

# Universality without Totality: Diagonal Constraints on Information, Coherence, and Holography

Andrei T. Patrascu

*FAST Foundation, Destin FL, 32541, USA*

*email: andrei.patrascu.11@alumni.ucl.ac.uk*

We argue that the black-hole information paradox is not a paradox of information loss, but a manifestation of a more general structural limitation: in sufficiently expressive, self-referential quantum systems, information *about the system* cannot be totally internalized as a single, globally coherent, and uniformly decodable code. By “information about the system” we mean semantic distinctions or predicates concerning the system’s states, observables, or histories, as opposed to the mere existence of physical correlations. While such information may exist and be preserved in correlations, diagonal arguments of Gödel-, Turing-, and Lawvere-type imply that no internal procedure can encode and decode all such distinctions uniformly within the same system once self-reference is present.

We formulate this limitation in the language of categorical holography by viewing bulk-to-boundary maps as encoding functors and show that any encoding sufficiently rich to internalize its own reconstruction cannot serve as a total classifier of bulk predicates. Consequently, holographic reconstruction can at best be universal in a relative sense, i.e. restricted to code subspaces, sectors, or contexts, but never total. We demonstrate that modern mechanisms such as quantum error-correcting code subspaces, entanglement wedge reconstruction, and the island prescription realize this non-totality physically through state- and context-dependent reconstruction rather than through information loss.

Finally, we connect this semantic non-totality to a physical mechanism, Higher Categorical Coherence Breakdown (HCCB), in which global linear and unitary descriptions fail to glue coherently across contexts despite local consistency. In this view, black holes represent maximally self-referential systems where the limits of total internal encoding become unavoidable and geometric, revealing a universal constraint on information in quantum theory rather than a pathology specific to gravity.

## I. INTRODUCTION

### A. From “information loss” to limits of decoding

The *black-hole information paradox* is often narrated as a tension between (i) unitary quantum evolution, (ii) semiclassical locality, and (iii) Hawking’s semiclassical computation of thermal radiation and ever-growing entanglement entropy. Hawking’s original conclusion was that gravitational collapse may lead to a breakdown of predictability and an apparent loss of information [3]. Subsequent work reframed the question in operational terms: if a black hole evaporates unitarily, what is the time-dependence of the information accessible in the Hawking radiation, and what is the *mechanism* by which it becomes accessible? A central benchmark in this discussion is the *Page curve* [4], which predicts that the fine-grained entropy of the radiation increases until roughly the Page time and then decreases back to zero in a globally pure evaporation process.

The modern landscape has sharpened the paradox in two complementary directions. On the one hand, the *firewall* argument [5] showed that demanding a single, globally consistent semiclassical description that simultaneously realizes unitarity, effective field theory at the horizon, and standard entanglement monogamy leads to an apparent inconsistency. On the other hand, developments in holographic entanglement and bulk reconstruction provided concrete tools for tracking quantum information in gravitational systems: the Ryu–Takayanagi prescription [6] and its covariant extension [7] relate boundary entanglement entropy to extremal surfaces in the bulk; quantum corrections [8] and the quantum extremal surface (QES) framework [9] incorporate bulk entanglement effects; and the operator–algebraic/quantum–error–correcting viewpoint of AdS/CFT [10] together with modular/relative entropy control [11] clarified why bulk reconstruction is naturally *subsystem–relative* rather than globally absolute.

A decisive recent step was the emergence of the *island* prescription and replica wormholes [14–16], which recover a Page–curve–consistent fine-grained entropy from semiclassical geometry by allowing the entanglement wedge of the radiation to include interior “islands”. In parallel, entanglement wedge reconstruction was used to articulate precisely when (and in what sense) interior operators can be reconstructed from the radiation [13], building on the older Hayden–Preskill insight that, after the Page time, rapid scrambling can make newly infalling quantum information recoverable from the radiation given sufficient control [12].

These developments strongly suggest that the “paradox” is not simply a conflict between unitarity and gravity, but a conflict between *certain totality assumptions* about *decoding* and the structure of self-referential quantum systems. The guiding thesis of this paper is that the black-hole information paradox is best understood as exposing a universal limitation:

*Information **about the system** may exist and be preserved in correlations, yet cannot always be totally internalized as a single, globally coherent, uniformly decodable code within a sufficiently expressive, self-referential system.*

The remainder of this introduction makes this statement precise enough to be assessed logically, and motivates why it naturally leads to (i) diagonal/undecidability constraints (Gödel/Turing/Lawvere) [17–19], (ii) “universality without totality” in categorical holography, and (iii) a physical realization in terms of Higher Categorical Coherence Breakdown (HCCB), in which global gluing of a linear/unitary description fails despite local consistency.

### B. The hidden assumption: total internal decoding

We begin by separating three distinct notions that are frequently conflated:

- (A) **Existence/preservation of correlations:** quantum states determine correlation functions and entanglement measures, and unitary dynamics preserves global purity.
- (B) **Information about the system:** semantic distinctions or predicates concerning the system’s states, observables, or histories (e.g. “which microstate?”, “which interior configuration?”, “which sector?”).
- (C) **Internal encodability/decodability:** the existence of an *internal* representation and a *uniform* decoding procedure which, from within the same system/theory, recovers the answers to *all* predicates of interest.

Diagonal arguments constrain (C), not (A). In particular, the claim that “information is not totally encodable” is *not* a claim of information destruction; it is a claim that no single internal coding/decoding structure can uniformly represent *all* semantic distinctions about a sufficiently expressive, self-referential system.

To formalize the hidden assumption, let  $\text{Sys}$  denote the “system” (e.g. the full quantum gravitational system describing an evaporating black hole plus radiation) and let  $\mathcal{Q}$  denote a class of *questions* about  $\text{Sys}$ . Concretely,  $\mathcal{Q}$  may be taken as a space of predicates on states, a family of bulk observables, or a family of operationally meaningful propositions that one would like to decide. We write an *answer set* as a set (or space)  $\mathcal{A}$ .

An *encoding* is any map (or functorial assignment)

$$\mathcal{E} : \text{Sys} \longrightarrow \text{Code}, \quad (1)$$

which associates to the system an internal *code object*  $\text{Code}$  (e.g. the radiation state, a boundary algebra, a family of reduced density matrices, a set of accessible correlators). A *decoder* is a procedure

$$\mathcal{D} : \text{Code} \times \mathcal{Q} \longrightarrow \mathcal{A}, \quad (2)$$

intended to output the answer to question  $q \in \mathcal{Q}$  using only the internally available code.

*Total internal decoding (informal definition).* We call  $(\mathcal{E}, \mathcal{D})$  a *total internal decoder* if:

- (i) **Totality:**  $\mathcal{D}$  is defined for *all*  $q \in \mathcal{Q}$  and all relevant system states (i.e. it halts/returns an answer uniformly).
- (ii) **Correctness:** for each  $q \in \mathcal{Q}$ , the value  $\mathcal{D}(\mathcal{E}(\text{Sys}), q)$  agrees with the “ground truth” semantics of  $q$  for  $\text{Sys}$  (e.g. correct expectation values, correct membership in a sector, correct interior predicate, etc.).
- (iii) **Internality and uniformity:**  $\mathcal{E}$  and  $\mathcal{D}$  are definable within the same theoretical framework describing  $\text{Sys}$  and do not rely on external oracles (no meta-theory privileged viewpoint).

Many informal arguments in the black-hole literature implicitly assume such a total decoder exists, at least in principle, for the Hawking radiation. The Hayden–Preskill protocol, for instance, assumes “unlimited control” over the radiation in order to analyze retrieval times [12]; and many versions of the paradox tacitly treat decoding as an abstract operation unconstrained by self-reference. However, the diagonal phenomena behind Gödel incompleteness [17] and the halting problem [18] teach that *total internal decoders are not generically available* once the system is expressive enough to encode statements about its own encoding.

### C. Diagonal logic: why total internal decoding fails

The underlying logical structure is classical: whenever a system can represent (enough of) its own semantics or operational behavior internally, universal “deciders” become impossible by diagonalization. In arithmetic, Gödel’s diagonal lemma produces a sentence asserting its own unprovability, implying that no consistent sufficiently expressive system is complete [17]. In computation, Turing’s diagonal construction shows that no program can decide halting for all programs [18]. In category theory, Lawvere gave a unifying formulation of these diagonal arguments in cartesian closed categories (CCC), showing that a suitable “evaluation + weak surjectivity” hypothesis forces fixed points for all endomorphisms—and hence yields contradictions in the presence of negation-like maps [19]. We will later give a detailed formulation adapted to holographic encoding functors, but the meta-lesson can already be stated:

*If a decoding scheme is rich enough to internalize its own action on encodings, then demanding totality (a uniform correct decoder for all predicates) forces a diagonal predicate/operator which the decoder cannot handle consistently.*

This is the sense in which self-reference imposes a *limit of total encodability*: the relevant “information” is information *about* the system (predicates, properties, and semantic distinctions), not the raw existence of correlations.

### D. Universality without totality in holography and islands

The AdS/CFT correspondence motivates the idea that boundary data encodes bulk physics. However, modern developments already indicate that reconstruction is intrinsically *relative*: bulk locality behaves like a quantum error-correcting code, and reconstruction is controlled on a code subspace rather than across the full Hilbert space [10]. Entanglement wedge reconstruction, guided by RT/HRT and its quantum refinements [6–9], makes reconstruction explicitly dependent on the boundary region and the state. The island rule and replica wormholes [14–16] go further: the *reconstructible region itself* can change discontinuously (a QES phase transition), yielding a Page curve consistent with unitarity while avoiding naive “total decoder” expectations. This picture is sharpened by analyses that tie the Page transition directly to entanglement wedge reconstruction of the radiation [13].

Our proposal is to interpret these facts through a single structural lens: *holography supports universality of encoding but forbids totality of decoding*. In other words, there can be universal *cores*—minimal structures through which all successful reconstructions factor—without the possibility of a single globally defined decoder for all bulk predicates. The island prescription then appears not as an ad hoc fix, but as a geometrized implementation of contextual (state-dependent) decoding consistent with diagonal constraints.

### E. Preview: Higher Categorical Coherence Breakdown (HCCB)

The diagonal discussion above is semantic: it constrains what can be *internally decided* or *uniformly represented* in a self-referential system. A separate but complementary question is physical: *how does the theory realize these constraints dynamically and operationally?* The guiding idea of Higher Categorical Coherence Breakdown (HCCB) is that the minimal consistency requirement in quantum theory is not “global symmetry” or even “global unitarity” as an absolute structure, but rather *coherence closure*—the ability to glue local/sectorial descriptions consistently. When higher coherence conditions fail, one expects the global description to cease to be a single linear/unitary object while retaining local consistency; physically this appears as sectorization, history dependence, and effective completely positive (CP)

dynamics on accessible subalgebras. These ideas are developed in detail in our HCCB framework [1, 2]. In this paper, we use the black-hole setting as a maximally self-referential regime where these issues are geometrically unavoidable, and we propose a unified narrative: diagonal non-totality provides the *logical* reason total decoders fail, while HCCB provides a *physical* mechanism by which the world implements coherence without totality.

*Roadmap.* Section I has isolated the hidden totality assumption. In subsequent sections we: (i) develop the diagonal constraint in a form suited to holographic encoding functors, using the Lawvere-style categorical diagonal mechanism [19] and its logical ancestors [17, 18]; (ii) reinterpret entanglement wedge reconstruction and islands as concrete realizations of “universality without totality” [6–9, 13, 14, 16]; and (iii) connect these semantic constraints to HCCB as a unifying physical mechanism.

## II. WHAT DO WE MEAN BY “INFORMATION”, “ENCODING”, AND “DECODING”?

### A. Why definitions matter: avoiding a category error

A large fraction of the confusion surrounding “information loss” in gravitational settings comes from conflating distinct layers of description: (i) *physical correlations* carried by quantum states, (ii) *semantic distinctions* (predicates) *about* the system, and (iii) the existence of a *uniform internal* procedure that can encode and decode all such distinctions from within the system. Diagonal arguments (Gödel/Turing/Lawvere) constrain the third layer, not the first [17–19]. Accordingly, we fix terminology precisely.

Throughout, “information” is understood in the operational/information-theoretic sense introduced by Shannon [20] and generalized to quantum systems by von Neumann’s entropy and its refinements [21, 22]. We distinguish this from “information *about* the system”, meaning semantic distinctions or predicates concerning states, observables, and histories. This section formalizes (i) what it means for such information to *exist* and be *preserved*, (ii) what it means to be *encodable* and *decodable*, and (iii) what “total” means in each case.

### B. Operational information: correlations as constraints

*States and observables.* Let  $\mathcal{H}$  be a (separable) Hilbert space and let  $\mathcal{B}(\mathcal{H})$  denote the bounded operators. A *state* is a density operator  $\rho \geq 0$  with  $\text{Tr}\rho = 1$ . Operational predictions are expectation values  $\langle O \rangle_\rho = \text{Tr}(\rho O)$  for observables  $O$  in an admissible algebra (often a von Neumann algebra in QFT contexts).

*Entropy and mutual information.* The von Neumann entropy

$$S(\rho) = -\text{Tr}(\rho \log \rho) \quad (3)$$

quantifies uncertainty of  $\rho$  and reduces to Shannon entropy on classical distributions [20, 21]. For a bipartite state  $\rho_{AB}$  with marginals  $\rho_A = \text{Tr}_B \rho_{AB}$ ,  $\rho_B = \text{Tr}_A \rho_{AB}$ , the mutual information

$$I(A:B)_\rho = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \quad (4)$$

measures total correlations. It can be written as a relative entropy:

$$I(A:B)_\rho = D(\rho_{AB} \parallel \rho_A \otimes \rho_B), \quad (5)$$

where  $D(\rho \parallel \sigma)$  is the (Umegaki) quantum relative entropy [23]. (Equation (5) follows directly from the definition  $D(\rho \parallel \sigma) = \text{Tr}(\rho(\log \rho - \log \sigma))$  and basic logarithm rules when  $\sigma = \rho_A \otimes \rho_B$ .) Relative entropy is the operational measure that controls distinguishability and obeys monotonicity under physical channels, a property crucial for any discussion of accessible information [21, 22].

*Monotonicity and coarse-graining.* A *quantum channel* (CPTP map)  $\Phi$  is a completely positive, trace-preserving linear map on density matrices. The monotonicity (data processing inequality)

$$D(\rho \parallel \sigma) \geq D(\Phi(\rho) \parallel \Phi(\sigma)) \quad (6)$$

expresses the fact that distinguishability cannot increase under physical processing [22, 23]. This is the precise mathematical form of “forgetting” or coarse-graining: channels destroy or hide information by reducing distinguishability. We emphasize that (6) is a statement about *accessible information* and correlation structure; it does *not* by itself impose any limitation on the existence of semantic truths about the system.

### C. Encoding and decoding as internal channels

We now formalize *encoding* and *decoding* in a manner suitable both for quantum information theory and for holographic/bulk–boundary settings.

*Encoding map.* An *encoding* is modeled as a channel (or functorial assignment)

$$\mathcal{E} : \mathcal{S}(\mathcal{H}_{\text{bulk}}) \longrightarrow \mathcal{S}(\mathcal{H}_{\text{code}}), \quad (7)$$

where  $\mathcal{S}(\mathcal{H})$  denotes the state space on  $\mathcal{H}$ . In holographic language,  $\mathcal{H}_{\text{code}}$  can be the boundary/radiation subsystem (or an algebra thereof), and  $\mathcal{E}$  is the physical map that produces the accessible data (e.g. restriction to a subalgebra, partial trace, or boundary encoding). For generality, we keep the channel formulation; it is standard that every channel admits a Stinespring dilation, i.e. can be realized as a unitary on a larger space followed by discarding an environment [21, 24]:

$$\mathcal{E}(\rho) = \text{Tr}_{\text{env}}(U(\rho \otimes |0\rangle\langle 0|)U^\dagger). \quad (8)$$

This is important conceptually: “encoding” is not mysterious; it is just a physical interaction plus discarding degrees of freedom.

*Decoding map.* A *decoder* is a channel

$$\mathcal{D} : \mathcal{S}(\mathcal{H}_{\text{code}}) \longrightarrow \mathcal{S}(\mathcal{H}_{\text{bulk}}), \quad (9)$$

intended to reconstruct bulk information from the code. In quantum error correction, one demands that  $\mathcal{D} \circ \mathcal{E}$  act as the identity on a specified *code subspace* (or on a set of states), not globally [21, 22]. This distinction (relative vs global decoding) will be central when we interpret holographic reconstruction and islands as *contextual* decoding mechanisms rather than total ones.

*Semantic information vs operational reconstruction.* To connect decoding to *information about the system*, introduce a family of predicates or queries  $\mathcal{Q}$ . Operationally, a query can be:

- an expectation functional  $q_O(\rho) = \text{Tr}(\rho O)$  for some observable  $O$ ,
- a sector label (superselection index),
- a statement about membership in a code subspace,
- or any decision problem definable from the available algebra of observables.

A decoder is said to *answer* a query if the reconstructed state reproduces the same operational statistics for the relevant family of observables/predicates. This makes explicit that “information about the system” is not merely  $S(\rho)$  but a structured family of semantic distinctions one would like to recover from the internal code.

### D. Total encoding/decoding: what “total” means

We now define the notion that will be *ruled out* by diagonal logic in the regimes of interest.

*Total decoding (definition).* Fix a class of queries  $\mathcal{Q}$  and an operational semantics mapping

$$\llbracket \cdot \rrbracket : \mathcal{Q} \times \mathcal{S}(\mathcal{H}_{\text{bulk}}) \rightarrow \mathcal{A}, \quad (10)$$

where  $\mathcal{A}$  is an answer space (numbers, distributions, labels, etc.). An *internal decoding scheme* for  $(\mathcal{E}, \mathcal{Q})$  is a decoder  $\mathcal{D}$  such that the composite  $\mathcal{D} \circ \mathcal{E}$  reproduces the semantics for all queries and all states in a specified domain  $\Omega \subseteq \mathcal{S}(\mathcal{H}_{\text{bulk}})$ :

$$\llbracket q, \rho \rrbracket = \llbracket q, (\mathcal{D} \circ \mathcal{E})(\rho) \rrbracket, \quad \forall q \in \mathcal{Q}, \forall \rho \in \Omega. \quad (11)$$

We call the decoder *total* (relative to  $\mathcal{Q}$  and  $\Omega$ ) if it is *uniformly defined* on  $\mathcal{E}(\Omega)$  and satisfies (11) for *all* queries in  $\mathcal{Q}$ .

*Total internal decoding (strengthened).* The qualifier “internal” means that the scheme  $(\mathcal{E}, \mathcal{D})$  is implementable within the same physical/theoretical framework that defines  $\text{Sys}$  and its semantics, without an external oracle. In particular, when  $\mathcal{Q}$  includes questions about the encoding/decoding process itself (self-reference), the demand that a *single* uniformly defined  $\mathcal{D}$  answer *all* such questions becomes the analogue of a “total halting decider” or “complete truth predicate” [17–19]. This is precisely the demand that will be shown to be untenable in sufficiently expressive self-referential regimes.

*Relative (non-total) decoding.* Importantly, nothing in physics requires totality in the above sense. In quantum error correction and in holographic reconstruction, one typically has:

$$(\mathcal{D} \circ \mathcal{E})(\rho) \approx \rho \quad \text{only for } \rho \in \Omega_{\text{code}}, \quad (12)$$

a restricted set of states (a code subspace, a sector, a semiclassical regime). This is *universality without totality*: reconstruction is robust within a controlled context but not globally uniform.

### E. Existence vs encodability vs decodability

We now explicitly separate the three notions that will recur throughout the paper.

(i) *Existence/preservation.* Information in the operational sense exists whenever the state has nontrivial correlations (e.g. nonzero mutual information (4)) and is preserved under unitary evolution on a closed system. The statement “information is preserved” is therefore, at base, a statement about the existence and evolution of correlations in  $\rho(t)$ .

(ii) *Encodability.* Encodability is the existence of an *internal representation* that makes a chosen class of semantic distinctions accessible from within a code. Formally, given  $\mathcal{Q}$  and  $\Omega$ , encodability asks whether there exists a code space and an encoding  $\mathcal{E}$  such that  $\mathcal{E}(\Omega)$  retains enough distinguishability to answer  $\mathcal{Q}$ .

(iii) *Decodability.* Decodability asks whether there exists a decoder  $\mathcal{D}$  satisfying (11) on the relevant domain. Total decodability is the strongest form: one decoder, defined uniformly, answers all queries on all states in the domain.

Diagonal arguments constrain (iii), not (i), and often only constrain (iii) once self-reference makes  $\mathcal{Q}$  sufficiently rich.

### F. What is *not* claimed

Because the paper’s claims can be misread, we state explicitly what is *not* asserted.

*No claim of information destruction.* We do *not* claim that information in the sense of correlations is destroyed. To the contrary, the modern semiclassical Page curve obtained via islands is consistent with global unitarity of the evaporation process (as reviewed in Section I).

*No claim of fundamental nonunitarity per se.* We do *not* claim that quantum dynamics must violate unitarity as a fundamental law. Our claim is more structural and strictly weaker: even if global unitarity holds, the demand for a *total internal decoding* of all semantic distinctions about a sufficiently expressive self-referential system is too strong and is obstructed by diagonal logic [17–19]. In later sections we will connect this semantic obstruction to HCCB as a physical mechanism that replaces global linear/unitary descriptions by coherence closure on accessible algebras [1, 2].

*Summary.* Information (correlations) may exist and be preserved, while information *about* the system may fail to be totally internalizable as a single globally coherent code admitting a uniform decoder. This is the precise sense in which “universality” need not imply “totality”, and it is the conceptual foundation for our subsequent diagonal and holographic analysis.

## III. SELF-REFERENCE AND THE LIMITS OF TOTAL ENCODING

### A. Goal and logical stance

The purpose of this section is to introduce, in a unified and logically transparent way, the *diagonal* mechanism that underlies the strongest impossibility results about “total internal decoding.” The main theme is simple but profound:

*Once a system is expressive enough to represent descriptions of its own behavior (or its own encodings) internally, any attempt to provide a **total** internal classifier/decoder for **all** semantic distinctions about the system generates, by diagonalization, a predicate that escapes classification.*

This is the shared logical skeleton behind:

- Cantor’s theorem (no set encodes all its subsets) [25],
- Gödel’s first incompleteness theorem (truth  $\neq$  provability) [17, 26],
- Turing’s halting theorem (behavior  $\neq$  decidability) [18, 27],
- Lawvere’s fixed-point theorem (categorical diagonalization) [19].

As emphasized in Section II, the obstruction concerns *totality of internal encodability/decodability of information about the system* (semantic predicates), not the existence of correlations or the possibility of restricted (relative) reconstruction.

### B. Self-reference as internal representability of descriptions

We now formalize the notion of *self-reference* at the level needed for diagonal arguments.

*Self-reference (operational form).* A system (or formal theory) is *self-referential* (for our purposes) if:

- (SR1) it admits an internal representation of descriptions of its own objects (e.g. encodings of predicates, programs, proofs, decoders),
- (SR2) it admits an internal mechanism to *apply* a description to an input (evaluation, execution, interpretation),
- (SR3) the class of admissible queries  $\mathcal{Q}$  (Section II) includes queries about this very representability/evaluation (i.e. the encoding is itself a valid subject of inquiry).

In logic this is realized by Gödel numbering (syntax becomes arithmetic) [17]; in computation by the existence of universal machines and program self-application [18]; and categorically by the availability of evaluation morphisms in a cartesian closed category [19].

*Why self-reference matters for “total decoding”.* Recall the notion of a *total internal decoder* from Section IID: a single, uniformly defined procedure that decides *all* semantic predicates of interest from an internal code. If (SR1)–(SR3) hold, the system can encode the statement “the decoder outputs 1 on the encoding of this predicate” and can form predicates that talk about (and flip) their own decoded value. This is the diagonal move.

### C. The diagonal argument as a general schema

We state an abstract schema that will be instantiated in Cantor/Gödel/Turing/Lawvere.

*Diagonal schema.* Suppose we have:

- a domain  $X$  of “names” or “codes”,
- a class of predicates on some space  $Y$  (often  $Y = X$ ),
- a map (an “enumerator” / “classifier” / “decoder”)

$$\Phi : X \longrightarrow \mathcal{P}(Y), \quad (13)$$

which claims to associate to each code  $x \in X$  a predicate  $\Phi(x) \subseteq Y$ , and

- a means of *self-application* allowing us to compare  $\Phi(x)$  with the value of  $\Phi(x)$  at  $x$  (this is the crucial self-reference step).

Define the diagonal predicate

$$D := \{ y \in Y : y \notin \Phi(y) \}. \quad (14)$$

Then  $D$  differs from every  $\Phi(x)$  at least at the point  $x$ , hence  $D \neq \Phi(x)$  for all  $x$ . Therefore  $\Phi$  is not surjective: the claimed classification misses at least one predicate.

This is the entire mechanism. The only nontrivial issue in each application is whether (SR1)–(SR3) allow the construction (14) internally and whether the meaning of  $\notin$  (negation) is available in the relevant setting.

### D. Cantor: no total powerset encoding

We begin with the cleanest diagonal theorem, which requires no computation or proof theory.

**Theorem III.1** (Cantor, diagonal form). *For any set  $X$ , there is no surjection  $f : X \rightarrow \mathcal{P}(X)$  onto its power set. Equivalently,  $|X| < |\mathcal{P}(X)|$ .*

*Proof.* Assume for contradiction that  $f : X \rightarrow \mathcal{P}(X)$  is surjective. Define the diagonal subset

$$D := \{x \in X : x \notin f(x)\} \in \mathcal{P}(X). \quad (15)$$

By surjectivity, there exists  $a \in X$  such that  $f(a) = D$ . Now ask whether  $a \in D$ . By definition (15),

$$a \in D \iff a \notin f(a).$$

But since  $f(a) = D$ , this becomes

$$a \in D \iff a \notin D,$$

a contradiction. Therefore no such surjection exists.  $\square$

*Interpretation for “encoding”.* Cantor’s theorem says: there is no *total* encoding of *all predicates on  $X$*  (i.e. all subsets of  $X$ ) by elements of  $X$  itself. This is the archetype of “no total internal classifier.” In later sections, the role of “subsets” will be played by predicates about states/observables, and the role of  $X$  by a space of codes/representations available internally. Cantor’s theorem thus provides the simplest intuition for why “totality” is dangerous in self-referential settings [25].

### E. Gödel: truth is not identical to provability

We next isolate the diagonal core of Gödel’s first incompleteness theorem.

*Setup.* Let  $T$  be a recursively axiomatized theory extending a weak fragment of arithmetic sufficient to represent primitive recursive functions (e.g. Robinson arithmetic  $Q$  suffices for the diagonal lemma; stronger hypotheses are used for standard incompleteness statements). Gödel’s key move is to arithmetize syntax: formulas and proofs are coded by natural numbers (Gödel numbers). This supplies (SR1).

*Provability predicate.* Because  $T$  is recursively axiomatized, there exists an arithmetic formula  $\text{Prov}_T(x)$  expressing “ $x$  is the Gödel number of a  $T$ -provable sentence”. Crucially,  $\text{Prov}_T$  is *internal* to arithmetic and therefore to  $T$  (SR1–SR2).

*Diagonal lemma (self-reference engine).* The diagonal lemma states that for any one-free-variable formula  $F(x)$  there exists a sentence  $G$  such that  $T$  proves

$$G \leftrightarrow F(\ulcorner G \urcorner), \quad (16)$$

where  $\ulcorner G \urcorner$  is the numeral coding the sentence  $G$ . This is the formal realization of self-application (SR2–SR3). A detailed presentation can be found in [26].

*Gödel sentence.* Choose  $F(x) := \neg \text{Prov}_T(x)$ . Then (16) yields a sentence  $G$  with

$$T \vdash G \leftrightarrow \neg \text{Prov}_T(\ulcorner G \urcorner). \quad (17)$$

Thus  $G$  asserts “ $G$  is not provable in  $T$ ”.

**Theorem III.2** (Gödel, first incompleteness (diagonal core)). *If  $T$  is consistent, then  $T \not\vdash G$ . Moreover, under standard mild additional hypotheses (e.g.  $\omega$ -consistency, or suitable soundness assumptions),  $T \not\vdash \neg G$  as well; hence  $T$  is incomplete.*

*Proof of  $T \not\vdash G$  under consistency.* Assume for contradiction that  $T \vdash G$ . Then  $T$  also proves  $\text{Prov}_T(\ulcorner G \urcorner)$ , since a proof of  $G$  is a witness to its provability. But (17) implies  $T \vdash \neg \text{Prov}_T(\ulcorner G \urcorner)$ . Thus  $T$  proves a contradiction. This contradicts the consistency of  $T$ . Hence  $T \not\vdash G$ .  $\square$

*Interpretation.* The theorem shows that the semantic notion “truth in the standard model” cannot be *totally* internalized as “provability in  $T$ ” once  $T$  is expressive enough to talk about its own proofs. In our language: “provability” is an attempted internal *decoder* for semantic truth about arithmetic; diagonalization produces a predicate ( $G$ ) that escapes total decoding. This is the direct analogue of Cantor’s  $D$  in the space of sentences rather than subsets.



### F. Turing: behavior is not identical to decidability

We now state the computational incarnation of the same diagonal mechanism.

*Setup.* Fix a standard model of computation (Turing machines). Let  $H(M, x)$  denote the predicate “machine  $M$  halts on input  $x$ ”.

**Theorem III.3** (Turing, halting problem). *There is no total computable function (algorithm)  $\text{Halt}(M, x) \in \{0, 1\}$  that returns 1 iff  $M$  halts on input  $x$  and returns 0 otherwise, for all  $(M, x)$ .*

*Diagonal proof.* Assume for contradiction that such a total decider  $\text{Halt}$  exists. Define a new machine  $D$  that, on input  $M$ , does the following:

1. computes  $\text{Halt}(M, M)$ ;
2. if  $\text{Halt}(M, M) = 1$ , then  $D$  loops forever;
3. if  $\text{Halt}(M, M) = 0$ , then  $D$  halts immediately.

This definition is effective if  $\text{Halt}$  is effective. Now evaluate  $D$  on its own code: consider  $D(D)$ . If  $\text{Halt}(D, D) = 1$ , then by definition  $D(D)$  loops forever, contradicting halting. If  $\text{Halt}(D, D) = 0$ , then by definition  $D(D)$  halts, contradicting non-halting. Thus  $\text{Halt}$  cannot exist.  $\square$

*Interpretation.* Here the “information about the system” is the semantic predicate “halts/does not halt.” The “internal decoder” is the hypothetical halting decider. Self-reference arises because programs can take programs as input, and because the decider is itself a program [18, 27]. The diagonal machine  $D$  plays the same role as Cantor’s  $D$  and Gödel’s  $G$ .

### G. Lawvere fixed points: diagonal logic in categorical form

Cantor/Gödel/Turing are often presented as separate theorems. Lawvere’s contribution was to exhibit a common categorical spine in cartesian closed categories (CCC) [19]. We present the conceptual statement and its proof in a form that makes the diagonal mechanism transparent and that will be adaptable later to categorical holography.

*Cartesian closed background.* Let  $\mathcal{C}$  be a CCC. For objects  $A, B \in \mathcal{C}$ , there exists an exponential object  $B^A$  and an *evaluation* morphism

$$\text{ev} : B^A \times A \rightarrow B, \quad (18)$$

which satisfies the usual adjunction  $\mathcal{C}(X \times A, B) \cong \mathcal{C}(X, B^A)$ .

*Weak point-surjectivity.* A morphism  $\phi : A \rightarrow B^A$  is called *weakly point-surjective* (wps) if for every morphism  $g : A \rightarrow B$  there exists an element  $a : 1 \rightarrow A$  (a “global point”) such that

$$g = \text{ev} \circ (\phi \circ a \times \text{id}_A). \quad (19)$$

Informally: every predicate/behavior  $g$  is realized by evaluating a *code*  $\phi(a)$  at its input. This is precisely the categorical form of “the system can internally represent all predicates.”

**Theorem III.4** (Lawvere fixed-point theorem (diagonal form)). *Let  $\mathcal{C}$  be a CCC, let  $A, B \in \mathcal{C}$ , and assume there exists a wps morphism  $\phi : A \rightarrow B^A$ . Then every endomorphism  $h : B \rightarrow B$  has a fixed point in the following sense: there exists a global element  $b : 1 \rightarrow B$  such that*

$$h \circ b = b. \quad (20)$$

*Proof.* Let  $h : B \rightarrow B$  be given. Consider the composite

$$A \xrightarrow{\Delta} A \times A \xrightarrow{\phi \times \text{id}_A} B^A \times A \xrightarrow{\text{ev}} B \xrightarrow{h} B, \quad (21)$$

where  $\Delta$  is the diagonal map. Denote this composite by  $g : A \rightarrow B$ :

$$g := h \circ \text{ev} \circ (\phi \times \text{id}_A) \circ \Delta. \quad (22)$$

By weak point-surjectivity of  $\phi$ , there exists a point  $a : 1 \rightarrow A$  such that

$$g = \text{ev} \circ (\phi \circ a \times \text{id}_A). \quad (23)$$

Now define  $b : 1 \rightarrow B$  by evaluating  $\phi(a)$  on  $a$ :

$$b := \text{ev} \circ (\phi \circ a \times a) : 1 \rightarrow B. \quad (24)$$

We compute  $h \circ b$ :

$$\begin{aligned} h \circ b &= h \circ \text{ev} \circ (\phi \circ a \times a) \\ &= (h \circ \text{ev} \circ (\phi \times \text{id}_A) \circ \Delta) \circ a \quad (\text{by definition of } \Delta \text{ and functoriality}) \\ &= g \circ a \quad (\text{by (22)}) \\ &= (\text{ev} \circ (\phi \circ a \times \text{id}_A)) \circ a \quad (\text{by (23)}) \\ &= \text{ev} \circ (\phi \circ a \times a) \\ &= b \quad (\text{by (24)}). \end{aligned} \quad (25)$$

Thus  $h \circ b = b$ , proving (20).  $\square$

*From fixed points to impossibility.* If the object  $B$  admits a “negation-like” endomorphism  $n : B \rightarrow B$  with no fixed point, then Theorem III.4 implies that no wps  $\phi$  can exist. This is precisely the categorical version of the diagonal contradiction: demanding total representability (wps) forces fixed points, which negation forbids [19]. In later sections, the role of “negation” will be played by self-referential predicates that flip the claimed decoder’s output, yielding non-totality.

## H. Key Proposition 1: the semantic no-go

We can now state, at an abstract level, the central meta-result we will carry into the holographic setting.

**Proposition III.5** (Semantic No-Go: no total internal decoder). *Let  $\text{Sys}$  be a system (physical or formal) admitting: (i) internal representations of predicates/decoders about  $\text{Sys}$ , (ii) internal evaluation/self-application, and (iii) a query class  $\mathcal{Q}$  rich enough to include predicates about the decoding process itself (self-reference). Then there is no single, uniform, internal decoding scheme that correctly decides all predicates in  $\mathcal{Q}$  for all states/instances in its intended domain. Equivalently, “information about the system” (semantic distinctions) cannot be totally internalized as a globally coherent, uniformly decodable code.*

*Proof sketch as a diagonal reduction.* Assume, for contradiction, that a total internal decoder exists for  $\mathcal{Q}$ . By (i)–(ii), predicates in  $\mathcal{Q}$  can be internally represented and evaluated on their own representations. By (iii), the system can form the diagonal predicate that asserts the negation (or complement) of the decoder’s prediction on the encoding of the predicate itself, in the same spirit as (14), (17), and Theorem III.3. This predicate is meaningful (it is a query in  $\mathcal{Q}$  by closure under self-reference) yet cannot be decided consistently by the supposed total decoder, yielding contradiction. Therefore no such total decoder exists.  $\square$

*Physics reading.* Proposition III.5 does *not* say that correlations do not exist or are destroyed. It says that once the decoding problem is internalized (the decoder is part of the system and can be queried), *totality* is too strong. This is exactly the structural pressure that appears in black-hole settings (Hawking radiation as an internal encoding; decoding as an internal physical process) and motivates why holographic reconstruction must be relative (code subspaces, sectors, state dependence) rather than global.

## IV. DEGREES OF SELF-REFERENTIALITY

### A. Goal: self-reference is not binary

Section III established a *semantic no-go*: once a system is sufficiently expressive and self-referential, *total internal decoding of information about the system* is obstructed by diagonal logic (Proposition III.5).

However, physical systems do not all sit at the same logical “distance” from the diagonal threshold. In this section we build a controlled *hierarchy of self-referentiality*, with the purpose of making three points precise:

1. Self-reference is a *graded structural property*: it strengthens as (i) internal representations of descriptions become richer, (ii) evaluation/self-application becomes more operationally unavoidable, and (iii) the set of admissible queries  $\mathcal{Q}$  becomes closed under reflection on the encoding/decoding process itself.
2. The strength of self-reference correlates with the necessity of *contextual* encoding/decoding: code-subspace dependence, sectorization, edge-mode extensions, and non-factorization are not pathologies but the expected ways in which physics avoids “totality”.
3. Black holes sit at (or near) the maximal end of this hierarchy because the horizon forces observer-splitting and because encoding/decoding procedures are inevitably internal to the same system (radiation, observers, and geometry are not separable). Complementarity [33] already anticipates this logic.

We will use this hierarchy to connect: (i) diagonal non-totality as a semantic constraint, (ii) holography/islands as a geometric realization of contextual decoding, and (iii) HCCB as a physical mechanism by which coherence closure replaces global totality.

### B. A structural yardstick: representation, evaluation, and closure of queries

To speak quantitatively, we isolate three ingredients whose presence can be graded in physical models. *(R) Internal representability.* How much of the system’s descriptive apparatus can be represented *within* the system? For arithmetic this is syntax encoded as numbers [17]; for computation it is programs as data [18]; for physics it is “states of the world” being represented inside subsystems (memory registers, records, radiation, boundary algebras).

*(E) Evaluation/self-application.* How directly can the system *apply* internal descriptions to internal data? In computation, this is execution; in category theory, it is evaluation (18) [19]; in physics, it is the operational ability of an observer-apparatus to act on recorded information and thereby change the joint state.

*(C) Closure of admissible queries.* Let  $\mathcal{Q}$  be the class of semantic predicates “about the system” that we insist must be decidable from a code (Section II). Self-reference strengthens when  $\mathcal{Q}$  is closed under:

- queries about *predictions* of the decoder itself,
- queries about *which code* is being used (choice of algebra/region/sector),
- queries about *consistency across observers* (Wigner-friend-type nesting).

In short: increasing closure pushes the system toward the diagonal regime.

These three axes allow us to define a hierarchy of self-referentiality levels relevant to physics.

### C. Level 0: external description (classical idealization)

*Definition (Level 0).* A system is Level 0 (non-self-referential) if:

- (R0) the system is described from an external vantage point (meta-level),
- (E0) no internal evaluation of its own descriptions is required,
- (C0) admissible queries  $\mathcal{Q}$  do not include predicates about the act of decoding/description itself.

*Interpretation.* Classical mechanics, treated as an externally modeled dynamical system, approximates this idealization: the “code” is an external mathematical description; the system is not asked to represent or decode itself. In this regime, total decoding is conceptually unobstructed because the decoder is external. Diagonal logic does not apply internally because (R)–(E)–(C) are not satisfied internally.

This level is not meant as a literal claim about nature, but as a baseline: it isolates what *disappears* once observers and encodings are internalized.

### D. Level 1: weak self-reference (quantum systems with external observers)

*Definition (Level 1).* A system is Level 1 (weakly self-referential) if:

- (R1) it admits internal records (measurement outcomes) but these are treated as external classical data for the purposes of decoding,
- (E1) evaluation of descriptions is performed externally (the observer/experimenters is not modeled as part of the quantum system),
- (C1)  $\mathcal{Q}$  does not include queries about the internal use of the theory by agents within the system.

*Interpretation.* This is the standard textbook regime in which quantum mechanics is applied to “the system” while the observer and measurement outcomes are taken as external. In this regime, decoding is effectively total for the chosen  $\mathcal{Q}$  because the decoder is not required to be internal. Many practical quantum information protocols live here, and it is also the regime where “unitarity” is typically treated as a globally valid dynamical principle.

### E. Level 2–3: strong self-reference (nested observers, gauge constraints, AQFT)

As soon as observers and subsystems are treated within the same quantum description, the system becomes strongly self-referential and diagonal pressures emerge. This happens in at least three major ways.

#### 1. Measurement chains and nested observers (Wigner-type self-reference)

Wigner’s friend scenarios make the logical structure explicit: an observer  $F$  measures a system, then a higher-level observer  $W$  treats  $(F + \text{system})$  as a quantum system. This introduces internal representability and evaluation of measurement records and their predictions. Wigner’s original discussion already highlights the tension between unitary evolution and collapse when observers are internalized [28]. Modern sharpenings, such as Frauchiger–Renner, show that if one insists on a single agent-independent quantum description of all agents reasoning about each other, then the theory “cannot consistently describe the use of itself” under a set of natural assumptions [29].

*Interpretation.* This is a clear physical instance of increasing (C): the query set now includes meta-queries about what other agents predict and infer. Total internal decoding becomes untenable in the strongest sense because the decoding map itself is part of the system that is being decoded (self-reference is operational, not just semantic).

#### 2. Gauge theories: non-factorization, centers, and edge data

Gauge constraints obstruct naive subsystem factorization. In lattice and continuum gauge theories, the gauge-invariant operator algebra associated to a spatial region typically has a nontrivial *center*, and different choices of algebra/“center prescription” correspond to different operational notions of localization [30]. Donnelly and Freidel show that defining local subsystems in gauge theory and gravity naturally requires introducing boundary/edge degrees of freedom to maintain gauge invariance and well-defined gluing [31].

*Interpretation.* This is strong self-reference in a structural (not psychological) sense: the attempt to define “the subsystem” and “its information” necessarily refers to the global constraints that define the theory itself. The query set  $\mathcal{Q}$  becomes contextual (depends on the chosen algebra/center and on boundary data), and therefore the notion of a *single total decoder* is already mathematically disfavored. Instead, one obtains sectorization and extensions (edge modes) as the correct language.

#### 3. AQFT and Type-III structure: locality without global density matrices

In Algebraic Quantum Field Theory (AQFT), the net of local von Neumann algebras is a primary object; local algebras in relativistic QFT are generically Type III, and thus do not admit a trace or

density-matrix description in the naive finite-dimensional sense [32]. This is a paradigmatic setting where “information” is naturally encoded in algebraic relations and relative entropy rather than in global state vectors of factorized subsystems.

*Interpretation.* AQFT provides an intrinsic form of strong self-reference: the operational content (local observables) and the global structure (net consistency, isotony, locality) are inseparable. The system is not “total” with respect to density-matrix-based encodings; instead, it forces a relational encoding language. This prepares the ground for later sections, where holography/islands will be interpreted as a further gravitational sharpening of this same phenomenon.

#### F. Level 4: maximal self-reference (black holes)

We now articulate why black holes sit at the extreme end of this hierarchy.

*Horizon-induced observer splitting.* A horizon divides observers into causally inequivalent classes. External observers cannot access interior operators directly, yet the Hawking radiation they measure is entangled with degrees of freedom that are naturally described as interior partners. Hence the “code” (radiation) is internal, and the semantics one wants to recover (interior information) concerns degrees of freedom that no single observer can access globally.

*Encoding and decoding are internal physical processes.* In an evaporating black hole, the encoding of interior information into radiation is not an external bookkeeping device: it is a physical process within the theory. Likewise, any decoding operation is a physical interaction performed by some subsystem (observer/apparatus) on the radiation, and therefore belongs to the same global quantum system it attempts to decode. This realizes (R)–(E)–(C) at maximal strength: the system contains observers, the observers attempt to encode/decode the system, and those operations feed back into the system itself. Black hole complementarity is precisely an early recognition that global simultaneity of descriptions is too strong and that consistency requires observer-relative descriptions [33].

*Interpretation.* Black holes therefore force the semantic no-go of Proposition III.5 into a geometric regime: “total decoding” becomes an operationally meaningful but impossible demand. This is the conceptual bridge to islands and entanglement wedge reconstruction: the theory avoids paradox not by destroying information, but by making decoding relative (code-subspace/sector/state/region dependent), i.e. by abandoning totality while preserving coherence in the minimal sense.

#### G. Relation to encodability and coherence (and to HCCB)

The above hierarchy can be read as a hierarchy of what the system demands from *coherence*:

- At Level 0–1, one may idealize global coherence as a single unitary linear semantics for all questions of interest (external decoder; low closure of  $\mathcal{Q}$ ).
- At Level 2–3, coherence must be formulated as *contextual gluing*: different agents/algebras/centers define different but overlapping domains of validity. The correct structure is not a single global code but a family with compatibility data.
- At Level 4 (black holes), the demand for global gluing becomes maximally inconsistent with the system’s internal expressivity. The only stable notion is *coherence closure*: local/sectorial consistency plus controlled failure of global totality.

This is exactly the conceptual place where Higher Categorical Coherence Breakdown (HCCB) enters: HCCB proposes that “global unitary linearity” is not the minimal requirement; rather, the minimal requirement is coherent gluing of local descriptions, and when higher coherence fails the effective dynamics becomes sectorized, history-dependent, and CP on accessible subalgebras [1, 2]. Black holes are not the only setting where this occurs, but they are the setting where the hierarchy reaches its sharpest geometric form.

*Takeaway.* Self-reference is graded. As it increases, “total encoding/decoding” is progressively replaced by relative encoding, sectorization, and coherence closure. Black holes sit at the extreme because the horizon makes the self-referential loop (system  $\rightarrow$  internal encoding  $\rightarrow$  internal decoding  $\rightarrow$  system) unavoidable.

## V. CATEGORICAL HOLOGRAPHY: UNIVERSALITY WITHOUT TOTALITY

### A. Goal and guiding question

The goal of this section is to translate the diagonal constraint of Section III into holographic language. Informally, “holography” suggests that boundary data encodes bulk physics, but Section III suggests that *total* internal decoding is incompatible with self-reference. Our task is therefore to formalize a notion of holographic encoding in which:

*the encoding may be universal (robust, representation-independent on an appropriate domain), while decoding cannot be total (globally uniform for all bulk predicates).*

We will express this in categorical terms and state a precise obstruction—our No-Total-Holography Proposition (Proposition V.1)—which is the holographic counterpart of Proposition III.5. Technically, we will only use standard categorical notions: functors, (co)limits, Yoneda-style probing, and the Lawvere diagonal mechanism [19, 34, 35].

### B. Bulk and boundary as categories of structures and probes

We begin by specifying the minimal structure needed to speak of “bulk”, “boundary”, and “encoding” in a representation-independent way.

*Bulk category  $\mathcal{B}$ .* The bulk category  $\mathcal{B}$  may be chosen according to the level of description:

- Objects may be bulk states (or state sectors), bulk effective theories, nets of algebras, or semiclassical geometries decorated with controlled excitations.
- Morphisms encode admissible maps between bulk descriptions: embeddings between regimes, coarse-grainings, restrictions to subregions, or structure-preserving maps.

We deliberately keep  $\mathcal{B}$  abstract, because the diagonal obstruction is structural.

*Probe/boundary category  $\mathcal{P}$ .* The boundary/probe category  $\mathcal{P}$  contains the accessible structures:

- boundary algebras, reduced states, modular data, detector response functionals, asymptotic charges, or other operationally meaningful observables,
- morphisms encode admissible transformations between probe contexts (inclusions of algebras, coarse-grainings, restriction/extension of accessible data).

*Encoding functor.* A categorical holography map is an *encoding functor*

$$U : \mathcal{B} \longrightarrow \mathcal{P}, \quad (26)$$

which assigns to each bulk object its accessible boundary/probe avatar and transports bulk morphisms to boundary/probe morphisms. This is the categorical abstraction of statements such as “the boundary algebra/state is determined by the bulk” or “radiation encodes interior information”.

*Universality vs totality (preview).* The key conceptual distinction is:

- **Universality:**  $U$  is robust under re-description (naturality), and captures a forced core structure (factorization property) on an appropriate domain.
- **Totality:**  $U$  would classify *all* bulk predicates/observables globally and uniformly from boundary data.

Diagonal logic does not forbid universality; it forbids totality once self-reference is present.

### C. Predicates, probes, and the Yoneda viewpoint

To speak meaningfully about “information about the system” (Section II) in categorical terms, we need a notion of *predicates* or *properties* of bulk objects.

*Predicates as subobjects (one standard choice).* If  $\mathcal{B}$  has a well-behaved subobject theory (e.g.  $\mathcal{B}$  is a topos or at least admits a stable notion of mono), one can define the predicates on an object  $X \in \mathcal{B}$  as its subobjects:

$$\text{Pred}_{\mathcal{B}}(X) := \text{Sub}_{\mathcal{B}}(X), \quad (27)$$

where  $\text{Sub}_{\mathcal{B}}(X)$  denotes the poset of subobjects of  $X$  up to isomorphism. This is a canonical way to represent “yes/no” properties of the object. (For a systematic development of subobject logic and its relation to categorical semantics, see [36, 37].)

*Predicates as functors of points (probe semantics).* Even when subobjects are subtle, the Yoneda perspective provides an alternative: an object  $X$  is determined (up to isomorphism) by its behavior against all probes [34, 35]. Concretely, for each  $X \in \mathcal{B}$  consider the presheaf

$$h_X := \text{Hom}_{\mathcal{B}}(-, X) : \mathcal{B}^{\text{op}} \rightarrow \mathbf{Set}. \quad (28)$$

A “predicate about  $X$  detectable by probes” can be modeled as a subfunctor or as a natural transformation into a truth-value object in a suitable topos of presheaves. We do not need the full internal logic here; we only need the guiding lesson:

*A would-be holographic encoding is “complete” only relative to a chosen class of probes.  
Changing the probe class changes what counts as a predicate.*

This is already the first appearance of *non-totality*: there is no canonical choice of “all probes” in physics once observers are internalized.

#### D. What “total holography” would require (and why it is too strong)

We now formalize a strong (and intentionally over-demanding) notion of “total holography” to isolate precisely what will be obstructed.

*A (too-strong) notion of total holography.* Fix a predicate assignment  $\text{Pred}_{\mathcal{B}}$  on  $\mathcal{B}$  (e.g. (27) or a probe-based substitute), and similarly  $\text{Pred}_{\mathcal{P}}$  on  $\mathcal{P}$ . A natural way to express “boundary data classifies bulk predicates” is the existence of a natural isomorphism

$$\Theta_X : \text{Pred}_{\mathcal{B}}(X) \xrightarrow{\sim} \text{Pred}_{\mathcal{P}}(U(X)), \quad \text{natural in } X \in \mathcal{B}. \quad (29)$$

This says: for every bulk object  $X$ , every predicate about  $X$  is representable as a predicate about its boundary encoding  $U(X)$ , in a manner stable under morphisms (naturality).

If one further insists that this classification holds for *all* objects, and that the predicate logic is closed under self-reference (queries about the encoding/decoding scheme itself; cf. (SR3) in Section III B), then one has demanded a *total internal decoder of all bulk predicates*.

*Why (29) is too strong in self-referential regimes.* There are two distinct reasons:

- (1) **Physics reason (contextual probes).** In holography and QFT, the set of operational probes is context-dependent (choice of region, algebra, sector, code subspace). Even before diagonal logic, expecting a single probe category  $\mathcal{P}$  to classify all bulk predicates is already a strong and often false idealization [10, 11].
- (2) **Logic reason (diagonal obstruction).** If  $\text{Pred}_{\mathcal{P}}$  is expressive enough to internalize predicates about (29) itself, then diagonal logic produces a predicate that cannot be consistently classified. This is the categorical counterpart of Cantor, Gödel, and Turing (Section III) [17–19, 25].

#### E. Diagonal obstruction in categorical holography

We now state the diagonal logic in the present functorial setting. To avoid unnecessary technicalities, we phrase the argument at the level of internal “predicate classifiers” rather than committing to a single formalism (topos, CCC, etc.).

*Self-reference requirement (holographic form).* The diagonal mechanism requires that the system be able to express predicates that reference the encoding/decoding operation. In categorical terms, this means (schematically):

- the probe category  $\mathcal{P}$  supports a sufficiently rich predicate logic (e.g. subobject classifier in a topos, or evaluation in a CCC, or an equivalent representability structure) [19, 36, 37];
- the encoding functor  $U$  is rich enough that the image  $U(X)$  carries internal representations of predicate claims about  $U$  itself (closure of  $\mathcal{Q}$ ).

Under these conditions, demanding total classification (29) forces a diagonal predicate, exactly as in (14).

*The essential idea.* If  $\Theta_X$  exists for all  $X$  and the predicate logic includes negation/complement in the relevant sense, then for each  $X$  one can define a “liar-style” predicate

$$D_X := \neg \Theta_X^{-1}(\text{“}D_X \text{ holds of the code that names } D_X\text{”}), \quad (30)$$

whose content is: “this predicate fails exactly when the boundary classifier says it holds of itself.” By construction,  $D_X$  cannot be consistently classified by  $\Theta_X$ . This is the same structure as Cantor’s  $D$  (Theorem III.1), Gödel’s  $G$  (Theorem III.2), and Turing’s diagonal machine (Theorem III.3).

The statement (30) is schematic because its fully formal expression depends on the precise internal logic adopted for predicates in  $\mathcal{P}$ ; the unifying categorical engine behind all such diagonal constructions is Lawvere’s fixed point theorem [19].

## F. Key Proposition 2: No-Total-Holography

**Proposition V.1** (No-Total-Holography). *Let  $U : \mathcal{B} \rightarrow \mathcal{P}$  be an encoding functor. Assume:*

- (H1) (Predicate expressivity) *The probe category  $\mathcal{P}$  supports a sufficiently rich internal predicate logic to represent predicates about objects in  $\mathcal{P}$ , including predicates about the action of  $U$  and its would-be reconstruction (e.g. via a topos/subobject classifier or via CCC evaluation as in Lawvere’s framework) [19, 36, 37].*
- (H2) (Self-reference) *The class of admissible queries  $\mathcal{Q}$  is closed under reflection on the encoding itself: predicates about the output of the would-be decoder applied to encodings are themselves admissible.*
- (H3) (Total classification demand) *There exists a natural family of maps  $\Theta_X$  as in (29) that purports to classify all bulk predicates from boundary predicates for all  $X$  in the intended domain.*

*Then the demand (H3) cannot hold simultaneously with (H1)–(H2): there is no encoding functor  $U$  that is a total classifier of bulk predicates in a self-referential regime. Equivalently, holography can be universal only relatively (restricted to contexts such as code subspaces/sectors/probe families), but not total.*

*Proof sketch (diagonal reduction).* Assume (H1)–(H3). By (H1) the probe logic can represent predicates about  $U(X)$  and, crucially, about the act of decoding/classifying itself. By (H2) such “meta-predicates” are admissible queries and are therefore subject to the classification map  $\Theta_X$ . Now form the diagonal predicate  $D_X$  whose truth value is defined to disagree with the classifier’s own prediction about  $D_X$  (as in the general diagonal schema (14) and the schematic liar predicate (30)). By construction,  $D_X$  cannot be consistently classified by  $\Theta_X$  without contradiction. Hence (H3) fails. The only way to avoid contradiction is to weaken totality: restrict the domain (code subspace), restrict the query class, or accept contextual/state-dependent classification.  $\square$

*Interpretation and physics reading.* Proposition V.1 is not an “anti-holography” statement. It says that, in self-referential regimes, boundary encoding cannot be promoted to a single globally defined decoder that decides all bulk predicates uniformly. Instead, one expects precisely the structures seen in modern holography:

- reconstruction valid on code subspaces (relative universality) [10],
- entanglement-wedge/state/region dependence [6, 9],
- algebra/center/edge-mode extensions in gauge/gravity localization [31],
- and, in evaporating black holes, island-induced reconstruction transitions [13, 14, 16].

All of these are consistent realizations of “universality without totality.”



### G. Why “universality” survives: abduction and factorization

The diagonal obstruction targets only totality, not universality. In the abductive (universal-property) sense discussed earlier, universality means: *there exists a minimal core structure through which all successful reconstructions factor*. Categorically, such a core is characterized by factorization/universal properties, not by total classification of all predicates. This distinction is the backbone of our narrative: holography can remain a powerful universal principle precisely because it does *not* promise total internal decoding.

## VI. ENTANGLEMENT WEDGE RECONSTRUCTION AS RELATIVE UNIVERSALITY

### A. Goal and relation to No-Total-Holography

Section V established that “total holography”—a single globally defined decoder/classifier for *all* bulk predicates—is too strong in self-referential regimes (Proposition V.1). The purpose of this section is to show that modern holographic reconstruction already implements the correct logic: it provides *relative universality* rather than totality.

Concretely, bulk reconstruction is best understood as a *quantum error correction* (QEC) phenomenon: bulk degrees of freedom are redundantly encoded in boundary degrees of freedom, and the existence of recovery maps depends on the *code subspace*, the *boundary region*, and (in gravitational settings) sometimes the *state* through which the semiclassical geometry is defined. This is not a defect; it is precisely the form of universality compatible with diagonal constraints.

### B. Quantum error correction perspective: encoding is physical, decoding is relative

The QEC viewpoint on AdS/CFT was articulated clearly in [10] and has become standard: a bulk effective Hilbert space (or algebra) is encoded into the boundary Hilbert space in a way analogous to a quantum code.

*Code subspace.* Let  $\mathcal{H}_{\text{CFT}}$  denote the boundary Hilbert space. A *code subspace*  $\mathcal{H}_{\text{code}} \subseteq \mathcal{H}_{\text{CFT}}$  models the subspace of states admitting a semiclassical bulk interpretation (low-energy excitations on a fixed background, or a controlled family of backgrounds). The inclusion isometry

$$V : \mathcal{H}_{\text{code}} \hookrightarrow \mathcal{H}_{\text{CFT}} \quad (31)$$

plays the role of an *encoding* (in the sense of (7)): for a code state  $\rho_{\text{code}}$ , the corresponding physical boundary state is

$$\rho_{\text{CFT}} = V \rho_{\text{code}} V^\dagger. \quad (32)$$

The crucial point is that  $V$  is not claimed to encode *all* of  $\mathcal{H}_{\text{CFT}}$  into a bulk; it encodes a *restricted* domain. This restriction is the simplest, most concrete manifestation of “universality without totality.”

*Subsystem recovery.* Let the boundary degrees of freedom be partitioned (at least conceptually) into a region  $R$  and its complement  $\bar{R}$  with  $\mathcal{H}_{\text{CFT}} \cong \mathcal{H}_R \otimes \mathcal{H}_{\bar{R}}$ . A central QEC question is: given access only to  $R$ , can one *recover* certain code observables? This is precisely the operational meaning of “bulk reconstruction on a boundary region.”

### C. Operator algebra quantum error correction and recovery criteria

The most robust formulation uses operator-algebra quantum error correction (OAQEC), which naturally matches holography’s emphasis on reconstructing *algebras of operators* rather than arbitrary states. A convenient entry point is the work relating boundary and bulk relative entropies [11] and the subsequent development of operator algebra reconstruction in the entanglement wedge framework [38].

*Algebraic reconstruction statement (informal).* Let  $\mathcal{A}_{\text{bulk}}(W)$  be the algebra of (low-energy) bulk operators supported in a bulk region  $W$  (typically an entanglement wedge), and let  $\mathcal{A}_R$  be the boundary

operator algebra on  $R$ . Entanglement wedge reconstruction asserts that, under appropriate conditions, for each  $O \in \mathcal{A}_{\text{bulk}}(EW(R))$  there exists an operator  $\tilde{O} \in \mathcal{A}_R$  such that

$$V^\dagger \tilde{O} V = O \quad \text{on } \mathcal{H}_{\text{code}}. \quad (33)$$

Equation (33) is the precise meaning of “ $O$  is reconstructible from  $R$ ” within the code subspace.

*Relative entropy control.* A powerful characterization of when reconstruction holds is via relative entropy. One formulation (in a suitable regime) is that relative entropy of reduced states on  $R$  equals relative entropy of corresponding bulk states in the entanglement wedge [11]. This “entropy equality” is a strong compatibility condition indicating that  $R$  contains exactly the information needed to distinguish code states as far as bulk observables in  $EW(R)$  are concerned. Because relative entropy is monotone under channels (6), equalities of this type are highly constraining and naturally express “recoverability”.

*OAQEC language (sketch).* In OAQEC one corrects an algebra  $\mathcal{A}$  rather than a full code space. The correctability condition can be phrased in terms of commutants and the action of the noise channel on the code (see e.g. [39]). In holography, the “noise” is the restriction (partial trace) to the accessible region  $R$ , and the algebra to be corrected is the bulk algebra associated to  $EW(R)$ . The key point is: *the existence of recovery depends on the algebra and the code.*

#### D. Why state- and region-dependence are not flaws

The dependence of reconstruction on region and state is often presented as puzzling. In our framework it is expected and conceptually necessary.

*Region-dependence.* In QEC, which subsystem can correct which information is a structural feature of the code. Different regions  $R$  have different error-correcting capabilities. Thus reconstruction being region-dependent is not a failure of holography; it is the precise content of holography-as-QEC [10, 38].

*State-dependence.* The code subspace itself is commonly defined relative to a reference semiclassical geometry. In gravitational settings, the map between bulk effective degrees of freedom and boundary operators can depend on the background state. This is not an “ad hoc” feature; it is a manifestation of working in a restricted domain where bulk effective field theory is valid. When one tries to enlarge the domain toward “all states,” diagonal obstructions appear (Section III) and totality must fail. Debates about state-dependent interior operators can be read as concrete instantiations of this general non-totally pressure.

*Relative universality.* The correct conceptual status of reconstruction is therefore:

*Universality holds **within** a context (code subspace, sector, region), not as a global, state-independent decoder for all bulk predicates.*

This is exactly what Proposition V.1 predicts should happen once self-reference is present.

#### E. Non-reconstructible operators as diagonal remainders

We now connect the QEC viewpoint to diagonal logic. Diagonal arguments guarantee that, beyond a certain expressive threshold, any attempt at total classification produces predicates/operators that escape. In holography, the analogue is:

*There exist bulk operators/predicates that are not reconstructible from a given boundary region (or not reconstructible in a single state-independent manner) because doing so would constitute a total internal decoder.*

*Two precise senses of “non-reconstructible.”*

- (i) **Outside the entanglement wedge:** if an operator is supported outside  $EW(R)$ , it is not expected to be representable on  $R$  without enlarging the accessible algebra. This is geometric non-reconstructibility.
- (ii) **Outside the code domain:** even if an operator is reconstructible on a small code subspace, there is no guarantee the same reconstruction exists on a much larger state space. Attempting to define a single reconstruction for *all* states is the totality demand obstructed by diagonal logic.

*Diagonal remainder as a structural prediction.* From our standpoint, “non-reconstructible operators” are not mysteries; they are the necessary remainder left when one imposes consistency constraints strong enough to avoid total internal decoding. This remainder may manifest as:

- superselection/center data (edge modes),
- contextual dependence of the reconstruction map,
- or intrinsic limitations on jointly measurable interior observables.

All of these appear across gauge theory, AQFT, and holography (Section IV).

## F. Takeaway

Entanglement wedge reconstruction, understood through quantum error correction, is an explicit realization of “universality without totality”:

- Bulk operators are reconstructible on a boundary region  $R$  *relative to* a code subspace and relative to an algebra of interest.
- Region- and state-dependence are not pathologies; they are structural necessities once self-reference rules out total decoding.
- Non-reconstructible operators are expected diagonal remainders—the precise place where the system refuses to close into a single global decoder.

This prepares the ground for Section VII, where islands will be interpreted as a geometric mechanism that adjusts the reconstructible region itself to maintain consistency in an evaporating black hole.

## VII. ISLANDS AS A PHYSICAL REGULARIZATION OF NON-TOTALITY

### A. Goal and logical stance

Sections III–VI developed a unified message: *self-referential regimes obstruct total internal decoding* (Proposition III.5), and modern holography already realizes this as *relative universality* via code subspaces and entanglement wedge reconstruction (Section VI) [10, 13, 38]. The purpose of this section is to show that the *island mechanism* is not merely a technical trick to reproduce the Page curve, but can be understood as a *physical regularization of non-totally*:

*islands implement, in semiclassical gravity, the only consistent alternative to “total decoding” by making reconstruction contextual and regime-dependent.*

This will be stated precisely as Proposition VII.1 below.

### B. The QES/island prescription: statement and meaning

We first review the quantum extremal surface (QES) framework and the island formula at the level needed for our logical narrative.

*From RT/HRT to quantum extremal surfaces.* For holographic CFT states with a semiclassical bulk dual, the Ryu–Takayanagi (RT) prescription [6] and its covariant Hubeny–Rangamani–Takayanagi (HRT) generalization [7] propose that the von Neumann entropy of a boundary region  $R$  is given, at leading order in  $1/G_N$ , by the area of an extremal surface  $\gamma_R$  homologous to  $R$ :

$$S(R) \approx \frac{\text{Area}(\gamma_R)}{4G_N}. \quad (34)$$

Quantum corrections were incorporated by Faulkner–Lewkowycz–Maldacena [8], leading to the *generalized entropy* functional

$$S_{\text{gen}}(R; \gamma) := \frac{\text{Area}(\gamma)}{4G_N} + S_{\text{bulk}}(\Sigma_\gamma), \quad (35)$$

where  $\Sigma_\gamma$  is a bulk region bounded by  $R \cup \gamma$  and  $S_{\text{bulk}}$  is the bulk entanglement entropy of quantum fields across  $\gamma$  in the semiclassical state.

Engelhardt and Wall refined this into the QES prescription [9]: the relevant surface is a *quantum extremal surface*, i.e. a surface  $\gamma$  that extremizes  $S_{\text{gen}}$ , and the entropy is given by evaluating (35) on the minimizing quantum extremal surface. Formally:

$$S(R) = \min_{\gamma \in \text{Ext}} \left[ \frac{\text{Area}(\gamma)}{4G_N} + S_{\text{bulk}}(\Sigma_\gamma) \right], \quad (36)$$

where Ext denotes the set of candidate (quantum) extremal surfaces.

*Islands.* In evaporating black hole settings, the region  $R$  of interest is typically a region of *radiation* (outside the black hole), and one allows the bulk region  $\Sigma_\gamma$  to include an “island” region  $I$  inside (or near) the black hole. The island formula is a specialization of (36):

$$S(R) = \min_I \text{ext} \left[ \frac{\text{Area}(\partial I)}{4G_N} + S_{\text{bulk}}(R \cup I) \right], \quad (37)$$

where  $I$  ranges over candidate island regions,  $\partial I$  is their boundary (a QES), and  $S_{\text{bulk}}(R \cup I)$  is the bulk entropy of quantum fields on  $R \cup I$ . In the modern derivations, (37) arises from gravitational replica computations and replica wormholes [14–16].

*A key conceptual point.* The optimization in (37) means that the entropy, and therefore the entanglement wedge of the radiation, is determined by a *variational principle*. Hence the *reconstruction region itself* becomes a *state- and regime-dependent output* rather than a fixed input. This is precisely the structural feature we need: non-totality is not patched; it is *regulated* by allowing the effective decoding region to adapt.

### C. Page curve from islands: the mechanism in one line

We briefly recall how islands reproduce Page behavior [4].

*Two competing saddles.* In an evaporating black hole, there are typically two relevant saddles for (37):

(i) **No-island saddle:**  $I = \emptyset$ , giving

$$S_{\text{no-island}}(R) = S_{\text{bulk}}(R), \quad (38)$$

which grows with time as Hawking quanta accumulate.

(ii) **Island saddle:**  $I \neq \emptyset$ , giving

$$S_{\text{island}}(R) = \frac{\text{Area}(\partial I)}{4G_N} + S_{\text{bulk}}(R \cup I), \quad (39)$$

which can remain bounded and eventually decrease, because the bulk entropy of  $R \cup I$  does not grow indefinitely when  $I$  captures the interior partners.

The min/ext prescription (37) selects the smaller of these, producing a Page-like rise-and-fall.

*Logical reading.* From our perspective, this competition is the physical manifestation of *non-totality*: the naive (no-island) decoding picture treats the radiation as a fixed subsystem from which one attempts total recovery; the island saddle corrects the notion of “what is encoded in the radiation” by making the reconstructible region state-dependent. Thus the Page transition is a *context switch* in reconstruction, not an information-loss event.

### D. Reconstruction region as a state-dependent output

Entanglement wedge reconstruction (Section VI) tells us that bulk operators in  $EW(R)$  are reconstructible from  $R$  (relative to a code subspace) [38]. The island prescription changes  $EW(R)$  itself.

*Entanglement wedge with islands.* In the presence of an island  $I$ , the entanglement wedge of the radiation includes  $I$ , so operators supported in  $I$  become reconstructible from the radiation algebra (again, relative to a suitable code domain). Penington emphasized this perspective early in the island era [13].

*Why this is not “cheating”.* The key is that the map “boundary region  $\rightarrow$  reconstructible bulk region” is not fixed by kinematics alone; it is selected by the variational principle (37), which depends on the state (via  $S_{\text{bulk}}$ ) and on the semiclassical geometry (via the area term). Thus, reconstruction is *not a single global functor* valid for all regimes; it is a *family* of reconstruction maps indexed by state/sector/context. This is exactly the pattern predicted by Proposition V.1.

### E. Monogamy, AMPS, and the pressure toward totality

The AMPS firewall argument [5] can be read as diagnosing an inconsistent simultaneous demand:

- purity/unitarity of the radiation after the Page time,
- semiclassical EFT entanglement across the horizon,
- and a single global factorization/decoding picture in which the same interior mode is simultaneously purified by both early radiation and near-horizon partners.

This demand is a form of *totality*: it implicitly assumes that the decoding map from radiation to interior can be treated as globally valid and simultaneously compatible with the semiclassical horizon description.

From the viewpoint of Sections III–V, this is exactly the regime where diagonal/self-reference constraints are expected to bite: the system is attempting to close on a single globally coherent decoder that works for all questions and all observers at once. Islands provide a concrete resolution by changing the effective reconstruction region so that one does not demand incompatible simultaneity of encodings. In other words, islands implement the abandonment of totality required for consistency.

### F. Key Proposition 3: Islands as a Diagonal Resolution

**Proposition VII.1** (Islands as a diagonal resolution of non-totality). *In semiclassical evaporating black hole settings where:*

- (I1) *the Hawking radiation  $R$  is treated as an internal code subsystem whose entropy is computed by a generalized entropy extremization principle (37), and*
- (I2) *entanglement wedge reconstruction is valid on an appropriate code domain (so that bulk operators in  $EW(R)$  can be represented on  $R$ ) [38],*

*the island mechanism provides a physical realization of the semantic non-totality constraints of Section III and Proposition V.1:*

Island formation implements a consistent alternative to total internal decoding by making the reconstructible bulk region a state- and regime-dependent output. Consequently, information about the system may be preserved and encoded in correlations without admitting a single global, uniform decoder valid across all regimes.

*Argument (structural, not a new replica derivation).* We do not re-derive the replica wormhole calculation; those derivations are in [14–16]. Instead we show the logical role of islands. Assume one insists on a fixed notion of reconstruction from the radiation (a fixed decoding region and a globally uniform decoder). In the post-Page regime, this fixed-decoder demand conflicts with the simultaneous semiclassical entanglement structure at the horizon (AMPS-type pressure) [5]. By Section III, such a demand is precisely the kind of “total internal decoder” assumption that becomes inconsistent in maximally self-referential settings.

The island formula (37) replaces the fixed-decoder demand by a variationally selected entanglement wedge  $EW(R)$  that can include an interior island  $I$ . Thus the effective decoding map is not global but contextually selected (state/regime dependent). This is exactly the weakening of totality prescribed by Proposition V.1. Therefore island formation constitutes a physical regularization of non-totality: it preserves unitarity-compatible Page behavior [4] without requiring a single global internal decoder of all interior predicates from the radiation alone.  $\square$

## G. Takeaway

Islands make precise what our diagonal/categorical narrative predicts:

- The radiation can encode information about the system and yield a Page curve consistent with unitarity, but this does *not* imply the existence of a single globally defined decoder.
- The reconstructible region is selected by a state-dependent extremization principle; hence reconstruction is inherently contextual and regime-dependent.
- This contextuality is not an ad hoc patch; it is the physically realized alternative to the impossible demand of total internal decoding in a maximally self-referential system.

In the next section we connect this to Higher Categorical Coherence Breakdown (HCCB) as a general physical mechanism by which coherence closure replaces global totality.

## VIII. HIGHER CATEGORICAL COHERENCE BREAKDOWN (HCCB)

### A. Goal: from semantic non-totality to a physical mechanism

Sections III–VII established a *semantic* constraint: in sufficiently expressive self-referential regimes, one cannot demand *total internal decoding* of all semantic predicates about the system (Proposition III.5), and holography/islands implement a consistent alternative by making reconstruction *contextual* and *regime-dependent* (Propositions V.1 and VII.1). The remaining question is physical:

*What dynamical/structural mechanism produces (and stabilizes) this non-totality in the actual physics, while preserving local consistency and operational predictability?*

Higher Categorical Coherence Breakdown (HCCB) is our proposed answer [1, 2]. In this section we (i) isolate the notion of *coherence closure* as the minimal consistency requirement, (ii) explain why demanding global *symmetry/unity/totality* is stronger than coherence and is generically obstructed, (iii) show how breakdown of higher coherence naturally yields *sectorization* and *history dependence*, and (iv) show why the operational effective dynamics becomes *completely positive* (CP) and typically *semigroup-like* on accessible algebras.

### B. Coherence vs symmetry vs totality

We first separate three notions that are often conflated:

*Coherence (minimal consistency).* In categorical language, coherence is the existence of consistent gluing/composition data: commuting diagrams and higher coherence conditions ensure that local identifications compose consistently [34]. Coherence is fundamentally about *closure under composition*: different ways of assembling a composite process yield canonically equivalent results.

*Symmetry (a sufficient but not necessary condition).* Symmetry is a stronger property: invariance under a group (or groupoid, or higher group) action. Symmetry typically *implies* coherence (there are canonical identifications), but coherence can exist without a global symmetry principle. This is precisely the guiding insight we have emphasized in our broader program: *symmetry is sufficient but not necessary for coherent gluing; coherence/closure is the minimal requirement.*

*Totality (global closure under all queries).* Totality is stronger still: it demands that a single global structure internalizes and decides *all* semantic predicates about the system (Sections II and III). Diagonal logic shows that in self-referential regimes, totality is generically impossible. Thus, the correct physical demand is not totality but coherence closure.

### C. Higher coherence and gluing: the mathematical core

We now state the structural idea of HCCB.

*Why “higher” coherence?* Ordinary categorical coherence (1-categorical) controls associativity/unitarity up to unique isomorphism (Mac Lane coherence) [34]. However, many physical constructions involve *families* of identifications and transformations between transformations: gauge choices, renormalization prescriptions, observer-dependent coarse-grainings, subsystem embeddings, and code-subspace reconstructions. These are naturally organized not in a mere category but in a *2-category* or higher category (bicategory /  $(\infty, 1)$ -category), where:

- objects are contexts (sectors, algebras, code subspaces, observers, scales),
- 1-morphisms are admissible transitions between contexts (coarse-grainings, embeddings, encodings, reconstructions),
- 2-morphisms (and higher) are *coherence data* (natural transformations between 1-morphisms, higher homotopies between transformations).

Standard references for coherence in higher categories include [40].

*Coherence closure vs coherence breakdown.* Coherence closure means that all higher associativity/compatibility constraints can be satisfied (possibly up to controlled equivalence) so that any composite of transitions is well-defined independent of parenthesization/path. HCCB posits that, in sufficiently rich physical settings, these higher coherence conditions can fail: there exist closed loops in context space whose composite coherence data does not reduce to an identity (or does so only up to a nontrivial defect). In the physics language, the defect is a *coherence curvature* or *holonomy gap* in the space of descriptions.

*Local consistency, global obstruction.* A key point is that coherence breakdown can occur even when each local piece is perfectly well-defined: each local patch of description is unitary/linear, but gluing them globally requires higher coherence that may be obstructed. This is the structural analogue of:

- gauge/gribov-type global obstructions,
- sheaf/descent obstructions (no global section),
- diagonal obstructions (no total classifier).

In other words, HCCB is the *physical* avatar of “universality without totality.”

#### D. From coherence defects to loss of global unitary/linear semantics

We now explain why coherence breakdown leads to an effective breakdown of global unitarity and linearity *as a single total semantics*, without claiming that local unitarity must fail in each patch.

*Patchwise unitarity.* Suppose each context  $c$  carries a Hilbert-space (or algebraic) description with a unitary evolution  $U_c(t)$ . If contexts are related by intertwiners (change-of-description maps)  $W_{c \rightarrow c'}$ , then global consistency would require higher coherence conditions such as

$$W_{c \rightarrow c''} \simeq W_{c' \rightarrow c''} \circ W_{c \rightarrow c'}, \quad (40)$$

and analogous higher relations for the intertwiners themselves. Coherence breakdown means that around some loop  $c \rightarrow c' \rightarrow c'' \rightarrow c$  one obtains a nontrivial defect:

$$W_{c \rightarrow c} := W_{c'' \rightarrow c} \circ W_{c' \rightarrow c''} \circ W_{c \rightarrow c'} \not\simeq \text{id}_c. \quad (41)$$

This defect obstructs the existence of a *single globally defined* unitary dynamics on a single Hilbert space that simultaneously represents all contexts.

*Why linearity breaks globally.* Linearity is a property of a chosen state space (vector space/Hilbert space) and a chosen evolution map acting linearly on it. If there is no globally valid identification between contexts (due to higher coherence defects), then there is no single globally valid linear state space on which all dynamics can be represented. Instead one has:

- a family of linear spaces (one per context) plus nontrivial gluing,
- and defects that appear as effective nonlinearities when pushed into a single chart.

Thus “breakdown of linearity” is not posited ad hoc; it is an emergent descriptor of attempting to force a global linear chart on a space that only admits local charts with nontrivial holonomy.

## E. Emergence of CP flows, sectorization, and history dependence

We now connect coherence defects to the operational appearance of CP dynamics and memory.

*Accessible algebras and coarse-graining.* In practice, observers access a subalgebra  $\mathcal{A}_{\text{acc}}$  of the full algebra (or a reduced state on a subsystem). The act of restricting to accessible degrees of freedom is a channel (Section II):

$$\rho \mapsto \rho_{\text{acc}} = \Phi(\rho), \quad \Phi \text{ CPTP.} \quad (42)$$

If coherence defects prevent a globally consistent unitary identification across contexts, then from the viewpoint of  $\mathcal{A}_{\text{acc}}$  the effective dynamics is generically non-unitary. The correct structural class for such reduced dynamics is *completely positive* maps, because complete positivity is precisely what guarantees consistency under extension by ancillas and avoids negative probabilities (Section II).

*Kraus/Stinespring representation.* Any CPTP map admits a Kraus representation

$$\Phi(\rho) = \sum_k K_k \rho K_k^\dagger, \quad \sum_k K_k^\dagger K_k = \mathbf{1}, \quad (43)$$

equivalently a Stinespring dilation (8) [24, 41]. This representation is not a model assumption; it is a theorem: CP evolution is exactly the class of physically admissible reduced evolutions.

*Semigroups and Lindblad form.* When the effective dynamics is Markovian (memoryless) on the accessible algebra, one expects a one-parameter semigroup of CPTP maps  $\{\Phi_t\}_{t \geq 0}$  with  $\Phi_{t+s} = \Phi_t \circ \Phi_s$  and continuity at  $t = 0$ . The general structure theorem (Gorini–Kossakowski–Sudarshan–Lindblad) states that the generator  $\mathcal{L}$  of such a uniformly continuous quantum dynamical semigroup has the Lindblad form

$$\frac{d\rho}{dt} = \mathcal{L}(\rho) = -i[H, \rho] + \sum_\alpha \left( L_\alpha \rho L_\alpha^\dagger - \frac{1}{2} \{L_\alpha^\dagger L_\alpha, \rho\} \right), \quad (44)$$

for some effective Hamiltonian  $H$  and noise (jump) operators  $L_\alpha$  [42, 43]. We emphasize: HCCB does *not* assume Markovianity; it predicts that coherence defects produce effective CP dynamics on accessible algebras, which may be history-dependent and non-Markovian. The semigroup/Lindblad case is a clean limit.

*Sectorization and superselection-like structure.* Coherence defects generically imply that not all contexts can be globally identified; the system decomposes into sectors labeled by the holonomy/obstruction class. Operationally this appears as superselection-like structure: some observables distinguish sectors, but dynamics within a sector is well-defined while transitions between sectors are forbidden or context-dependent.

*History dependence.* If the effective map on  $\mathcal{A}_{\text{acc}}$  depends on the path taken in context space (e.g. which coarse-graining sequence, which reconstruction choice, which observer chain), then the reduced dynamics is *non-Markovian*. In HCCB language, memory is the operational trace of higher coherence holonomy. This matches the “directionality/history dependence” logic we previously emphasized in your work on operational geometry and related contexts.

## F. Key Proposition 4: Physical realization of diagonal non-totality

**Proposition VIII.1** (Physical realization via HCCB). *In sufficiently expressive self-referential quantum systems, diagonal logic forbids a total internal decoder of all semantic predicates about the system (Proposition III.5). HCCB provides a physical realization of this non-totality as follows:*

- (P1) *Local descriptions remain consistent and can be unitary/linear within contexts (patchwise semantics).*
- (P2) *Higher coherence conditions required to glue these contexts into a single global unitary/linear semantics fail, producing nontrivial coherence defects/holonomy.*
- (P3) *When restricted to accessible algebras or operational probes, these coherence defects manifest as intrinsically CP reduced dynamics, typically with sectorization and history dependence, rather than as a single global decoder.*

*Thus diagonal non-totality is realized dynamically as higher categorical coherence breakdown, replacing “global unitarity as total semantics” by coherence closure as the minimal consistency requirement [1, 2].*



*Proof sketch (structure-to-dynamics).* Proposition III.5 forbids total internal decoding once self-reference is present. Attempting to enforce totality requires globally consistent identifications of all contexts and reconstructions. In HCCB, these identifications are encoded as higher morphisms and coherence data. If higher coherence fails (nontrivial holonomy (41)), then no single global unitary/linear semantics exists that can serve as a universal decoder.

Operational observers access reduced descriptions; reduction is necessarily CPTP (Stinespring/Kraus) [24, 41], and therefore coherence-defect-induced context dependence yields CP effective dynamics on accessible algebras. In regimes where the reduction is approximately memoryless, the GKLS theorem gives the Lindblad generator (44) [42, 43]. The sectorization and history dependence follow from the existence of distinct holonomy classes and path dependence of gluing. This is exactly the dynamical/operational imprint of non-totally.  $\square$

## G. Takeaway

HCCB supplies the missing physical bridge:

- diagonal logic explains why total internal decoding is ill-posed in self-referential regimes;
- categorical holography and islands show how gravity realizes *relative* decoding;
- HCCB explains how the world maintains local consistency while abandoning global totality: global unitarity/linearity may fail as a *single total semantics* because higher coherence cannot be glued, and the operational remnant is CP dynamics with sectorization and memory.

In Section IX we return to black holes as maximally self-referential systems where these mechanisms become unavoidable and geometric.

## IX. BLACK HOLES AS MAXIMALLY SELF-REFERENTIAL SYSTEMS

### A. Goal and synthesis statement

We now synthesize the logical and physical threads developed so far.

- Section III established the diagonal obstruction to *total* internal decoding of *information about the system* (Proposition III.5) using the shared Cantor/Gödel/Turing/Lawvere mechanism [17–19, 25].
- Sections V–VI showed how holography naturally implements *universality without totality* via encoding functors, code subspaces, and entanglement wedge reconstruction [10, 11, 38].
- Section VII interpreted islands as a physical regularization of non-totally, making the reconstructible region a state-dependent output of a variational principle [13–16].
- Section VIII introduced HCCB as the physical mechanism realizing semantic non-totally through higher coherence breakdown and CP/sectorial operational dynamics [1, 2].

The remaining step is to explain why black holes are the regime in which these issues are *maximally forced*: the horizon converts self-reference from a merely semantic option into an unavoidable geometric/operational feature.

### B. Horizon as irreversible observer separation

*Causal separation and observer classes.* An event horizon partitions spacetime into regions that are causally inaccessible to one another for certain observers. An exterior observer cannot access interior degrees of freedom directly. This is not a matter of experimental limitation but of spacetime causal structure.

*Irreversibility in the operational sense.* From the viewpoint of an exterior observer, matter falling through the horizon is not recoverable by any local operation performed outside. Even if the global evolution is unitary (as expected in a consistent quantum gravity theory), the *operational* description available to the outside observer is effectively open and irreversible, because it is a restriction to an accessible algebra or subsystem. This is precisely the regime where reduced dynamics is naturally CPTP (Section VIII) and where “total decoding” becomes a meaningful but over-demanding requirement.

*Complementarity as an early recognition of non-totality.* Black hole complementarity [33] anticipates that no single global description simultaneously accessible to all observers can exist without contradiction: one must allow observer-relative descriptions that are individually consistent but not jointly realizable as a single globally factorizing account. This viewpoint is naturally aligned with our “universality without totality” framework.

### C. Hawking radiation as internal encoding

Hawking’s semiclassical result [3] shows that black holes radiate. In modern unitary-compatible interpretations, the Hawking radiation is not merely thermal noise: it is the *physical carrier* of whatever information about the initial state becomes accessible to the exterior.

*Encoding is not local, but relational.* A central conceptual point is that the encoding of information into radiation is not a localized event “at the horizon” in the naive sense. Rather, it is distributed in the pattern of correlations among:

- the early radiation,
- the late radiation,
- near-horizon degrees of freedom (often effectively modeled by a stretched horizon),
- and interior partner modes in semiclassical descriptions.

Thus the encoding is *nonlocal* and *relational*: information about the system is present in correlations even when no single Hawking quantum carries it in isolation.

*The Page criterion as a statement about correlations.* Page’s calculation [4] provides the benchmark that, under global unitarity, the fine-grained entropy of the radiation must follow a rise-and-fall (the Page curve). Islands supply a semiclassical mechanism that reproduces this behavior (Section VII) [14, 16]. This supports the statement:

*information about the system can be preserved and encoded in correlations without being totally and uniformly decodable.*

### D. Decoding as a physical process inside the same system

The step that pushes black holes to maximal self-reference is that decoding is not an abstract mathematical inverse; it is a physical operation performed by an observer-apparatus acting on the radiation.

*Internality of decoding.* Any decoder is implemented by a physical subsystem (an agent with a memory and an apparatus). That subsystem is part of the global quantum system (black hole + radiation + environment). Thus, decoding is itself an internal dynamical process; it is not an external oracle. This is exactly the setting of Proposition III.5: the system contains the means to represent and act on descriptions of itself.

*Self-reference loop.* The black-hole situation realizes the maximal loop:

$$\text{system} \longrightarrow \text{radiation encoding} \longrightarrow \text{decoder action} \longrightarrow \text{system},$$

because the decoder acts on radiation produced by the system, and thereby alters the global state whose “information” it aims to decode. This is self-reference in an operational, not merely semantic, form.

### E. Why total decoding leads to paradox

We can now articulate precisely what is meant by “total decoding” and why it is too strong.

*Total decoding demand (black-hole form).* A strong form of the information-paradox demand can be phrased as:

*There exists a single, globally valid, state-independent decoding map from the Hawking radiation to all interior semantic predicates/observables, such that all information about the black hole interior becomes uniformly recoverable from the radiation alone.*

This is a totality demand in the sense of Section II D.

*Diagonal and monogamy pressure.* Total decoding is precisely what diagonal logic forbids in self-referential regimes: if the decoding map is internal and the query class is closed under self-reference, one can construct diagonal predicates that escape any total classifier (Section III) [19]. In the black-hole literature, the same pressure appears operationally as incompatibilities between different entanglement requirements (AMPS) [5]: attempting to maintain a single global factorized decoding picture leads to conflicts with monogamy and with semiclassical horizon entanglement structure.

*Interpretation.* From our viewpoint, the paradox is not a paradox of information destruction but a paradox of *totality*: it arises from demanding a total internal decoder in a maximally self-referential system. This demand overreaches what coherence can globally glue.

## F. Why relative decoding resolves it

The consistent alternative is to abandon totality while preserving coherence closure.

*Relative universality.* Entanglement wedge reconstruction already implements relative decoding: operators are reconstructible from a boundary region  $R$  only relative to a code subspace and relative to the entanglement wedge  $EW(R)$  [10, 38]. This matches Proposition V.1.

*Islands as contextual reconstruction.* In evaporating black holes, islands implement contextuality at the geometric level: the reconstructible region  $EW(R)$  becomes a state-dependent output of the QES extremization (37) [13, 14]. This avoids the total-decoder demand while reproducing the Page curve and maintaining unitarity-compatible entropy evolution (Section VII).

*HCCB as the physical mechanism.* Finally, HCCB provides a structural reason why global totality fails while local consistency survives: higher coherence obstructions prevent a single global unitary/linear semantics from gluing across all contexts, and the operational remnant is sectorized CP dynamics and history dependence on accessible algebras [1, 2]. Thus the “resolution” is not to abandon physics principles, but to formulate the minimal consistency principle correctly: *coherence closure* rather than global totality.

## G. Takeaway

Black holes sit at the extreme end of the self-reference hierarchy (Section IV) because:

- the horizon enforces irreversible observer separation and restricts accessible algebras,
- Hawking radiation is a physical internal encoding of information about the system,
- decoding is an internal physical process that feeds back into the system,
- demanding total decoding forces contradictions (diagonal/monogamy),
- while relative, contextual decoding (EWR + islands) resolves the tension without information loss.

Thus black holes reveal a universal constraint on information in self-referential quantum systems: information about the system may exist and be preserved in correlations, yet cannot be totally internalized as a single globally coherent uniformly decodable code.

## X. UNIVERSALITY BEYOND BLACK HOLES

### A. Goal and organizing principle

The black-hole setting is a maximally self-referential regime where the limits of total internal decoding become geometrically unavoidable (Section IX). However, the logical mechanism behind these limits is

not specific to gravity: it is the diagonal obstruction associated with self-reference (Section III). The purpose of this section is to show that the same “universality without totality” pattern appears across quantum theory whenever:

1. descriptions and decoders are internalized (self-reference),
2. subsystem notions are constrained (non-factorization, centers),
3. or operational access is a coarse-graining (CPTP reduction).

We emphasize: these are not separate phenomena but different physical realizations of the same semantic constraint (Proposition III.5) and its coherence-based physical mechanism (Proposition VIII.1).

### B. AQFT Type-III algebras: locality without global density matrices

A paradigmatic example where “information” is present but not totally internalizable in the naive Type-I (density-matrix) sense is Algebraic Quantum Field Theory (AQFT).

*Local algebras and Type-III structure.* In relativistic QFT, the physically natural objects are nets of local von Neumann algebras  $\mathcal{O} \mapsto \mathcal{A}(\mathcal{O})$ , satisfying isotony, locality, and covariance. A central structural fact is that local algebras in interacting QFT are generically Type III [32]. Type III algebras have no normal trace and admit no density matrix representation for restrictions of global states in the way familiar from finite-dimensional quantum mechanics.

*Interpretation: “information” is relational.* The absence of a trace/density matrix does not mean that information is absent; it means that the naive notion “state = density operator” is not globally applicable. Operationally meaningful quantities are instead formulated via relative entropy and modular theory (we will not develop modular theory here, but see [32]). This is a sharp instance of our general distinction:

- correlations and semantic distinctions exist,
- but there is no single globally uniform encoding (Type-I density matrix) that captures *all* local restrictions as objects of the same form.

Thus AQFT exhibits universality of local quantum physics without totality of naive encoding.

### C. Gauge theories: edge modes, centers, and the necessity of extensions

Gauge theories provide another canonical setting where totality fails for structural reasons.

*Non-factorization and centers.* In gauge theories, the gauge-invariant operator algebra associated to a region often has a nontrivial center, and different prescriptions for the local algebra correspond to different operational notions of localization and entanglement. Casini–Huerta–Rosabal analyzed this in the context of entanglement entropy for gauge fields, highlighting the role of centers and the ambiguity of naive subsystem factorization [30].

*Edge modes and extended Hilbert spaces.* Donnelly and Freidel showed that to define consistent local subsystems in gauge theory and gravity one must often enlarge the phase space/Hilbert space with boundary degrees of freedom (edge modes), restoring a gluing/factorization structure in an extended setting [31]. From our viewpoint, this is exactly how physics avoids totality contradictions: one cannot demand a globally factorizing, universally applicable notion of subsystem without introducing additional structure that keeps track of boundary/center data.

*Interpretation.* Gauge theories are therefore a non-gravitational exemplar of Proposition V.1: naive “total” localization and decoding fail; relative universality survives once one accepts sectorization/edge extensions.

### D. Quantum measurement: nested observers and self-referential consistency

Measurement theory provides a direct operational arena for self-reference.

*Wigner’s friend.* Wigner’s original observation [28] is that treating an observer plus measured system as a quantum system leads to a tension between unitary evolution and the experience of definite outcomes. The essential feature is that the observer’s record is itself part of the quantum state; thus, descriptions about the system include descriptions about the observer’s description.

*Frauchiger–Renner.* Frauchiger and Renner sharpened this into a no-go statement: under natural assumptions about the universal validity of quantum theory and consistency of agents’ reasoning, quantum theory cannot consistently describe the use of itself [29]. This is a direct operational manifestation of the diagonal self-reference constraint: demanding a single agent-independent total semantics for all nested descriptions is too strong; contextual/relative descriptions are unavoidable.

*Interpretation.* Measurement chains thus mirror the black-hole story at a different scale: information about the system exists (records, correlations), but total internal decoding across all observer contexts leads to contradiction. Coherence closure requires contextual semantics, exactly as in our framework.

### E. Open quantum systems: CP dynamics as the natural operational language

Finally, open quantum systems provide the most concrete operational arena where non-totality becomes a daily working principle.

*CPTP maps are unavoidable under restriction.* Whenever one restricts to an accessible subsystem or coarse-grains over an environment, the reduced dynamics is CPTP and admits a Kraus/Stinespring form [21, 22, 24, 41]. This is not a choice but a theorem: complete positivity is the minimal requirement for consistency under extension by ancillas.

*Semigroups and Lindblad generators.* In Markovian limits, one obtains quantum dynamical semigroups and GKLS/Lindblad generators [42, 43]. Even when non-Markovianity is present, the key point remains: operational descriptions are not globally unitary on the accessible algebra. This provides a clean physical analogue of “coherence closure”: the world is consistent, but not all information is accessible/decodable under a single global unitary on the reduced description.

*Interpretation.* Open quantum systems show, in a non-gravitational and experimentally ubiquitous setting, that “information preservation” and “total decodability” are distinct notions. The reduced description is consistent (CP), predictive, and experimentally adequate, yet it does not admit a total internal decoder of the full system’s semantic distinctions.

### F. Information as relational and not totally encodable: unified takeaway

Across AQFT, gauge theory, measurement chains, and open systems, we observe the same structural theme:

*Information **about** the system exists and is preserved in correlations, but cannot in general be totally internalized as a single globally coherent and uniformly decodable code once self-reference, constrained factorization, or operational restriction is present.*

Black holes are not exceptional in kind; they are exceptional in degree: they are maximally self-referential and therefore force the non-totality boundary to become geometric (horizon separation) and reconstructive (islands). This universality beyond gravity supports our central thesis: the black-hole information paradox is a particularly vivid manifestation of a general constraint on information in self-referential quantum systems.

## XI. DISCUSSION AND OUTLOOK

### A. From “information paradox” to “totality paradox”

A central claim of this paper is conceptual and can be stated succinctly:

*The black-hole information paradox is, at its core, a **totality paradox**: it arises from demanding a total internal decoder/classifier of all semantic predicates (information about the system) in a sufficiently expressive, self-referential quantum system, a demand obstructed by diagonal logic.*

This reframing preserves the empirical and theoretical achievements of semiclassical gravity and holography while clarifying which assumption is overreaching. Specifically:

- Hawking radiation exists and carries correlations [3].

- Unitary evaporation (in a suitable UV-complete theory) implies a Page curve for the fine-grained radiation entropy [4].
- Islands and replica wormholes provide a semiclassical mechanism reproducing Page-like behavior while maintaining consistency [14–16].
- Entanglement wedge reconstruction and the QEC perspective explain why reconstruction is *relative* (code-subspace/region/state dependent) rather than globally total [10, 13, 38].

What changes is the interpretation: the “paradox” is not evidence for information loss, but evidence that *total internal decoding* is not a legitimate demand in maximally self-referential regimes.

### B. Coherence-first principle as the minimal consistency requirement

Diagonal logic (Cantor/Gödel/Turing/Lawvere) explains why totality is unstable under self-reference [17–19, 25]. The physical question is what replaces totality. Our proposal is a *coherence-first principle*:

*The minimal requirement for a physical theory is not global symmetry or global totality, but **coherence closure**: local/sectorial descriptions must glue consistently on overlaps, while global totality may fail in the presence of higher coherence obstructions.*

This principle is both weaker and more robust than demanding a single globally valid decoder. It is weak enough to survive diagonal constraints and strong enough to constrain physics nontrivially:

- It predicts contextuality and relative reconstruction as structural features rather than defects (Sections VI–VII).
- It suggests that “symmetry” should be treated as a sufficient, often emergent, but not necessary condition for coherent gluing.
- It motivates Higher Categorical Coherence Breakdown (HCCB) as the dynamical mechanism by which the world maintains coherence without totality [1, 2].

### C. Implications for quantum gravity

(i) *Reconstruction is inherently relative.* The No-Total-Holography proposition (Proposition V.1) suggests that attempts to make bulk reconstruction globally uniform are conceptually misdirected. Instead, the correct object is a *family* of reconstructions indexed by context: code subspaces, sectors, boundary regions, and semiclassical regimes [10, 38].

(ii) *Islands are not a patch but a structural necessity.* Interpreting islands as a physical regularization of non-totally (Proposition VII.1) frames island formation as a geometrized mechanism that avoids total-decoder contradictions while maintaining unitarity-compatible entropy evolution [9, 14]. This suggests a general research strategy: *identify where totality would be demanded, and look for the corresponding geometric/coherence mechanism that replaces it.*

(iii) *“Interior” is a coherence-relative notion.* In maximally self-referential settings, interior observables are not expected to admit a state-independent global representation on a fixed boundary algebra. Rather, interior reconstruction is coherence-relative and may jump across regimes (islands/QES transitions). This does not diminish AdS/CFT; it clarifies which statements are well-posed.

### D. Implications for foundations of quantum mechanics

(i) *Nested observers and contextual semantics.* Wigner-type scenarios and their sharpenings (e.g. Frauchiger–Renner) show that demanding a single globally consistent description of all agents using quantum theory about each other is too strong [28, 29]. Our framework places this into the same class as black holes: both are self-referential regimes where total internal decoding is ill-posed.

(ii) *Collapse vs unitarity is not the central dichotomy.* From the coherence-first viewpoint, the correct dichotomy is:

$$\textit{total global semantics} \quad \text{versus} \quad \textit{coherence closure across contexts}.$$

HCCB suggests how intrinsic CP dynamics and history dependence can emerge from coherence defects without denying local unitary behavior [1].

### E. Implications for information theory

(i) *“Information” vs “decodability.”* Information-theoretic quantities (entropy, mutual information, relative entropy) quantify correlations and distinguishability [20, 22]. They do not guarantee the existence of a *total* internal decoder for all semantic predicates about the system. This distinction becomes operationally essential in self-referential regimes.

(ii) *Relative entropy as a universal control variable.* The role of relative entropy in holography (e.g. equality of boundary and bulk relative entropies) highlights that recoverability and reconstruction are naturally formulated in terms of monotonicity/data processing and its saturations [11]. This suggests that “universality without totality” can be quantified by identifying which relative-entropy equalities hold in which contexts.

### F. Open problems and research directions

We list open problems whose resolution would sharpen the framework and test it against both mathematical rigor and physical applications.

#### 1. Quantitative measures of self-referentiality

In Sections IV and IX we presented a qualitative hierarchy. A quantitative theory should assign a measure  $\Sigma$  of “self-referentiality” to a system/context, capturing:

- the richness of internal representability (R),
- the operational availability of self-application/evaluation (E),
- the closure of query classes under reflection (C).

One possible route is to define  $\Sigma$  via a hierarchy of internal languages (fragments of a topos/CCC internal logic) and measure the extent to which a total classifier would have to exist to decide those fragments. Another route is operational: quantify the degree of observer-internalization and back-action required for decoding.

#### 2. Coherence-defect entropy

If higher coherence defects (holonomy gaps) are the physical manifestation of non-totally, they should admit an information-theoretic characterization. We propose the concept of a *coherence-defect entropy*: an entropy-like invariant measuring the minimal additional structure (edge/sector/context data) needed to restore coherence closure. This notion should be compatible with:

- edge-mode/center contributions in gauge theory [31],
- Type-III algebraic obstructions in AQFT [32],
- island-induced wedge transitions in gravity [15].

Developing such an invariant would make HCCB quantitatively predictive.

### 3. Links to complexity and computation

Diagonal logic is historically tied to computation (Turing) and incompleteness (Gödel) [17, 18]. In the black-hole context, decoding is often argued to be computationally complex (e.g. the Hayden–Preskill decoding problem) [12]. A promising direction is to relate:

- *computational hardness* of decoding (complexity barriers),
- *semantic non-totality* (diagonal barriers),
- and *coherence defects* (HCCB barriers),

as complementary manifestations of the same underlying limitation: the system cannot fully internalize and execute a total decoder of its own semantic content. A rigorous bridge here would significantly strengthen the conceptual unification proposed in this paper.

## G. Closing perspective

The main message is a shift in what we ask from fundamental theory. Rather than demanding global totality (a single internal decoder for all predicates), we should demand coherence closure (consistent gluing across contexts) and accept sectorization and contextual reconstruction as intrinsic where self-reference is unavoidable. Black holes are the place where this becomes geometrically sharp, but the phenomenon is universal across quantum theory.

## XII. CONCLUSION

### A. Summary of results

We conclude by summarizing the logical structure of the paper and the main results, with emphasis on the distinction between (i) the existence/preservation of correlations and (ii) total internal decodability of semantic predicates about the system.

(1) *Definitions: what is (and is not) being constrained.* In Section II we separated:

- *operational information* as correlations and distinguishability (entropy, mutual information, relative entropy) [20, 22],
- *information about the system* as semantic distinctions/predicates concerning states, observables, or histories,
- *encodability/decodability* as the existence of a uniform internal representation and decoder for a specified query class.

Diagonal constraints target the last notion (total internal decoding), not the first (existence of correlations).

(2) *Diagonal logic: the semantic no-go.* In Section III we presented the shared diagonal mechanism behind Cantor, Gödel, and Turing and its categorical unification via Lawvere fixed points [17–19, 25]. We formulated the resulting semantic obstruction as Proposition III.5: sufficiently expressive self-referential systems do not admit a single total internal decoder/classifier of all semantic predicates.

(3) *Categorical holography: universality without totality.* In Section V we translated diagonal constraints into holographic language by modeling bulk-to-boundary encoding as a functor  $U : \mathcal{B} \rightarrow \mathcal{P}$  and showing that “total holography” (a total classifier of all bulk predicates) is too strong in self-referential regimes. This was crystallized in Proposition V.1.

(4) *Modern holography already implements the correct logic.* In Section VI we showed that entanglement wedge reconstruction, understood through the quantum error correction viewpoint, is intrinsically *relative*: it holds on code subspaces and depends on region and sometimes state [10, 11, 38]. Non-reconstructible operators are therefore not anomalies but expected “diagonal remainders.”

(5) *Islands as physical regularization of non-totality.* In Section VII we interpreted islands and the QES prescription as a geometric implementation of contextual, regime-dependent reconstruction that avoids the over-demand of total decoding and reproduces Page-curve behavior in semiclassical gravity [9, 13, 14, 16]. This was formalized as Proposition VII.1.



(6) *HCCB: a physical mechanism.* In Section VIII we proposed Higher Categorical Coherence Breakdown (HCCB) as a dynamical/structural mechanism by which semantic non-totality is realized physically: global unitary/linear semantics may fail to glue coherently across contexts even when local descriptions remain consistent, yielding sectorization, history dependence, and CP effective dynamics on accessible algebras [1, 2]. This was summarized in Proposition VIII.1.

(7) *Black holes as maximally self-referential systems.* In Section IX we argued that black holes sit at the extreme of the self-reference hierarchy: horizons enforce observer separation, Hawking radiation is an internal encoding, and decoding is an internal physical process. As a result, total decoding demands generate paradox, while relative decoding resolves it without information loss. Finally, in Section X we showed that the same “universality without totality” pattern appears beyond gravity in AQFT, gauge theory, measurement chains, and open quantum systems.

## B. Final perspective

The conceptual shift advocated here is that what fails in the black-hole setting is not information itself, but the *totality demand* that information about the system be internalizable and uniformly decodable from within the same self-referential system. Once one replaces this demand by the weaker and more robust requirement of *coherence closure*, the modern picture of holography, entanglement wedge reconstruction, and islands fits naturally into a single logical narrative.

*Final sentence.* *Black holes do not destroy information; they expose a universal limit of information itself—namely, that in sufficiently self-referential quantum systems, information about the system can exist and be preserved in correlations without ever being totally decodable by a single internal, globally coherent, and uniformly valid decoding scheme.*

- 
- [1] A. T. Patrascu, “Higher-order categorical coherence breakdown: a geometric framework for nonlinear quantum mechanics and its applications to strongly correlated electron systems,” *Eur. Phys. J. B* **98** (2025) 210. doi:10.1140/epjb/s10051-025-01062-6.
  - [2] A. T. Patrascu, “Higher Categorical Coherence Breakdown as a Quantum Process: Measurement without Classicality and Its Dual Role in Quantum Computing Limitations and Algorithmic Advantages,” Zenodo preprint (v1, published Aug. 2025). doi:10.13140/RG.2.2.11749.72168.
  - [3] S. W. Hawking, “Breakdown of Predictability in Gravitational Collapse,” *Phys. Rev. D* **14** (1976) 2460–2473. doi:10.1103/PhysRevD.14.2460.
  - [4] D. N. Page, “Average Entropy of a Subsystem,” *Phys. Rev. Lett.* **71** (1993) 1291–1294. doi:10.1103/PhysRevLett.71.1291. arXiv:gr-qc/9305007.
  - [5] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, “Black Holes: Complementarity or Firewalls?,” *JHEP* **02** (2013) 062. doi:10.1007/JHEP02(2013)062. arXiv:1207.3123 [hep-th].
  - [6] S. Ryu and T. Takayanagi, “Holographic Derivation of Entanglement Entropy from AdS/CFT,” *Phys. Rev. Lett.* **96** (2006) 181602. doi:10.1103/PhysRevLett.96.181602. arXiv:hep-th/0603001.
  - [7] V. E. Hubeny, M. Rangamani and T. Takayanagi, “A Covariant Holographic Entanglement Entropy Proposal,” *JHEP* **07** (2007) 062. doi:10.1088/1126-6708/2007/07/062. arXiv:0705.0016 [hep-th].
  - [8] T. Faulkner, A. Lewkowycz and J. Maldacena, “Quantum corrections to holographic entanglement entropy,” *JHEP* **11** (2013) 074. doi:10.1088/1126-6708/2013/07/074. arXiv:1307.2892 [hep-th].
  - [9] N. Engelhardt and A. C. Wall, “Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime,” *JHEP* **01** (2015) 073. doi:10.1007/JHEP01(2015)073. arXiv:1408.3203 [hep-th].
  - [10] A. Almheiri, X. Dong and D. Harlow, “Bulk Locality and Quantum Error Correction in AdS/CFT,” *JHEP* **04** (2015) 163. doi:10.1007/JHEP04(2015)163. arXiv:1411.7041 [hep-th].
  - [11] D. L. Jafferis, A. Lewkowycz, J. Maldacena and S. J. Suh, “Relative entropy equals bulk relative entropy,” *JHEP* **06** (2016) 004. doi:10.1007/JHEP06(2016)004. arXiv:1512.06431 [hep-th].
  - [12] P. Hayden and J. Preskill, “Black holes as mirrors: quantum information in random subsystems,” *JHEP* **09** (2007) 120. doi:10.1088/1126-6708/2007/09/120. arXiv:0708.4025 [hep-th].
  - [13] G. Penington, “Entanglement Wedge Reconstruction and the Information Paradox,” arXiv:1905.08255 [hep-th].
  - [14] A. Almheiri, R. Mahajan, J. Maldacena and Y. Zhao, “The Page curve of Hawking radiation from semiclassical geometry,” arXiv:1908.10996 [hep-th].
  - [15] A. Almheiri, N. Engelhardt, D. Marolf and H. Maxfield, “The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole,” *JHEP* **12** (2019) 063. doi:10.1007/JHEP12(2019)063. arXiv:1905.08762 [hep-th].
  - [16] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, “Replica Wormholes and the Entropy of Hawking Radiation,” *JHEP* **05** (2020) 013. doi:10.1007/JHEP05(2020)013. arXiv:1911.12333 [hep-th].

- [17] K. Gödel, “Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I,” *Monatshefte für Mathematik und Physik* **38** (1931) 173–198.
- [18] A. M. Turing, “On Computable Numbers, with an Application to the Entscheidungsproblem,” *Proc. Lond. Math. Soc.* **2** **42** (1936) 230–265; erratum **2** **43** (1937) 544–546. doi:10.1112/plms/s2-42.1.230.
- [19] F. W. Lawvere, “Diagonal arguments and cartesian closed categories,” *Theory Appl. Categ.* **15** (2006) 1–13; reprint of *Lecture Notes in Mathematics* **92** (1969) 134–145. Available at <https://www.tac.mta.ca/tac/reprints/articles/15/tr15.pdf>.
- [20] C. E. Shannon, “A Mathematical Theory of Communication,” *Bell System Technical Journal* **27** (1948) 379–423, 623–656.
- [21] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (2000).
- [22] M. M. Wilde, *Quantum Information Theory*, 2nd ed., Cambridge University Press (2017).
- [23] H. Umegaki, “Conditional expectation in an operator algebra. IV. Entropy and information,” *Kodai Mathematical Seminar Reports* **14** (1962) 59–85.
- [24] W. F. Stinespring, “Positive Functions on  $C^*$ -Algebras,” *Proc. Amer. Math. Soc.* **6** (1955) 211–216.
- [25] G. Cantor, “Über eine elementare Frage der Mannigfaltigkeitslehre,” *Jahresbericht der Deutschen Mathematiker-Vereinigung* **1** (1891) 75–78.
- [26] R. M. Smullyan, *Gödel’s Incompleteness Theorems*, Oxford University Press (1992).
- [27] M. Sipser, *Introduction to the Theory of Computation*, 3rd ed., Cengage Learning (2012).
- [28] E. P. Wigner, “Remarks on the Mind-Body Question,” in I. J. Good (ed.), *The Scientist Speculates*, Heinemann, London (1961).
- [29] D. Frauchiger and R. Renner, “Quantum theory cannot consistently describe the use of itself,” *Nature Communications* **9** (2018) 3711. doi:10.1038/s41467-018-05739-8. arXiv:1604.07422 [quant-ph].
- [30] H. Casini, M. Huerta and J. A. Rosabal, “Remarks on entanglement entropy for gauge fields,” *Phys. Rev. D* **89** (2014) 085012. doi:10.1103/PhysRevD.89.085012. arXiv:1312.1183 [hep-th].
- [31] W. Donnelly and L. Freidel, “Local subsystems in gauge theory and gravity,” *JHEP* **09** (2016) 102. doi:10.1007/JHEP09(2016)102. arXiv:1601.04744 [hep-th].
- [32] R. Haag, *Local Quantum Physics: Fields, Particles, Algebras*, 2nd rev. and enl. ed., Springer, Berlin (1996).
- [33] L. Susskind, L. Thorlacius and J. Uglum, “The stretched horizon and black hole complementarity,” *Phys. Rev. D* **48** (1993) 3743–3761. doi:10.1103/PhysRevD.48.3743. arXiv:hep-th/9306069.
- [34] S. Mac Lane, *Categories for the Working Mathematician*, 2nd ed., Springer, New York (1998).
- [35] S. Awodey, *Category Theory*, 2nd ed., Oxford University Press (2010).
- [36] P. T. Johnstone, *Sketches of an Elephant: A Topos Theory Compendium*, Oxford University Press (2002).
- [37] B. Jacobs, *Categorical Logic and Type Theory*, Elsevier (1999).
- [38] X. Dong, D. Harlow and A. C. Wall, “Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality,” *Phys. Rev. Lett.* **117** (2016) 021601. doi:10.1103/PhysRevLett.117.021601. arXiv:1601.05416 [hep-th].
- [39] C. Bény, A. Kempf and D. W. Kribs, “Generalization of the Knill–Laflamme conditions for quantum error correction,” *Phys. Rev. Lett.* **98** (2007) 100502. doi:10.1103/PhysRevLett.98.100502. arXiv:quant-ph/0608071.
- [40] T. Leinster, *Higher Operads, Higher Categories*, Cambridge University Press (2004).
- [41] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory*, Lecture Notes in Physics **190**, Springer (1983).
- [42] G. Lindblad, “On the generators of quantum dynamical semigroups,” *Commun. Math. Phys.* **48** (1976) 119–130. doi:10.1007/BF01608499.
- [43] V. Gorini, A. Kossakowski and E. C. G. Sudarshan, “Completely positive dynamical semigroups of  $N$ -level systems,” *J. Math. Phys.* **17** (1976) 821–825. doi:10.1063/1.522979.