The Imagination Machine I: A View from Somewhere

Epistemic Closure, Physical Law, and Entropy Embedded in a Block Universe

Mark Tracy
Boston University
mrktracy@bu.edu

Abstract

This paper develops a minimal formal framework for epistemology under the constraint that epistemic systems are embedded within the world they attempt to model. Because such systems lack access to an external vantage point, knowledge cannot be defined by correspondence with an independently accessible reality. Instead, epistemic coherence must arise from internal structural consistency.

Observations generate world models through an inference map, while world models generate canonical observational profiles through an implication map. Together these maps form an inference–implication loop that induces an operator on model space. Self-consistent world models appear as fixed points of this operator: models whose own implied observational profiles, when reinterpreted through inference, reproduce the models themselves. Each model therefore acts as a compression of the observation space, inducing a classifier and a corresponding quotient representation of observations.

A key structural feature of the framework is that classifiers themselves belong to the observation space. This follows from the conditions of self-representation: any system capable of epistemic reasoning must be able to encounter and revise its own acts of classification. As a result, the evaluative processes that guide model selection—valuation and will—also appear as observable elements subject to the same representational compression.

Within a given model, empirical regularities emerge as relational invariants in the induced quotient space, while entropy arises as a measure-theoretic quantity associated with the same compressive structure. The framework therefore characterizes scientific theories as stable representational compressions of observational structure for agents embedded within the environments they model.

1 Introduction

Embedded epistemic systems cannot access the universe from outside. Observations, models, classifiers, and their relations therefore exist as structures within the same universe. No external vantage point is available from which to define correspondence between representation and world.

The guiding constraint is:

> An embedded epistemic system can at most classify the ways in which it classifies the world, within the world itself.

Rather than describing temporal learning, we treat the universe as a single relational structure containing observations, models, and consistency relations between them. Within such a framework

coherence must be defined internally, as the closure of the inference–implication loop rather than as external correspondence.

This paper forms the first part of a four-paper series titled *The Imagination Machine*. The present paper develops the formal epistemic framework for embedded observers and the structure of representational closure. Companion papers develop complementary aspects of the framework: *The Imagination Machine II: Systems* analyzes agent–environment representational dynamics, *The Imagination Machine III: Toy Model of Predictive Classification* provides a minimal computational environment in which predictive agents recover relational invariants, and *The Imagination Machine IV: Institutional Intelligence* examines how such epistemic processes extend across communities and institutions.

This position is closest in spirit to Hilary Lawson's closure theory of the world. Lawson argues that openness—raw, unstructured reality—is fixed as "something" only through interventions he calls closures, and that no closure fully captures the openness beneath it. The present framework formalizes a version of this picture. The inference–implication loop is the closure mechanism; the fixed points of the operator it induces are the stable closures; a quotient space Q_w is the closed texture through which an embedded system encounters the world under the model w. The crucial point is that what a model implies is not best understood as a single isolated observational consequence, but as a canonical observational profile internal to that closure: a structured way the world shows up for a life situated within the model.

But the framework adds something to Lawson's account that his descriptive language leaves implicit: the acts of will and valuation that select among possible closures are not external to the representational structure. Because classifiers are themselves observations—for reasons derived in Section 3 rather than merely asserted—valuation is interior to the system it animates. This is the structural heart of the paper.

Two clarifications are important at the outset. First, the framework is not a form of coherentism in which any internally consistent system of representations counts as knowledge. The structure of observations within the universe constrains admissible models through the probability measure introduced below. Closure of the inference–implication loop occurs only relative to this observational structure. Second, the framework does not deny the existence of an external world. It instead observes that embedded epistemic systems cannot compare representations with that world directly. The problem addressed here is therefore structural rather than metaphysical.

A further clarification concerns model-relativity. Different self-consistent models may in principle induce different quotient spaces and therefore different families of laws. This does not imply arbitrariness. Models must compress the same observational distribution and remain stable under their own implications. The resulting plurality, if it occurs, is constrained plurality.

The aim of the framework is not to replace empirical science or traditional epistemology, but to describe the structural constraints under which an epistemic system embedded within the universe must operate.

A word on what the framework does and does not claim. The formal architecture precisely locates three problems that resist full resolution from within any closure: the problem of will, the problem of distinguishing genuine from merely apparent epistemic openness, and the problem of the criterion by which a system recognises new observations as demanding refinement. The paper argues that locating these problems with formal precision is itself a contribution—that a framework which shows exactly where explanation runs out is preferable to one that conceals those limits behind descriptive fluency.

2 Relation to Existing Approaches

The framework developed here sits at the intersection of several existing lines of research, while differing from each in its formal treatment of embeddedness, representational closure, and model-relative structure.

Most directly, it formalises central commitments of Hilary Lawson's closure theory. Lawson argues that the world as encountered is always a world fixed by closure, that openness underlies and escapes every closure, and that the question of which closures to adopt is therefore irreducibly evaluative (Lawson, 2001). The present framework gives these claims a precise structural expression: the inference–implication loop is the closure mechanism, $\mathcal{W}^*$ is the space of stable closures, the quotient space is the closed texture, and the inclusion $C \subseteq D$ is the formal statement that evaluation is interior to the representational structure rather than prior to it. The analysis of institutions and refinement extends this picture by showing that the evaluative dimension of closure is not merely a feature of individual systems but is transmitted, compressed, and potentially lost across generations.

The account also bears comparison with predictive and Bayesian approaches in contemporary philosophy of mind and cognitive science. Predictive processing models treat cognition as the continuous generation of predictions that are compared with incoming sensory signals, with discrepancies driving model revision (Clark, 2016; Friston, 2010). The inference–implication loop introduced here has a related structure: observations generate models through the inference map F, while models generate observational implications through the map g. However, the present framework differs from predictive-processing accounts in one crucial respect: both observations and models are treated as structures internal to a single universe rather than as elements of an external inference problem. The framework therefore addresses not only how models are updated, but how coherence is to be defined for an epistemic system that has no access to an external vantage point. A minimal computational environment in which predictive agents recover relational invariants from structured observations is developed in *The Imagination Machine III: Toy Model of Predictive Classification*.

In philosophy of science, the view developed here is also close in spirit to structural realism. Structural realists argue that scientific knowledge concerns the relational structure of the world rather than the intrinsic nature of unobservable entities (Worrall, 1989; Ladyman, 1998). In the present framework, relational structure appears in an explicitly model-relative mathematical form. Each self-consistent world model w induces a classifier $\pi_w : D \to Z_w$ that partitions the observation space, and empirical regularities arise as relational invariants in the quotient space $Q_w = D/\!\sim_w$ determined by that partition. What embedded observers identify as physical laws are therefore relational structures within a representational quotient induced by the model. In this sense the framework provides a formal account of how structural knowledge arises from representational compression.

This model-relative account of law also bears comparison with relational approaches in physics. Rovelli's relational quantum mechanics emphasises that physical properties are defined relative to interactions rather than to absolute external states (Rovelli, 1996). Physical laws in the present framework are likewise relational invariants, though the present argument grounds their model-relativity in epistemological rather than specifically physical considerations.

The entropy measure introduced here connects the framework to statistical mechanics and information theory. Shannon introduced entropy as a logarithmic measure of expected surprisal associated with a probability distribution (Shannon, 1948). Jaynes later interpreted statistical mechanics as inference over probability distributions subject to informational constraints (Jaynes, 1957). The present framework recovers entropy as a consequence of representational compression rather than positing it as primitive: the classifier π_w partitions the observation space into equivalence classes, and the entropy $H(w)$ measures the expected surprisal of those classes. The

framework does not claim identity between this quantity and thermodynamic entropy; rather, it argues for a structural convergence between them, grounded in their shared dependence on the partitioning of a probability space.

In biology and systems theory, Maturana and Varela described cognition as arising from operational closure within self-referential systems (Maturana and Varela, 1980; Varela et al., 1991). The self-consistency condition $T(w) = w$ is a formal analogue of operational closure, with the additional feature that the closed system contains its own evaluative structure as classified content. Read in the present terms, closure is reproduced not merely from isolated outputs but from the structured observational profile a model makes possible from within. The dynamical structure of agent–environment interaction underlying such representational frameworks is analyzed in *The Imagination Machine II: Systems*.

Finally, the social extension of the framework places it in conversation with social epistemology. Longino and Kitcher have both argued, in different ways, that knowledge is constitutively social and that the norms governing inquiry are sustained and revised by communities rather than by isolated individuals (Longino, 1990; Kitcher, 1993). The institutional analysis developed here is consistent with this emphasis while grounding it in the formal architecture of the framework. The distinction between generative and compressed inheritance corresponds, at the social level, to the difference between communities that transmit the capacity for inquiry and communities that merely conserve its prior outputs. A fuller treatment of institutional knowledge generation and transmission is developed in *The Imagination Machine IV: Institutional Intelligence*.

3 The Block Universe and the Derivation of $C \subseteq D$

Let Ω denote the universe. Define the following subsets:

$$D \subseteq \Omega \quad \text{(the set of observations)}$$

$$\mathcal{W} \subseteq \Omega \quad \text{(the set of world models)}$$

$$C \subseteq \Omega \quad \text{(the set of classifiers)}.$$

We argue for, rather than merely stipulate, the inclusion

$$C \subseteq D \subseteq \Omega.$$

The argument proceeds from the conditions of self-aware representation. Consider what distinguishes an epistemic system—a genuine subject—from a mere transducer. A thermostat classifies temperature, but its classification is not available to it as an object of experience. It cannot encounter its own sorting activity as something that could have been otherwise. An epistemic system, by contrast, is one whose classificatory acts are themselves accessible to it: it can attend to how it is attending, sort its ways of sorting, and in principle revise the dispositions that govern its encounter with the world.

This reflexive accessibility is not an optional feature added to an otherwise complete epistemic system. It is the condition that makes a system epistemic in the first place. A system that cannot encounter its own classifiers cannot recognise itself as one possible closure among others, cannot doubt its own representations, and therefore cannot be said to know in any sense that involves the distinction between appearance and reality. Cartesian doubt is only possible for a system whose classificatory acts are elements of its observation space.

The inclusion $C \subseteq D$ is therefore a transcendental condition: any system that satisfies the minimal criterion for being an epistemic subject must satisfy it. The formal apparatus of this paper applies precisely to systems meeting that criterion.

Remark 1 (Reflexivity Without Vicious Regress)**.** *The condition $C \subseteq D$ means that the system can classify its own classifiers. One might worry that classifying a classifier requires a further classifier, which requires a further classifier still, generating an infinite regress. This regress does not arise in the block universe framing because that framing is atemporal: all observations, including observations of classifiers, are simultaneous elements of the single relational structure Ω. The self-consistency condition $T(w) = w$, developed in Section 10, is a fixed-point condition rather than a termination condition. What matters is not that the regress terminates in a foundation but that the loop closes on a stable fixed point.*

4 World Models and Classification

Each world model $w \in \mathcal{W}$ induces a classifier

$$\pi_w : D \to Z_w$$

where $Z_w \subseteq D$. Thus a model compresses observations by mapping them to representative observational states. Because $C \subseteq D$, the domain of π_w includes classifiers themselves. A world model therefore classifies not only raw observational content but also the evaluative and selective dispositions of the system that holds it.

Remark 2 (Representational Witness)**.** *The condition $Z_w \subseteq D$ ensures that every abstract class induced by π_w is instantiated by at least one observational state. The representative is not assumed to be unique or privileged; it merely witnesses the existence of the class.*

Definition 1 (Model-Induced Equivalence Relation)**.** *For $d_1, d_2 \in D$ define*

$$d_1 \sim_w d_2 \quad \textit{iff} \quad \pi_w(d_1) = \pi_w(d_2).$$

Definition 2 (Equivalence Class)**.** *For $d \in D$, define*

$$[d]_w = \{d' \in D \mid \pi_w(d') = \pi_w(d)\}.$$

The classifier therefore induces a partition of the observation space. When d is itself a classifier—that is, when $d \in C$—its equivalence class $[d]_w$ groups together all observational states that the world model treats as equivalent ways of sorting the world. Different valuations may thus collapse into the same equivalence class under a given model, or be distinguished by a more refined one.

5 Valuation and Will as Interior Observations

The inclusion $C \subseteq D$ has a consequence that deserves explicit statement before the formal development continues.

Valuation—the assignment of significance to observations—and will—the selective pressure that drives a system toward one closure rather than another—are traditionally treated as standing outside epistemological frameworks. They appear as boundary conditions: given that a system values certain outcomes, what can it know? The present framework does not dissolve this exterior status so much as restate it with formal precision.

If the acts by which a system evaluates and selects are themselves classifiers, and if classifiers are observations, then valuation and will are elements of D. They are subject to the same measure μ_D, the same quotient structure induced by π_w, and the same representational compression as any

other observation. A self-consistent world model does not merely organise perceptual content; it also classifies the evaluative structure through which the system engages the world.

This does not reduce will to mechanism, nor does it claim to resolve the problem of agency. What it establishes is more modest and more precise: will appears within D, is partially compressed by every model, and yet is not exhausted by any compression. This is not because will is supernatural or causally unconstrained, but because it is the condition under which the world becomes held as anything at all—the potentiality that precedes and exceeds any particular representation of it. The formal loop determines the space of stable closures $\mathcal{W}^*$, but the selection of a particular element from that space is precisely what the framework locates as irreducible. Willing is not explained away; it is what remains when the inference–implication loop has done everything it can do—not a gap in the framework, but the condition the framework must include without being able to absorb.

Metaphysical closure is therefore prevented not by any deficiency of the representational apparatus, but by what the apparatus must include: the very acts of valuation that animate it. The framework's contribution here is not resolution but precision—knowing exactly where the limit lies is different from not knowing where to look.

6 Statistical Structure

Assume the observation space carries a probability structure

$$(D, \Sigma_D, \mu_D)$$

where Σ_D is a σ-algebra and μ_D a probability measure.

The measure μ_D is the principal way in which observational structure constrains closure. It prevents the framework from collapsing into the view that any self-supporting classificatory system is epistemically on a par with any other. Models partition one and the same observational space, and the measure of those partitions is not up to the model alone.

Proposition 1 (Measurable Partition)**.** *If each π_w is measurable and Z_w carries a σ-algebra in which singletons are measurable, then the equivalence classes $[d]_w$ form a measurable partition of (D, Σ_D, μ_D).*

Proof. Since π_w is measurable, the preimage of each singleton in Z_w lies in Σ_D. But

$$[d]_w = \pi_w^{-1}(\{\pi_w(d)\}),$$

so each equivalence class is measurable. The classes partition D by construction. □

Lemma 1 (Probability of Classes)**.** *For any model w,*

$$\sum_{[d]_w \in Q_w} \mu_D([d]_w) = 1.$$

Proof. The sets $[d]_w$ form a measurable partition of D. Since μ_D is a probability measure on D, the total measure of the partition equals $\mu_D(D) = 1$. □

Remark 3 (Origin and Calibration of the Observational Measure)**.** *The probability measure μ_D represents the empirical distribution of observations across the observation space D. Conceptually it may be understood in several compatible ways.*

First, it may represent the long-run frequency distribution of observations generated across the ensemble of observers embedded in Ω. Since D contains the observations of all observers, the measure aggregates the empirical structure encountered throughout the block universe. This need not be understood as arbitrary sampling from an undifferentiated flux. In many natural settings, observers are embedded in environments structured by stable but incommensurate dynamical cycles whose relative phases continually drift without exact repetition. Under such conditions, sequential observation repeatedly samples a structured signal that is neither perfectly periodic nor wholly unconstrained. The result is an empirical distribution over observational states: enough recurrence for stable frequencies to emerge, enough phase drift for novelty to persist. On this view, μ_D arises from the statistical structure induced by the dynamical environment in which embedded observers occur.

Second, μ_D may be interpreted inferentially. Following the information-theoretic programme associated with Jaynes, probability distributions can be understood as representations of incomplete knowledge subject to constraints. Under this interpretation μ_D encodes the informational constraints under which an embedded epistemic system performs inference.

These two readings are compatible. A structured observational environment gives rise to stable empirical frequencies, while inference treats those frequencies as constraints on admissible closure. The framework therefore does not require commitment to probability as either purely objective or purely epistemic. What matters structurally is that all world models compress the same observational distribution. This shared measure prevents the space of self-consistent closures from collapsing into arbitrary coherent systems.

However, the compatibility of these two readings is itself a condition that can be satisfied or failed. Call this condition calibration*: the alignment between a system's inferential μ_D—the weights it brings to inference—and the actual empirical distribution of observations in its environment. Calibration is an achievement rather than a default. It can fail in at least two ways. A system may be* miscalibrated*: its inferential weights systematically diverge from actual observational frequencies, producing self-consistent closures that are stable relative to the wrong measure. Such a system refines willingly and generates genuine laws—but laws of a distribution that does not reflect the environment it inhabits. Miscalibration is therefore distinct from both dogmatism and ordinary error: the closure is open to refinement, yet refinement proceeds against a distorted image of the world. Calibration can also fail under* distributional shift*: in genuinely novel environments, a system's inferential μ_D is an extrapolation from past frequencies into regions where those frequencies no longer apply. The alignment between the two readings breaks down precisely where epistemic pressure is greatest.*

Miscalibration thus constitutes a third structural location of epistemic risk, alongside dogmatic refusal to refine and the irreducible remainder of will. The framework diagnoses all three as failures at different levels of the hierarchy $(F, g) \to T \to \mathcal{W}^* \to \pi_w \to Q_w \to R_w$: dogmatism is a failure at the level of (F, g); miscalibration is a failure at the level of μ_D itself, prior to the construction of any particular closure; and will names the underdetermination that persists even when both are functioning well.

7 Representational Quotient

Each model induces a quotient space

$$Q_w = D/\!\sim_w .$$

The elements of Q_w represent observational states modulo the classification performed by the model. This is the closed texture through which the world is encountered: not the world as it is prior to closure, but the world as fixed by the representational intervention of π_w.

Because $C \subseteq D$, the quotient space Q_w contains equivalence classes of classifiers alongside equivalence classes of other observations. The closed texture therefore includes, within itself, the compressed image of the evaluative structure of the system that produced it.

To collect these model-relative quotient spaces into a single ambient codomain, define

$$Q := \bigsqcup_{w \in \mathcal{W}} Q_w,$$

the disjoint union of all quotient spaces induced by world models in $\mathcal{W}$. Thus each Q_w is canonically embedded in Q, while remaining distinguished from $Q_{w'}$ when $w \neq w'$.

8 Implication

For each model $w \in \mathcal{W}$, let

$$\Gamma_w$$

denote the set of canonical observational profiles induced by w, where each such profile is structured in the quotient space Q_w. These profiles are not single isolated observations, but model-relative patterns of observational life: structured ways the world becomes legible from within the closure determined by w.

Define the ambient profile space

$$\Gamma := \bigsqcup_{w \in \mathcal{W}} \Gamma_w.$$

World models produce canonical observational profiles through a map

$$g : \mathcal{W} \to \Gamma$$

such that, for each model $w \in \mathcal{W}$,

$$g(w) \in \Gamma_w \subseteq \Gamma.$$

Thus g assigns to each world model a model-relative observational profile internal to the closure induced by that very model.

9 Inference

Canonical observational profiles generate world models through

$$F : \Gamma \to \mathcal{W}.$$

10 The Consistency Loop

The system is governed by the pair of maps

$$\Gamma \xrightarrow{F} \mathcal{W} \xrightarrow{g} \Gamma.$$

Define the induced operator

$$T = F \circ g : \mathcal{W} \to \mathcal{W}.$$

Definition 3 (Self-Consistent World Model)**.** *A model w is self-consistent if $T(w) = w$.*

Define

$$\mathcal{W}^* = \{w \in \mathcal{W} \mid T(w) = w\}.$$

Self-consistent models reproduce themselves when inference is applied to their own implied observational profiles. In Lawson's terms, they are stable closures: the system's representational intervention reproduces itself under the loop of implication and re-inference. More precisely, a self-consistent model is one whose implied observational profile, when re-submitted to inference, regenerates the same model.

A natural worry is that the fixed-point condition may be too weak: if the maps F and g are unconstrained, perhaps trivial fixed points proliferate. That worry is legitimate in the abstract. The framework does not claim that every fixed point is equally significant. Its claim is that any epistemically admissible closure must at least satisfy this condition, and that the observational measure μ_D together with the refinement structure developed below provides a basis for distinguishing empty stability from informative stability.

Remark 4 (Existence of Fixed Points)**.** *The framework defines epistemically admissible closures as fixed points of the operator $T = F \circ g$. The formal development does not assume that fixed points exist for arbitrary choices of F and g. Rather, the framework identifies a structural condition that any stable closure must satisfy if it exists.*

In many natural settings fixed points arise under mild assumptions. For example, if $\mathcal{W}$ is endowed with a compact topology and T is continuous, Schauder's fixed-point theorem ensures the existence of at least one $w^* \in \mathcal{W}$ such that $T(w^*) = w^*$.

In algorithmic or statistical settings the operator may instead be interpreted as an iterative update rule whose empirical convergence defines the effective closure.

The present framework therefore does not claim that all conceivable inference– implication structures admit stable closures. It instead provides the formal characterisation that any such closure must satisfy when it occurs. In this sense the framework is generative: it specifies meta-structural constraints that a world model must satisfy in order to reproduce itself under the inference–implication loop.

Remark 5 (Plurality of Stable Closures)**.** *Nothing in the framework requires $\mathcal{W}^*$ to be a singleton. Multiple incompatible self-consistent models may coexist as elements of $\mathcal{W}^*$. This plurality is not a defect. It corresponds directly to Lawson's insistence that no single closure is metaphysically privileged. The operator T determines the space of possible stable closures, but it does not determine which element of $\mathcal{W}^*$ is instantiated.*

11 Relational Structure

For each integer $i \geq 1$ define

$$K_i(Q_w) = Q_w^i,$$

the i-fold Cartesian product of Q_w with itself. Thus an element of $K_i(Q_w)$ is an ordered tuple

$$\tau = ([d_1]_w, [d_2]_w, \ldots, [d_i]_w).$$

Let

$$K(Q_w) = \bigsqcup_{i=1}^{\infty} K_i(Q_w) = \bigsqcup_{i=1}^{\infty} Q_w^i,$$

the disjoint union of all finite Cartesian powers of Q_w, collecting relational tuples of every arity into a single set.

Elements of $K(Q_w)$ represent finite relational configurations among equivalence classes of observations, together with their arities. A relational classifier

$$R_w : K(Q_w) \to Q_w$$

assigns canonical relational consequences within the quotient space.

12 Physical Law

Definition 4 (Relational Equivalence)**.** *For $\tau_1, \tau_2 \in K(Q_w)$ define*

$$\tau_1 \sim_{R_w} \tau_2 \quad \textit{iff} \quad R_w(\tau_1) = R_w(\tau_2).$$

Definition 5 (Physical Law)**.** *A physical law under a model w is a relational equivalence class*

$$L = [\tau]_{R_w}$$

for some $\tau \in K(Q_w)$.

Physical laws appear as relational structures within the quotient representation induced by a self-consistent world model. They are stable patterns in the closed texture, not features of an independently accessible world. Different elements of $\mathcal{W}^*$ may induce different quotient spaces and therefore different relational invariants; which laws appear depends on which closure is sustained.

This model-relativity should not be confused with arbitrariness. Any such law is still a law of one and the same observational world as compressed under a particular stable closure. If multiple closures persist, they persist under the constraint of the same D and the same μ_D.

13 Entropy

The classifier π_w compresses the observation space. In this section we assume that the partition $Q_w = D/\!\sim_w$ is finite or countable, so that the sums below are well defined.

Definition 6 (Class Measure)**.**

$$M_w(d) = \mu_D([d]_w).$$

Definition 7 (Model-Relative Surprisal)**.**

$$S_w([d]_w) = -\log \mu_D([d]_w).$$

Definition 8 (Model-Relative Entropy)**.**

$$H(w) = -\sum_{[d]_w \in Q_w} \mu_D([d]_w) \log \mu_D([d]_w).$$

The quantity $S_w([d]_w)$ measures the probability mass of the equivalence class $[d]_w$, which is the fiber of the projection $\pi_w : D \to Q_w$. The quantity $H(w)$ is the expected surprisal induced by the partition defined by π_w and therefore measures the representational compression associated with the model.

Because classifiers are elements of D, both surprisal and entropy assign measure-theoretic weight not only to equivalence classes of perceptual content but also to equivalence classes of valuations. A valuation that is rare in D carries high surprisal. A coarse model that collapses many distinct

valuations into a single class yields low surprisal for that class and lowers the effective distinguishability of evaluative structure. Entropy is therefore not merely a feature of perceptual content; it also measures the coarseness with which a model distinguishes the system's evaluative dispositions.

A note on scope is warranted. The entropy $H(w)$ defined above is a Shannon-type quantity derived from representational compression. The framework does not claim identity between this quantity and thermodynamic entropy. It claims structural convergence: both quantities arise from the same underlying operation of partitioning a probability space, and Jaynes' programme of deriving statistical mechanics from inference over probability distributions subject to informational constraints suggests that this convergence is not superficial. The precise conditions under which model-relative entropy and thermodynamic entropy coincide are left for subsequent work.

14 Representational Refinement

Definition 9 (Refinement)**.** *A model w_2 refines w_1 if $[d]_{w_2} \subseteq [d]_{w_1}$ for all $d \in D$.*

Theorem 1 (Monotonicity of Surprisal)**.** *If w_2 refines w_1, then*

$$S_{w_2}([d]_{w_2}) \geq S_{w_1}([d]_{w_1}).$$

Proof. Refinement implies $[d]_{w_2} \subseteq [d]_{w_1}$, so $\mu_D([d]_{w_2}) \leq \mu_D([d]_{w_1})$. Applying $-\log$ reverses the inequality. □

Theorem 2 (Entropy Equality for Equivalent Observations)**.** *If $d_1 \sim_w d_2$, then $S_w([d_1]_w) = S_w([d_2]_w)$.*

Proof. If $d_1 \sim_w d_2$ then $[d_1]_w = [d_2]_w$, so $\mu_D([d_1]_w) = \mu_D([d_2]_w)$ and the definition of S_w yields the result. □

14.1 Refinement as Dilution

It is a common intuition that refinement—the transition from w_1 to w_2 where $[d]_{w_2} \subseteq [d]_{w_1}$—represents a "narrowing in" on a point-like truth. However, the framework suggests the inverse. As the partition becomes finer, the measure μ_D associated with each class decreases, and the surprisal increases.

If we consider the individuals within each class as the unobserved territory, refinement actually produces a *coarser coverage* of the underlying openness. Each refined class $[d]_{w_2}$ holds fewer observational states, but the density of the unknown within our representation increases. We do not converge toward an external object; rather, we dilute our representational density, creating the very space in which the constitutive remainder manifests as surprisal. Refinement is not the elimination of uncertainty, but the formal expansion of the system's capacity to be surprised.

15 Interpretation

The hierarchy of structure is

$$(F, g) \to T \to \mathcal{W}^* \to \pi_w \to Q_w \to R_w.$$

The condition $C \subseteq D$ runs through every level of this hierarchy. Classifiers enter the observation space as observations, are compressed by π_w, appear in the quotient space Q_w, figure in relational

tuples in $K(Q_w)$, and carry surprisal under S_w. Valuation is not a parameter set from outside the system; it is a structural feature of the observation space that the system's own representational apparatus must absorb, compress, and partially lose.

The implication map now makes explicit that closure is reproduced not from an atomized observational residue but from a structured observational profile. A stable closure is therefore a view from somewhere in the strict sense: a model whose own internally generated way of inhabiting the world, when reinterpreted through inference, yields the same model again.

15.1 Backing into the Future

The atemporal nature of the block universe framing suggests a reinterpretation of the experience of time. If Ω is a static relational structure, then "the future" is not a state that has not yet occurred, but a region of the block universe toward which the system's current stable closure w^* has not yet been extended.

The system does not move into the future; rather, it *backs into it* according to past patterns. The inference–implication loop $T(w) = w$ is a consistency condition derived from the existing weight of observations. When the system encounters the unmapped regions of the block, it projects its current relational invariants (L) as a structural expectation.

On this reading, the experience of time is the process of this projection being stressed by the constitutive remainder. Because refinement increases surprisal, the future feels ultimately unpredictable not because it is non-existent, but because our attempt to know it more finely—to refine our "backing" movement—necessarily dilutes our coverage. We encounter the future as a growing coarseness of classes, where the patterns of the past are the only machinery available to navigate the increasing openness of the territory ahead. This unpredictability is the formal address of will: the necessity of choosing a closure in a territory that the model can never fully exhaust.

16 Institutions as Intergenerational Compression

The framework developed so far treats an epistemic system as a single relational structure. But embedded systems are not isolated. They exist within communities of systems that share, contest, and transmit closures across time. This section extends the framework to that social dimension, focusing on the role of institutions. A fuller institutional formalization is developed in *The Imagination Machine IV: Institutional Intelligence*.

The central observation is this: no individual knower transmits a closure to a successor by reproducing the full observation space D that gave rise to it. What is transmitted is always a compression—a residue of the inferential work that produced a given $w^* \in \mathcal{W}^*$. Institutions are the mechanisms by which this intergenerational compression is stabilised.

More precisely, what passes between generations is not the loop itself—the maps F and g that generated the fixed point—but a projection of the implied observational profile and the quotient structure it presupposes into the observation space of the successor generation. The successor receives the closed texture without necessarily receiving the closure mechanism. Institutions are the structures within Ω that perform and stabilise this projection, re-embedding the inherited profile of closure as observations in the successor's D, making it available for classification by the successor's own π_w.

This framing carries an immediate consequence. A successor generation may inherit a stable closure without inheriting the capacity to regenerate it under pressure from new observations. The quotient structure arrives, but the inferential machinery that produced it does not.

We distinguish two modes of institutional transmission. *Compressed inheritance* transmits the closed profile alone: the successor can apply the inherited partition but cannot update it. *Generative inheritance* transmits F and g alongside that profile: the successor can regenerate the closure from within, extend it, and revise it when new observations demand a finer partition.

The distinction matters because the observation space D does not stand still. New observations enter D in every generation, and a partition that was self-consistent under an earlier μ_D may fail to remain so as the measure shifts. A generatively inherited closure can meet this pressure; a compressedly inherited one cannot. The institution that transmits only the quotient structure is therefore more fragile—not because it contains false beliefs, but because it has lost the capacity to refine.

Note that institutions may also transmit miscalibrated measures. A community that inherits both F and g alongside a systematically distorted μ_D possesses the machinery for refinement while lacking accurate observational weights on which to exercise it. Generative inheritance is therefore necessary but not sufficient for epistemic health: the inferential measure must also track the environment it purports to represent.

17 Knowledge, Dogma, and the Structure of Refinement

A natural question arises from the plurality of stable closures established in Section 10: if $\mathcal{W}^*$ may contain many incompatible elements, and the framework provides no external criterion for preferring one over another, how does it distinguish knowledge from dogma? Both are self-consistent. Both survive the inference–implication loop. Both can be institutionally transmitted.

The answer is that the distinction does not require an external criterion. It falls out of the structure already in place, specifically from the relationship between a closure and its behaviour under refinement.

Recall that a model w_2 refines w_1 when $[d]_{w_2} \subseteq [d]_{w_1}$ for all $d \in D$. Refinement always costs higher surprisal: a finer partition assigns lower probability mass to each class and therefore higher S_w to each observation. A closure disposed toward knowledge is one that remains willing to pay this cost—one whose inference–implication loop, when supplied with observations that increase the consistency gap under the current partition, responds by generating a finer π_w rather than forcing the new observations into existing classes.

Dogmatic closure is precisely the refusal to pay this cost. A dogmatic model maintains its self-consistency not by genuinely absorbing new observations but by compressing them into existing equivalence classes regardless of their character. New elements of D are mapped by π_w to existing elements of Z_w even when a more faithful compression would require extending Z_w. The partition is held fixed; the observations are bent to fit it.

Miscalibration, introduced in Remark 3, constitutes a distinct failure mode. A miscalibrated closure may be fully open to refinement—willing to extend Z_w whenever the consistency gap demands it—and yet refine systematically against a distorted image of the observational world. Where dogmatism is a failure of disposition at the level of (F, g), miscalibration is a failure of the measure μ_D itself, prior to any particular act of closure. The two failures are formally separable: a closure can be dogmatic without being miscalibrated, or miscalibrated without being dogmatic, or both simultaneously.

A clarification is required here. The criterion just stated relies on a notion of stable absorption that is not itself fully decidable from within a single closure. Determining whether a new observation d genuinely strains the existing partition or is legitimately compressed into it requires assessing the consistency gap, and different closures may assess that gap differently. The framework does

not resolve this from outside; it rather establishes the vocabulary within which the question can be precisely posed and contested. The distinction between knowledge and dogma is therefore best understood as identifying a structural disposition—the preparedness to extend Z_w under pressure—rather than as a decision procedure that can be applied mechanically from within any single closure. Crucially, this question is available to any system satisfying $C \subseteq D$, since such a system can observe its own classificatory behaviour and the consistency of its loop.

Several further consequences follow. First, the distinction is not binary but gradational. A closure may be refinable with respect to some regions of D while dogmatic with respect to others. Institutions that transmit F and g alongside the inherited profile of closure preserve the capacity for refinement, but may do so selectively—maintaining the inferential machinery for some domains while suppressing it for others.

Second, the surprisal cost of refinement explains a persistent feature of actual epistemic communities. Dogmatic compression avoids this cost by refusing to see new observations as genuinely new. Coarser models assign lower surprisal to the observations they assimilate, and lower surprisal feels, from within the closure, like greater understanding. The framework thus provides a structural account of why the pressure toward dogmatic closure is not merely psychological but has a measure-theoretic basis.

Third, because $C \subseteq D$, the distinction applies to evaluative structure as well as perceptual content. A closure that refuses to refine its classification of classifiers—that compresses distinct valuations into the same equivalence class regardless of the observational pressure to distinguish them—is dogmatic about value in precisely the same structural sense. The framework does not treat these as different in kind.

Returning to the hierarchy established in Section 15, the distinction between knowledge and dogma lives at the level of (F, g) rather than at the level of $\mathcal{W}^*$. Two closures may be indistinguishable as fixed points—equally self-consistent, equally stable—while differing fundamentally in whether the loop they instantiate remains open to refinement. Stability is not the same as openness, and it is openness to refinement—the disposition to pay the surprisal cost when the consistency loop demands it—that the present framework identifies as the structural mark of what distinguishes knowledge from its appearance.

18 Conclusion

Embedded epistemic systems cannot appeal to external correspondence as their standard of coherence. Coherence appears instead as internal closure of the inference–implication loop under the statistical structure of observations. Self-consistent world models arise as fixed points of the operator this loop induces, and each such model compresses the observation space into a quotient representation whose relational invariants constitute physical law and whose measure-theoretic multiplicity constitutes entropy.

The structural feature that distinguishes this framework from earlier accounts is the inclusion $C \subseteq D$: classifiers are observations. This inclusion is not stipulated but derived—it is the transcendental condition on any system capable of Cartesian doubt, any system that can recognise itself as one possible closure among others. This means that valuation and will—the dispositions that select among possible closures—are interior to the representational architecture. They appear in the observation space, are subject to compression, and leave their trace in the quotient structure. Yet they are not exhausted by any compression. The formal loop determines the space of stable closures, but not which closure is instantiated. This remainder is not a gap in the framework; it is the constitutive openness that the inference–implication loop must encompass but cannot exhaust.

The implication map clarifies the form of that closure. What a model implies is not merely an isolated consequence but a canonical observational profile internal to the model itself: a structured way the world appears from somewhere. A self-consistent world model is therefore one whose own implied profile of observational life, when reinterpreted through inference, reproduces that same model. Stable theory and stable world-profile co-arise.

The social extension of the framework yields two further results that follow from the same architecture without requiring external normative imports. Institutions are the mechanisms by which stable closures are transmitted across generations, but they transmit closures in two structurally distinct modes: generative inheritance conveys the inferential machinery alongside the fixed point, while compressed inheritance conveys only the inherited profile of closure. And the distinction between knowledge and dogma reduces, within the framework, to the distinction between closures that remain open to refinement and those that hold their partition fixed against the pressure of new observations—a difference that identifies a structural disposition rather than a decision procedure applicable from outside any particular closure.

The framework thus diagnoses three irreducible structural locations of epistemic risk. Dogmatism is a failure of disposition at the level of (F, g): the loop exists but refuses to refine. Miscalibration is a failure at the level of μ_D: the loop refines willingly but against a distorted image of the world. And will names the underdetermination that persists even when both are functioning well—the necessity of choosing a closure in territory no model can fully exhaust. Together these three constitute the complete formal topology of epistemic failure for an embedded system.

Epistemic closure, physical law, entropy, and the social conditions of knowledge therefore emerge as successive consequences of a single embedded representational architecture. What prevents metaphysical closure—what keeps the system in relation to the openness beneath its representations—is the evaluative structure that the architecture must include but cannot fully exhaust.

References

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11:127–138.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106(4):620–630.

Kitcher, P. (1993). *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford University Press.

Ladyman, J. (1998). What is structural realism? *Studies in History and Philosophy of Science Part A*, 29(3):409–424.

Lawson, H. (2001). *Closure: A Story of Everything*. Routledge.

Longino, H. E. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.

Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel.

Rovelli, C. (1996). Relational quantum mechanics. *International Journal of Theoretical Physics*, 35(8):1637–1678.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Worrall, J. (1989). Structural realism: The best of both worlds? *Dialectica*, 43(1–2):99–124.