sam@maddoxcp.com

Observer Theory and the Ruliad: An Extension to the Wolfram Model

Sam A. Senchal, Independent Researcher.

Abstract

This paper presents an extension of Observer Theory within the context of the Ruliad, using a mathematically rigorous formalisation with category theory as the unifying framework.

This paper demonstrates how the interaction between Observers and the Ruliad can give rise to hierarchical domains of reality through sampling constraints, information integration, and entropy reduction.

By formalizing the Observer as an active agent that samples and integrates information across different domains, we provide a novel approach to understanding consciousness, causation, and the nature of reality within a computational framework.

The framework reconciles discrete computational structures with apparently continuous Observer experiences, addresses the causal relationship between different domains of reality, and provides a foundation for understanding qualia through information integration within the Ruliad.

We conclude with empirical predictions and applications across physics, cognitive science, and artificial intelligence.

Reading Roadmap

Who is this paper for?

Given the multidisciplinary nature of this work, different readers will naturally gravitate to different aspects. Here, I provide a roadmap to guide readers from various backgrounds on how to approach the paper:

- For Physicists: *Focus on Sections 1 and 5.* Section 1 introduces the Ruliad and its formal definitions. Section 5 on *Multi-Level Causation* (especially 5.1 *Cross-Domain Causal Pathways* and 5.2 *Formal vs Efficient Causation*). This will be of interest, as it addresses how causal relationships can manifest across different scales and domains. Physicists may also want to skim Section 3 (Observer-Constructed Continuity) to see a formalism that reconciles discrete models with continuous spaces – connected to debates about digital physics vs. continuity. The technical details in Section 2 (category theory and domain hierarchies) can be referenced as needed, but the big picture in 1, 3, and 5 will give you the physical intuition of the framework.
- For Computer Scientists and AI Researchers: *Focus on Sections 2, 4, and 7.* Section 2 *Hierarchical Informational Domain Structure* lays out a layered architecture (P, V, S, M domains) that will feel familiar to those in AI (it parallels multi-layer networks or cognitive architectures). It's essentially the "system design" of our Observer model in computational terms. Section 4 *Information Integration and Consciousness* is key for researchers interested in AI cognition: 4.1 introduces integrated information and qualia in a formal way – connecting to machine consciousness or the integration of multi-modal data (think of $\mathbf{I}(\mathrm{F}_0)$ as how an AI might fuse vision and language, and $\mathbf{Q}(x)$ as the emergent "understanding" of a concept across its knowledge base). If you work in ML or AGI, Section 4.2 on *Observer Sampling and Persistent Structures* describes a learning process akin to how an agent discovers stable features (it parallels how AI systems find

latent variables that remain stable – effectively our Observer reduces entropy similarly to how machine learning reduces prediction error). Section 7.1-7.2 (Applications) explicitly discusses potential AI implementations and cognitive science experiments –it enumerates how one might simulate an Observer, test for integrated information correlates, or apply the framework to AI alignment (e.g. ensuring an AI's R_O is aligned with ours). The paper is heavy on category theory language, but these suggested sections contain a 'translation' into AI terms (the glossary is designed to help match terms like "Observer functor" to "sensor mapping," though it is, by its nature a rough correspondence!).

- For Mathematicians: *Focus on Sections 1.1, 2.2, and Appendix B.* Section 2.1 *The Ruliad and Category Theory* and the latter part of Section 1 (Definitions) will give the categorical foundation –the Ruliad is defined as a category and Observers as functors. That's your entry point: it sets up objects, morphisms, functoriality, and references the ∞-groupoid structure. Then jump to Section 2.2 *Recursive Categorical Nesting*, where we define functors $S_{i,j}$ between domain categories and list their properties (preservation of composition, etc.) – this is a sketch of a 2-category simplification of this idea. Appendix B (Formal Developments and Proofs) contains more rigorous propositions. If you're inclined, read Appendix B after 2.2 for the nitty-gritty details and sketched proofs that aren't in the main body. You might also find Section 6 interesting, where we introduce True Infinity (TI) as a terminal object to avoid infinite regress – it's a categorical way to cap off the hierarchy, which could spark thoughts about universality and limits in category theory.

- For Philosophers of Mind: *Focus on Sections 4 and 5, with background from 1.* We recommend starting with Section 1.2, which defines the Observer and Observer-Sampled Ruliad in concept (you can read it qualitatively, skipping equations if needed, to get the idea of "reality as Observer-dependent" and the triad of boundedness-persistence-relevance constraints on perception). Then delve into Section 4 *Information Integration and Consciousness*. This section is written to bridge to cognitive science: 4.1 discusses how consciousness emerges from integrated information across domains, explicitly referencing and extending IIT – this should align well with philosophical theories of consciousness (we provide a new twist by rooting it in a fundamental computational universe). Additionally, the definition of qualia, $\mathbf{Q}(x)$ and the discussion around it – we address the hard problem is what it "feels like" for information to be integrated across all of an Observer's ontological levels. Section 4.2 on persistent structures is essentially about how habits or stable perceptions form – a concept very relevant to philosophy of mind (e.g. how do stable mental objects arise from flux?) – we formalize "persistent mental structures" as fixed points of an entropy-reduction process. Section 5 *Multi-Level Causation* will interest those thinking about downward causation or mental causation: 5.1 shows how events in one domain (say, a decision in the mind domain) can causally influence another (the body in physical domain) through what we call constraint cascades. This is our answer to the classic mind-body causation problem – framed in category theory but conceptually accessible (we use terms like formal vs efficient cause in 5.2, directly from philosophy). By reading 4 and 5, with the foundational understanding from 1, you'll get a full picture of how our model approaches consciousness, free will, and the mind-body relationship in a computational context. Section 7.1's conceptual hypotheses might also be of interest (we propose e.g. that altering an Observer's capacity could change their perceived reality – a nod to thought experiments about expanded consciousness or altered phenomenology).

- For General Readers: If you're coming from a broad interest in "physics and consciousness" or "computation and philosophy" without a deep technical background, we suggest the following: read the Introduction, skim Section 1 to pick up the key definitions in words (don't worry if the notation is dense – the summaries at the end of each section will help). After that, you might jump to Section 4.1, which is high-level and discusses what this formalism implies for our understanding of consciousness in relatively accessible terms, and Section 5's beginning to see the big questions we tackle (causation across domains). The summaries at the end of each section are designed to be plain language explanations – make use of them to build intuition. Finally, the Conclusion and Section 7 are written in a less formal tone – those sections will solidify your understanding without the math. And remember to consult the *Glossary* whenever a term like "Observer functor" or "entropy reduction functor" appears – they are plain-English equivalents with rough correspondence to terms across different related disciplines there (note that this is, by its nature, imperfect). This roadmap should allow a non-specialists to grasp the essence of Observer Theory.
- One final point. Do I believe that computation is all there is? In short, I'm not as confident as Stephen is. I don't know, and given my constraints, I'll never know for sure. However, computation has utility. It is an abstract structure that allows Observers like us to access the largest possibility space (i.e. it contains the most possible information), given our current technological progress. In time we may invent new abstractions that enable us to talk about things that are not computational or are hypercomputational. If that is the case, and those new abstractions allow Observers like us to process more information, then it is highly likely that this formalism will be superseded by whatever that new thing is.

sam@maddoxcp.com

Introduction

The relationship between Observers and reality represents one of the most fundamental questions in both physics and philosophy.

From quantum mechanics' measurement problem to the problems of consciousness [6], the role of the Observer appears fundamental yet lacks a formal explanatory framework—especially given that the Observer is part of the very system it measures.

Recent developments in computational models of physics, particularly the concept of the Ruliad as developed by Wolfram [1] and formalized by Gorard et al [2, 3], offer a promising framework for addressing these questions.

This paper builds upon these foundations to extend the ideas of Wolfram [14], Rickles, Elshatlawy and Arsiwalla [3] including their recent Minimal Observer Model [15], which can be thought of as a potential mechanic for the categorical structures described in this paper.

It proposes that Observers sample regions of the Ruliad based on computational constraints, creating hierarchical sampling domains that exhibit distinct informational properties, determined by the rules that bound each informational domain (i.e. these rules must be 'on').

Through this lens, I aim to address longstanding questions about the emergence of causality and the nature of consciousness given Wolfram's position of a computational universe.

A central aim of this work is to establish a formalism that incorporates what we describe as 'value' and 'meaning' into our understanding of reality.

By grounding subjective experience, purpose, and value within category theory, we seek to bridge the divide between objective description and subjective meaning that has characterized much of modern science.

This approach offers a potential unification of the "two cultures" - the scientific pursuit of objective description and the humanistic pursuit of meaning and value.

Our approach is distinguished by several key innovations:

1. A category-theoretic formalization of the Observer sampling processes
2. A hierarchical model of the observable information domains in the Ruliad, with a recursive nesting of those domains
3. An Information Integration framework / function for the Observer that gives an initial explanation for the emergence of qualia
4. A potential resolution to the causality problem
5. A mathematical account of how discrete computational processes could give rise to apparently continuous Observer experiences

1. Background

1.1 The Ruliad and Category Theory

Definition 1 *Ruliad:* A meta-structural domain that encompasses every possible rule-based system, or computational eventuality, that can describe any universe or mathematical structure. It acts as a theoretical space wherein the boundaries between map and territory blur, functioning as the ground for the possibility of multi-computation. Within the Ruliad, every conceivable physical and mathematical system can be situated, but their accessibility or

sam@maddoxcp.com

meaningfulness is determined by the specific Observer-related frames or constructs. The Ruliad is thus a pre-physical framework, and its utilization in physics is to pinpoint the exact rule-based system that corresponds to our observed reality.

A computational definition is as follows:

Let $\boldsymbol{R}$ be the space of all possible computational (or rewriting) rules. We refer to $\boldsymbol{R}$ as the "Rulial space". Consider $r \subseteq \boldsymbol{R}$. r may represent either a single rule or a collection of rules capable of generating a computational universe $\mathbf{U}_r$ (the outcome of all possible computations following a given rule set). The Ruliad, denoted by $\boldsymbol{R}$, is the collection of all computational universes.

That is, $\boldsymbol{R} = \{ \mathbf{U}_r \mid \mathbf{r} \in \boldsymbol{R} \}$.

Furthermore, an Observer, **O** can interpret or access a subset of $\boldsymbol{R}$ based on their specific observational constraints. The observability of any $\mathbf{U}_r$ is contingent on the Observer's **computational boundedness, persistence** and their view of the **relevance** of the rules they sample. In this paper we formally define those subsets of $\boldsymbol{R}$, *where* $\boldsymbol{R_0}$ is the limit of what **O** can sample and $\mathbf{F}_0$ is the part of $\mathbf{R}_0$ they are actively sampling now.

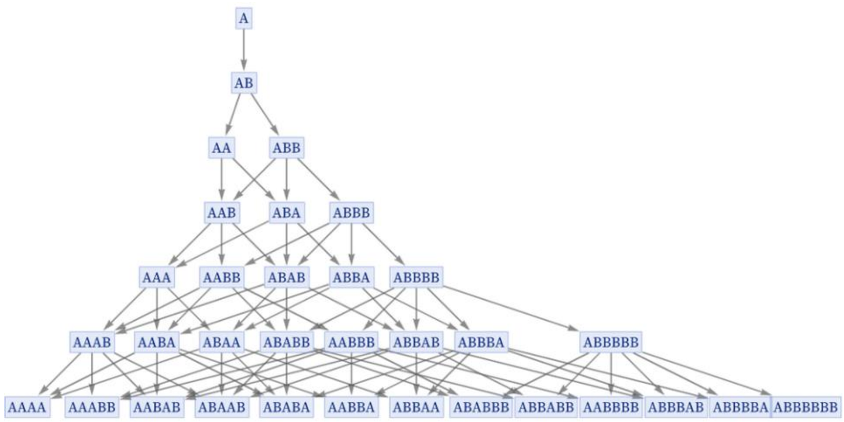
Stated simply, the Ruliad represents the entangled limit of all possible computational processes [1]. It can be formalized as a category $\boldsymbol{R}$ whose:

- Objects are computational states
- Morphisms are computational transitions
- Composition reflects causal structure

In Gorard's formalization [2], the Ruliad is modelled as an **∞-groupoid**, representing the limiting structure of multiway rewriting systems with all possible rules.

Definition 2 *Ruliad (Category Theoretic)*: The Ruliad $\boldsymbol{R}$ is a category whose objects are computational states and whose morphisms are computational transitions generated by all possible rules, with $\boldsymbol{R}$ forming an **∞-groupoid** in the limit.

Figure 1 *Visualisation of Multiway State Graph*

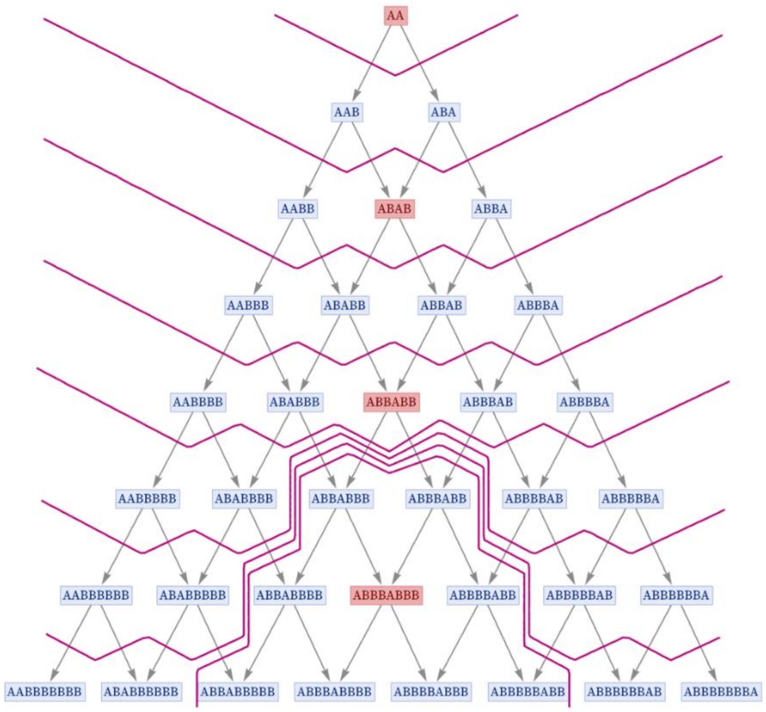


This figure provides a baseline multiway system visualization, depicting branching computational histories without any particular Observer filtering. It shows the state-space evolution for a specific set of rules (A → AB and BB → A in this example). The top node (A) evolves into two distinct states (AA and ABB) at the next step (blue and red arrows indicate which rule applied), demonstrating a branch in possible outcomes. Importantly, some branches later merge back (for instance, different sequences leading to the same ABABB state). This merging corresponds to different histories yielding an identical configuration, a common feature in multiway systems. The multiway graph thus encodes all possible paths of evolution: each unique path from the root to a given node represents one possible history. In the full Ruliad (with all possible rules), such a graph becomes vastly complex – effectively an "entangled" network of every computation. This diagram shows how the Ruliad's structure contains branching (multiple futures from one state) and merging (identical outcomes via different routes), lay the groundwork for concepts like causal invariance and Observer-independent structure.

© Sam A. Senchal, May 2025
sam@maddoxcp.com

1.2 Observer and Observer-Sampled Ruliad

Figure 2 *Observer Sampling in a Multiway System*



This figure illustrates how an Observer picks out a subset of the Ruliad. The full structure (gray nodes and arrows) is a multiway system – a directed acyclic graph of all possible state evolutions from a given starting condition.

The Observer (conceptually represented by the red path) samples one path through this graph, effectively following a specific sequence of states and treating alternative branches as un-realized possibilities.

The magenta curves slicing across the graph depict "moments" of the Observer's foliation (akin to consistent time steps), and the red nodes (e.g. AA → AAB → ... → ABBABB) form the unique trajectory that the Observer considers "actual."

In other words, out of the vast entangled network of computational histories, the Observer's internal rules select a single consistent history. This visual emphasizes how subjective reality can emerge from the Ruliad's objective multitude: by sampling only one branch, an Observer sees a definite sequence of events (analogous to a single outcome in the Many-Worlds interpretation of quantum mechanics [13]). All other potential branches are filtered out by the Observer's constraints, remaining inaccessible "what-ifs."

Definition 3 *Observer*[1]: An Observer **O** is defined as an entity that samples the Ruliad, $\boldsymbol{R}$ by means of a functor. We denote this functor by $\boldsymbol{S_O}$: $\boldsymbol{R} \to \boldsymbol{R_O}$. Here $\boldsymbol{R}$ is the full Ruliad category and $\boldsymbol{R_O}$ is the image (accessible subset) of the functor, i.e. the part of the Ruliad, $\boldsymbol{R}$ accessible to Observer, **O**. Formally, $\boldsymbol{R_O} = \mathrm{Limit}(S_O)$

In plain language, the Observer, **O** has certain internal rules or limitations that restrict the portion of Ruliad it can sample. The sampling functor $\boldsymbol{S_O}$ formalizes this by mapping each state and process in the Ruliad to the corresponding state/process as experienced by **O**. This functor "forgets" or compresses information that **O** cannot detect, and it coarse-grains many rule states into what **O** considers the same state. The result is that **O** perceives an Observer-dependent subset $\boldsymbol{R_O} \subset \boldsymbol{R}$ as its reality.

[1] Note that we maintain Gorard's [2] formal distinction that the Ruliad, R, is itself, Observer agnostic, whereas any concrete description of physics implicitly assumes an Observer (through the sampling functor). Thus, $\boldsymbol{S_O}$, encapsulates those assumptions, making our Observer-dependence explicit. In the notation we use $R_O = S_O(R)$, to denote the Observable Ruliad.

Critically, **all** Observers ultimately sample the Ruliad, but each samples it in their own unique way (for example, a human Observer samples the Ruliad in a way that yields our familiar physics; an alien or AI might sample a different slice, but any consistent Observer still picks out some subset $\boldsymbol{R_O}$ of the Ruliad).

The Observer-accessible subset of the Ruliad, $\boldsymbol{R_O}$ is constrained by:

1. **Definition 3a** *Boundedness* $\boldsymbol{B(x) > \beta}$
   $\boldsymbol{x}$ must lie within $\boldsymbol{O}$'s bounded computational capacity (i.e. not too large or detailed to measure). *Where* $\boldsymbol{\beta}$ *represents some minimal threshold required for the Observer to measure or sample* $\boldsymbol{x}$
2. **Definition 3b** *Persistence* $\boldsymbol{P(x) > \gamma}$
   $\boldsymbol{x}$ must persist for a minimum number of hypergraph updates such that $\boldsymbol{O}$ can sample it. $\boldsymbol{\gamma}$ is a threshold ensuring that ephemeral states are filtered out of the Observer's sampling
3. **Definition 3c** Relevance rules $\boldsymbol{R(x)}$ = true
   Where $\boldsymbol{R(x)}$ represents a Boolean predicate, that is true and only true if $\boldsymbol{x}$ carries information that is useful or meaningful to **O**. It encodes the ideas that on Observer only notices patterns that are related to its function of integrating information about $\boldsymbol{R_O}$

Definition 4 *Observer Field*: Not everything in $\boldsymbol{R_O}$ is observed simultaneously – an Observer has a further restriction to what it is actively sampling at a given moment or context. We define the **Field of Observation,** $\boldsymbol{F_O}$ as the subset of $\boldsymbol{R_O}$ that the Observer is 'focusing on' or 'integrating information about' over at a given number of hypergraph updates:

$$F_O = \{x \in R_O \mid B(x) > \beta, P(x) > \gamma, R(x) = \text{true}\}$$

In simpler terms, F_O is what the Observer actually perceives or pays attention to. For instance, your visual system's F_O includes the features of the scene you're currently looking at (within your capacity B, focusing on persistent objects P, and relevant stimuli R like movement or familiar faces).

Mathematically, F_O is analogous to a "filtered output" of R_O

1.3 Information Structure and Entropy

Definition 5 *Information Content*: The informational content of a state x is defined as:

$$\phi(x) = H(x) - H_{\text{red}}(x)$$

Where:

- $H(x)$: Initial entropy (in the Shannon [5] / Algorithmic Complexity sense) associated with state x
- $H_{\text{red}}(x)$: Entropy reduced by Observer interaction (the computationally reducible part of the information content. This represents the portion of x that is predictable, or already accounted for by simpler components (e.g. correlations / regularities that O already captures / has already reduced computationally to something predictable)
- Thus $\phi(x)$ measures **net new information content.** This notion is analogous to surprise or novelty for an Observer **O**, and it aligns with Friston's Free-Energy Principle in neuroscience [11] (though our formulation generalizes this beyond an anthropomorphic case)

- If $\phi(x) = 0$ the Observer gains no new information from its sampling of R_0. By contrast, if $\phi(x)$ is significantly larger than average (relative to a typical result from a foliation of R_0), then the state x provides a great deal of novel information to the Observer.
- This aligns closely with prior work on Information theory, specifically the idea of Bayesian Brains in cognitive science (though this is, again, a generalisation of that idea)

This is the first extension of the Wolfram Physics Project formalism for Observers.

We posit that entropy, as defined here, for an Observer, corresponds to the concept of *computational irreducibility* within the Ruliad formalism.

We posit that Observers are aiming to integrate information. In practice this means that rule-based Observation within the Ruliad must decrease uncertainty, creating pockets of computational reducibility, here a reduction of Informational Entropy within $\boldsymbol{R_0}$.

Our conjecture is that all Observers are teleological, they have a 'target' or 'goal' within the Ruliad, namely, to reduce entropy (or create pockets of computational reducibility) by integrating more information, so that less computational resources are utilised in exploring their accessible part of Rulial space, $\boldsymbol{F_0}$ to the limit of $\boldsymbol{R_0}$**.**

Or, in other words, to 'complete' that function in as few computational steps as is possible, given their constraints.

Definition 6 *Information Integration*: For a set of states (e.g. the Observer's field, $\boldsymbol{F_0}$), we define the integrated information content over those states as:

$I(F_0) = \sum_{x \in F_0} \phi(x) = \sum_{x \in F_0} (H(x) - H_{\mathrm{red}}(x))$

In other words, $I(F_0)$ is the total useful information an Observer has integrated from its accessible field of observation, $\boldsymbol{F_0}$ **within** $\boldsymbol{R_0}$, its ultimate observable limit at state / universal time, **t**

This function takes the field of Observation, F_0, within R_0 and determines how much initial entropy (or alternatively some measure of computational irreducibility) is in that 'area' of R_0 and then samples different rules accessible to it (morphisms between states / categories) to reduce entropy and increase computational reducibility in that part of F_0

This metric directly parallels concepts from the Integrated Information Theory (IIT) of consciousness but generalises them to any class of Observer within the Ruliad. This generalisation outputs the quantum of computationally irreducible information in the system's state or the degree to which information, accessible in $\boldsymbol{F_0}$, is not reducible to independent parts.

A high $I(F_0)$ result suggests that the Observer is sampling a lot of information across each accessible informational domain (see section 3). We posit that this type of result would correspond, anthropomorphically, to a conscious state analogous to human experience. A low $I(F_0)$ result means that the information is fragmented and is harder for the Observer to derive utility from.

This quantitative notion measures how much coherent and persistent structure the Observer is extracting from the Ruliad.

Summary
Section 1 introduces the conceptual backbone of our framework. The Ruliad as the "entangled limit of all computations," a sort of ultimate mathematical cosmos containing every possible universe. We then formalized what an Observer is in this context: essentially a filter or functor that selects (via something called foliation) a sliver of that cosmos to call its reality. We broke down the Observer's constraints into

boundedness (limited computational capacity), persistence (needing stable patterns), and relevance (focusing on what matters).

These constraints determine which events in the Ruliad the Observer can actually experience. Think of the Observer as both a sculptor and a spectator – carving out a personal world from the grand block of marble that is the Ruliad.

By the end of Section 1, we have a toolkit of definitions: the Ruliad (all possible states and rules), R_O (the Observer's accessible portion), F_O (what the Observer actively sees at a given time), and measures of information and integration (ϕ, I, Q) to quantify experience. These form the language we'll use to discuss everything from physics to consciousness in the upcoming sections. Essentially, we now have a common formal ground to talk about how "the world of an Observer" emerges from "the world of all possible worlds."

2. Hierarchical Informational Domain Structure

2.1 The Four Fundamental Domains

We propose that the Observer-sampled Ruliad, $\boldsymbol{R_O}$, is composed of four stratified, fundamental domains characterized by different degrees of informational constraint.

Though it is, by its nature, speculative, we focus on four domains for concreteness (drawing from our own cognitive layers), the framework, in principle, could accommodate a different domain decomposition appropriate to other classes of Observers.

Here, informational constraints are the number of rules that the Observer has to sample within that domain to produce a reduction in the entropy (as defined above) within that system to make a stable Observation such that $\phi(x)$ is invariant under each hypergraph update.

When $\phi(x)$ is invariant, the Observer needs to use less of its computationally bounded resources, $\boldsymbol{B(x)}$, leaving more of its computational resources to explore new morphisms (or sample new rules) between states.

1. **Physically Constrained Information (P)**: These are states that are maximally constrained by physical laws e.g. strict locality, causal invariance, conservation laws etc.
    a. This corresponds to our physical universe: information patterns here require many ***specific*** rules for an Observer to maintain invariance and persistence through time. Consequently, they are always relevant (as defined above)
    b. This domain can be thought of as classical physical reality

2. **Valuationally Constrained Information (V)**: These are states with strong path dependence, or required causal order, organised around goals (or attractors). This involves information shaped by value, utility or biological drivers.
    a. As a qualitative example, this could include emotional states or beliefs – patterns that persist because they're reinforced by feedback loops (attractors).
    b. Fewer constraints are absolute compared to **P**, but there are still constraints that mean certain states cannot be simultaneously experienced / integrated by the Observer (these could, for Observers like us, include contradictory rules that cannot be experienced simultaneously within that informational domain, e.g. total despair and total elation, in their extremes are structurally and affectively incompatible).

3. **Symbolically Constrained Information (S)**: States composed of symbolic structures (language, abstract concepts, mathematics) governed by syntactic or logical rules. Information here is structured into discrete symbols or tokens with combinatorial rules (e.g. a proof in math, or a sentence in English at different orders of logical construction). The constraints are the symbolic grammar or logic that keep the patterns meaningful (relevant) to the Observer.

4. **Minimally Constrained Information (M):** Information patterns with minimal local constraints (i.e. fewest rules required for invariance), allowing non-local connections and high degrees of freedom.
    a. One could, anthropomorphically, analogise this to 'Supra-verbal imagination' or highly theoretical structures where almost anything goes unless ruled out by laws of consistency or identity.
    b. Patterns in this domain can have long-range entanglements in the Ruliad
    c. Necessarily, information that cannot be integrated in the domains P, V and S could be integrated within this domain.

Formally, each domain is defined as:

Definition 7 *Information Domains*: For each domain $D_i \in \{M, S, V, P\}$:

$D_i = \{x \in R_O \mid C_i(x) > \tau_i\}$

Where:

1. C_i: Constraint function for domain i
2. τ_i: Threshold of rules sampled for domain i

Here $C_i(x)$ measures how much rule constraint (of the type for the informational domain **i***)* applies to state x, and $\boldsymbol{\tau_i}$ is the boundary condition for that domain, **i**

For example, if **i = P**

$C_p(x)$ could quantify how strictly physical laws (i.e. the world of physics) determines x (where a high τ_p result means only states subject to many Physical rules qualify as belonging to domain P)

In contrast, $C_M(x)$ might be very low for most results of x, since this informational domain accepts states with only a few rules enforcing Observational consistency.

Thus, each domain consists of states requiring at least a certain level of structured constraint (regularised rule sampling) to be stable or meaningful in that domain.

Conceptually, **M requires the Observer sampling the fewest rules for consistency and identity (highest freedom) and P the most rules**, with S and V as the intermediate domains.

Note these intermediate domains are speculative, but structurally, this hierarchy would apply to any domain distinction a researcher would choose, subject to domains M and P always being the top and bottom of the hierarchy.

This speculative position is inspired by anthropomorphic informational **hierarchy**: physical sensations (P), emotional (V), symbolic / logic (S), and generalized sense of identity / super consciousness / archetypes (M).

We will use these domain distinctions throughout the paper to discuss how Observers operate. Though a different type of Observer may actively sample different intermediate domains based on the constraints of that Observer and it's 'position' in the Ruliad (though this could be adapted and generalised to any type of 'possibility space').

2.2 Recursive Categorical Nesting

Definition 8 *Nested Domains*: Although we distinguish between P, V, S & M for analytic clarity, in this model these domains are assumed to overlap and embed within each other.

For instance, a piece of information from the physical domain (P) can carry symbolic meaning (S) — i.e. it is expressible in words or numbers, a lower informational 'resolution.' It can also impact value-oriented goals in V. For example, Alice needs energy to maintain herself as an Observer, so she eats food (extracting relevant information from P). This produces a feeling of satisfaction (a value-laden state in V), which she can then describe in words or numbers (using S):

Formally, we utilise functorial embeddings between domains:

For any two domains D_i and D_j there is an embedding functor $\mathbf{S}_{i,j}$, mapping the states in D_j, to D_i.

Where $S_{i,j}$ is the embedding functor:

$S_{i,j}: D_j \rightarrow D_i$

Specifically, $S_{i,j}$:

1. Maps objects (states) from domain j to corresponding states in domain i
2. Maps morphisms (transitions) from domain j to corresponding transitions in domain i
3. Preserves composition of morphisms and identity morphisms
4. Adapts the mapped structures to conform to the constraint profile of domain i

We define nested domains as follows:

$D_i^{\text{nested}} = D_i \cup \{S_{i,j}(x) \mid x \in D_j, j \neq i\}$

Linguistically, this function D_i^{nested} contains domain *i*, itself, and the images of all other domains *j* embedded in *i.*

This can be understood with qualitative examples that utilise the domains we posit in this paper.

- The embedding functor $S_{P,S}$ embeds symbolic structure into a physical state i.e. encoding a theorem (an S domain information structure) into a configuration of atoms (a P domain information structure)
- Similarly, reversing this, the embedding functor, $S_{S,P}$, could be interpreted as an Observer extracting symbolic value from a physical pattern (seeing shapes in a Rorschach blot test or seeing faces / shapes in clouds during a psychedelic experience)

These functors are defined to preserve structure. They map morphisms across informational domains that preserve compositions and identity under Gorard's and Arsiwalla's n-fold categorical structure of the Ruliad. This ensures that the categorical structures and relationships remain consistent across the informational domains.

This enables the domains to form a single coherent categorical framework that is consistent with the infinite-groupoid structure implicit in a Wolfram multi-way hypergraph.

This multi-domain structure is a fibered category or a lattice of subcategories within the Ruliad, rather than unrelated disjoint domains that break the implied Grothendieck topology implicit in the infinite groupoid structure.

This recursive embedding structure explains phenomena that can span multiple domains, such as embodied cognition, psychosomatic effects, and symbolic representations of physical processes.

Summary
Section 2 revealed a hierarchical picture of Observer-accessible reality. We saw that an Observer's world R_0 isn't monolithic – it can be stratified into different domains of information (we posit four; Physical, Valuational, Symbolic, Minimal), each with its own degree of constraint. This gives a structured way to think about "levels of description": for example, physical laws operate in domain P, while thoughts and meanings operate in domain S or M.

Crucially, we learned that these domains are nested and interconnected – none operates in isolation. Using functors $S_{i,j}$ we can map phenomena from one domain into another (e.g. interpret a physical event in symbolic terms). The hierarchy is like a layered map of reality, with each layer embedding into the next. This formalism explains how something like a mind can influence a body: they are different domains in the same structured reality and there are translation maps between them.

For the reader, the key takeaway is that our framework doesn't see physics, psychology, etc., as separate realms but as different slices of one information structure. By organizing R_0 into a lattice of domains, we can talk about cross-domain interactions rigorously in Section 6.

Section 2 has set up an ontology – a "world of the Observer" divided into layers but tied by mathematics – which we will use to tackle complex phenomena (like embodied cognition or symbolic reasoning) in a unified way.

If the definitions of functors and category nesting were a bit heavy, remember this image: reality as experienced by an Observer is like an onion – layered, but with all layers ultimately connected at the core.

3. Observer-Constructed Continuity

3.1 Discrete Sampling and Continuous Perception

The Ruliad and its evolutions through time are fundamentally discrete as they are made of distinct states of the Wolfram Hypergraph and update events on said Hypergraph.

An Observer's qualitative experience appear continuous at macroscopic scales (the flow of time and smooth motion are two obvious manifestations). This seems contradictory to the Ruliad's discrete nature and highlights a tension with theories like General Relativity, which assume a continuous space time.

Observer Theory resolves this apparent contradiction by positing that continuity is an emergent construct of the Observer (they foliate / sample the Ruliad in a 'coarse grained' fashion).

We define a **continuous Observer field** (F_0^{cont}) as the result of mathematically integrating an Observer's discrete samples of F_0. When an Observer samples the Ruliad at a coarse-grained resolution, it creates the *impression* of continuity.

In other words, by focusing only on broad, persistent patterns (and ignoring fine, momentary details), the Observer perceives a smooth experience. This emergent continuity has practical relevance for the Observer, as it aids in entropy reduction and information integration within R_0:

sam@maddoxcp.com

Definition 9 *Observer-Constructed Continuity*: The continuous perception constructed by an Observer is:

$$F_O^{\text{cont}} = \int_{x \in F_O} \phi(x)\, d\mu(x)$$

Where:

- $\phi(x)$: Information content at state x

- $\mu(x)$: Result of Observer-specific measure function for integrating discrete samples. Here $\mu(x)$ is the Observer-specific persistence bias (formally defined in Definition 10 below).

4.2 Observer-Dependent Measure Spaces

Each class of Observers comes equipped with a measure, μ_θ which formalises how the Observer "weights" information both from different domains and within the domain they are measuring. For example, a human may implicitly assign more weight to visual information at the centre of their gaze than the periphery, devoting more of their bounded computational resources to that point in a distribution.

Definition 10 *Observer-Dependent Measure*: For each Observer type θ:

$$\mu_\theta: F_O \to \mathbb{R}+$$

This function defines a unique measure that determines how discrete samples are integrated into continuous perception.

These Observer-specific measures constitute functors from the category of discrete Ruliad samples to the category of continuous Observer experiences, varying in how they interpolate between the discrete sampling points and which relationships they have the ability to preserve during the mapping process.

This measure can also be extended through time, where the Observer averages across all their observations of a given measurement, (discounting past observations with some discounting factor).

In this formalism μ_θ is used to define F_O^{cont} to parameterise the Observer's ranking of different informational states (colloquially akin to the Observer's sensitivity both within and between measurement types, or the set of possible measurements they can make when they sample or foliate the Wolfram Hypergraph within F_O).

We do not specify μ_θ for generality, as different types of Observers and different observational regimes within the same class of Observers (conjectured to be derived, within their class, by their causally invariant histories) can drastically alter the 'shape' of the Observer's perceived reality, even if their accessible subset of the Ruliad, $\boldsymbol{R_O}$ is equivalent. Mathematically we can posit that this space, $\boldsymbol{(R_O, \mu_\theta)}$, is a kind of measure space, which would allow the use of tools from Measure Theory and probability theory to be applied to Observer Theory.

This aligns with Gorard's work and develops a formal model of the foliation he described. It provides a framework for how an Observer generates their view of 'reality' in the Ruliad.

This measure μ_θ, is a proposal for a more fine-grained specifier of that sampling dynamic and has explanatory value in the context of Observer's with equivalent computational boundedness and persistence, in a sense, 'seeing things differently', based on their differing causal graphs through the Wolfram Hypergraph update events.

Summary

Section 3 addressed the apparent paradox between a discrete underlying reality (the Ruliad's bits and branches) and the continuous world we seem to observe. We introduced the idea that continuity is essentially an illusion created by the Observer's processing. By integrating discrete events (using the measure μ_θ and constructing F_0^{cont}), an Observer fills in gaps and produces a smooth narrative – much like our brain turns flickering still images into the continuous motion of a movie.

Analogies to make this intuitive. The formal definition is abstract, but its meaning is simple: it's the Observer's continuous approximation of a chunk of the Ruliad. The important consequence is that what we call "space" or "spacetime" might not be fundamental, but emergent from Observer sampling.

Thus, Section 3 closes the gap between discreet hypergraph updates and our continuous experience of it, suggesting that continuity, differentiability, and even geometry could come from Observer-mediated stitching of discrete elements.

This aligns with Wolfram's idea around 'emes', or 'atoms of space' that are not yet observed in experiments.

For the rest of the paper, this means we are free to use continuous mathematics (calculus, measure, etc.) for the Observer's perspective, knowing it's grounded in an underlying discrete sampling.

sam@maddoxcp.com

4. Information Integration and Consciousness

4.1 Qualia and Integrated Information

Finally, we provide a conjecture on **qualia,** the subjective quality of a given experience. Hoffman uses the analogy of the "experience of the colour red" to illustrate subjective experiences beyond mere neural firings [9].

This formalism utilises the multi-domain structure described in section 3 to connect these ideas to the extended Observer Theory described in this paper.

This approach generalizes Hoffman's **Multimodal User Interface (MUI)** theory [9] and the ideas from Seth's and Tononi's **Integrated Information Theory (IIT)** [10][4] across all classes of possible Observers, treating them as special cases of a broader computational Observer model.

Here, Qualia is associated with an integration information pattern across all the domains detailed in Section 3.

We have attempted an early formalisation of this process through the category theory formalism that underpins the Wolfram Hypergraph Model:

Definition 11 *Qualia*: For an information pattern x, we define its associated qualia $\mathrm{Q}(x)$ as the integrated information across all domains:

$$\mathrm{Q}(x) = \sum_{d \in \{P,V,S,M\}} \mathrm{I}(x, \mathrm{d})$$

Here $\mathrm{I}(x, \mathrm{d})$ denotes the informational content of x when projected into domain d, and the sum (or integral) runs over the set of all domains (in this model, Physical, P, Valuational, V, Symbolic, S and Minimally Constrained, M).

Essentially, we take x, which itself is a composite state spanning multiple domains and evaluate how much information x is present in each domain. For each pair of domains $(\mathrm{d}_i, \mathrm{d}_j)$, we can define an integration functor;

$$\mathrm{I}_{ij}: \mathrm{d}_i \times \mathrm{d}_j \rightarrow \mathrm{Int}(\mathrm{d}_i, \mathrm{d}_j)$$

Which merge two domain specific measures into an integrated description. By composing all such functors for all domain pairs (in this conjectured model, all possible pairings of P, S, V and M), we get a comprehensive integration $\mathrm{Q}(x)$ that spans the entire sets of domains, or more generally defines the limit of the R_0, the Observer's sampleable part of the larger Ruliad, R.

Qualia emerge through the composition of these integration functors across all domains:

$$\mathrm{Q} = \mathrm{I}_{MP} \circ \mathrm{I}_{MS} \circ \mathrm{I}_{MV} \circ \mathrm{I}_{SP} \circ \mathrm{I}_{SV} \circ \mathrm{I}_{VP}$$

Note that the sequential order here is shown for clarity, we do not prejudice a certain order over another, at this stage, though an idea for future work, is to determine whether there is a strict ordering required in this integration function to output recognisable Qualia to Observer's like us.

We posit that if:

$$\mathrm{Q}(x) > \Psi$$

Where Ψ is some threshold, above which we associate that sensation with the linguistic term, Qualia. We posit that threshold's exist for a range of conscious experiences across all possible classes of Observer.

In simple terms, $\mathrm{Q}(x)$ measures how widely x is 'spread' across the informational domains and how much the Observer had sampled and integrated of x across the domains.

For example, anthropomorphically, if an experience involves a simultaneous measurement of all four domains, it is likely to constitute something that we understand as Qualia. Conversely, if information is confined to a single domain (say purely physical, a fundamental force acting on a particle) $\mathrm{Q}(x)$ would be low, meaning that while it is still technically a measured Observation under this formalism, it would not be recognisable as anthropomorphic qualia.

Using the abstraction of the Ruliad, this definition generalises the ideas of IIT.

Qualia are what we get when an Observer maps an information state through all of the domain lenses utilising all it's computational resources to tie it to an irreducible, indivisible whole.

4.2 Observer Sampling and Persistent Structures

Observers reduce entropy (or generate 'pockets' of computational reducibility) through interaction with the Ruliad, sampling persistent structures:

Definition 12 *Persistent Structure*: A state $x \in R_O$ is persistent when:

$H_{\text{red}}(x) = H(x)$

At this point, x becomes invariant under repeated sampling, enabling Observers to build stable models of reality. In category-theoretic terms, persistent structures form fixed points under the entropy reduction functor:

ER: $R_O \to R_O$

where ER maps states to their entropy-reduced, maximally computational reduced forms through Observer measurement.

5.3 Growth through Morphism Discovery

As Observers identify persistent structures, they unlock new morphisms in the categorical structure of the Ruliad:

Definition 13 *Unlocked Morphisms*: For a persistent structure x:

$\mathrm{Morphisms}(x) = \{\gamma\colon x \to \mathrm{y} \mid H_{\text{red}}(\gamma) \to 0, \mathrm{y} \in R_O\}$

These new morphisms expand the Observer's field, $\mathrm{F_O}$:

$F_O \to F_O \cup \bigcup \{x \in F_O\}\ \mathit{Morphisms}(x)$

The discovery of morphisms corresponds to the process of learning and understanding, where the Observer, **O**, gains access to new relationships between states in the Ruliad.

These newly discovered morphisms form a subcategory of the Ruliad that becomes accessible to the Observer through entropy reduction and information integration over Hypergraph updates and changes between each foliation / functorial sampling an Observer makes of the Ruliad.

Summary
Section 4 took us to the heart of how conscious experience might arise in our framework. We proposed that consciousness is tied to the integration of information across the different domains of an Observer's reality. This is not a new idea and reflects a generalisation of Hoffman, Seth's and Tononi's work. However, we formalise it in a general schema. By defining measures like $I(F_O)$ (total integrated information in the Observer's field) and the qualia function $Q(x)$, we gave mathematical form to ideas like "the whole is more than the sum of its parts."

We argued that when an Observer integrates physical sensations, with emotional context, with symbolic interpretation, etc., into one unified state, that's when a qualitative experience "lights up." This takes ideas from IIT and grounds them in the multi-domain Ruliad setting – it's not just the brain integrating, it's the Observer's entire interface with reality integrating.

We also introduced Persistent Structures – essentially the idea that Observers create stable patterns (like concepts, memories, or physical predictions) by repeatedly encountering similar information until it's fully expected (entropy-reduced / computationally maximally irreducible).

This explains learning and the formation of stable perceptions: e.g. you learn a new word by hearing it in context enough times that it stops surprising you and becomes a persistent piece of your reality. In our formal terms, it becomes a fixed point under the entropy reduction functor. Finally, we saw that as Observers discover persistent structures, they effectively unlock new transformations (morphisms) in their world, this aligns with historic work by the inventors of the Ruliad Categorical Theoretic formalism – it is akin to how scientific discovery opens up new possibilities of interaction (once you understand electromagnetism, you can build a radio; in our terms, you've unlocked a new morphism in the Ruliad that was always there but now is in your F_O).

The takeaway is that consciousness and knowledge are not static; they grow as the Observer actively shape F_O by integrating information.

Section 4 bridges the gap between a formal information measure and the lived experience of an Observer. It suggests that what an Observer deems "real" and "meaningful" is exactly those patterns that have been woven into a holistic web (high Q), and it explains growth of understanding as expansion of the Observer's accessible category. Armed with this, we can tackle how cause and effect work in such a multi-level Observer-driven reality (next, in Section 6).

5. Multi-Level Causation

5.1 Cross-Domain Causal Pathways

Our framework resolves the apparent paradox of cross-domain causation through the concept of constraint cascades in the categorical structure of the Ruliad:

Definition 14 *Cross-Domain Causation*: For a state t in domain D_i and a state p in domain D_j:

$\text{Causal}(t, p) = \exists$ path γ: $t \to p$ in R_O

The causal path $\gamma = (t, s_1, s_2, ..., s_n, p)$ represents a sequence of constraint changes that connect the domains.

This causal structure can be formalized categorically through a process category, **Process(R_O)**, whose:

- Objects are states in R_O
- Morphisms are processes connecting these states
- Composition represents sequential causal influence

Cross-domain causal pathways then correspond to morphisms in Process(R_O) that span multiple domains:

$\gamma \in \text{Hom}_{\text{Process}}(R_O)(t, p)$ where $t \in D_i$, $p \in D_j$, $i \neq j$

5.2 Formal vs. Efficient Causation

This framework distinguishes between different types of causal influence within the categorical structure:

- Formal Causation: Represented by morphisms from less constrained (i.e. fewer rules) to more constrained domains (i.e. more rules), where higher domains shape the possibility space of lower domains through constraint imposition
- Efficient Causation: Represented by morphisms from more constrained to less constrained domains, where lower domains provide feedback that updates higher-level structures

Mathematically, these causal types can be expressed as subsets of morphisms in Process(R_0):

$\text{Formal}_{\text{Cause}} = \{\gamma \in \text{Hom}_{\text{Process}}(R_0)(t, p) \mid \text{C}(t) < \text{C}(p)\}$

$\text{Efficient}_{\text{Cause}} = \{\gamma \in \text{Hom}_{\text{Process}}(R_0)(t, p) \mid \text{C}(t) > \text{C}(p)\}$

Where $\text{C}(x)$ denotes the constraint level of state x.

This categorical formulation of causation aims to explain how thoughts can influence physical reality (and vice versa) without violating physical causal closure, as all causal morphisms exist within the unified structure of the Ruliad.

Summary
Section 5 dealt with causation when events span multiple domains.

We introduced the concept of constraint cascades to explain cross-domain causation. In plain terms, a cause in one domain (say a mental decision) can manifest as an effect in another domain (a physical action) because the domains are interlinked by the Observer's integrated structure. Formalizing this, we defined cross-domain causation in terms of relationships within the Ruliad's categorical structure – essentially showing that a morphism (cause → effect) might involve a chain that goes through different domains but ultimately commutes in the overall category.

We differentiated formal causation (the upstream informational structures enabling something – like the design of a circuit enabling an electrical event) and efficient causation (the direct trigger – like the actual current flow causing a light to turn on). In our framework, formal causes often live in higher domains (the plan, the pattern, the intention) and efficient causes in lower (the material execution), yet both are connected through functors.

The key insight from this section is that the age-old philosophical riddle of mind-body causation or top-down causation finds a natural place here: since all domains are part of one structure, a cause can loop through domains without violating any laws. Observers just perceive it differently at each level. For example, kicking a ball involves mental domain (decision), symbolic domain (calculating the kick), physical domain (leg motion) all in one causal sequence – our model can accommodate that by tracing the morphisms through M to S to P functorial mappings.

The practical upshot is that nothing "mystical" is needed to explain mind influencing matter – it's built into the multiway causal graph of the Ruliad (given the right Observer mapping).

We also noted that from an external view, an Observer imposing their will is like selecting paths in the Ruliad that align with their internal goals – consistent with how we earlier said Observers create pockets of reducibility (order) aligned with what they find relevant.

Section 6 thus provides a satisfying resolution to multi-level causation: every domain speaks the language of cause and effect, and Observers translate between them.

This sets the stage for Section 6, where we clean up the framework's boundaries (to avoid infinite regress), and Section 7, where we consider what all this implies for predicting and testing in the real world.

sam@maddoxcp.com

6. The Terminal Object and First Mover

To avoid an infinite regress (colloquially, a 'tower of turtles') within this formalism we introduce **True Infinity (TI)** as both a categorical terminal object and the generative source of rules that structure the Ruliad.

This object is an axiom of this formalism and is unprovable for an Observer within **R**:

Definition 15 *Terminal Object*: **TI** is the terminal object in our category, with:

$\forall X \in \mathrm{ob}(\mathrm{T}), \exists! f: X \to \mathrm{TI}$

This says that every object in the Ruliad has a unique morphism taking it to the terminal object, TI.

The terminal object is more than a formal mathematical convenience—it reflects the foundational role of TI in grounding the entire categorical structure of reality (and of the Ruliad).

Definition 16 *Rule Generation*: The rules governing the Ruliad emerge from TI:

$\mathrm{Rules}(R) = \mathrm{T}\_\infty(\mathrm{TI}) \cap R$

As the terminal object, **TI** grounds the constraint structures from which all domains emerge. Through these constraint structures, value emerges in the interactions of Observers with the Ruliad.

Thus, while **TI** is not itself directly "valuable" in an anthropocentric sense, it is the categorical ground that enables value to exist at all.

This relationship can be formalized through the constraint generation functor:

$\mathrm{CG}: \{\mathrm{TI}\} \to \mathrm{Constraints}$

Where Constraints are the category of constraint structures that define the domains D_i (i.e. the limit of the sampleable information structures an Observer has access to in R_0).

The emergence of value can then be viewed as a functor from constraints to value structures:

$\mathrm{Value}: \mathrm{Constraints} \to \mathrm{ValueStructures}$

The composition Value ∘ CG thus traces the emergence of value from the terminal object, **TI**.

This addresses the bootstrapping problem by identifying **TI** as the ultimate categorical ground that enables both the existence of the Ruliad and the emergence of value through Observer interactions with the Ruliad's constraint structures.

Summary
Section 6 confronted the issue of infinite regress in our Observer hierarchy ("it's Observers all the way up?") by positing a Terminal Object TI – an effective "first mover" or ultimate Observer perspective that caps the chain. We introduced TI (True Infinity) as a conceptual top of the lattice, an object in the category to which all other objects have a unique morphism. In everyday terms, TI plays a role analogous to a theoretical omniscient viewpoint or a final unification of all information (one could poetically call it the "mind of God" in a metaphoric sense, though we treat it mathematically).

By having TI in our model, we avoid loops or never-ending climbs up domains: there is always an "end of the line" that everything feeds into. This is satisfying because it means our hierarchy of nested Observers does not result in paradox – there is a completion point.

sam@maddoxcp.com

Additionally, TI provides a fixed reference for absolute comparisons, which we might liken to an absolute frame or a ground truth reality (though no real Observer attains it).

The First Mover aspect suggests that TI could be seen as initiating the dynamics (since morphisms from TI to other objects could represent "God's hand" setting initial conditions, in metaphorical language).

However, in this formalism, it's more about completeness than theology: it ensures the category of Observer domains has a terminal object and thus all diagrams seeking a limit actually have one. The casual reader can think of TI as a useful book-keeping device: an assumption that there is an ultimate consistency or unity to the entire structure (so we don't spin in circles asking, "who observes the Observer?" indefinitely).

With Section 6, our theoretical edifice is complete and self-consistent. This paves the way for Section 7, where we step back and discuss what predictions or applications emerge now that the framework is in place.

7. Empirical Predictions and Applications

7.1 Conceptual Hypotheses

Our framework generates several conceptual hypotheses that could guide future empirical research:

1. Information Integration Hypothesis: Systems with higher measured information integration across domains should report richer subjective experiences. This suggests that consciousness correlates with the degree to which a system integrates information across the four domains we've identified.
2. Domain Transition Hypothesis: Activities that cross domain boundaries (meditation, artistic creation, psychedelic experiences) should show distinctive patterns of neural activity and information processing. Specifically, these states should exhibit increased connectivity between neural systems normally associated with different processing domains.
3. Persistent Structure Formation Hypothesis: Learning and skill acquisition should follow a pattern where initial entropy reduction is followed by morphism discovery, measurable through information-theoretic analysis of neural activity during learning.
4. The Boundedness and Persistence Hypothesis: If an Observer's boundedness, B is changed (e.g. giving humans technological extensions like faster sensors or memory aids), their effective R_0 should grow – they might observe phenomena previously too subtle or complex. This could be as simple as using slow-motion cameras (increasing persistence, P) to reveal events (like a hummingbird's wings) that we normally can't integrate. Our framework would describe that as raising γ to include those events in F_0

These hypotheses, which have been explored in aspects of neuroscience, represent conceptual predictions derived from our formalism, pointing to directions for future computational modelling rather than specific computational implementations.

7.2 Applications

The framework has potential applications across multiple disciplines:

- Physics: Offers a perspective on the measurement problem in quantum mechanics by formalizing the Observer's role in sampling and constructing reality
- Cognitive Science: Provides a formal, non-idealised model for embodied cognition, the emergence of consciousness, and the relationship between different modes of cognition

sam@maddoxcp.com

- Artificial Intelligence: Suggests design principles for systems that can integrate information across multiple domains, potentially leading to more human-like artificial general intelligence
    - As a speculative example, one could create a multi-agent simulation where each agent has tunable boundedness (processing power / memory) and persistence (frame-rate of observation) and see how different settings lead to different 'perceived' environments. Do agents with stricter bounds (lower processing power) fail to notice certain high-complexity phenomena? This could be tested in a virtual physics environment by comparing an agent with high sampling rate against one with low sampling rate, testing their ability to detect fast events. This would directly test the B and P constraints.

In each application domain, our framework points to new conceptual approaches rather than prescribing specific computational implementations, which will be developed in future work as the mathematical formalism is refined and extended.

Summary

Section 7 took our theoretical model and pointed it toward the real world, outlining how we might test and apply the Observer-Ruliad framework. We proposed several conceptual hypotheses – for instance, the idea that varying an Observer's boundedness (like increasing human working memory via neural tricks or AI augmentation) could let them perceive phenomena normally invisible (maybe slight violations of assumed physics, or entirely new patterns).

We also suggested experiments like altering an Observer's relevance filter (through training or altered mental states) and seeing if their observed reality's "laws" appear to change (for example, a highly trained physicist literally sees different aspects of motion than a layperson – anecdotally true, formally interesting here).

In terms of applications, we listed interdisciplinary uses: in physics, treating different reference frames as different S_0 functors could help resolve discrepancies between quantum and classical viewpoints; in AI, building agents that are aware of their own Observer bias could improve alignment by targeting it's F_0 to match with ours. We pointed to cognitive science: our model provides a way to quantify integration (which could correlate with EEG measures of consciousness – a testable angle). We even touched on perception studies – for example, this framework predicts certain optical illusions or gestalt effects occur because the brain's R_0 is trying to impose persistence and relevance, and we could design new illusions to validate that.

In AI simulations, we can implement toy models of the Ruliad and toy models of Observers and see if the Observers indeed perceive stable emergent "physics" – which, if successful, lends credence to our interpretation of our own physics.

Overall, Section 7 is a roadmap for moving from theory to practice: it identifies where to look for evidence (in neuroscience, in AI experiments, in maybe even particle physics anomalies) and how adopting this framework could open up new technology (like Observers with tunable parameters). The conclusion of Section 7 – and thus the paper – is that while ambitious, our Observer-Ruliad model is not just metaphysics; it yields concrete, testable ideas that can unify understanding across domains.

If an overarching theme is to be stated: by acknowledging the Observer as part of the system, we unlock explanations and innovations that remain hidden under traditional siloed viewpoints. Section 8 leaves us with both a challenge and an invitation to the scientific community: test these ideas, refine them, and explore reality not just as it is, but as it is observed.

8. Philosophical Implications

Our category-theoretic framework for Observer Theory gives philosophical implications that extend beyond the formalism. By grounding subjective experience within a formal computational model, we provide new perspectives on several enduring philosophical questions.

8.1 The Nature of Value and Meaning

In traditional scientific frameworks, value and meaning have been relegated to the realm of the subjective, considered beyond the reach of formal description.

Our formalism provides a potential resolution by identifying value with information integration across domains.

When an Observer reduces entropy (or increases computational reducibility, to the limit defined by the facets of Observation detailed in this paper) and discovers new morphisms, they are simultaneously creating value by expanding their field of observation (and the field of Observation for Observers that share the same properties) and integrating information in increasingly complex ways.

This perspective aligns with philosophical traditions that view meaning as arising from interconnection and context. As Wittgenstein noted, meaning emerges from use within a system of relationships.

In this framework, meaning emerges naturally as Observers integrate information across domains, by creating persistent structures and discovering new morphisms.

8.2 Free Will and Agency

The problem of free will has traditionally been framed as a conflict between determinism and freedom. Our framework offers a novel perspective by locating agency in the Observer's capacity to reduce entropy or increase computational reducibility (or in lay-language making F_0 approach the limit R_0 in as few hypergraph updates as possible, through the choice of morphisms between states).

While the Ruliad itself contains all possible computational paths, the Observer's sampling and entropy reduction represent genuine creativity.

This view resonates with compatibilist positions in philosophy, suggesting that freedom is not the absence of causation but rather the capacity for complex information integration and the discovery of new morphisms.

An Observer's choices are simultaneously determined (within the ultimate structure of the Ruliad) and free (as the selection of morphisms between states is determined by the Observer in the system, subject to their constraints; Boundedness, Persistence and Relevance).

8.3 The Mind-Body Problem

Our framework provides a fresh approach to the mind-body problem by situating both mental and physical phenomena within a unified categorical structure with different constraint profiles.

Rather than positing a fundamental divide between the mental and physical, we demonstrate how these domains emerge from different patterns of constraint on information.

This perspective echoes aspects of dual-aspect monism and neutral monism, where a single underlying reality manifests different aspects depending on how it is observed. In our framework, mind and matter are not separate substances but different constraint patterns within the same categorical structure.

9. Future Work

While this paper has established a formal, abstract framework for Observer Theory, several directions for future work promise to extend and apply these ideas:

9.1 Computational Implementation

To make this framework more empirically tractable, future work will focus on developing computational implementations. One toy-model approach involves representing the Ruliad as a simplicial set, where:

- 0-simplices represent computational states
- 1-simplices represent transitions between states
- Higher simplices represent homotopies between paths

This approach would preserve the essential topological structure of the ∞-groupoid while providing a pathway to practical computation and modelling.

9.2 Empirical Studies

The framework suggests several empirical research directions:

- Studies of information integration in neural systems during different states of consciousness
- Analysis of entropy reduction patterns in learning and creative problem-solving which should lead to more accurate predictions post learning
- Investigation of cross-domain causation in psychosomatic phenomena and embodied cognition

These empirical directions could provide validation for key aspects of the theoretical framework.

9.3 Applications to Artificial Intelligence

The framework has significant implications for AI design. Current AI systems typically operate within narrow information domains and lack the integration across domains that characterizes human consciousness. Future work could explore:

- Architectures for cross-domain information integration in AI systems
- Methods for implementing entropy reduction and morphism discovery as learning mechanisms
- Approaches to embedding value and meaning in computational systems

9.4 Mathematical Refinement

Further mathematical work is needed to fully formalize certain aspects of the framework:

- More precise definitions of constraint functions and measures for each domain
- Rigorous formulation of the embedding functors between domains
- Exploration of the categorical structure of persistent patterns and their morphisms

sam@maddoxcp.com

10. Conclusion

This paper has attempted to present a rigorous and significant extension of Observer Theory within the context of the Ruliad, utilising the category-theoretic formalism established by Gorard et al.

By formalizing the Observer as an entity that samples and integrates information across hierarchical domains, I have attempted to provide a framework to probe fundamental questions about the nature of consciousness, causation, and reality itself, under a single model.

The framework explanatory power is focused on the following areas:

1. How discrete computational processes could give rise to continuous experiences
2. How causal influence can exist across different domains without violating causal closure
3. How consciousness could emerge through information integration
4. How Observers expand their sense of agency through entropy reduction and morphism discovery

This work, which straddles Philosophy, Philosophy of Physics and Computation attempts to bridge the divide between objective description and subjective meaning by creating a formal framework where anthropomorphic value and purpose can emerge naturally from the information and category theory.

It echoes and formalises Wheeler's 'it from bit' conjecture, which posited that information underlies physical reality [7], and aligns with Deutsch's constructor theory approach that emphasizes information-centric laws of physics [12].

By providing a new categorical ground for meaning, we open new possibilities for understanding our place in the universe.

In summary, we have developed a formalism that posits:

(1) The Ruliad, from Wolfram's prior work, describes every possible computational universe
(2) The Ruliad, from Wolfram's prior work, can be consistently constrained by an Observer to produce a 'reality'
(3) The Ruliad, from this work, is a structure that contains a hierarchy of informational domains that map into each other via functors
(4) Conscious experience corresponds to information that spans these hierarchies (i.e. an Observer integrates information across their accessible informational domains), and
(5) Cross-domain causation does not violate underlying computational rules, by viewing higher-level causes as functorially propagating constraints to lower-level dynamics.
(6) Taken together, it constructs a framework that extends existing models (Wolfram's physics, IIT, etc.), increasing their operational utility, whilst also making testable predictions.

Note

In June 2024, I attended Stephen Wolfram's summer school as a visiting scholar. My interest was around the theological, metaphysical and philosophical implications of the Ruliad. During that time I discussed conceptual outlines of these extensions with Stephen. I was also lucky enough to spend lots of time with Hatem Elshatlawy and James Wiles. Without them this work would not have been possible.

The first draft of this paper to them in July 2024. It wasn't fantastic. Over the past year I've spent time integrating the mathematics of the Ruliad, particularly Jonathan Gorard and Xerxes Arsiwalla's category theory formalism. This paper is the result of that work.

This paper was sent, in draft, to Wolfram's team April 2025 via email and was published in draft on the Wolfram Institute discord, for review, in April 2025. It was also shared with Joseph Natal, Ali Ozbay, Erkan Kaya, Daniel Rowe and Eli Katz who gave me constructive feedback to improve it. Thank you for your help.

sam@maddoxcp.com

12. References

[1] **Wolfram, S.** (2021). *The concept of the Ruliad*. *Stephen Wolfram Writings*.

[2] **Gorard, J., & Arsiwalla, X. D.** (2021). *Pregeometric spaces from Wolfram model rewriting systems as homotopy types*. arXiv:2111.03460.

[3] **Rickles, D., Elshatlawy, H., & Arsiwalla, X. D.** (2023). *Ruliology: Linking computation, Observers and physical law*. arXiv:2308.16068.

[4] **Tononi, G.** (2008). *Consciousness as integrated information: a provisional manifesto*. **Biological Bulletin, 215**(3), 216–242.

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

[6] **Chalmers, D. J.** (1995). *Facing up to the problem of consciousness*. **Journal of Consciousness Studies, 2**(3), 200–219.

[7] **Wheeler, J. A.** (1990). *Information, physics, quantum: The search for links*. In **W. Zurek (Ed.)**, *Complexity, Entropy, and the Physics of Information*. Westview Press.

[8] **Lakoff, G., & Johnson, M.** (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. Basic Books.

[9] **Hoffman, D. D., Singh, M., & Prakash, C.** (2015). *The interface theory of perception*. **Psychonomic Bulletin & Review, 22**(6), 1480–1506.

[10] **Seth, A. K., Barrett, A. B., & Barnett, L.** (2011). *Causal density and integrated information as measures of conscious level*. **Philosophical Transactions of the Royal Society A, 369**(1952), 3748–3767.

[11] **Friston, K.** (2010). *The free-energy principle: a unified brain theory?* **Nature Reviews Neuroscience, 11**(2), 127–138.

[12] **Deutsch, D., & Marletto, C.** (2014). *Constructor theory of information*. arXiv:1405.5563.

[13] **Everett, H.** (1957). *"Relative state" formulation of quantum mechanics*. **Reviews of Modern Physics, 29**(3), 454–462

[14] **Wolfram, S**. (2023). Observer Theory. *https://writings.stephenwolfram.com/2023/12/observer-theory/*

[15] **Rickles, D., Elshatlawy, H., Blum, A. & Arsiwalla, X. D.** (2025). Towards a Generalized Theory of Observers. arXiv:2504.16225

sam@maddoxcp.com

Appendix A: Toy Model of an Observer-Ruliad System

To concretely demonstrate the principles, construct a minimal **toy model**.

Consider a very simple "universe" (Ruliad subset) consisting of a few states and rules, and an Observer with strict limits. This will show how the Observer's functor samples the universe step-by-step.

Toy Ruliad: Let the foundational set of possible information be $V_0 = \{E, F, G, H\}$ (think of these as basic observable states). Define a few possible **rules** (rulial relations) r_1, r_2 which dictate how states can lead to others (like laws of nature):

- Rule r_1: E to F (state E leads to F in one step), and F to G.
- Rule r_2: E to H, and H to G.

In a multiway graph view, from start E_0 at time 0, the universe can evolve to either F_1 or H_1 at time 1 (depending on whether r_1 or r_2 is applied), and by time 2 both paths converge to G_2.

So the Toy Ruliad here includes the history paths: E_0 to F_1 to G_2 and E_0 to H_1 to G_2.

Assume G is an end state.

Now consider an **Observer** with very limited capacity: it can only follow one rule at a time (boundedness **B** is 1), and it has perfect persistence (it doesn't forget within the short time considered).

At t=0, the Observer's knowledge set is X={E} (it perceives state E). At t=1, two new states F and H are possible in the universe. However, the Observer's *boundedness constraint* says it can only integrate something that is related and not more complex.

Both F and H are directly related to E (since r(E,F,1) is true via r_1, and r(E,H,1) is true via r_2). Suppose, however, that because of some internal bias or limitation, the Observer only tracks one rule – say it "chooses" rule r_1 (this could be random or based on habit). Thus it integrates E and not H.

In formal terms: there existed an x in X (x=E) such that R(x,F,1), so F passes relevance; and F isn't more unbounded than E (both are single symbols, so fine); so F is integrated.

H also had R(E,H,1) is true, but the Observer fails to integrate H because it had already committed its resources to F – this could be seen as the Observer's relevance filter being effectively exclusive (it can't follow two divergent branches due to its boundedness B=1).

Now at time 1 the Observer's state is X={E, F}, meaning it thinks the world went from state E to state F.

At t=2, the universe, regardless of path, is at state G.

For the Observer, from its state {E,F}, what can happen?

r(F,G,2) is true (via rule r_1) and also R(H,G,2) is true (via r_2), but the Observer never integrated H.

So from its perspective, G is reachable (relevant) because $\exists F \in X$ with $R(F, G, 2)$.

Also, G is a simple symbol no more complex than F, so the boundedness test passes. Thus the Observer integrates G.

End result: the Observer experienced a continuous path E to F to G. It never noticed the alternative rule 2-related H route.

This toy scenario illustrates key points:

- The **multiway branching** (E to F vs. E to H) is resolved by the Observer's boundedness; it couldn't follow both, effectively *choosing a single path*. This models how an Observer perceives a single outcome / thread of time even if multiple paths were possible.
- At the convergence G, the Observer's and the universe's histories re-align – G is perceived normally since it lies along the Observer's path (and indeed along all paths here).
- Had the branches not converged (suppose E led to F and H led to a new informational state, **L** different from **G**), our Observer would only ever witness G, and not realize **L** was a possible outcome. Another Observer might have followed E to H to L (using the new path) and swear **L** happened and not **G**. In the Ruliad both happened, but each Observer sees a consistent single timeline.

Let's compute some of the toy functions:

At t=0,

$B(X=\{E\},0) = |\{\text{rules } r \mid \exists z \in \{E\}, w \in V_0: r(z, w, 0)\}|$.

There are 2 rules applicable to E (r_1 and r_2), so B({E},0) =2.

Our Observer being bounded means it effectively acts as if it can only use one rule at a time.

At t=1, X={E,F}, we might say B(X,1) would again be 2 (since from F one rule applies and from E another, though E, at t=1, might be moot).

The Observer only actually applied one rule.

Persistence P({E,F},1)=2 (i.e. the Observer exists at t=0 and t=1).

For Relevance, we can list: R(E,F,1)= true, R(E,H,1)= true; the Observer sampled the first rule as it was relevant.

This simplified model can be visualized as a small branching graph with an "eye" following one branch.

It demonstrates how **discrete possibilities can yield a single stream of experience** when filtered by an Observer.

One can imagine an even simpler toy Observer (say with persistence, P=0 meaning it has no memory – akin to a goldfish-like Observer).

It would perceive flickers without continuity: in our example, at t=1 it might see F, but by t=2 it forgets E and F, so seeing G would not be integrated; G might appear as a sudden unrelated event (a surprise).

That Observer would not experience a coherent thread of time, highlighting the necessity of P>0 for Observer continuity of experience.

This toy model, while trivial, demonstrates the core idea: **the Observer's rules of operation actively shape which part of the Ruliad is realized as "actual" or "real"** for that Observer.

In more complex models, we could add internal states for the Observer and more elaborate rules, but the principles scale up from this toy model: relevance gates perception, boundedness limiting parallel tracking, and persistence provides continuity.

sam@maddoxcp.com

Appendix B: Formal Developments and Proofs

B.1 Nested Informational Domains as a Lattice:

Let's formalize the **hierarchy of domains** (R_0, R_1, R_2, R_3) introduced in Section 3.

Each R_i can be thought of as a set (or σ-algebra) of information states at level i.

We assume: $R_0 \subseteq R_1 \subseteq R_2 \subseteq R_3 \subseteq \Omega$,
where Ω represents the total information content of the universe (and in the limit, Ω tends toward True Infinity, TI).

Inclusion of $R_i \subseteq R_{(i+1)}$ means any state or fact observable at level **i** is also representable at level **i+1** (higher levels have at least as much computational capacity as the lower levels).

Each inclusion is **proper** (higher level contains strictly more possibilities / accessible morphisms than lower ones). For example, R_1 (physical states) include all R_0 (Observer's accessible states) plus more (e.g., unobserved microscopic states); R_2 includes all physical configurations and adds branching superpositions or multiple possibilities, etc.

We can equip each R_i with a **topology** reflecting the distinguishable states at that level.

R_0's algebra might identify many different Ω -states as the same if the Observer can't tell them apart.

Formally, the Observer functor F: $\Omega \rightarrow R_0$ induces a partition on Ω

Likewise, natural projection maps;

$p_{i,(i-1)}$: $R_i \rightarrow R_i - 1$ exist

(e.g., a conceptual state projects to a resulting physical state, a physical state projects to an Observer's state if observed).

Definition: We can say the chain (R_0, ..., R_3) is **integrative** if each $R_{(i+1)}$ is the *smallest* superset of R_i that is closed under whatever new operations characterize level **(i+1)**.

For instance, going from R_1 to R_2 we close under "simultaneous branching histories" and obtain a larger set.

Going R_2 to R_3 we close under "abstraction/integration of information".

This integrative property ensures there are no gaps – it's a kind of completeness: if a certain combination of information is possible in principle, it will appear by the time we reach R_3 (short of TI).

The **Nested Domain Theorem** (conceptual): *The four domains form an increasing chain of σ-algebras such that R_3 is maximal consistent information for the Observer, and any further extension would require moving to TI.* (In other words, R_3 is like an asymptote – you can approach **TI** but not have more domains without effectively reaching for the absolute.)

B.2 Functorial Mapping and Entropy Reduction

We now formalize the Observer as a functor and show how it relates to entropy. Let **R** denote the category (or possibility-space) of the full Ruliad (or a relevant subset, Ω), and let **O** denote the category of the Observer's perceived states (essentially equivalent to R_0 in the main body of the paper but structured with morphisms representing observed transitions between states).

The **Observer functor** F: R→ O maps each object/state in R to an object in O, and each morphism (state transition) in R to a morphism in O (if perceptible).

By definition of a functor, if $X \to^{r} Y$ in R (state X leads to Y via rule, r), then $F(X) \to^{\bar{r}} F(Y)$ in O, for some (perhaps non-unique) $\bar{r}$ representing the observed transition. Many different rules in R might map to the *same* $\bar{r}$ in O if the Observer can't distinguish between them (this is the essence of coarse-graining).

Because F is **many-to-one** in general (i.e., not injective), the mapping loses information.

In information-theoretic terms, F induces a **partition** on R's state space: several distinct R-states correspond to one O-state.

Any partition like that is associated with an entropy reduction.

Formally, consider a probability distribution P over states of R. The Observer will perceive a pushforward distribution P^F over states of O: $P^F(o) = P(\{x \in R \mid F(x) = o\})$.

It is a standard result (related to the **Data Processing Inequality** in information theory) that coarse-graining cannot increase total information sampled.

In fact, the **entropy** of P^F will be *lower* than the entropy of P on R, once we condition on the fact that the mapping is many-to-one. More directly:

Proposition (Entropy Reduction by Observation): *Let x be a random variable taking values in R with entropy H(x) (Shannon entropy). Let O = F(x) be the induced random variable taking values in observed states O. Then H(O)≤H(x), with equality if and only if F is one-to-one on the support of P*

This inequality captures that an Observer sees a world with entropy (uncertainty) not greater than the underlying world's uncertainty.

In many cases, it's *much* smaller – e.g., the microstates of molecules in a room have enormous entropy, but our observation of the room (macrostate like temperature, pressure) has far lower entropy (we ignore the molecular fine-grained details).

The difference, H(x)-H(O), can be interpreted as the **information filtered out** (or "ignored") by the Observer. It doesn't vanish – often it's effectively stored in the Observer's **ignorance**.

(N.B. In thermodynamics, an Observer not tracking molecular motions means that information is treated as hidden, leading to the second law of thermodynamics from the Observer's perspective.)

Proof Sketch: The mapping F defines a partition of the outcome space of x. It's a basic result that the entropy of a partitioned random variable satisfies H(x) = H(O) + H(x | O), where $H(x \mid O) \geq 0$

Here $H(x \mid O)$ is the entropy of the underlying state given the observed state – essentially the uncertainty an Observer has about the exact microstate.

Since $H(x \mid O)$ is non-negative, we have H(O) = H(x) - H(x |O) $\leq$ H(x).

(N.B. Equality implies H(x |O)=0, meaning the observation fully determines the state – F was one-to-one, so no there is no information loss.)

Thus, **entropy does not increase under the Observer functor**.

This formalizes the claim that Observers "reduce entropy" or "minimize surprise" by their very nature – not by magically altering the world's entropy, but by considering a coarse-grained space where the description requires fewer bits.

This is analogous to how in the Friston's Free Energy Principle, an organism's internal states model the environment and thereby effectively encode only a summary of sensory inputs

Our result is generalised version of this statement.

B.3 Additional Theorems

We list two important propositions that add rigor to multi-level interactions:

- **Theorem (Consistency of Multi-Level Observation):** *Let $F_{1,0}:R_1 \to R_0$ be the physical-to-Observer functor (mapping physical states to observed states), and $F_{2,1}:R_2 \to R_1$ be the abstract-to-physical functor (mapping multiway superstates to effective physical outcomes).*

  The composition $F_{1,0} \circ F_{2,1}: R_2 \to R_0$ is equivalent to the direct functor $F_{2,0}: R_2 \to R_0$ one would obtain by treating the Observer as mapping abstract states directly to their perception.

  In other words, $F_{1,0} \circ F_{2,1} = F_{2,0}$
  Similarly for $F_{3,2} \circ F_{2,1} = F_{3,1}$ for as many domains as your model allows

Proof Sketch: In category theory terms, this requires that our functors form a commutative diagram.

By construction, an Observer perceives the end result of a process, not the intermediate branching (unless they have specific measurement devices directed to measure branching).

The functor mapping is defined such that whether the branching of Domain 2 collapsed to a physical outcome first, or whether we imagine the Observer perceiving the abstract state and conflating outcomes themselves, the final distribution on R_0 is the same.

Essentially, if something is unobservable at physical level, it's equally unobservable at abstract level.

This guarantees *no double-counting or contradictions*: multi-level emergence doesn't confuse the Observer.

(Akin to saying: it doesn't matter if quantum wavefunctions collapse before we see it, or if we treat it as collapsing upon seeing – either way, we never observe a superposition directly.)

- **Proposition (Convergence of Iterated Integration):** *Consider the iterative information integration operator. Under reasonable conditions (finite state spaces or discounting of new information over time), the sequence of Observer states $X(0),X(1),X(2),\ldots$ will converge to a fixed point $X(n)$* or cycle, representing an equilibrium of knowledge

Explanation: Intuitively, an Observer continuously gathering information in a stationary environment will eventually learn all there is to learn (reach a fixed X), or if the environment has repeating patterns, the Observer's state might cycle through a set of knowledge states (e.g., always forgetting and re-learning periodic events).

Formally proving convergence requires assumptions (like contraction mappings in the integration operator in a metric space of knowledge states). In many models, **Information Integration** might be monotonic (knowledge non-decreasing until saturation), in which case it will approach a limit (it can't increase forever due to Observer boundedness – though, n.b. humans can and have invented technologies to decrease our natural computational boundedness within F_o, examples include things like sensors, computers, and calculators).

That limit could be the full information of the environment *accessible* given the constraints. This aligns with ideas of *equilibrium* in thermodynamics or steady-state in inference (eventually the Observer's posterior = prior when no new evidence).

The proof is beyond the scope of this paper, but it relies on the finiteness or compactness of state space and the Observer's limits.

Overall, these support the qualitative claims: our nested structure is well-defined and complete up to **TI**, Observers inevitably simplify (reduce entropy of) the description of reality, and repeated observation tends to fill the Observer's knowledge to a stable point (given no new surprises).

More elaborate theorems could be developed (for example, relating functor categories to a topos and showing **TI** behaves like a terminal object or an inaccessible cardinal, etc.), but those go beyond the scope of this paper.

sam@maddoxcp.com

Appendix C: On True Infinity (TI) – Justification and Context

True Infinity, **TI**, is a central but abstruse concept in our framework – essentially the "container of all that is". Here I clarify its role and necessity and relate it to mathematical and philosophical precedents.

C.1 Category-Theoretic and Topos-Theoretic Perspective

In category theory, one often encounters the issue that the "category of all sets" is too large to be a set itself (leading to Russell-like paradoxes).

The usual resolution is to work in a **Grothendieck universe** or assume an inaccessible cardinal – effectively positing a very large set that contains all sets one needs yet is not maximal.

By analogy, **TI** plays the role of a **universe of discourse** so large that all other domains (sets of states, categories of processes) are contained within it, but **TI** itself is not a member of any manageable domain.

This is akin to treating **TI** as a **proper class** in set theory (something that is not a set but a collection too vast to handle directly).

In a topos-theoretic view, one might think of **TI** as analogous to the class of all objects in an Infinite Topos of structures – it's not an object within the topos, but the topos itself or its limit.

We could attempt to formalize TI as the *colimit* (union) of the ascending chain (if one imagined more levels) or as the *limit* of some inverse system. In simpler terms, if you had an infinite sequence of increasingly comprehensive descriptions, **TI** would be the hypothetical end description containing all information.

Category theory countenances such "ideal elements" often (like adding a point at infinity to compactify a space). Here, TI is treated as an idealized limit that *completes* the structure of informational domains – ensuring there is an ultimate context in which all partial contexts reside.

One might ask: can't we do without **TI** and just say everything is relative to some context?

In mathematics, if one doesn't assume a universe or an infinite cardinal, one runs into limits on what one can discuss (e.g., one can't uniformly talk about "all groups" as a set, only as a proper class, which complicates certain proofs).

Similarly, without TI, our framework would;

a. have an open-ended hierarchy (what lies beyond Domain 3? Domain 4, 5, … ad infinitum?), or
b. have to arbitrarily cut off at some highest domain.

Option (a) might lead to an **infinite regress** – always a larger context beyond any given one, which could be conceptually troubling and practically unwieldy (Observers would never have a well-defined ultimate reality to approach, just more layers).

Option (b) raises "why stop here?" and begs the question of what justifies that maximum. By positing an axiomatic **TI**, I assert that there is a maximal element (even if unattainable) that grounds the entire structure.

It is analogous to how physicists use the concept of an absolute but unattainable zero temperature or the speed of light as an unreachable limit – these provide boundary conditions that make the theory complete, even if one never actually reaches them.

C.2 Derived Necessity vs. Axiomatic Postulate

Is **TI** something we can *derive* must exist, or are we simply *postulating* it for convenience?

This touches on deep philosophical arguments. One line of reasoning (inspired by metaphysics) is a **principle of plenitude**: anything that can exist, does exist (in some form, somewhere). If one assumes this, then indeed the "set of all possibilities" – an everything – exists. Another reasoning: for any Observer or system, one can imagine a larger context that includes it (like a meta-Observer considering it).

If one keeps doing this, logically one either goes forever or one ends at a comprehensive context that includes all.

If one appeals to some version of Ockham's razor or elegance, one might prefer the existence of a single Absolute Infinity rather than an endless tower (the famous 'it's turtles all the way down' metaphor).

Thus, one could argue TI is a **logical necessity** if we believe reality (in the broadest sense) is self-contained and not endlessly deferred.

On the other hand, formal axiom systems (like ZF set theory) often *axiomatize* something akin to **TI** (not directly, but through axioms of infinity, power set, etc., which guarantee arbitrarily large sets exist).

We can analogize that here: **TI** can be treated as an **axiom** – assume an all-encompassing reality – and explore consequences. It's not currently derivable within physics or empirical science (since by definition it's beyond any observation), but it functions as a guiding postulate, telos or ultimate attractor.

If future theoretical work showed contradictions arising unless an absolute context is assumed, that would bolster it as a derived necessity as opposed to an Axiomatic predicate.

C.3 Historical Philosophical Parallels

The idea of an infinite, unbreakable whole has many echoes:

- **Spinoza's "Substance":** Baruch Spinoza in the 17th century argued for a single substance (which he identified with God or Nature) that is infinite and contains all attributes. Everything that exists (modally) is a mode of that one substance. This maps well to TI – an infinite substance containing all properties and realities. Spinoza's substance is also unchanging in itself (though its modes change), similarly TI is envisioned as an absolute that doesn't evolve because it already *is* everything (change is a relation in subordinate domains).
- **Peirce's "Unlimited Continuity" and Final Logical Interpretant:** Charles Sanders Peirce, a philosopher and logician, held that the universe is continuous and that truth is the limit of inquiry. He had the notion that if investigation were pushed to infinity, the belief it would end in is the truth (sometimes dubbed the "final opinion"). This resonates with treating **TI** as the *limit of all inquiries/Observers*. No matter how far knowledge progresses (in principle), **TI** is the ideal final knowledge containing all truths. Peirce also dabbled with the idea of an 'Absolute Mind' or a cosmic consciousness in which all ideas subsist; again analogous to **TI** acting like a platonic realm of all information.

- **Kabbalistic Ein Sof:** In Lurianic Kabbalah (16th century Jewish mysticism), **Ein Sof** is the term for the infinite, endless divine essence. Ein Sof is completely incomprehensible and contains all of reality in potential. According to that tradition, at creation Ein Sof underwent **Tzimtzum** – a self-contraction – to leave a "void" in which a finite world could exist, with only a tiny spark of Ein Sof's light (the **Kav**) entering that void to create and sustain the universe, leaving an infinitesimal shard that seeded our finite reality. That's essentially a secular retelling of Tzimtzum: **TI** is something akin to an Ein Sof; the initial condition is like the Kav; and our Observer-accessible universe is the "void" that is illuminated in a controlled way so as not to overwhelm the Observer. The comparison isn't just metaphorical – it highlights that to get a finite experience out of an infinite totality, some process akin to selection or contraction is needed, which our model implements via the nested domains and Observer constraints.
- **Leibniz's Plenitude and "Best of all possible worlds":** Leibniz argued God considered all possibilities (an infinite set) and chose to actualize the best. While we don't dive into theological framing, the idea of all possibilities existing at least in concept is similar to **TI**. One might say the Ruliad is like the set of all possible worlds; our particular universe is a branch or section through it.

These historical ideas show that **TI** is not an ad-hoc invention of our theory but sits in a lineage of thinking that reality in total must be boundless and all-inclusive. Each of those philosophies grappled with how the finite world relates to an infinite foundation – similar to how we use TI versus the finite Observer.

C.4 TI as Equilibrium and Attractor

I described TI as an "equilibrium boundary condition" guiding Observers.

What this means is: in the space of all possible Observer states or knowledge, the state corresponding to **TI** (having complete information) is an **attractor** – a state that, if it could be reached, would be an absorbing state (no more change, maximum coherence, maximum entropy reduction in our formalism).

Observers only maximisation function in this framework is the integration of more information in as few hypergraph updates as possible (via entropy reduction in their observed reality, as per Appendix B) which is like climbing a hill whose unreachable summit is **TI**.

They can get arbitrarily close (improve knowledge, unify physics theories, build faster and more data intensive computers etc.) but **TI** is at infinity, never reached.

Still, it *guides* the process like a telos (goal).

In control theory or dynamics, one sometimes stabilizes a system by adding a potential function that has minimum at the desired equilibrium; one might view **TI** as providing a "direction" for integration processes – the direction of increasing integration (consistency and coherence are other words we could use).

As long as **TI** is consistent (assume it is – it contains all truths so it cannot be internally contradictory in any ultimate sense, or else no reality could exist), then moving toward it yields fewer contradictions or in the parlance of the Ruliad, the maximal computational reducibility for the Observer within $\mathbf{R_0}$.

An Observer that somehow drifted away (losing information or introducing inconsistency) would be pushed back by the requirement of coherence (this is speculative, but one could imagine formulating Observer learning as a gradient flow in an information landscape with a global minimum at **TI**).

In summary, I justify **True Infinity** on both pragmatic and philosophical grounds: it completes our theoretical scaffold (preventing infinite regress of contexts), and it resonates with a long tradition of positing an Absolute.

It is not directly observable, but its **effects** are felt. It is the unseen whole that gives meaning to the partial views, much like an entire puzzle picture guides the fitting of individual pieces.

One may choose to take **TI** as a regulative ideal (in the Kantian sense) – something we assume for the sake of theory coherence, whether or not one thinks it literally "exists."

Our framework remains operational for finite Observers either way, but assuming **TI** grants a coherence and unity that, in our view, any theory should strive for.

*One final note: One might worry **TI** is too abstract. But note, even in mathematics, treating infinities carefully leads to powerful results – e.g., the existence of inaccessible cardinals can't be proven in ZF set theory, but if assumed, it doesn't lead to contradiction and allows one to talk about larger collections safely.*

*Similarly, assuming **TI** doesn't create a logical contradiction in our framework; rather it provides a horizon that all possible Observers share irrespective of their domain composition.*

sam@maddoxcp.com

Appendix D: Additional Examples Illustrating the Framework

To build intuition, we present a variety of real-world examples showing Observer Theory concepts in action.

D.1 Thermostat (Simple Cybernetic Observer)

A home thermostat is an example of a minimal Observer.

Its sensor measures temperature (one type of information), and it has a target setting. It perceives the environment (room temperature) and acts (turn heater on/off) to minimize the discrepancy – very much like Friston's free-energy minimization but in a simple device.

In our terms, the thermostat's boundedness, **B** is extremely low: it only registers a single variable (temperature, possibly with limited precision).

Its relevance function **R** is trivial: it only "cares" about temperature readings (nothing else in the room matters to it).

Its persistence **P** is effectively encoded in its circuitry – it "remembers" the setpoint and the last measured value until updated.

So the thermostat integrates information over time: if the temperature is below target (observation at time, **t**), it infers heater should be on, leading to a new state at **t+1** with increased temperature, which it then measures.

The thermostat thus achieves a stable equilibrium (room at set temperature) – an empirical example of an Observer driving the world toward a desired macro-state.

It doesn't "know" anything about molecules or even that it's heating air; it works purely with coarse-grained information (i.e. it is heavily bounded in the rules it samples) and in doing so it reduces surprise (temperature error).

This exemplifies how even unconscious systems fit Observer Theory.

D.2 Human Vision and Blind Spot Filling

The human visual system has a literal blind spot (where the optic nerve exits the retina). Observers (us) aren't aware of a dark void in our vision because our brains *fills in* the gap using surrounding information.

This is a live example of the Information Integration function **I**: the visual Observer (brain) takes the incomplete retinal map and *integrates* it with prior knowledge of the scene (consistency, continuity) to create a continuous experience.

Relevance, **R**, plays an important role: the brain deems the blank spot "relevant" to fill since it's surrounded by structured image, and boundedness plays a role in *how* it fills (it assumes no sudden unknown object lurks exactly in the blind area, which is generally a valid assumption).

One can test this by a simple experiment: close one eye, stare at a point on a paper such that a small nearby X falls on the blind spot – the X disappears, and you'll see the background as if X isn't there.

The Observer (brain) edited reality for coherence. This everyday example mirrors how our framework allows that Observers don't get raw data; they get processed, integrated data – an interface view. It's benign usually (helping us see smoothly), but it reminds us that our perceived continuity is an active construct.

sam@maddoxcp.com

D.3 Scientific Community as a Collective Observer

Zooming out, consider the entire scientific community as a kind of meta-Observer of the universe.

Different scientists and instruments act as sub-Observers gathering data (from particle colliders, telescopes, etc.), and their findings are integrated into a collective body of knowledge (journals, theories).

Over centuries, this community's "integrated information" has grown tremendously. We can see paradigm shifts as the moments when the information integration function (under our framework) grew dramatically (e.g., relativity providing new science and explanations where Newtonian physics couldn't).

The persistence function here is documentation and memory – science builds on past results (high **P** across generations).

Boundedness is reflected in technology limits – e.g., we can only observe up to certain energy scales or distances so far (that's a "boundedness" of our current collective Observer – hence we have unanswered questions).

Relevance is managed via frameworks like theories that tell scientists what data is pertinent (you ignore some noise or unrelated phenomena, focusing on tests of your hypothesis).

In effect, the scientific endeavour is Observer Theory writ large: it samples the Ruliad (all possible computational phenomena) in a constrained way to build an increasingly integrated model of reality to reduce surprise and increase prediction ability. Predictions that are accurate reduce the computational resources required by an Observer, i.e. there is less entropy and the state of the hypergraph is more reducible.

The anthropic principle is at play too: scientists find themselves in a universe that is comprehensible enough to yield to such integration, otherwise the process would fail.

So this example illustrates multi-Observer integration and the gradual approach to **TI** (as final truth) as an ideal.

D.4 Cell as an Observer in Development

During an animal's embryonic development, cells differentiate and move to form organs. Each cell is like a little Observer receiving chemical signals (morphogens, cell-cell contacts), persisting some internal state (gene expression memory), and responding (growth, apoptosis, etc.).

The collective behaviour yields the correct anatomy. If we take one cell's perspective: its boundedness is limited to local signals (it has no idea of the whole body plan, just what neighbours broadcast).

Its relevance function is tuned by evolution: e.g., a certain gradient is highly relevant (telling it "you're near the head, become a neuron"), other signals are irrelevant noise.

Its persistence is in how stable its gene regulatory network is (some cell "decisions" are irreversible, maintaining state through cell divisions).

Now, despite each cell's local view, the entire system achieves a globally coherent outcome (the organism).

How? It's as if a higher-level Observer (the tissue or the whole embryo) exists, but really it's an emergent property of all cells interacting (as per Levin's nested agents).

Assembly theory comes in: the embryo's complexity is assembled step by step, with each cell's actions building on prior ones.

If something disturbs a cell, often neighbouring cells compensate (they observed the deviation and adjusted – "downward causation" from tissue pattern to cells).

This example concretely shows multi-level integration: signals form a pattern (Observer at tissue level integrating cell info) and pattern directs cells (downward causation). Experiments in developmental biology, such as splitting an embryo (as in identical twins forming) or grafting cells to new positions, demonstrate the resilience and re-adjustment capabilities – Observer Theory would attribute that to the remaining cells updating their perceived reality and finding a new coherent plan (twins each think they have a full embryo's signals and develop two bodies; grafted cells re-interpret their fate based on new neighbours, etc.).

Thus, even in biology, thinking of parts as Observers that integrate information has significant explanatory power.

D.5 Maxwell's Demon (Information and Entropy)

Maxwell's Demon is a famous thought experiment: a tiny demon purportedly observes molecules in two chambers and only allows fast ones to go one way, slow ones the other, seemingly decreasing entropy.

The resolution of the paradox is that the demon's observation and memory increases entropy (or free energy usage) such that overall the second law holds – essentially the demon can't cheat thermodynamics because information is substantiated physically.

In our framework, the demon is an Observer with very high unboundedness (it can observe individual molecules – far beyond human capacity!).

Its Relevance is all molecules' speeds, and it acts to segregate them. It appears to create a temperature difference (lower entropy in gas distribution) – but Landauer's principle says erasing its memory will dissipate that entropy elsewhere.

Observer Theory insight: the demon+gas as a whole is a closed system; the demon's acquired information allowed it to perform work (like an engine).

This exemplifies how *observation is a physical act* that can be converted to work, reinforcing that the act of observing (gathering info) is part of physical processes.

If an Observer could reduce entropy in their perceived world, they must pay with entropy increase in unobserved degrees of freedom (here, the demon's memory tape).

This is consistent with our Appendix B proposition: the demon's observation coarse-grained the gas microstates (reducing entropy of its observed subsystem), but the missing info is in its memory (unobserved by others) which increased entropy.

This balances out. Maxwell's Demon thus concretely illustrates the interplay of boundedness (demon needed to measure each molecule – a daunting task requiring a delicate apparatus), persistence (demon must remember info until it uses it), and how **entropy reduction in an observed subsystem** comes at a cost per Observer Theory (the total entropy including the Observer's "hidden" entropy doesn't decrease).

This aligns this model with thermodynamics – a tested domain.

D.6 Interface of a Smartphone (User vs Underlying Mechanism)

A modern smartphone presents a slick GUI with icons, while under the hood there are binary circuits and complex operations. The phone's user is like an Observer limited to the interface. From the user's perspective, opening a photo gallery is tapping a picture icon – they don't see the file system traversal, the decompression algorithms, etc. This echoes Donald Hoffman's Interface Theory.

The user sees the "desktop" not the circuits. In our terms, the *user as Observer* has boundedness that doesn't extend to electronic signals; those are not relevant (R, in the function in the paper, is false for raw circuit voltages).

Instead, relevance is tuned to meaningful symbols (apps, images).

The phone's OS acts as an intermediary Observer too – it monitors hardware states and updates the screen accordingly (with its own B, P and R geared to hardware events).

This layered observation allows a human to effectively manipulate billions of transistors by simple intuitive actions.

It's a mundane example but drives home the point: different Observers (user vs. phone's firmware) carve reality at different joints, each integrating information appropriate to their level.

Where the interface is well-designed, the user's lower boundedness doesn't matter – they can achieve complex tasks without ever perceiving the complexity. In less ideal cases (say, a glitch), the user might briefly see the "truth" (an error code, raw data) that means nothing to them – demonstrating why interfaces evolve to hide complexity for fitness or usability.

This shows the complementary roles of multiple Observers: one Observer's output (the OS's processed info) becomes the world of the higher Observer (user). It's analogous to how our perceptual systems process photons into visual objects which then our conscious mind deals with as reality; the intermediate "truth" (neural firings) we never directly sense.

These examples – from machines to cells to people to society – illustrate key points: Observers filter and simplify reality (thermostat, smartphone UI), integrate discrete bits into wholes (vision filling blind spot), operate in nested hierarchies (cells in tissues, OS for user), and obey physical laws relating information to energy (Maxwell's demon).

They demonstrate the versatility of Observer Theory in describing phenomena across disciplines in a unified way.

sam@maddoxcp.com

Appendix E: Addressing Potential Criticisms

Finally, we anticipate and respond to some criticisms and questions that readers might raise about Observer Theory and the necessity of its components:

- **E.1 "True Infinity seems overly complex and unnecessary."**

  Critics might argue that introducing **TI** – an unobservable absolute – complicates the theory without empirical payoff.

  Why not just talk about very large but finite systems or leave it at "potentially infinite" without identifying it as TI?

  The response is that TI provides a **conceptual closure** that avoids vagueness about "and so on…".

  It's analogous to how mathematicians talk about the limit to infinity – one never reaches it, but having infinity in the vocabulary allows concise reasoning (like "this series converges / diverges).

  Likewise, TI lets us discuss the integration of all information without always couching it in limiting processes. TI is a boundary condition.

  You could formulate all our arguments with "in the limit as knowledge grows without bound…" instead of TI and get the same practical outcomes; I chose TI as a stand-in for that limit state to simplify discourse.

  Historically, similar constructs (absolute space, absolute time, the continuum of real numbers) were sometimes doubted but have proved extremely useful. If one is uncomfortable, one can view TI as a convenient fiction – an "ideal" that needn't be physical.

  However, many philosophical traditions (as noted in Appendix C) indicate that assuming an Absolute can be explanatory.

  As for complexity, adding TI doesn't actually introduce new free parameters or equations.

  In fact, it *reduces* conceptual complexity by providing a single unifying object rather than implying an endless open hierarchy. Thus, we argue TI is an elegant inclusion that brings coherence, even if it is metaphysical.

  Should there be an alternative framework that operates strictly within finitary bounds and still accounts for everything, TI might be dispensable – but thus far, every attempt to explain reality ultimately invokes something unbounded (be it an infinite multiverse, an indefinitely large phase space, etc.).

  I've just made that explicit.

- **E.2 "Couldn't observed phenomena be explained by simpler, separate theories?"**

  One might say: we already have quantum physics for micro-level, classical emergence for macro, neuroscience for cognition, etc. Why layer them under the Observer Theory umbrella? Are we unnecessarily inventing a new lexicon?

  My stance is that Observer Theory is not discarding those successful theories – it's **connecting and contextualizing them**.

Each of those fields deals with a level of description, often holding others fixed or treating the Observer externally.

By having a functor-based approach, we can discuss how classical reality emerges from quantum (example, a functor formalizing decoherence), or how neural firings lead to conscious perception (a functor taking you from physical to Observer reality), in one consistent language.

This cross-talk between domains is where many puzzles lie (quantum measurement problem, mind-brain problem, etc.).

Traditional separate theories don't solve those; they simply bracket them off. This framework is an attempt to dissolve those puzzles by showing they are natural outcomes of considering the constraints of observation.

It is more general and abstract, but that generality is its strength: it shows the common structure behind phenomena in physics, biology, cognition, and computation. In essence, we aren't competing with domain-specific theories on their turf; we're taking the meta-theory of the Ruliad and extending it to provide an overarching scaffold they all fit into.

This unification can lead to new insights (e.g., drawing analogies between how coarse-graining works in thermodynamics and in information processing by brains).

As for simplicity, while the framework is broad, each component (functor, B,P,R, domains) is conceptually simple.

The alternative – having separate unrelated models – might seem simpler locally but is more complex globally (lots of independent axioms versus one coherent set). Think of how the Newtonian gravitational law seemed simpler than Einstein's spacetime curvature concept initially; but Einstein's framework resolved anomalies and encompassed more in one go.

Similarly, Observer Theory is ambitious but potentially rewarding in its explanatory power.

- **E.3 "Is this falsifiable or testable?"**

  This is a crucial challenge for any new proposal.

  Parts of Observer Theory – especially the TI concept or the full Ruliad – are untestable directly (they are "too big" to prove).

  However, the framework yields **many testable sub-theories and predictions** as we outlined in **Section 7**.

  For example, the idea that an Observer's boundedness affects what laws they perceive could be tested by simulations or experiments altering cognitive load or information availability (to see if perceived causality or patterns change).

  The notion that physical laws emerge from Observer consistency could be indirectly tested by searching for phenomenological signatures of simple rule systems (Wolfram's project aims for this).

  More concretely, the interplay of information and thermodynamics (Maxwell's demon reasoning) is testable and has been validated in experiments that have measured

Landauer's bound of *energy dissipated per bit erased*, supporting the link between information and entropy.

Our framework embraces those principles, so it's consistent with those tests. Another testable aspect: if multi-level causation resolution is right, then interventions at higher levels should have predictable effects on lower levels without violation.

In neuroscience, for instance, if someone consciously intends (high level) to do something, we predict we should find a corresponding pattern in neural data that isn't random but follows the constraints (some preliminary work in brain-machine interfacing does find distinct patterns corresponding to intentions, aligning with them being real causal variables).

In short, while Observer Theory as a whole is broad (like how the theory of evolution is broad), it produces many specific hypotheses that can be and have been tested in pieces.

If results came in that fundamentally contradicted the idea that Observers adhere to B,P and R constraints (for instance, if someone demonstrated perception of phenomena that by all physical accounts should be beyond their boundedness – akin to "seeing" individual atoms with naked eyes – that would challenge the framework), we'd have to revise or abandon it.

Thus, it's falsifiable in principle: it forbids certain things (like an Observer perceiving completely random high-entropy microstates as structured – if someone routinely did Maxwell's demon feats with no entropy cost, that'd break our theory as much as it breaks thermodynamics).

- **E.4 "Anthropic reasoning is circular."**

  The use of a generalized anthropic principle might raise eyebrows. Critics say anthropic arguments (we observe this because if not, we wouldn't be here to observe) can be tautological or unfalsifiable.

  Our use of it is not to hand-wave away fine-tuning, but to guide the **selection of explanatory structures**.

  We identified four domains not arbitrarily but by asking what structure must reality have such that Observers (like us) can exist and not immediately reach paradoxes or limits.

  This is a heuristic, not a strict derivation. However, it's akin to a design principle. It doesn't replace empirical checking; it suggests what to look for.

  If the anthropic rationale is wrong, empirical evidence will show a different structure (maybe there are more than four layers, or fewer).

  For example, if we found complex life that doesn't require something like conceptual integration, that would challenge our claim that Domain 3 is necessary. But all known intelligent life relies on thinking, so that's supportive.

  The anthropic principle in cosmology has mild predictive power (like expecting to find ourselves in a universe with conditions at least marginally stable for life, which is trivial, but also predicting certain statistical distributions of constants if there's a multiverse selection effect). In our case, it predicted that an Observer will perceive layered realities – which is what we do see (different sciences for different scales etc.).

sam@maddoxcp.com

So while not rigorously falsifiable on its own, anthropic reasoning here is used carefully to shape the theory, which then yields falsifiable statements.

We also ensure not to abuse it: whenever possible, we seek mechanistic accounts (e.g., *how* Domain 3 arises from Domain 2 interactions, not just "if it didn't, we wouldn't be here").

- **E.5 "Comparative advantage – aren't other theories like Free Energy or Integrated Information sufficient?"**

  To summarize the outcome: Observer Theory is more encompassing but less specific.

  The Free Energy Principle (FEP) explains brain function elegantly but doesn't directly address why physical laws are what they are; Wolfram's Ruliad gives a abstract multi-computational picture but does not provide a detailed account of mind or agency; Integrated Information Theory (IIT) quantifies an idealistic consciousness but doesn't tie it into physics (the process Tononi describes can't be completed in our universe) or deal with the emergence of it (i.e. the move from non-life objects to life).

  Our theory bridges gaps: it could accommodate IIT as a 'real world' version of Tononi's ideal. It could produce a way to measure Boundedness or Integrated Information, and it aligns with FEP's idea of entropy (surprise) minimization as shown in Appendix B, and it is based on the established formalism of Wolfram's Ruliad.

  In that sense, it's not in competition, but in synthesis – it aims to be *the framework that includes the others as special cases*.

  We strive to stay on the right side of the analytic philosophy line by keeping things mathematically and logically defined (functors, sets, etc.) and pointing to concrete tests.

- **E.6 "Empirical Falsifiability Revisited – what would disprove Observer Theory?"**

  As a framework, it's not as easily falsified by a single experiment as a specific hypothesis would be. But there are conceivable outcomes that would undermine it.

  If, for instance, the universe turned out to be fully deterministic and the concept of "possible computations" irrelevant (Wolfram's notion of the Ruliad would be moot, since only one computation actually happens), that would weaken the foundation of our approach (we rely on multiple possibilities to give meaning to boundedness – if only one path existed, an Observer's limits are trivial, they just see that path).

  Another disproof would be if consciousness or observation had demonstrably nothing to do with physical processes – e.g., if we found that removing an Observer (in a controlled quantum experiment) had no effect, implying observation truly has no role (current evidence in quantum physics actually affirms the opposite: observation – or at least interaction with a measuring device – does affect outcomes).

  Additionally, if a future theory showed that the universe's laws are exact and closed under all scales (no new emergent info at higher scales, everything derivable from microscopic physics alone), one would conclude Observer Theory's multi-domain approach is unnecessary.

  However, evidence points to genuine novelty at higher scales (you can't trivially derive biology from particle physics, etc., due to complexity and emergent structures).

Finally, I've utilised categorical structures that have been designed and proved by others. The use of their category theory formalism and the functors described herein could be both inexact and off-base – if it turns out reality cannot be described in these terms, one might ditch that formalism (though category theory is quite general – if something is describable in set/relational terms, it should be describable in categorical terms).

But that would be a modification of the mathematics, not the conceptual core.

In summary, while no single datapoint will shout "Observer Theory is wrong!", the framework lives or dies by its *usefulness and explanatory power*: if it organizes known data well and predicts new connections, it will thrive; if phenomena routinely defy its structure, it will either evolve or be set aside.

I have tried to make it falsifiable by tying it to many existing measurable principles (information entropy etc.), any of which if conclusively violated in their domain would cause trouble for the assumptions.

In conclusion, Observer Theory is a bold but reasonable attempt to address a part of the scientific discourse (namely Philosophy of Physics and Observation) that is typically brushed under the carpet by materialists, physicalists and post-modernist philosophers.

It does not throw out the successes of past science or philosophy – it builds on them, asserting that the common thread is the role of information and observation.

By directly addressing potential criticisms – complexity (we argue it's warranted for completeness), anthropic circularity (we use it cautiously), testability (we propose concrete experiments) – we aim to show that this framework is not just metaphysical musing but a scaffold with firm footings in known science that provides a clear path to future discoveries.

sam@maddoxcp.com

Glossary and Analogues from Other Disciplines

Term in this paper	Physics analogue	AI / ML Analogue	Information Theory Analogue	Other Comparable Notation
Ruliad (entangled limit of all computations), ***R***	Ultimate Multiverse (all possible physical laws)	The space of all possible algorithms (cf. phase space of all models)	Related to universal Turing machine's 'tape' of all possible programs	∞-groupoid of all computational structures Similar to, but not the same as, Tegmark's Level IV Multiverse
Multiway System	Many-worlds branching in quantum mechanics (Feynman path integrals as summing over these branches)	Branching search tree in AI planning	Ensemble of all possible message decodings	Directed Acyclic Graph
Observer, *O*	Reference frame in relativity / measuring apparatus that collapses the wave function in quantum (selection of outcome)	The **Agent** that filters data	A **Channel** or filter that selects certain information from a source	Functor S_0 to R_0 mapping a larger category to an accessible sub-category
Observer's Field, F_0	The **past light cone** for a physical Observer (with causal contact). Parallels a 'Heisenberg cut' – i.e. what is actually measured vs. unmeasured	Attention of Focus of an AI at a given time (the subset of inputs / database it actively processes to achieve a result)	A chosen subset of bits from a larger data stream	The subset defined by B(x), P(x) and R(x)
Boundedness, B(x)	Finite measurement resolution	Finite compute and memory	Finite channel capacity or bit-rate limit	A resource bound, comparable to a complexity constraint
Persistence, P(x)	The need for events to have a given duration to be observable (e.g. flicker-fusion threshold)	Sampling rate	Cutoff in signal processing	n/a
Relevance, R(x)	Comparable to "anthropic" constraints in cosmology (i.e. only certain universes matter to us)	Goal-driven attention	Information gain criteria – i.e. only sampling data that reduces uncertainty about a relevant variable is retained by the system	This is a Boolean indicator, a true / false
Entropy, H(x) and Reducible Entropy H_{red}(x)	H(x) has equivalence to statistical mechanics measure of entropy **$S=k_B \ln\Omega$** (a count of microstates) H_{red}(x) is analogous to the correlations that make physical laws predictable	H(x) is equivalent to 'surprise' for a predictive model H_{red}(x) is equivalent to pre-trained knowledge an AI already has access to when it samples it's search space	H(x) has some equivalence to Shannon Entropy, measuring bits of uncertainty in a given state. H_{red}(x) is related to but not perfectly analogous to Redundancy of certain bits of information	n/a
Information Content, *ϕ(x)*	Quantum Information Theory	Linked to Prediction Error	Kolmogorov Complexity of new information	n/a
Integrated Information, I(F_0)	Analogous to total entropy absorbed by a physical Observer (e.g. total bits registered by a measurement apparatus) Analogous in cosmology to 'Information Content of the Observable Universe'	Total learning by an AI from a given dataset (sum of 'surprises' equivalent to informational gain)	Mutual Information between an Observer's internal state and the environment	Reminiscent of Tononi's ***ϕ*** *(although note Tononi talks about this at the system level and his measure is irreducible, whereas this sums the irreducible parts across domains)*
Domains P, V, S & M	Levels of organisation in nature; P, Physical (matter, energy, forces) V, Value (evolutionary related goals) S (logic, math, text) M (Platonic structures)	Different modules of cognition: Sensorimotor (P), Reward (V), Cognitive/Symbolic (S), High-Level Abstract (M)	n/a	n/a

sam@maddoxcp.com

Term in this paper	Physics analogue	AI / ML Analogue	Information Theory Analogue	Other Comparable Notation
Embedding Functor, $S_{i,j}$	Correspondence to the idea that any higher-level phenomena must have a realisation in the physical domain e.g. in physics-to-biology, how atomic / quantum processes manifest as life (an example of a $P \rightarrow V$ embedding)	Corresponds weakly to Representation Learning, mapping of low-level data into a higher-level conceptual mapping e.g. a network that translates pixel data (P) into labels (S)	Related to decoding and recoding	In category theory this is functor that preserves elements of the underlying structure
True Infinity, TI	In cosmology analogous to an ultimate limit / boundary e.g. the point at infinity that 'completes' spacetime	Likened to a point of conceptual closure in an AI search space (a hypothetical, unreachable point, where an AI has considered and weighted all possible outcomes from it's dataset in every possible way)	The limit of increasing complexity, an unattainable endpoint containing all information	In category theory, a terminal object that means there is a unique morphism from any object into it. A formalisation of a 'ultimate point' that everything maps to