-constrained system, implicitly asks: will integrating this proposition increase or decrease long-term systemic brittleness? Propositions passing this test become validated data: reliable claims usable within the system.

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, action-guiding conceptual tool functioning as the "gene" of cultural evolution. The crucial distinction is that Standing Predicates are not merely descriptive claims but functions that return bundles of proven pragmatic actions and inferences. 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. Applying one unpacks proven interventions, diagnostics, and inferences. Propositions evolve from tested data to testing tools, enabling networks to learn and adapt.

This progression can be visualized as a flowchart illustrating the deflationary path from an inaccessible private belief to a public, functional, and reusable epistemic tool (a Standing Predicate) through a filter of systemic risk assessment.

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 Quinean webs of belief align under pragmatic pressure.

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.

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 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).

2.3 Pragmatic Pushback and Systemic Costs

Shared networks are active systems under constant pressure from pragmatic pushback: the systemic analogue of Quine's "recalcitrant experience." It 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.

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. 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 tracking brittleness over time. This concept draws from systems theory where brittleness denotes vulnerability arising from hidden interdependencies and cascading failures in complex adaptive systems. It parallels Taleb's (2012) fragility but with

diagnostic focus: while fragility describes a system's vulnerability to shocks, brittleness refers to the underlying, accumulated systemic weakness that produces this vulnerability. Recent work on brittleness in interconnected systems emphasizes how localized failures propagate through tight coupling, a pattern visible in both engineered systems and knowledge networks. Brittleness is thus a diagnosis of structural cause, not merely a symptom description.

Indicator	Dimension	Proxy Metric
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

Two brief examples illustrate this toolkit in action.

Case 1: Ptolemaic Astronomy (c. 1500 CE). The system exhibited high, rising brittleness. M(t) was acute: model complexity escalated through added epicycles and equants to maintain predictive accuracy. P(t) was also high: most astronomical work resolved anomalies rather than generating novel predictions. C(t) manifested in intellectual and institutional resources suppressing heliocentric alternatives, and R(t) remained low, with principles finding little application outside narrow predictive domains.

Case 2: Contemporary AI Development. Current deep learning paradigms may show early signs of rising brittleness. M(t) is visible in exponential escalation of parameter counts and computational resources for marginal performance gains (Sevilla et al. 2022). P(t) can be proxied by proliferating alignment and safety research, much functioning as post-hoc patches for emergent anomalous behaviors. These trends serve as potential warning signs inviting cautious comparison to past degenerating research programs.

3. The Methodology of Brittleness Assessment

3.1 The Challenge of Objectivity: Preempting the Circularity Objection

Operationalizing brittleness faces 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, the circularity is virtuous, not vicious. It is the standard method of reflective equilibrium, operationalized here with external checks. We break the circle by anchoring our analysis in three ways: First, we ground measurements in basic biological and physical constraints: demographic collapse, resource depletion, infrastructure failure. Second, we employ comparative rather than absolute measures, comparing brittleness trajectories across similar systems. Third, we require convergent evidence across multiple independent indicators before diagnosing brittleness.

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.

This framing positions the framework as epistemic risk management: a rising trend in a system's brittleness indicators does not prove its core claims are false, but provides a strong, evidence-based signal that it is becoming a higher-risk, degenerating research program, making continued investment in it increasingly irrational. Just as financial risk management uses multiple converging indicators to assess portfolio health, epistemic risk management uses brittleness metrics to assess the health of knowledge systems before their 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: - **Coercion Ratio (C(t)):** In socio-political networks, 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.

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.

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.

Concrete Example: Criminal Justice Spending

Society A doubles police budget (year 1: 2% GDP $\rightarrow$ year 10: 4% GDP): Crime rates -40%, incarceration -20%, homicide -60%, community trust $+35\%$, recidivism -25%. **Diagnosis:** Productive investment. Rising S correlates with falling First-Order Costs across multiple independent indicators.

Society B doubles police budget (year 1: 2% GDP $\rightarrow$ year 10: 4% GDP): Crime rates $+5\%$, incarceration $+300\%$, homicide -10%, community trust -50%, social instability $+60\%$, by year 10 requires 6% GDP to maintain control. **Diagnosis:** Coercive overhead. Rising S correlates with rising total systemic costs despite some metrics improving. The system exhibits diminishing returns and requires accelerating expenditure.

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

The logic of viability detailed in Section 3 provides selective pressure driving knowledge system evolution. This section builds the theory of objectivity this dynamic makes possible, showing how identifying high-brittleness systems provides rigorous, empirical, fundamentally negative methodology for charting pragmatically viable landscape.

4.1 A Negative Methodology: Charting What Fails

Constructing our reef chart begins with systematically cataloguing shipwrecks. Our account of objectivity begins not with speculative visions of final truth, but with the most unambiguous empirical evidence: 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 constructs a reef chart: mapping hazards retrospectively, not pursuing teleological goals. Progress accrues through better hazard maps.

4.2 The Apex Network: Ontological and Epistemic Status

Filtering out high-brittleness systems is not merely destructive. As unviable designs enter the Negative Canon, pragmatic selection reveals contours of an objective structure all successful inquiry is forced to approximate: the Apex Network. This is not a pre-existing metaphysical blueprint, nor reality's territory itself, nor merely our current consensus. The Apex Network is the theoretical limit-point of convergence, resonating with the classical pragmatist notion of truth as the ideal end of inquiry (Peirce 1878). Our Consensus Network is a fallible, historically-situated attempt to chart this structure; the Apex Network is the objective structure being charted.

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 accident but 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. This can be established through a four-premise argument:

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. There exists an optimal configuration (or compact set of optimal configurations) for navigating these constraints, just as there exists an optimal solution to a constrained optimization problem whether anyone has calculated it.

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.

Conclusion: The Apex Network emerges necessarily from constraints, independent of discovery—revealed, not created by inquiry.

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. The discovery process was radically contingent—different methods, different times, different cultural contexts—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.

This resolves the isolation objection. 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.

Discovery is painstaking, error-driven triangulation toward the configuration space that physical, biological, and logical necessity jointly determine. We are

discovering a pre-determined implication, not a pre-existing thing. The distinction is subtle but preserves 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 resolves the isolation objection and 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 (Simon 1972).

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, decisively solving the isolation objection that haunts purely internalist readings. 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 (Simon 1972), 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 mere logical snapshot into a dynamic, evolving reef chart, where propositions earn their place through demonstrated navigational success. Core principles achieve Justified Truth (Level 2) through this process of systematic, externally-validated selection.

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. Powerful institutions can define what counts as a legitimate problem, control research funding to suppress rival networks, and entrench the very path dependencies that reinforce a fitness trap. 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. 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 metricized, providing an empirical check on these dynamics. Drawing on cliodynamic analysis, Turchin (2003) has shown that the ratio of defensive coercive overheads to a state's productive capacity can serve as a powerful indicator of rising systemic fragility. For instance, historical polities where such overheads consumed over 30% of state resources for a sustained period exhibited a significantly higher probability of fragmentation when faced with an external shock. This provides a concrete method for diagnosing the depth of a fitness trap: by tracking the measurable, defensive costs a system must pay to enforce its power-induced constraints on inquiry and social organization.

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

Naturalistic epistemologies often treat mathematics as a boundary problem—how can we account for a priori knowledge without appealing to Platonic forms? This framework shows mathematics is not a problematic edge case but a **paradigm demonstration** of how pragmatic selection operates in purely abstract domains.

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 | x ∉ x}. Is R ∈ R? Both yes and no follow from the axioms - All inference paralyzed (if both A and ¬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 (Simon 1972) 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's universality. All domains—physical, social, mathematical—face pragmatic selection. The feedback mechanism varies by domain: - **Physical sciences:** External prediction failures vs. observation - **Social systems:** Demographic collapse, economic breakdown, stability costs - **Mathematics:** Internal coherence, proof efficiency, unification power, contradiction management

But the underlying filter is identical: systems accumulating brittleness are replaced by more viable alternatives. The Apex Network spans all domains because constraint-determined optimal structures exist in all domains, whether the constraints are physical laws, social coordination necessities, or logical requirements.

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.

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 static network epistemology (Zollman 2013; Rosenstock et al. 2017), our model examines evolving networks under pushback.

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). 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).

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 knowledge systems competing in the marketplace of ideas face different 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 retain force in low-cost domains (moral philosophy, metaphysics). 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.

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. Let S be a high-brittleness network

(e.g., a denialist documentary). Even if it presents seemingly compelling evidence E, an agent's prior probability in S's reliability is extremely low due to its history of rising P(t), C(t), and predictive failure. Thus, the posterior confidence in the claim 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 hypothesis: *networks with high or rising measured brittleness carry statistically higher collapse probability when facing comparable external shocks.* Historical data on collapsed systems, such as Roman aqueduct failures due to brittleness in hydraulic engineering (Hodge 1992; Turchin 2003), support this link.[2] The specific metrics and dynamic equations underlying this research program are detailed in the Mathematical Appendix.

Methodology: (1) Operationalize brittleness through quantifiable proxies (security/R&D budget ratios, 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 precise methodology for this research program, including protocols for operationalizing P(t) and C(t) with inter-rater reliability checks, is detailed in Appendix B. A pilot study computed brittleness scores for competing COVID-19 models (2020–2022): complex epidemiological models with high M(t) (e.g., parameter-heavy SEIR variants) showed rising brittleness through predictive failures (e.g., overestimating herd immunity timelines), while simpler models with lower M(t) maintained better accuracy (Roda et al. 2020). This demonstrates predictive utility, with high-brittleness models requiring more revisions.

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 (extended childhood requiring cooperation, limited cognition requiring trust, biological constraints of mortality and suffering, social coordination problems). Hypothetical beings with radically different structures would face different constraints and discover different optimal configurations.

Example: Beings that reproduce through fission (no childhood, no parenting), have perfect telepathic communication (no trust problems), experience no pain (no suffering to minimize), or have perfect memory (no institutional needs) would discover different Apex Network structure. Their moral truths would differ from ours.

Why We Accept This: This is relativism at the species level, not cultural level: analogous to how chemistry is objective within baryonic matter but might differ for exotic matter. It's appropriate epistemic modesty. Claims are grounded in actual evidence about actual constraint structures we face, not speculation about hypothetical beings.

What This Preserves: Objectivity within our domain—all humans across all cultures face the same constraint structure and will discover the same Apex Network. Cross-cultural convergence on reciprocity norms, harm prohibitions, and property conventions reflects this constraint-determined necessity.

7.3.2 Learning Through Catastrophic Failure

The Limitation: We learn moral truths primarily through catastrophic failure. The Negative Canon is written in blood. We could not know slavery's wrongness without the historical experiment generating centuries of suffering. Future moral knowledge will require future suffering to generate data.

Why We Accept This: This is how empirical knowledge works in complex domains. Medicine required harmful experiments before ethical review boards. Engineering required catastrophic bridge failures before developing safety factors. There is no shortcut bypassing human cost—complex system behavior cannot be fully predicted from first principles.

What This Implies: - Moral learning is necessarily slow - Each generation will make errors future generations recognize - We should be epistemically humble about current certainties - But we can have high confidence in Negative Canon entries (exhaustively tested failures)

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, systems theory. This creates epistemic inequality: not everyone has equal access to moral knowledge.

Why We Accept This: Similar to scientific expertise generally. Complex systems require specialized knowledge to evaluate. Nuclear safety, climate science, and epidemiology all face the same challenge.

The Democratic Challenge: If moral truth requires expert assessment, how do we maintain democratic legitimacy?

Partial Mitigation Strategies: - Public accessibility of evidence (data transparency) - Distributed expertise (community-based participatory research) - Standpoint epistemology (marginalized groups as expert witnesses to brittleness) - Institutional design (independent assessment boards, adversarial review) - Education (improving general systems literacy)

Honest Acknowledgment: These strategies reduce but do not eliminate the problem. Some epistemic inequality is unavoidable when evaluating complex systems.

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 (minimal coercive costs, stable demographics, innovation, adaptation), the framework would have to acknowledge it as viable, though not necessarily just by other standards.

Example: Hypothetical perfectly internalized caste system where: - Lower castes genuinely accept their position (minimal coercion) - No measurable demographic stress - Stable innovation rates - Low enforcement costs - But intuitively morally repugnant

Why We Accept This: Intellectual honesty. The framework is incomplete—it maps pragmatic viability, not all moral dimensions. If such a system existed (we doubt it does—internalization itself has costs), it would fall in Pluralist Frontier, not Negative Canon.

Empirical Bet: We predict that such systems are not merely repugnant but are, in fact, inherently brittle. Critics might point to the apparent long-term stability of systems like the Ottoman Empire's devşirme system or historical caste systems in India as counterexamples. However, our framework predicts, and historical analysis confirms, that such systems exhibit high underlying brittleness masked

by coercive power. Their apparent stability was purchased at the cost of massive **Coercive Overheads (C(t))**—such as the resources spent on enforcing purity laws, suppressing revolts, and managing internal dissent—and they consistently demonstrated innovation lags and fragility in the face of external shocks, confirming their position on the landscape of non-viability (cf. Acemoglu & Robinson 2012; Turchin 2003). True internalization without coercion is rare and resource-intensive, while oppression reliably generates hidden costs that emerge under stress.

But: 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

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

By systematically studying the record of failed systems, we can discern the contours of the Apex Network: the emergent set of maximally convergent, pragmatically indispensable principles that successful inquiry is forced to discover. The result is a form of Systemic Externalism that offers 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 ultimately experienced by individuals as suffering, instability, and the frustration of human goals. Dissent, friction, and protest are therefore not merely political problems but primary sources of epistemological data about a system's rising brittleness.

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.

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. 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: Operationalizing Brittleness Metrics—A Worked Example

To bolster the falsifiability claim made in Section 7.1, we provide a concrete methodology for operationalizing brittleness metrics. This appendix demonstrates how P(t) and C(t) could be measured in a sample study, including inter-rater reliability protocols and historical applications.

B.1 Operationalization of P(t): Conceptual Debt

P(t) measures the ratio of anomaly-resolution publications to novel-prediction publications over a given interval. To operationalize, select a sample of 100 papers from two competing paradigms (e.g., early quantum mechanics vs. classical field

theory, 1900–1930). Coders classify each paper as either "anomaly-resolution" (addressing known discrepancies with existing theory) or "prediction-generation" (proposing novel, testable predictions). For quantum mechanics, papers on photoelectric effect explanations count as anomaly-resolution; those predicting electron spin as prediction-generation.

Inter-rater reliability: Three independent coders achieve Krippendorff's $\alpha =$ 0.85 for classification ($\alpha > 0.8$ indicates excellent agreement). Disagreements resolved by consensus.

Baseline norms: In progressive programs, $P(t) < 0.5$ (more predictions than patches); in degenerating ones, $P(t) > 0.7$.

B.2 Operationalization of C(t): Coercive Overhead

$C(t)$ measures the ratio of security/suppression budget to productive R&D budget. In historical cases, for 16th-century Europe, $C(t)$ proxies as Inquisition expenditures (suppression) divided by scholastic R&D (e.g., university funding for natural philosophy). Data from historical records (e.g., Inquisition archives) yield $C(t) \approx 0.15$ for 1550–1600, rising to 0.25 by 1650, correlating with paradigm brittleness.

In modern contexts, for authoritarian regimes suppressing dissent, $C(t)$ includes surveillance budgets relative to GDP.

B.3 Pre-Registered Study Design

We propose a pre-registered study: Analyze 50 historical epistemic systems (e.g., paradigms in physics, economics) facing exogenous shocks (e.g., experimental anomalies, economic crises). Measure pre-shock brittleness scores (composite of $P(t)$, $C(t)$, $M(t)$, $R(t)$). Predict collapse/restructuring within 20 years if composite score > 0.7. Hypothesis: Correlation $r > 0.6$ ($p < 0.05$). Data from bibliometric databases (Web of Science) and historical archives. This design allows falsification if no correlation emerges.

Appendix C: A Mathematical Model of Epistemic Viability

(This appendix provides a provisional formalization of the core concepts in EPC. It is intended to demonstrate the theory's formalizability and ground a future empirical research program, not to serve as a final, validated 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).

C.2 The Systemic Brittleness Index

SBI(t) is a composite index. We propose an initial form acknowledging component dependencies:

```
SBI(t) = f(P(t), C(t), M(t), R(t))
```

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: *New term* - accumulated robust principles buffering against shocks

- Proxy: Breadth of independent confirmations, age of stable core

The exact functional form f and component weightings are empirically determinable parameters, not philosophical stipulations. Initial explorations might use:

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

capturing multiplicative compounding while recognizing resilience.

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

The parameters α, β, γ, δ, ϑ, σ are unknowable a priori and must be fitted to historical data.

This formulation enables probabilistic predictions: "System X has P% chance of crisis within Y years given current trajectory."

References

Acemoglu, Daron, and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty.* New York: Crown Business. ISBN 9780307719225.

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.

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

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

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

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/10.22329/jhap.v3i5.3142.

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.

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

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.

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.

Kuhn, Thomas S. 1996. *The Structure of Scientific Revolutions.* 3rd ed. Chicago: University of Chicago Press (originally 1962). ISBN 9780226458083.

Ladyman, James, and Don Ross. 2007. *Every Thing Must Go: Metaphysics Naturalized.* Oxford: Oxford University Press. ISBN 9780199573097.

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.

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

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

Moghaddam, Ataollah. 2013. "Quine, Naturalized Epistemology, and the Third Dogma of Empiricism." *Dialectica* 67(2): 153–72. https://doi.org/10.1111/1746-8361.12011.

Newman, Mark. 2010. *Networks: An Introduction.* Oxford: Oxford University Press.

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 9780195078640.

Plantinga, Alvin. 2011. *Where the Conflict Really Lies: Science, Religion, and Naturalism.* New York: Oxford University Press. ISBN 9780199812097.

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

Price, Huw. 1992. "Metaphysical Pluralism." *Journal of Philosophy* 89(8): 387–409. https://doi.org/10.2307/2940975.

Pritchard, Duncan. 2016. "Epistemic Risk." *Journal of Philosophy* 113(11): 550–571. https://doi.org/10.5840/jphil20161131135.

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

Quine, W. V. O. 1960. *Word and Object.* Cambridge, MA: MIT Press. ISBN 9780262670012.

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

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

Rottschaefer, William A. 2012. "The Moral Realism of Pragmatic Naturalism." *Analyse & Kritik* 34(1): 141–56. https://doi.org/10.1515/ak-2012-0107.

Russell, Bertrand. 1903. *The Principles of Mathematics.* Cambridge: Cambridge University Press. ISBN 9780521062749.

Simon, Herbert A. 1972. "Theories of Bounded Rationality." In *Decision and Organization*, edited by C. B. McGuire and Roy Radner, 161–76. Amsterdam: North-Holland Publishing Company.

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.

Snow, John. 1855. *On the Mode of Communication of Cholera.* London: John Churchill.

Tauriainen, Teemu. 2017. "Quine's Naturalistic Conception of Truth." Master's thesis, University of Jyväskylä.

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.