Self-Consistency of Embedded Classification in a Block Universe

A Formal Model of Model-Relative Epistemic Closure

Mark Tracy

Abstract

Epistemic systems are embedded within the world they attempt to understand. Because they have no access to an external vantage point, the traditional notion that knowledge consists in correspondence between representation and an independent world cannot serve as their standard of coherence. This paper develops a formal framework for epistemology under this constraint. Observations, models, and classifiers are treated as structures within a single universe. World models compress observations by inducing partitions of the observation space. They also generate canonical observational implications. An observation is consistent with a model when it falls within the same equivalence class, under the model's own classification, as the observation implied by that model.

This structure produces a fundamental loop between observation and model formation. Self-consistent models appear as fixed points of this loop: models that reproduce themselves when their own implications are reinterpreted as observations. Under this view, epistemic coherence for an embedded system is not correspondence with an external world but closure of the inference–implication cycle within the system's own classification structure. The framework thereby provides a minimal formal account of model-relative epistemic closure in a block-universe setting, where observations, models, and their consistency relations are themselves internal features of the universe being modeled.

1 Preface

The theory presented in these pages addresses a problem that is prior to most of what passes for epistemology: not how we come to know things, but what it means for a knowing system to be coherent when that system is itself part of what it is trying to know. The author does not approach this problem from outside. No such outside is available. The framework is built entirely from within the situation it describes, and this is not a limitation of the work but its central thesis.

The reader will find here a modest number of definitions and a small collection of structural claims. The mathematical apparatus is intentionally minimal. The objects considered are sets equipped, where needed, with measurable structure, and the maps between them are ordinary functions satisfying basic compatibility conditions. But the reader who moves quickly through the definitions in search of theorems will miss what the work is doing. The definitions *are* the contribution. To define, precisely, what it means for an observation to be consistent with a model under that model's own classification—and then to define self-consistency as a fixed point of the loop between inference and implication—is to do something that looser treatments of the same problem have not done. The precision is structural rather than technical.

What the work proposes, at bottom, is this: an embedded epistemic system cannot appeal to correspondence with an external world as its standard of coherence, because the notion of external is unavailable to it. The only coherence on offer is internal closure—the condition under which a model's implied observation falls within the same equivalence class, under the model's own partition, as the observation that generated the model. This is the fixed point. It does not guarantee truth in any classical sense. It guarantees something more modest and, in the circumstances, more honest: that the system is not contradicting itself by its own lights.

The philosophical tradition offers several predecessors to this move. Kant argued that the conditions of possible experience are not themselves given in experience but are its preconditions. Wittgenstein, in his earlier work, traced the limits of what can be said by pressing language against those limits until they showed. Maturana and Varela asked what structural properties a system must have in order to be a knowing system at all, and found their answer in the notion of closure. The present work belongs to this lineage without announcing the fact. It is more compressed than any of these predecessors, and more formally explicit, though it shares their essential orientation: the knowing subject is not a spectator but a participant, and the theory of knowledge must account for this from the start.

One comparison may be especially illuminating. Darwin's theory of natural selection is, among other things, a theory that applies to itself. It describes a process that produced, eventually, the minds capable of formulating it. Once formulated, the theory ramified through human practice—in medicine, agriculture, philosophy—but each of these developments is itself an instance of the process the theory describes. The theory did not escape its own domain. It became more deeply embedded in it. The present framework has the same structure. Its account of embedded classification is itself an act of embedded classification. The universe it describes is one that contains, as one of its internal structures, the theory through which it is being understood.

This does not make the theory circular in any vicious sense. It makes it self-consistent in precisely the sense the theory defines. Whether it has reached the fixed point it describes, the reader cannot determine from outside—and the theory predicts exactly this. The most the reader can do is ask whether, in following the argument, they find themselves in the situation the argument describes.

2 Introduction: Solipsism in its Place

I invert solipsism from a metaphysical claim to an epistemic constraint.

Classically, solipsism asserts that only the self can be known to exist. Properly interpreted, however, it establishes something weaker but more precise: the self has access only to its own representations.

Once this constraint is accepted, the problem of knowledge changes form. Knowledge cannot consist in direct comparison between representation and an external world, since such comparison is unavailable to the system making the judgment. Instead, knowledge becomes the refinement of representations on their own terms in light of the latent structure governing that refinement.

Because all mathematical and physical structures are encountered only through representation, the same process of representational refinement can reveal both mathematical regularities and the physical structures of the world.

We consider epistemic systems embedded within the universe they attempt to model. Such systems cannot classify the world from an external vantage point. Observations, models, classifiers, and the relations between them all exist as structures within the same universe.

The guiding principle is:

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

In this formulation no appeal to temporal evolution is required. Instead, the universe is treated as a single global relational structure whose internal relations include observations, models, classifications, and consistency relations between them.

3 The Block Universe

Let Ω be a set, interpreted as the universe.

We define the following subsets.

$$\mathcal{D} \subseteq \Omega$$

denotes the set of possible observations.

Individual observations satisfy

$$d \in \mathcal{D}.$$

We also define

$$\mathcal{W} \subseteq \Omega$$

as the set of world models and

$$\mathcal{C} \subseteq \Omega$$

as the set of classifiers.

We assume that classifiers themselves may appear as observations, so that

$$\mathcal{C} \subseteq \mathcal{D}.$$

World models represent latent explanatory structures and are not assumed to lie in the observation space.

Thus the containment structure is

$$\mathcal{C} \subseteq \mathcal{D} \subseteq \Omega, \qquad \mathcal{W} \subseteq \Omega.$$

4 World Models and Classification

Each world model $W \in \mathcal{W}$ induces a classification of observations through a function

$$\pi_W : \mathcal{D} \to \mathcal{Z}_W$$

where the representation set satisfies

$$\mathcal{Z}_W \subseteq \mathcal{D}.$$

Thus a world model compresses observations by mapping them to representative observations within the observation space itself. The set $\mathcal{Z}_W$ may be interpreted as the set of representative observations retained by the model after compression.

Remark 1. *The requirement $Z_W \subseteq D$ ensures that every abstract class induced by a model is instantiated by at least one observational state. The representative of a class induced by π_W is not assumed to be unique or privileged; it merely witnesses the existence of an observational instance of the class.*

Definition 1 (Model-Induced Equivalence Relation)**.** *For a fixed model W, define*

$$d_1 \sim_W d_2 \quad \iff \quad \pi_W(d_1) = \pi_W(d_2).$$

Definition 2 (Equivalence Class)**.** *For $d \in \mathcal{D}$,*

$$[d]_W = \{d' \in \mathcal{D} \mid \pi_W(d') = \pi_W(d)\}.$$

These equivalence classes represent the compression induced by the model.

Definition 3 (Data Modulo a World Model)**.** *Define*

$$d \bmod W := [d]_W.$$

Remark 2. *Each model W therefore induces a partition of the observation space $\mathcal{D}$. The sets that remain observationally distinguishable under the model are unions of these equivalence classes.*

5 Statistical Structure of Observations

We assume that the observation space carries a probability structure

$$(\mathcal{D}, \Sigma_{\mathcal{D}}, \mu_{\mathcal{D}})$$

where $\Sigma_{\mathcal{D}}$ is a σ-algebra on $\mathcal{D}$ and $\mu_{\mathcal{D}}$ is a probability measure.

The measure $\mu_{\mathcal{D}}$ represents the distribution of observations present in the universe.

In a block-universe interpretation this measure does not represent temporal randomness but rather the structural distribution of observational states within the universe.

Remark 3. *The statistical structure of $\mathcal{D}$ represents the latent regularities that models attempt to capture. A model proposes a partition of $\mathcal{D}$ through its equivalence classes, and these classes reveal regularities in the observational distribution $\mu_{\mathcal{D}}$.*

Remark 4 (Measurable Partition)**.** *If each map π_W is measurable and each set $\mathcal{Z}_W$ carries a σ-algebra for which singletons are measurable, then each equivalence class $[d]_W$ lies in $\Sigma_{\mathcal{D}}$. In this case the family $\{[d]_W\}$ forms a measurable partition of $(\mathcal{D}, \Sigma_{\mathcal{D}}, \mu_{\mathcal{D}})$.*

6 Classifiers

A classifier is a function

$$\pi : \mathcal{D} \to \mathcal{Z}$$

where

$$\mathcal{Z} \subseteq \mathcal{D}.$$

The set of classifiers is denoted

$$\mathcal{C}.$$

Remark 5. *Each world model induces a classifier through*

$$W \mapsto \pi_W.$$

Thus a world model may be regarded as a structure that generates a classifier.

7 Implications of Models

Models determine a canonical observational implication through a function

$$g : \mathcal{W} \to \mathcal{D}.$$

For a model W, the element

$$g(W)$$

represents the canonical observation structurally implied by the model.

8 Consistency

Definition 4 (Model Consistency)**.** *An observation d is consistent with a model W if*

$$d \in [g(W)]_W.$$

Equivalently,

$$\pi_W(d) = \pi_W(g(W)).$$

Thus the universe answers models with a binary response relative to the model:

$$\text{consistent} \qquad \text{or} \qquad \text{inconsistent.}$$

9 Model Inference

Observations generate models through a function

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

10 The Consistency Loop

The central structural loop of the framework is the pair of maps

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

For an observation d we infer the model

$$W = F(d).$$

Consistency requires

$$d \in [g(W)]_W.$$

11 Probability on Model Space

If $\mathcal{W}$ is equipped with a σ-algebra $\Sigma_{\mathcal{W}}$ and F is measurable, the observational distribution induces a probability measure on model space.

Definition 5 (Pushforward Measure)**.**

$$\nu = F_* \mu_{\mathcal{D}}$$

defined by

$$\nu(A) = \mu_{\mathcal{D}}(F^{-1}(A))$$

for measurable sets $A \subseteq \mathcal{W}$*.*

The measure ν represents how frequently observational data lead to different models.

12 Recursive Classification

Because classifiers may themselves appear as observations,

$$\pi^{(2)} : \mathcal{C} \to \mathcal{Z}_{\mathcal{C}}$$

defines higher-order classification of classification procedures.

13 Self-Consistency

Let $T : \mathcal{W} \to \mathcal{W}$ denote the composite operator

$$T = F \circ g.$$

Definition 6 (Self-Consistent Model)**.** *A model* W *is self-consistent if*

$$W = F(g(W)).$$

Equivalently, a self-consistent model is a fixed point of the operator T.

14 The Universe as a Self-Classifying Structure

All relations described above are substructures of the universe Ω. The universe therefore contains

- observations
- classifiers
- models
- consequences
- consistency relations

as internal structures.

Apparent learning corresponds to internal systems attempting to approximate the structural regularities present in the observation space.

15 Conclusion

World models classify observations through compressive mappings. Models also generate observational implications, and observations are compared with those implications through the model's own classification.

The central consistency condition is

$$d \in [g(F(d))]_{F(d)}.$$

Thus observations generate models whose implications are indistinguishable from the observations under the model's own classification.

The pair of maps F and g together constitute an interpretive and implicative structure through which observations generate models and models organize the probabilistic distribution over possible observations. One might regard this structure as analogous to what philosophical traditions have called logos: a principle through which the world becomes intelligible.

The universe may therefore be interpreted as a relational structure containing the classifiers through which it is understood, a timeless recursive procession of interpretation and action on the world.