

# Appendix D: From Information to Consciousness: The Foundation

## Overview

This appendix establishes the information-theoretic foundations that underpin the framework’s naturalistic account of mind, truth, and reality. While the main text treats Standing Predicates as “Markov Blankets” and systemic brittleness as “information leakage,” here we develop the full mechanistic story connecting raw information processing to conscious awareness, computational closure to emergent ontological levels, and evolutionary selection to objective truth.

The central thesis: **Consciousness is what it feels like from the inside of a hierarchical information-compression system that has achieved sufficient computational closure to become aware of its own predictive processes.** This is not eliminativism but mechanistic naturalism—showing how phenomenology emerges from, rather than reduces to, information geometry.

**Key Insight:** Standing Predicates are linguistic handles on successful Markov blankets—boundary configurations that have achieved computational closure and survived pragmatic selection. The process from vague notion (proto-Markov blanket) to validated Standing Predicate (successful blanket) to Apex Network (optimal blanket configuration) is the same process that creates biological entities (cells via membranes) and physical systems (thermodynamic equilibria via phase boundaries), just operating at the cultural-epistemic scale.

## 1. Information as Fundamental Substrate

### 1.1 The Primacy of Information

All physical systems process information. A rock in sunlight absorbs photons (information input) and radiates heat (information output). Most systems, however, are informationally transparent: information flows through without creating persistent internal structure. The rock does not build a model of sunlight patterns or predict tomorrow’s weather. It simply responds mechanically to immediate inputs.

Living and cognitive systems differ fundamentally. They compress information, building internal models that predict future sensory states. This compression is literal in Shannon’s sense, not metaphorical: reducing surprise by encoding regularities into reusable patterns.

Existence as a bounded entity requires information processing. A “thing” exists to the extent it maintains statistical boundaries distinguishing its internal states from external states.

### 1.2 The Free Energy Principle

Karl Friston’s Free Energy Principle provides the mathematical foundation:

$$\text{Variational Free Energy} = \text{Surprise (unpredicted sensory input)} + \text{Divergence (model complexity)}$$

All self-organizing systems minimize free energy by: 1. **Updating beliefs** to better predict sensory input (perceptual inference) 2. **Changing the world** to match predictions (active inference) 3. **Optimizing model structure** to reduce complexity while maintaining accuracy (structural learning)

This is not teleological but thermodynamic: systems that fail to minimize free energy dissipate. Those that succeed persist as bounded entities.

**Connection to Epistemic Brittleness:** Systemic brittleness is accumulated free energy. When a knowledge system's predictions consistently fail (information leakage), it must either patch the model with ad-hoc additions, suppress disconfirming evidence through coercion, or accept falsification and revise. The brittleness metrics developed in the main text (patch velocity, coercive overhead, model complexity, resilience reserve) measure these information-theoretic costs directly.

### 1.3 Dispositions as Compression Algorithms

Returning to the Quinean foundation: a disposition to assent is a compressed encoding of regularities.

After encountering many dogs, an organism develops a “dog-detecting” disposition—a neural pattern that fires when dog-relevant features appear. This disposition compresses thousands of observations into a single reusable pattern, predicts future behavior (minimizing surprise when encountering new dogs), enables efficient action (approach friendly dogs, avoid aggressive ones), and stores mutual information across sensory streams (visual, olfactory, auditory).

Better compression means deeper understanding. A child learning “all dogs bark” has simple but lossy compression. An ethologist understanding canine communication has complex but high-fidelity compression. The compression ratio—how much information is preserved with how few parameters—tracks what we intuitively recognize as understanding.

### 1.4 Information Complexes and Mutual Information

Mutual information measures how much knowing one variable tells you about another. Dispositions storing high mutual information across multiple domains form information complexes: stable attractors in the space of possible compressions.

**Example: The “Fire” Complex** - Visual (flames, light patterns) - Thermal (heat sensation) - Olfactory (smoke, burning) - Auditory (crackling, roaring) - Social (warnings, stories about danger) - Practical (cooking, destruction, tool-making)

These information streams share latent structure: they co-vary reliably. A disposition that compresses their mutual information (the concept “fire”) achieves massive compression efficiency. This is why “fire” feels like a unified thing: it is a genuine compression joint in reality’s information structure.

Not all compressions are equal. Some compressions are artifacts (rain dances cause rain), others track genuine causal structure (dry wood causes fire). Reality selects for the latter through differential brittleness.

## 1.5 Two Types of Patterns: Statistical Regularities and Structural Coherence

Not all patterns require the same evidential basis to be validly recognized. This distinction becomes crucial for understanding how knowledge can arise from limited or even singular encounters with phenomena.

**Statistical Regularities** emerge from repetition and frequency. They are discovered through observing that certain patterns recur reliably across many trials. A child learns that “dogs bark” by encountering many dogs and observing the correlation between the visual pattern (dog-shape) and the auditory pattern (barking). These regularities are fundamentally frequency-dependent—the strength of the compression relies on sample size.

**Structural Regularities**, by contrast, involve components that mutually constrain each other through necessary relationships. These patterns can be recognized even in singular instances because the components aren’t merely correlated but necessarily linked through causal or logical dependencies.

**Example: Fire** When Robinson Crusoe first encounters fire, he doesn’t need hundreds of observations to form the valid belief that “fire produces heat.” The relationship between combustion, heat, and light isn’t merely a statistical correlation but a thermodynamic necessity. The components constrain each other:

- Combustion releases energy
- Energy manifests as heat and light
- Heat propagates to nearby objects
- The process requires fuel and oxygen

These aren’t separate facts that happen to co-occur; they’re aspects of a unified causal process. A mind encountering this pattern even once can recognize its structural integrity—the internal coherence that makes it a genuine compression joint rather than an accidental correlation.

**Contrast with Pure Statistical Learning:** - “Hot stoves burn skin” (structural—recognized from single encounter, thermodynamic necessity) - “Dogs bark at strangers” (statistical—requires multiple observations, behavioral tendency) - “ $F=ma$ ” (structural—mathematical necessity once the concepts are understood) - “Swans are white” (statistical—inductively generalized from frequency, famously failed)

Recent work in phase epistemology provides formal treatment of this distinction. Ayvazov (2025) distinguishes between classical probability (frequency-based likelihood) and what he terms “improbabilistic coherence” (structural integrity that exists independent of repetition), defining it as “the generative condition for epistemic emergence.” While Ayvazov proposes a speculative quantum-mechanical formalism for this distinction, we employ it here purely as an epistemic category without committing to his physical interpretation.

**Implications for Information Compression:** - Statistical compressions require large ensembles to stabilize (high sample complexity) - Structural compressions can achieve validity from minimal data when the pattern exhibits internal constraint - The brain appears capable of detecting both types, but conscious reasoning particularly engages with structural patterns - Innovation often involves recognizing structural coherence before statistical validation

Structural patterns are not subjectively imposed but constrained by reality. You cannot validly infer that “ice produces heat” from a single encounter because thermodynamics forbids this relationship. The structural constraints are objective, even if recognizable from limited data.

This distinction becomes essential for understanding how notions (proto-Standing Predicates) can form before extensive empirical testing, and why some singular experiences carry immediate epistemic weight while others require statistical accumulation.

## 1.6 The Framework as Conceptual Scaffolding

Before proceeding, we must clarify the epistemic status of the information-theoretic language employed throughout this appendix.

**Metaphor or Mechanism?** Information theory provides conceptual scaffolding for understanding cognitive and epistemic processes, but we need not claim the brain literally computes Shannon entropy or that Standing Predicates are implemented as explicit  $\varepsilon$ -machines. The framework's value is primarily conceptual—it captures functional relationships and constraints that appear to govern how knowledge systems operate, regardless of specific implementation details.

**Analogy from Economics:** Consider how economics productively uses “utility maximization” to model decision-making without claiming that neurons actually calculate utility functions. The model captures something real about choice behavior under constraints, even if the mechanistic implementation differs from the formal apparatus. Similarly, our information-theoretic framework captures something real about how compressions succeed or fail, how boundaries form and persist, and how systems minimize prediction error—whether or not these processes literally instantiate Shannon’s equations.

**The Philosophical vs. Empirical Claims:** We can distinguish three levels of commitment:

1. **Weak Claim (Conceptual):** Information-theoretic language provides a coherent framework for thinking about dispositions, predicates, and truth. It clarifies what success and failure look like for knowledge systems.
2. **Moderate Claim (Functional):** Cognitive and epistemic systems behave AS IF they are minimizing information-theoretic costs. The functional constraints we describe (computational closure, information leakage, brittleness accumulation) genuinely constrain which belief systems persist.
3. **Strong Claim (Mechanistic):** Brains literally implement variational free energy minimization; Standing Predicates are actually encoded as Markov blankets in neural tissue; the Apex Network exists as a definite information structure.

**Our Position:** This appendix primarily makes claims at levels 1 and 2. Whether the strong mechanistic claims (level 3) are literally true remains an empirical question for neuroscience and cognitive science. Our philosophical insights about Standing Predicates, compression, and truth don’t depend on resolving this empirical question. Even if the brain’s actual mechanisms differ significantly from Free Energy minimization, the functional analysis of what makes a predicate successful (low brittleness, high compression, computational closure) retains its philosophical force.

**Connecting Functional Constraints to Normative Claims:** Readers may wonder how later sections derive substantive claims about objective truth (Section 6) and ethics (Section 8) from what we’ve framed as “conceptual scaffolding.” The answer: these claims rely on level 2 (functional) constraints, not level 3 (mechanistic) implementation. The argument is not “brains compute

Shannon entropy, therefore truth exists,” but rather “whatever systems successfully compress reality must satisfy certain functional constraints (computational closure, minimal information leakage, strong lumpability), and these constraints determine which predicates persist.” The Apex Network is “objective” not because it exists as a Platonic form or neural structure, but because the functional requirements for successful compression are determined by reality’s constraint structure, not by our beliefs about them. Similarly, the claim that coercion generates brittleness doesn’t require literal free energy calculations—it requires only that systems refusing to model agents’ autonomous responses must bear higher coordination costs. The normative force comes from functional necessity, not mechanistic implementation.

**Preserving the Insights:** When we say “dispositions are compression algorithms” or “Standing Predicates are Markov blankets,” read these as capturing functional roles rather than ontological identities. A disposition functions like a compression algorithm—it reduces redundancy, encodes regularities, enables prediction. Whether it literally performs Huffman encoding or some neurally-implemented equivalent doesn’t affect the philosophical point about what makes it successful or brittle.

**Empirical Openness:** This framework generates empirical predictions that could be tested. If consciousness really does track hierarchical compression processes, we should find neural signatures that correlate with compression improvements. If brittleness really does measure information leakage, we should be able to quantify it in failing belief systems. But even if specific empirical predictions fail, the conceptual framework for thinking about knowledge, truth, and existence retains value as a philosophical contribution.

With this methodological clarification in place, we can proceed to develop the framework confident that its philosophical insights don’t collapse even if particular empirical implementations prove incorrect.

## 2. Markov Blankets: The Architecture of Existence

### 2.1 What Is a Markov Blanket?

Consider a living cell. Its membrane separates “inside” (genes, metabolism) from “outside” (hostile chemistry). The membrane has sensors (receptors detecting nutrients) and actuators (secretions affecting environment). Crucially, the cell’s internal processes depend only on what crosses the membrane, not directly on the external world. This is a Markov blanket: a statistical boundary creating conditional independence.

Formally, for a system with states partitioned into: - Internal states ( $\mu$ ): The “inside” of the entity - External states ( $\eta$ ): The “outside” world - Sensory states ( $s$ ): Detecting external changes - Active states ( $a$ ): Affecting the external world

A Markov blanket exists when:  $P(\mu | s, a, \eta) = P(\mu | s, a)$

Internal states depend only on the blanket (sensory and active states), not directly on the external world. This creates conditional independence—the hallmark of autonomous existence.

## 2.2 Blankets Create “Things”

Markov blankets are not discovered but **enacted**—they emerge when certain configurations of matter successfully maintain statistical boundaries against entropic dissolution.

### Examples Across Scales:

Scale   Entity	Markov Blanket   Internal States	Sensory/Active States
———— ———— ———— ———— ———— ————	Molecular   Cell	
Phospholipid membrane   Genes, metabolism, proteins   Ion channels, receptors, secretions   Neural   Brain region   Synaptic connections   Local processing circuits   Axonal inputs/outputs   Cognitive   Concept   Attentional filter   Compressed representation   Pattern recognition triggers, behavioral outputs   Social   Institution   Bureaucratic procedures   Internal decision-making   Public-facing policies, enforcement   Epistemic   Standing Predicate   Definitional boundaries   Compressed causal model   Recognition criteria, licensed inferences		

**The Radical Implication:** What “exists” is not mind-independent but blanket-relative. Cells exist for systems with cell-detecting blankets. Quarks exist for systems with quark-detecting blankets. Gods exist for systems with god-detecting blankets.

### Defending Blanket-Relative Ontology:

This claim may seem to collapse into pure relativism or idealism, but it doesn’t. The key is understanding that while blankets are enacted rather than discovered, not all enactments succeed. Reality imposes severe constraints on which blanket configurations achieve computational closure.

Consider three cases:

1. **Cells** (successful blanket): The phospholipid membrane creates genuine conditional independence. Internal metabolic dynamics can be predicted from membrane states alone, without tracking every external molecule. The blanket achieves computational closure—it works. This is why cells persist across billions of years and countless environments.
2. **Phlogiston** (failed blanket): Attempts to draw a blanket around “phlogiston content” fail catastrophically. You cannot predict combustion outcomes using only phlogiston-level variables—oxygen levels, molecular structure, and thermodynamic conditions leak through constantly. The blanket never closes. This is why phlogiston was abandoned.
3. **Quarks** (successful but scale-dependent blanket): For particle physicists, quarks form a viable blanket—the Standard Model achieves computational closure at that scale. For ecologists studying predator-prey dynamics, quarks are irrelevant; the blanket is drawn at the organism level. Both blankets work for their respective purposes.

The crucial insight: **blanket-relativity is not arbitrariness**. You cannot successfully draw a blanket anywhere you please. Reality’s constraint structure determines which coarse-grainings achieve closure and which leak information. The cell membrane works because lipid bilayers genuinely create conditional independence in aqueous environments—this is a fact about chemistry, not about our beliefs. Phlogiston fails because combustion genuinely requires oxygen—this is a fact about thermodynamics, not about our preferences.

**Ontological Pluralism with Objective Constraints:** Different purposes require different blankets (quarks for physics, organisms for ecology, institutions for sociology), but within each domain, reality ruthlessly selects which blankets persist. This is neither naive realism (there is no single correct ontology) nor pure relativism (most attempted blankets fail). It is constrained pluralism: multiple viable ontologies exist, but viability is determined by reality's structure, not by our choices.

**The Biological-Epistemic Isomorphism:** The parallel between biological and cultural-epistemic Markov blankets is not metaphorical but structural:

Biological Example | Cultural-Epistemic Example | Shared Mechanism |

|—————|—————|—————| **Cell membrane**  
(phospholipid bilayer) blankets the interior from hostile chemistry outside → new causal level emerges (genes, metabolism, reproduction) | “...is an infectious disease” draws blanket around pathogen-host interactions, insulating public health reasoning from miasmas, humors, spirits → new causal level emerges (transmission chains, sterilization protocols, vaccines) | Both are coarse-grainings that minimize prediction error/free energy/brittleness at the higher level | Cell receptor proteins = sensory states detecting nutrients/threats | Recognition criteria = sensory states detecting instances (“has pathogen?”) | Both detect relevant features across blanket boundary | Cell secretions/flagella = active states affecting environment | Licensed inferences/interventions = active states affecting world | Both enable action based on internal model | Homeostasis = maintaining internal states despite external fluctuations | Functional entrenchment = maintaining predicate despite anomalies | Both resist dissolution through active maintenance |

Once the blanket is in place, **you no longer reason from first principles every time**. Saying “COVID-19 is an infectious disease” instantly inherits isolation protocols, PCR testing, ventilation engineering—just as a cell membrane instantly inherits billions of years of evolved receptor/secretion machinery. This is what a Markov blanket achieves: it lets the interior evolve under its own (much simpler) dynamics.

## 2.3 Computational Closure: When Emergence Succeeds

**Computational closure** occurs when coarse-grained states at the higher level form a complete, self-contained dynamical system—when you can predict future macro-states using only current macro-states, without tracking micro-details.

**Lumpability:** Can we group micro-states into macro-states such that the macro-dynamics are: - **Deterministic:** Same macro-state always transitions to same next macro-state - **Markovian:** Next state depends only on current state, not history - **Causally closed:** Macro-variables shield internal implementation from external observation

When all three conditions hold, **emergence has succeeded**—a new causal level exists.

Rosas et al. (2024) formalize this intuition by comparing two optimal predictors. Imagine you want to predict how temperature will change. The  $\epsilon$ -machine (epsilon-machine) uses only current temperature and pressure—macro-level data. The  $\upsilon$ -machine (upsilon-machine) tracks every molecule's position and momentum—full micro-level access.

When these two predictors perform equally well, something remarkable has happened: the macro-level has become causally autonomous. You have discovered a level of organization that needs no substrate information to continue operating. The macro-level is “running code” rather than merely describing patterns in the substrate. This is computational closure.

This equivalence admits degrees of robustness. **Weak lumpability** holds when macro-dynamics work for specific initial conditions; **strong lumpability** holds when macro-dynamics work regardless of the underlying micro-state distribution. Strong lumpability represents genuine substrate independence—the macro-pattern persists across different physical realizations.

#### Examples:

**Temperature (Successful Closure):** - Micro: Positions and momenta of  $10^{23}$  molecules - Macro: Single scalar (temperature) - The macro-variable (temperature) predicts thermodynamic behavior without tracking individual molecules - Causal closure: Heating water increases temperature increases pressure—the mechanism is shielded

**Phlogiston (Failed Closure):** - Attempted macro-variable: “Phlogiston content” - Failed lumpability: Cannot predict combustion outcomes without knowing oxygen levels, molecular structure, etc. - Information leaks through: Every new experiment reveals the blanket is porous - Brittleness accumulates: High  $P(t)$  from constant patches

**Connection to Standing Predicates:** “...is an infectious disease” achieves computational closure. Once you apply the predicate, you can reason about transmission, quarantine, sterilization—the higher-level causal dynamics—without tracking viral proteins, immune responses, etc. The predicate creates a new causal level.

## 2.4 $\epsilon$ -Machines: Optimal Blanket Constructors

**$\epsilon$ -machines** (epsilon-machines) are the mathematically optimal predictors: they compress past experience into the minimal set of causal states needed to predict the future.

**Formal Definition:** An  $\epsilon$ -machine is a hidden Markov model that: 1. Maximally compresses the past (minimal number of states) 2. Maximally predicts the future (no lossless compression possible) 3. Achieves causal shielding (states are indistinguishable to external observer)

#### Concrete Analogy: Chess Positions

Imagine two chess players trying to predict the outcome of a game:

- **The  $\nu$ -machine player** has access to the complete history: every move made, how long each player thought, what openings they studied, their heart rates, neuron firings in their brains. This player uses all micro-level information to predict the next move.
- **The  $\epsilon$ -machine player** sees only the current board position—the macro-state. No history, no neural data, just the pieces and their locations.

When both players predict equally well, the board position has achieved computational closure. The macro-level (piece arrangement) contains all the information needed to predict future macro-states (subsequent positions). The history of how you arrived at this position doesn’t matter—only the current

configuration does. The game has “detached” from its substrate (the players’ brains, their training, their moods) and runs purely on positional logic.

This is what Rosas et al. (2024) formalize: when your  $\epsilon$ -machine (macro-only predictor) performs as well as the  $\nu$ -machine (micro-informed predictor), you have discovered a level of organization that needs no substrate information to continue operating. The macro-level has become causally autonomous—it is “running code” rather than merely describing patterns in the substrate.

**Why This Matters:** - Dispositions are cognitive  $\epsilon$ -machines—they compress experience into causal states (notions, beliefs) - Standing Predicates are cultural  $\epsilon$ -machines—they compress collective experience into reusable causal tools - The Apex Network is the ultimate  $\epsilon$ -machine—the minimal compression of reality’s constraint structure

**The Search Process:** Organisms, communities, and species are  $\epsilon$ -machine explorers, trying different compressions. The ones that achieve genuine computational closure while minimizing brittleness survive. This is not random but hill-climbing on the landscape of viable blanket configurations.

## 2.5 Pragmatic Ontology: Same Information, Different Blankets

**The Hot Dog Paradox:** A perfect illustration of how Markov blankets enact ontologies rather than discovering them.

**The Information (Constant):** - Bread-like substance encasing protein filling - Condiments, preparation method, consumption context - Physical/chemical properties unchanged

**Different Markov Blankets (Variable):**

Community	Blanket Boundary	Rationale	Ontological Commitment
Tax regulators	“Bread + filling = sandwich”	Consistent classification for revenue codes	Hot dog IS a sandwich
Culinary purists	“Requires two separate bread pieces”	Preserving fine-grained distinctions	Hot dog is NOT a sandwich
Structural engineers	“Continuous base with vertical walls”	Engineering load distribution	Context-dependent

**Key Insight—Epistemic Equifinality:** In Systems Theory, *equifinality* describes how different structural configurations can achieve the same steady state. The Hot Dog Paradox illustrates **epistemic equifinality**: different Markov Blankets (definitions) can achieve comparable levels of computational closure depending on the system’s goal (taxation vs. cuisine vs. engineering). Each community draws the boundary where it reduces brittleness for their purposes. The information hasn’t changed—the coarse-graining has. This is not arbitrary (each blanket faces pragmatic testing) but it is pluralistic (multiple viable configurations exist).

**Connection to Truth:** The Apex Network doesn’t dictate “the one true hot dog ontology” but rather the set of boundary-drawing strategies that achieve genuine

computational closure with minimal brittleness. Different purposes require different closures. The Pluralist Frontier of the Apex Network is the zone where the constraint landscape is flat enough to support multiple, equally viable coarse-grainings—regions where equifinality holds.

**Implication:** Ontological disputes often aren't about facts but about which coarse-graining serves which purpose. The universe doesn't care if a hot dog is a sandwich, but food safety inspectors might need to draw that blanket for regulatory coherence.

## 2.6 From Notion to Standing Predicate: The Blanket Formation Process

### 2.6.1 Notions as Proto-Markov Blankets

Before a belief crystallizes into explicit form, it exists as a tentative boundary-drawing attempt—what we call a notion. A notion is the brain's exploratory effort to see if a statistical boundary can be successfully maintained around certain patterns.

**The Robinson Crusoe Insight:** Standing Predicates don't require social coordination to form—they require only multimodal integration within a single agent navigating a constraint-rich environment.

### 2.6.2 Two Pathways to Valid Notions

The distinction between statistical and structural regularities (Section 1.5) explains why some notions can achieve validity from limited evidence while others require extensive testing:

**Statistical Path (Requires Repetition):** 1. **Multiple encounters:** Animal repeatedly shows certain behaviors 2. **Pattern extraction:** Brain notices correlation (this shape → barking sound) 3. **Gradual strengthening:** Each additional instance reinforces the compression 4. **Threshold crossing:** After sufficient trials, disposition stabilizes 5. **Example:** “Dogs are friendly” requires many positive encounters to overcome individual variation

**Structural Path (Can Succeed from Singular Instances):** 1. **Multimodal coherence detection:** Brain recognizes mutually constraining relationships 2. **Structural integration:** Components aren't just correlated but necessarily linked 3. **Immediate validity:** The pattern's internal coherence validates it without requiring repetition 4. **Example:** Touching fire once → permanent valid belief “fire burns”

### 2.6.3 Why Singular Instances Can Suffice

When Crusoe first encounters fire, he doesn't need hundreds of trials because fire exhibits structural coherence: - Combustion necessarily releases energy (thermodynamic constraint) - Energy necessarily manifests as heat/light (physical constraint) - Heat necessarily transfers to touching skin (causal constraint) - The process necessarily requires fuel (chemical constraint)

These constraints are mutually reinforcing. Recognizing any subset activates expectations about the others. This is why even a child touching a hot stove once forms a lasting, valid compression—the pattern has structural integrity independent of frequency.

**Contrast: Why Statistics Sometimes Required:**

Not all patterns have this structural character. “Ravens are black” is a statistical regularity without structural necessity—there’s no thermodynamic or logical reason ravens couldn’t be white. Such compressions require extensive sampling to distinguish genuine patterns from accidents of limited experience.

#### 2.6.4 The Solitary Agent Formation Process

1. **Sensory encounter:** Pattern presents across multiple modalities
2. **Pattern type detection:** Brain assesses whether components show:
  - **Mere correlation** (statistical) → requires repetition
  - **Mutual constraint** (structural) → can validate from limited data
3. **Blanket formation attempt:** Brain draws tentative boundary around pattern
4. **Pragmatic testing:** Actions based on proto-blanket face reality
5. **Validation assessment:**
  - **Structural patterns:** If internal coherence detected, blanket strengthens immediately
  - **Statistical patterns:** If predictions work across multiple trials, blanket strengthens gradually
6. **Functional entrenchment:** Successful blanket becomes automatic disposition

#### Innovation Through Structural Recognition:

This explains how genuine innovations arise: A thinker detects structural coherence in a novel configuration before statistical validation. Newton seeing falling apple → universal gravitation didn’t require thousands of falling objects because he recognized the structural relationship between terrestrial and celestial motion. The mathematical constraints (inverse-square law, conservation principles) exhibited internal coherence that could be validated through theoretical analysis before extensive empirical testing.

**The Network Begins Inside:** Even Crusoe alone has a “network”—his visual, tactile, olfactory, and thermal subsystems must agree on “fire.” The Standing Predicate emerges when these internal streams achieve computational closure around a shared boundary. For structural patterns, this closure can succeed rapidly; for statistical patterns, it requires iterative refinement.

**Social Transmission Accelerates Both Pathways:** When language allows, successful blankets can be transmitted as linguistic handles: - “Fire burns” (structural) → transmitted with immediate credibility - “Mushrooms with red caps are poisonous” (statistical) → transmitted with caution, requires verification

But the fundamental process—drawing boundaries, testing them against reality, keeping those that close—operates identically in solitary and social contexts. The difference is whether the pattern’s structure permits rapid validation or demands extensive statistical accumulation.

### 3. From Information Processing to Consciousness

#### 3.1 The Phenomenology of Compression

**Core Hypothesis:** Consciousness is the subjective experience of high-level information compression happening in real-time, with particular salience for patterns exhibiting structural coherence rather than mere statistical regularity.

**Central Distinction:** Consciousness appears particularly engaged when the brain detects or attempts to detect structural coherence—when patterns present themselves as having internal necessity rather than mere correlation. Unconscious processing handles statistical pattern matching efficiently in the background, while conscious attention engages when structural relationships demand explicit reasoning.

##### Two Modes of Pattern Processing:

Not all information processing reaches conscious awareness. We can distinguish between processes that operate primarily unconsciously and those that engage conscious attention:

**Unconscious Processing (Statistical Pattern Matching):** - Operates through frequency-based pattern recognition - Builds implicit compressions through repeated exposure - Examples: Walking, driving familiar routes, recognizing faces, grammatical intuitions - Phenomenology: Largely transparent to introspection—"I just know" - Information structure: Statistical correlations extracted from large samples - Requires: Multiple trials, gradual refinement, practice

**Conscious Processing (Structural Pattern Recognition):** - Engages when encountering patterns with internal constraint structure - Particularly active with novel patterns, contradictions, or structural relationships - Examples: Solving puzzles, understanding arguments, recognizing causal necessity - Phenomenology: Explicitly felt—"I see why," "this must be so," "something doesn't fit" - Information structure: Mutual constraints between components - Can operate: From limited data when structure is detected

##### The Conscious/Unconscious Boundary:

This distinction maps roughly onto the statistical/structural divide from Section 1.5. Consciousness appears particularly engaged when the brain detects or attempts to detect structural coherence—when patterns present themselves as having internal necessity rather than mere correlation.

**Examples:** - Learning to ride a bike (initially conscious structural analysis → becomes unconscious statistical refinement) - Recognizing a logical contradiction (conscious—structural incoherence demands attention) - Reading familiar words (unconscious—statistical pattern matching) - Understanding a proof (conscious—following chain of structural necessities)

##### Why This Matters for Phenomenology:

If consciousness tracks structural pattern recognition while unconscious processing handles statistical regularities, this explains several features of conscious experience:

1. **Novelty salience:** New structural patterns demand conscious attention (potential compression improvement)

2. **Contradiction salience:** Structural incoherence can't be ignored (breaks assumed constraints)
3. **Aha! moments:** Sudden recognition of structural relationships (Section 3.1 table below)
4. **Automation through practice:** Once structural understanding achieved, execution becomes statistical refinement (unconscious)

### Phenomenological Correspondences (Revised):

Conscious Experience	Information-Theoretic Process	Pattern Type
“Aha!” moment	Sudden recognition of structural coherence—components snap into mutually constraining relationship	Structural
Confusion	Inability to find structural pattern—high prediction error persists without coherent explanation	Mixed
Understanding	Achieving structural compression—grasping why components must relate as they do	Structural
Certainty	Structural necessity recognized—pattern exhibits internal constraint	Structural
Doubt	Statistical evidence conflicts or structural coherence unclear	Mixed
Boredom	Maximal compression achieved—no new structural insights available (perfectly predictable)	Both
Curiosity	Hints of structural pattern not yet grasped—optimal zone for compression improvement	Structural
Flow state	Structural understanding guides action while statistical refinement operates unconsciously	Both
Anxiety	High prediction error without structural explanation—cannot find pattern that makes sense	Mixed
Familiarity	Statistical pattern matching sufficient—no structural analysis needed	Statistical

## 3.2 Hierarchical Compression and Meta-Awareness

Consciousness requires not just compression but **meta-compression**—compression of the compression process itself.

**Three Levels:** 1. **First-order processing:** Sensory data → compressed representations (largely unconscious) 2. **Second-order monitoring:** Awareness of dispositions—recognizing that you have a pattern-detector active 3. **Third-order reflection:** Thinking about thinking—modeling your own modeling process

**Example: Recognizing Bias** - First-order: “This person is untrustworthy” (disposition active) - Second-order: “I feel distrust toward this

person” (aware of the disposition) - Third-order: “My distrust might be biased by their accent” (modeling the disposition’s origins)

**Only humans (as far as we know) achieve third-order regularly.** This is meta-blanket formation: constructing a Markov blanket around your own Markov blankets, allowing self-modification.

**The Self as User Interface:** Following Rosas et al. (2024), the “Self” is not a ghost in the machine but a user interface—the brain’s own lossy compression of its massive, distributed neural activity. Just as a computer operating system represents billions of transistor states as a single “folder” icon, the brain compresses its complex somatic and cognitive states into a single variable: “I”. This variable is an  $\epsilon$ -machine state—a simplified causal token that allows the system to predict its own future actions without tracking the firing of every individual neuron. The Self is not the neural hardware but the computational software running on that hardware, the minimal effective theory required to predict the organism’s future behavior. This is not an illusion but a computational necessity: the brain must coarse-grain itself to operate at human-relevant timescales. The experience of being a unified “I” is what it feels like from inside this compression process, maintaining computational closure while billions of neural events churn beneath conscious awareness.

### 3.3 Emotions as Variational Free Energy Signals

Emotions are not irrational disruptions but **dashboard readings** of the epistemic engine’s state:

**Negative Emotions = High Free Energy:** - **Anxiety:** Persistent prediction error without identified cause - **Frustration:** Model predicts actions should work, but outcomes consistently differ - **Shame/Guilt:** Social model predicts acceptance, but feedback signals rejection - **Confusion:** Cannot compress incoming information into existing categories

**Positive Emotions = Low Free Energy:** - **Joy:** Predictions confirmed, model vindicated - **Relief:** Expected high free energy avoided - **Pride:** Social model predicts acceptance, feedback confirms - **Satisfaction:** Goals achieved as predicted

**Motivational Emotions = Free Energy Gradients:** - **Curiosity:** Moderate prediction error signaling compressible patterns (explore!) - **Fear:** High prediction error signaling danger (freeze/flee!) - **Anger:** Obstacle blocking expected state (remove/attack!)

From this perspective, emotions are not bugs but features—they make the costs of misalignment consciously accessible, motivating revision toward lower-brittleness configurations.

### 3.4 Engaging the Hard Problem

**Traditional Formulation:** Why is there “something it is like” to process information? Why aren’t we zombies? This is Chalmers’ hard problem of consciousness—the explanatory gap between physical processes and subjective experience.

**Correlates, Not Causes:** A crucial clarification. We have identified a functional correlate of consciousness—structural pattern recognition appears to track the phenomenological boundary. But correlation is not causation. The explanatory

challenge remains: why should detecting mutual constraints feel like *understanding* while detecting statistical correlations feels like nothing (or mere *familiarity*)? We have narrowed the space of functional properties to examine, but we have not explained why this particular functional property should generate subjective experience. The distinction between conscious and unconscious processing is functional, but the explanandum—why there is “something it is like” to perform one function but not the other—remains experiential and unresolved.

### **Our Contribution—A Working Hypothesis:**

We propose that consciousness relates to detecting and representing structural coherence rather than merely tracking statistical correlations. This isn’t a solution to the hard problem, but it identifies a functional distinction that may map onto the phenomenological boundary between conscious and unconscious processing:

**The Distinction:** - **Unconscious processing:** Statistical pattern matching—extracting correlations through frequency - **Conscious processing:** Structural pattern recognition—detecting necessary relationships between mutually constraining components

### **Why This Might Matter:**

If phenomenology relates to structural pattern recognition, this would explain several otherwise puzzling features:

1. **Why singular experiences feel meaningful:** Structural patterns carry intrinsic relationships that don’t require repetition to validate
2. **Why understanding feels different from familiarity:** Grasping structural necessity (conscious) vs. recognizing statistical pattern (unconscious)
3. **Why consciousness engages with novelty and contradiction:** Both demand structural analysis
4. **Why expertise becomes automatic:** Structural understanding converts to statistical refinement

### **The Phenomenological Texture of Structural Recognition:**

Consider the qualitative difference between: - **Recognizing a face** (unconscious statistical matching—no sense of “why”) - **Understanding why a proof works** (conscious structural analysis—sense of necessity)

The second has phenomenological character precisely because it involves representing constraint relationships—grasping that components must relate in certain ways. Perhaps phenomenology is what representing structural constraints feels like, while mere statistical correlation tracking proceeds unconsciously.

### **What This Doesn’t Explain:**

We must be clear about the limits of this account:

- **Why structural detection feels like anything at all:** The hard problem remains—why any functional process has subjective character
- **The specific quality of qualia:** Why red looks like *this* rather than something else
- **The unity of consciousness:** How distributed structural recognitions bind into unified experience
- **The possibility of zombies:** Whether systems implementing these functions necessarily have phenomenology

### **Our Position:**

This framework identifies a functional distinction (statistical vs. structural pattern recognition) that appears to track the conscious/unconscious boundary. If consciousness really does engage with structural coherence specifically, this narrows the explanatory target—we're not asking “why does any information processing feel like something?” but rather “why does detecting mutual constraints between components feel like something?”

That's still a hard problem, but it's a more precise one. And it suggests consciousness isn't an arbitrary add-on to cognition but tracks a genuine functional distinction in how patterns can be recognized.

#### **Not Eliminativism, Not Mysterianism:**

We're not denying consciousness exists (eliminativism) nor claiming it's forever inexplicable (mysterianism). We're offering a naturalistic framework that respects both the reality of phenomenology and the difficulty of explaining it.

Consciousness may be the subjective signature of structural pattern recognition—what mutual constraint detection feels like from inside the system performing it.

Why detecting structural constraints should feel like anything remains an open question this framework doesn't fully answer. We've identified a relevant functional distinction; we haven't solved the hard problem.

### **3.5 Agency and Free Will as Information-Driven Variation**

**Determinism Concern:** If beliefs are compressions shaped by information, where is agency?

**Resolution:** Agency is not freedom from causation but **self-caused variation within constraint space**.

#### **Three Senses of Agency:**

1. **Metabolic Agency (Cells):** Active inference—changing environment to match predictions
  - The cell extends pseudopods toward nutrients (active state)
  - This reduces prediction error (sensory state matches “food here” prediction)
2. **Behavioral Agency (Animals):** Exploration of action space
  - Try different behaviors, keep those that reduce free energy
  - Learning is  $\epsilon$ -machine construction through trial-and-error
3. **Cognitive Agency (Humans):** Mental simulation and deliberate blanket modification
  - Imagine counterfactuals (“what if I revise this belief?”)
  - Deliberately test alternative compressions
  - **Meta-blanket agency:** Modify your own information-processing architecture

**Free Will Recovered:** The capacity to generate novel compressions (new functional propositions, heresies) before reality tests them. Humans can: - Propose new blanket configurations (“what if disease is caused by invisible organisms?”) - Mentally simulate outcomes - Deliberately adopt high-cost positions to test them - Bear brittleness costs to explore the compression landscape

**The Variation Engine:** Individuals are variation generators; reality is the selection mechanism. The Apex Network is discovered through distributed exploration, not centrally imposed.

**Important:** This isn't libertarian free will (causally uncaused choices) but compatibilist agency (self-generated variation within causal constraints).

## 4. Logic and Mathematics as Necessary Compression Structures

### 4.1 Why Logic Occupies the Core

**Traditional View:** Logic is a priori, transcendentally necessary, or conventionally chosen.

**Information-Theoretic View:** Logic is the minimal compression structure required for any system capable of error-correction.

#### The Transcendental Argument:

Any system capable of error-correction must be able to distinguish success from failure. This requires recognizing when A and not-A cannot both be true (non-contradiction).

Chains of inference require transitivity: if believing A leads you to believe B, and believing B leads you to believe C, then believing A should lead you to believe C. Otherwise, your compressions fragment into isolated islands.

Together, these minimal requirements—non-contradiction and transitivity—form the core of classical logic. Logic is not metaphysically necessary in some Platonic sense, but functionally prerequisite: any system that learns (compresses experience, updates on prediction error) must implement logical structure. Logic is not selected by systems; it is the operating system of selection itself.

**Information-Theoretic Grounding:** - Non-contradiction: Same input cannot compress to contradictory outputs - Excluded middle: Compression requires binary decision boundaries - Modus ponens: Compression chains propagate information - Identity: Compression requires stable reference

**Revising Logic:** Would require dismantling the error-correction mechanism itself. This generates infinite brittleness—the system would have no way to evaluate whether the revision succeeded or failed.

### 4.2 Mathematics as Optimal Compression

Mathematical truths are not Platonic forms but **maximally efficient compressions of structural regularities**.

#### Examples:

**$\pi$  (Pi):** - Compresses infinite information (circle's circumference/diameter ratio) - Into finite symbol with infinite precision - Necessarily determined by Euclidean geometry's constraint structure - Discovered independently across cultures (Babylonians, Greeks, Indians) because constraint structure is objective

**Prime Numbers:** - Compress information about multiplicative structure - Their distribution compresses deep regularities in arithmetic - Riemann Hypothesis (if true) would be ultimate compression of prime distribution

**Group Theory:** - Compresses symmetries across domains (crystals, particles, equations) - One framework compresses structure in chemistry, physics, music, cryptography - Unreasonable effectiveness because it captures genuine compression joints

**Connection to Apex Network:** Mathematics is part of the Apex Network—the maximally compressed representation of structural constraints that any sufficiently thorough compression must discover.

### 4.3 The Unreasonable Effectiveness of Mathematics

**Wigner's Puzzle:** Physicist Eugene Wigner famously asked why mathematical structures discovered purely abstractly apply to physical reality with uncanny accuracy. Why should group theory, developed to study symmetries in abstract algebra, perfectly describe particle physics?

**Information-Theoretic Answer:** Mathematics and physics are exploring the same compression landscape from different angles: - Physics: Compress experimental observations - Mathematics: Compress structural necessities - They converge because both face the same constraint structure

**Example: General Relativity** - Einstein: “Find simplest equations describing gravity” - Mathematicians: “What’s the geometry of curved spaces?” - Same compression achieved from different starting points - Convergence reveals objective structure (the Apex Network of geometry)

## 5. Emergence Through Computational Closure

### 5.1 The Mechanism of Emergence Revealed

**Classical Mystery:** How do qualitatively new properties (liquidity, life, consciousness) emerge from mere rearrangement of parts?

**Computational Closure Answer:** Emergence succeeds when a Markov blanket configuration achieves: 1. **Lumpability:** Micro-states group into stable macro-states 2. **Markovianness:** Macro-dynamics depend only on current macro-state 3. **Causal Shielding:** Macro-level is informationally closed from micro-implementation

**Not Just Useful Fiction:** When computational closure succeeds, the emergent level is as real as the base level—it has autonomous causal dynamics.

Rosas et al. (2024) formalize this as **causal decoupling**: when the macro-level  $\varepsilon$ -machine and the micro-informed  $\nu$ -machine achieve equivalence, the macro-level becomes causally autonomous. The emergent pattern is not merely a convenient description—it genuinely causes its own future states. High-level concepts like “temperature,” “infectious disease,” or “recession” are causally real precisely because they achieve this decoupling. The macro-level runs its own causal dynamics, making the substrate details causally irrelevant (though they remain constitutively necessary).

**Temperature Example Revisited:** - Base:  $10^{23}$  molecules with positions, momenta - Emergent: Single scalar variable (temperature) - **Real Causation:**

“High temperature causes ice to melt” - This is NOT shorthand for molecular dynamics - It’s a genuine macro-level causal law operating in the emergent ontology

## 5.2 Failed Emergence = Information Leakage

When computational closure fails, information leaks: - Macro-predictions require micro-details (lumpability fails) - History matters (Markovianness fails) - Internal mechanism observable from outside (causal shielding fails)

**Examples:**

**Phlogiston (Failed Closure):** - Attempted compression: “Phlogiston content” explains combustion - Lumpability failed: Cannot predict outcomes without oxygen levels, molecular structure - Information leaked: Every experiment revealed porous blanket - Result: Brittleness accumulated until abandonment

**Élan Vital (Failed Closure):** - Attempted compression: “Life force” explains biological organization - Lumpability failed: Cannot predict outcomes without metabolism, genetics, evolution - Information leaked: Every discovery revealed mechanistic processes - Result: Brittleness accumulated until abandonment

**Connection to Brittleness:** Information leakage IS systemic brittleness: -  $P(t)$  increases: More patches needed as closure fails -  $M(t)$  increases: Complexity inflates trying to maintain failing blanket -  $C(t)$  increases: Must suppress disconfirming observations -  $R(t)$  decreases: Predictions fail across domains

This brittleness can be understood as the divergence between the  $\epsilon$ -machine (macro-only model) and the  $\nu$ -machine (micro-informed predictor). When a system’s macro-level model requires constant supplementation with substrate details to maintain predictive accuracy, computational closure has failed. The greater this divergence, the more brittle the emergent structure. Systems exhibiting large  $\epsilon/\nu$  divergence show characteristic brittleness signatures: they require increasingly complex auxiliary hypotheses, their predictions fail across contexts, and they eventually collapse as the computational burden of maintaining the failing closure becomes unsustainable.

## 5.3 Standing Predicates as Successful Closures

**Claim:** Standing Predicates are linguistically-encoded successful computational closures.

**“...is an infectious disease” achieves:** 1. **Lumpability:** Groups diverse pathogens (bacteria, viruses, prions) under unified macro-category 2.

**Markovianness:** Can predict epidemiological outcomes using only current disease-state (SIR models) 3. **Causal Shielding:** Enables intervention (quarantine, vaccination) without tracking molecular details

**This Is Why Standing Predicates Feel “Real”:** They carve reality at genuine compression joints—configurations where computational closure naturally succeeds.

**Historical Progression:** - Miasma Theory: Failed closure (information leaked everywhere) - Germ Theory: Successful closure (created new causal level) - Molecular Biology: Deeper closure (but germ theory remains valid at its scale)

**Nested Closures:** Higher-level closures don’t replace lower ones but provide coarser-grained compressions for different purposes.

## 6. The Apex Network as Ultimate $\epsilon$ -Machine

### 6.1 Synthesizing Information, Compression, and Truth

**The Apex Network** is the complete set of Standing Predicate configurations that achieve minimum systemic brittleness—the intersection of all maximally viable compression structures. In information-theoretic terms, it represents the ultimate  $\epsilon$ -machine: the minimal causal-state representation that compresses reality’s constraint structure with theoretical minimum information leakage.

Rosas et al. (2024) demonstrate that all valid coarse-grainings of a system form a mathematical **lattice**—a hierarchical structure of nested compression levels, where each node represents a different way to group micro-states into macro-states. Not all coarse-grainings are equally robust: some achieve only weak lumpability (working only for specific initial conditions), while others achieve strong lumpability (preserving macro-dynamics regardless of substrate details). The Apex Network corresponds to the optimal path through this lattice—the set of strongly lumpable coarse-grainings that maximize causal autonomy while minimizing computational complexity. Reality allows many valid maps (the full lattice), but the Apex Network represents those compressions that achieve genuine substrate independence.

**Terminological Note:** While the mathematical structure is technically a lattice (a partially ordered set with strict hierarchical derivation), we use “Apex Network” to emphasize the socially shared and collectively discovered nature of these optimal compressions. The term “network” highlights that this structure emerges through distributed exploration by communities of knowers, rather than individual deduction. Both terms capture important aspects: “lattice” emphasizes the rigorous mathematical relationships between valid compressions, while “network” emphasizes the epistemic process by which we discover them.

#### Ontological Status—Structural Emergent, Not Metaphysical Blueprint:

The Apex Network is not a pre-existing Platonic form or cosmic blueprint awaiting discovery. Rather, it is a **structural emergent**: the pattern that necessarily crystallizes from the interaction between information-processing systems and environmental constraints. Its existence is the existence of a determined pattern, not a transcendent entity.

**Modal Determinacy:** Given our universe’s actual constraint structure (thermodynamics, logical consistency, biological requirements), the Apex Network is the necessary optimal configuration—modally necessary relative to these constraints. However, in a universe with different fundamental physics or logical laws, a different Apex Network would emerge. There is no super-cosmic structure independent of physical reality itself.

**Analogy to Mathematical Necessity:** Just as  $\pi$  is not “somewhere” waiting to be found but is a necessary consequence of Euclidean geometry’s constraint structure, the Apex Network is a necessary consequence of reality’s pragmatic constraint structure. Ancient Babylonians, Greeks, and Indians discovered  $\pi$  independently not through cultural transmission but because geometric constraints determine its value. Similarly, independent cultures converge on similar low-brittleness principles (reciprocity norms, property conventions, harm prohibitions) because these are structurally necessary for viable coordination, determined by objective pragmatic constraints.

**The constraints exist first; the optimal structure they determine is a necessary implication.** Historical filtering is how we discover this structure, not how we create it.

#### **Maximum Computational Closure as Thermodynamic Minimum:**

In information-theoretic terms, the Apex Network represents the thermodynamic attractor in the phase space of possible compression systems—the configuration where information leakage is theoretically minimized. This is the limit state where Markov blankets achieve maximum alignment with environmental causal structure:

**The Limit State:** Maximum computational closure occurs when the internal model predicts the external environment with such accuracy that information leakage approaches zero. Not because the Markov blanket boundary vanishes, but because the compression achieves such high fidelity that the enacted boundary perfectly tracks genuine causal joints in reality.

**Analogy:** A perfect mirror doesn't eliminate the boundary between object and reflection, but the information crossing that boundary is transmitted with such fidelity that operationally, the distinction disappears. Similarly, the Apex Network is where conceptual boundaries (Markov blankets) achieve such high-fidelity compression that they mirror reality's constraint structure exactly.

#### **Plateau, Not Necessarily Single Peak:**

The Apex Network should not be understood 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 (basic thermodynamics, core logic), others may permit constrained pluralism of equally low-brittleness systems (hot dog taxonomy, aesthetic frameworks). Convergence is away from vast valleys of failure (the Negative Canon) and toward this resilient plateau of viable solutions.

#### **Ontologically Real, Epistemically Regulative:**

A crucial distinction: The Apex Network is ontologically real—the objective, mind-independent structure of viability that exists whether we correctly perceive it or not, determined by constraints rather than our beliefs. However, epistemically it remains a regulative ideal. We can never achieve final confirmation that our Consensus Network perfectly maps it; our knowledge is necessarily incomplete and fallible.

This dual status grounds realism (there is an objective structure) while preserving fallibilism (we cannot claim certainty about fully capturing it). The Apex Network exists as  $\pi$  exists—determined by constraints, counterfactually stable across possible histories, discoverable through systematic exploration. But unlike a Platonic form, it is an immanent pattern: the negative space revealed when systematic pragmatic filtering eliminates unviable alternatives.

#### **Formal Characterization (With Appropriate Caveats):**

We can characterize the Apex Network as the intersection of all maximally viable world-systems:

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

Where  $A$  = Apex Network,  $W_k$  = possible configurations of Standing Predicates,  $V(W_k)$  = viability function (inversely related to brittleness metrics), and  $\cap$  = intersection (common structure across all viable systems).

This formalism captures the concept but should not be mistaken for literal metaphysics. It represents the structural pattern that emerges from constraint-driven selection, not a pre-temporal mathematical object.

## 6.2 Truth as Successful Computational Closure

**Redefining Truth:** A proposition is true (Level 1) if its predicates are part of the Apex Network—the optimal computational closure configuration. In Rosas et al.'s (2024) terms, objective truth corresponds to **strong lumpability**: the predicate holds regardless of underlying substrate or initial micro-state distribution. A weakly lumpable predicate works only for specific conditions—it may be locally useful but not objectively true. A strongly lumpable predicate works across all valid realizations—it has achieved genuine substrate independence and thus qualifies as objective truth. Truth is not arbitrary social construction but achievement of maximal causal autonomy in the compression lattice.

**Three Levels Revisited Through Information Theory:**

Level	Information-Theoretic Characterization	Phenomenology
Level 3 (Coherence)	Internal consistency within a local compression scheme	Feels true within the system
Level 2 (Justified)	Compression validated by low brittleness in practice	Rationally confident it's true
Level 1 (Objective)	Part of the ultimate $\epsilon$ -machine—optimal compression of constraint structure	Would be true even if we never discovered it

**Example: Heliocentrism** - Level 3: Coherent within Copernican framework (even before validation) - Level 2: Justified once observations confirmed lower brittleness than geocentrism - Level 1: Objectively true because it's part of optimal compression of gravitational constraints

## 6.3 Convergence as Information-Geometric Necessity

**Why Different Cultures Converge on Similar Truths:**

Not because of: - Shared biology alone (though this constrains) - Social agreement (though this accelerates) - Divine revelation - Platonic access

But because: - **Same constraint structure generates same compression optima** - Independent  $\epsilon$ -machine explorers facing identical landscape - Selection pressure eliminates high-brittleness compressions - Thermodynamic attractors in the space of possible blanket configurations

**Mathematical Analogy:** Just as  $\pi$  is discovered independently (Babylonians, Greeks, Indians, Chinese) because Euclidean constraints determine it, scientific truths are discovered independently because physical constraints determine them.

**Pluralism at the Frontier:** Multiple viable compressions may exist (Pluralist Frontier), but catastrophically brittle ones (Negative Canon) are eliminated across all cultures.

## 7. Integrating With the Main Framework

### 7.1 Brittleness Metrics as Information Leakage Measures

**P(t) - Patch Velocity:** - Information-theoretic: Rate of local compression failures requiring ad-hoc fixes - Mechanistic: Blanket porosity increasing, closure failing - Phenomenology: Constant “but wait...” moments as predictions fail

**M(t) - Model Complexity:** - Information-theoretic: Compression efficiency decreasing (more parameters, worse predictions) - Mechanistic: Failed lumpability forcing micro-tracking - Phenomenology: “It’s complicated...” (unable to compress into simple story)

**R(t) - Resilience Reserve:** - Information-theoretic: Number of independent information streams successfully compressed - Mechanistic: Breadth of computational closure across domains - Phenomenology: Confidence from multi-source convergence

### 7.2 Coercion as Information Blindness

**C(t) - Coercive Overhead** deserves special attention:

Information-theoretic: Energy spent suppressing disconfirming information—creates information blindness

Mechanistic: Maintaining rigid blanket against thermodynamic gradient while severing the error signal

Critical insight: Coercion is not just energetically costly but epistemically catastrophic. It eliminates the feedback loop needed to update the Markov blanket. By suppressing dissent (the primary data stream signaling misalignment), the system goes blind to reality’s gradient, guaranteeing eventual collapse regardless of available resources.

Phenomenology: Effortful belief maintenance (“I must avoid thinking about X”), defensiveness when challenged

### 7.3 Pragmatic Pushback as Thermodynamic Necessity

**Information Can’t Be Suppressed Indefinitely:** Systems that maintain blankets misaligned with reality face thermodynamic costs: 1. Prediction errors accumulate (free energy increases) 2. Actions based on false compressions fail 3. Resources wasted compensating for misalignment 4. Eventually: Catastrophic collapse or forced revision

**This Is Not Social Construction:** It’s physics. A bridge designed with false material-strength compressions will collapse regardless of social consensus.

**The Ratchet Effect:** Once a better compression is found (lower brittleness), reverting becomes thermodynamically unfavorable—you’d have to re-pay all the information costs the compression solved.

### 7.4 The Negative Canon as Compression Failure Archive

Every entry in the Negative Canon represents a failed computational closure: - Phlogiston: Combustion blanket leaked - Miasma: Disease blanket leaked -

Lamarckian Inheritance: Evolution blanket leaked - Luminiferous Aether: Light-propagation blanket leaked

**Educational Value:** Studying failed compressions teaches the shape of constraint space—where the cliff edges are in the landscape of viable blankets.

## 7.5 Individual Agency Recovered Through Meta-Blankets

**How Free Will Emerges:** 1. **First-order blankets:** Automatic dispositions (reflexes, habits) 2. **Second-order blankets:** Awareness of dispositions (can observe own patterns) 3. **Third-order blankets:** Deliberate modification of compression strategies

**Humans Construct Meta-Blankets:** We build Markov blankets around our own information-processing, allowing: - Conscious evaluation of dispositions (“Is this bias?”) - Deliberate exploration of alternative compressions (“What if I’m wrong?”) - Willingness to bear short-term brittleness to test long-term improvements

**Agency = Variation Engine:** Individuals generate novel compressions; reality selects via differential brittleness. Neither pure determinism (we generate genuinely new patterns) nor libertarian freedom (causally constrained by information structure).

# 8. Implications and Open Questions

## 8.1 For Philosophy of Mind

**A Naturalistic Framework for Consciousness (Not a Complete Solution):**

This framework offers a functional account of consciousness that identifies relevant distinctions without claiming to eliminate the hard problem:

- **Phenomenology:** May relate to what structural pattern recognition feels like from inside
- **Conscious vs. Unconscious:** Maps roughly onto structural vs. statistical pattern processing
- **Qualia:** Potentially compression gradients rendered in subjective space, though why these have qualitative character remains unexplained
- **Self-awareness:** Meta-blanket formation (blanket monitoring its own blankets)
- **Unity of consciousness:** Integrated information across blanket hierarchies

**The Explanatory Gap Narrows But Doesn’t Close:** We’ve identified a functional distinction (statistical vs. structural pattern recognition) that appears to track the phenomenological boundary. This narrows the explanatory target: not “why does any information processing feel like something?” but “why does detecting mutual constraints feel like something?” That’s progress, but the hard problem—why any functional process has subjective character—remains open.

## 8.2 For Epistemology

**Knowledge Redefined:** - Not justified true belief (Gettier problems) - But optimized compression validated by low brittleness

**Justification Naturalized:** - Internal coherence (Level 3) necessary but insufficient - External validation (pragmatic testing) required - Truth tracks optimal compression, not correspondence to pre-existing propositions

## 8.3 For Metaphysics

**Ontology Enacted, Not Discovered:** - What “exists” = what blankets successfully compress - Different blanket configurations = different ontologies - But not arbitrary: reality constrains which blankets close

**Emergence Mechanized:** - New causal levels arise from successful computational closure - Not mysterious or epiphenomenal but thermodynamically real - The universe is layered compression hierarchies all the way up

## 8.4 For Ethics: Evil as High-Entropy Sociology

**The Markov Blanket View of Moral Agency:** If Standing Predicates are successful Markov blankets, then ethics concerns how we draw boundaries around the agency of others.

### Methodological Note—The Is/Ought Boundary:

Before proceeding, we must acknowledge a crucial limit. This framework describes how certain social configurations generate higher or lower thermodynamic costs, but it does not—and cannot—directly derive moral obligations from these descriptive facts. The move from “X generates high brittleness” to “therefore, X is morally wrong” crosses Hume’s is/ought gap.

What we can legitimately claim: Certain moral intuitions have information-theoretic grounding. The recognition that denying others’ agency generates catastrophic social costs helps explain *why* such behaviors are unsustainable and *why* moral progress often tracks toward lower-brittleness configurations. But whether we *should* care about minimizing brittleness, or whether thermodynamic efficiency has moral relevance, remains a normative question this framework doesn’t fully resolve.

With this caveat in place, we can explore how the information-theoretic perspective illuminates ethical phenomena without claiming to have derived ethics from thermodynamics.

### Information-Theoretic Analysis of Agency Denial:

Rosas et al. (2024) demonstrate that causally closed systems can be efficiently controlled through macro-level interventions—engaging with their computational closure rather than manipulating their substrate. This insight provides a mechanistic account of moral interaction: when we engage with another agent’s reasons, beliefs, and goals (their  $\epsilon$ -machine), we interact with their “software.” When we bypass their agency to force their physical body or manipulate their circumstances (intervening on the “hardware”), we breach their causal closure.

### Evil as Closure Breach = Bypassing the $\epsilon$ -Machine to Manipulate the Substrate

When a system (individual, institution, ideology) treats other agents as mere objects—as parts of the external environment to be manipulated rather than as causally closed entities with autonomous  $\epsilon$ -machines—it commits a specific information-theoretic error:

**Failed Closure Recognition:** - **Moral agents:** Achieve computational closure (autonomous  $\epsilon$ -machines, internal goals, reactive capacities) - **Objects:** Lack computational closure (can be freely manipulated without resistance) - **Evil:** Treating agents as objects (closure breach—forcing the substrate rather than engaging the software)

### Thermodynamic Consequences:

Moral Configuration	Information Structure	Brittleness Signature
<b>Recognition of Closure</b>	Engaging with others' $\epsilon$ -machines → arguments, persuasion, negotiation work through their computational closure	Low $C(t)$ : Coordination via understanding; Low $P(t)$ : Predictable responses when modeling their goals/beliefs
<b>Breach of Closure</b>	Bypassing $\epsilon$ -machines to manipulate substrate → coercion, violence, deception force the hardware while ignoring the software	High $C(t)$ : Massive coercion needed to suppress autonomous $\epsilon$ -machine responses; High $P(t)$ : Constant resistance as closed systems fight substrate manipulation

**Why Closure Breach Generates Brittleness:** When you bypass an agent's  $\epsilon$ -machine (their will, reasoning, goals) to force their substrate (their body, circumstances), you lose the predictive benefits of their internal model. You must now manage every micro-variable yourself, constantly suppressing their autonomous responses. The agent's computational closure actively resists your interventions, generating persistent prediction errors and requiring escalating coercion costs.

**Metastability Through Parasitic Endurance:** A critical qualification: coercive systems can achieve local stability (metastability) by successfully suppressing agency through overwhelming force or breaking the oppressed population's capacity for resistance. Historically, slavery, totalitarian regimes, and colonial systems have persisted for generations, not mere moments. The thermodynamic constraint operates on longer timescales than immediate collapse.

The key mechanism is **parasitic endurance**: these systems survive not through structural viability but by extracting surplus energy to pay the massive coercion costs. They appear stable only because they burn external resources—extracting surplus from the oppressed, exploiting resource windfalls (oil wealth, foreign aid), or cannibalizing their own future—to mask the entropy generated by their internal friction. This is energetic insolvency maintained through extraction, not genuine thermodynamic efficiency.

The claim is not that evil systems fail instantly, but that they accumulate higher brittleness costs than recognition-based alternatives—they require more resources to maintain, adapt more slowly to environmental changes, and face higher vulnerability to perturbations. These systems are **metastable, not viable**—appearing stable only while the subsidy lasts. Eventually, when external resources are exhausted or resistance accumulates beyond suppression capacity, the

thermodynamic gradient asserts itself. But “eventually” can span centuries. This is a long-run structural constraint, not a guarantee of swift moral justice.

Systems that refuse to engage with others’ computational closure accumulate elevated brittleness through parasitic endurance: - **Slavery:** Survived not through structural stability but by extracting surplus labor to pay the massive enforcement costs ( $C(t)$ ), though this created persistent resistance ( $P(t)$ ) and eventually collapsed when extraction could no longer cover coercion costs -

**Totalitarianism:** Achieves “zombie stability” by burning external resources (oil revenue, international subsidies, or the cannibalization of internal capital) to fund surveillance states while denying citizens’ agency—thermodynamically insolvent but can maintain metastability while resources last - **Genocide:** Ultimate closure denial—erasing agents entirely when the energetic cost of modeling their agency becomes perceived as exceeding the extraction capacity

**Not Moral Relativism:** Different cultures may draw different boundaries around “who counts as an agent” (children? animals? ecosystems?), but systems that catastrophically misalign with the actual distribution of agency in their environment pay thermodynamic costs. The Apex Network includes the recognition that other humans are agents—not because of moral axioms but because any other blanket configuration generates unsustainable brittleness.

**Connection to Expansion of Moral Circle:** Historical moral progress often involves recognizing previously-objectified groups as agents (women, slaves, colonized peoples). This isn’t just “being nicer”—it’s discovering that modeling these groups as having Markov blankets dramatically reduces social brittleness (abolishing slavery eliminates massive  $C(t)$  costs of enforcement).

## 8.5 Open Challenges

**The Integration Problem:** How do separate blankets integrate into unified experience? What determines which compressions “bind” into single qualia?

**The Novelty Problem:** How do genuinely new compressions arise? Is all creativity just recombination, or can systems generate truly novel blanket configurations?

**The Value Problem (Partially Addressed):** Section 8.4 shows how evil can be understood as high-entropy sociology—denying others’ Markov blankets. But open questions remain: Can all moral truths be reduced to thermodynamic efficiency? What about irreducibly normative dimensions (beauty, meaning, sacred values) that resist compression-theoretic analysis? These questions require further development beyond this appendix’s scope.

**The Limits Problem:** Are there hard limits to what can be compressed? Gödel’s incompleteness theorems suggest some truths resist finite compression. Exploring the implications for the Apex Network framework remains important future work.

## 9. Conclusion: A Naturalistic Framework (With Acknowledged Limits)

This appendix has developed an information-theoretic framework connecting raw information processing to conscious awareness to objective truth through compression, blanket formation, and thermodynamic selection:

1. **Information** is processed by all physical systems
2. **Compression** creates dispositions (minimal encoding of regularities)
3. **Markov Blankets** emerge when compressions achieve statistical boundaries
4. **Computational Closure** succeeds when blankets create autonomous causal levels
5. **Consciousness** may relate to meta-blanket hierarchies and structural pattern recognition
6. **Standing Predicates** are culturally-transmitted successful closures
7. **Brittleness** measures information leakage when closures fail
8. **Pragmatic Selection** eliminates high-brittleness compressions
9. **The Apex Network** is the constraint-determined optimal compression structure
10. **Truth** is not correspondence to static propositions but alignment with the Apex Network—the state where a system’s enacted boundaries map the environment’s causal constraints, achieving maximal computational closure with minimal information leakage

### **Conceptual Scaffolding, Not Dogmatic Mechanism:**

This framework employs information-theoretic language as conceptual scaffolding for understanding epistemic and cognitive phenomena. Whether brains literally implement variational free energy minimization or Standing Predicates are literally encoded as Markov blankets in neural tissue remains an empirical question. The philosophical insights—about what makes predicates successful, how knowledge systems fail, and why inquiry converges—retain their force even if specific mechanistic claims require revision.

### **What Remains Unexplained:**

We’ve identified relevant functional distinctions without solving all foundational problems: - **The hard problem of consciousness:** Why structural pattern recognition feels like anything remains open - **The is/ought gap:** Why thermodynamic efficiency should matter morally isn’t derived from the framework - **The plurality question:** How much genuine pluralism exists versus forced convergence remains empirical - **The limits of compression:** Whether some truths resist finite compression (incompleteness theorems)

### **What the Framework Achieves:**

This framework provides: - A mechanistic account of how notions form and validate through statistical vs. structural pattern recognition - An explanation of why singular experiences can carry immediate epistemic weight - A naturalistic grounding for cross-cultural convergence in knowledge systems - A functional account of consciousness that narrows the explanatory gap - A framework for understanding truth as constraint-determined structure rather than correspondence to Platonic forms - An information-theoretic analysis of why coercion generates systemic brittleness

**Phenomenology Preserved:** Consciousness, agency, and truth remain real within this framework—not eliminated or reduced away, but understood as

emerging from information processing under constraint. The framework is naturalistic without being eliminativist.

**Integration Complete:** The main paper's claims about Standing Predicates as Markov Blankets, brittleness as information leakage, and the Apex Network as thermodynamic attractor now have theoretical grounding in information theory, computational closure, and constraint-driven selection.

We are not passive observers of a pre-existing Platonic reality but active participants in discovering the constraint structure of our universe—exploring the landscape of viable compressions, mapping the Apex Network through systematic elimination of configurations that generate unsustainable brittleness.

## References

Ayvazov, Mohammad. 2025. “Toward a Phase Epistemology: Coherence, Response and the Vector of Mutual Uncertainty.” *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5250197>.

Rosas, Fernando E., Bernhard C. Geiger, Andrea I Luppi, Anil K. Seth, Daniel Polani, Michael Gastpar, and Pedro A.M. Mediano. 2024. “Software in the natural world: A computational approach to emergence in complex multi-level systems.” arXiv preprint arXiv:2402.09090.