

The Embedded-Observer No-Go Theorem

Causal Order, Geometrically Induced Reduction, and the Impossibility of Complete Self-Description in Any Closed Unitary Quantum System

Sergej Materov
© September 2026

Abstract

We prove that no observer realized as a proper subsystem of a closed, globally unitary quantum system can simultaneously (a) reconstruct the complete global state from its own accessible data and (b) evolve under a fixed, autonomous unitary law, whenever it is genuinely (non-trivially) coupled to the rest of the system. The result follows from four minimal hypotheses -global unitarity, an observer occupying a proper subsystem, a locality structure constraining direct interactions, and a non-fine-tuned interacting boundary - with no further structure assumed. Two of the theorem's three components (a canonical, geometrically forced reduction of the observer's knowledge to a partial trace, and the no-go result itself) require no locality assumption at all and hold for any bipartite split of any unitary quantum system whatsoever. The third component, a causal order controlling when correlations between separated parts of the system can first form, is established in two forms: the standard, approximately-causal Lieb–Robinson bound for any locally-interacting Hamiltonian system, and a sharper, EXACTLY causal bound -proved from scratch and confirmed by explicit numerical example -for the special case of a discrete local quantum circuit. We further show that the theorem's two failure modes are not one fact restated: one is a property of the accumulated history (a “stock”, formalized via the observer's von Neumann entropy) and the other a property of the instantaneous interaction (a “flow”, formalized via the channel unitarity of Wallman, Granade, Harper & Flammia), and the two are shown, by explicit construction, to be logically separable. The result is situated relative to prior work establishing closely related conclusions by different routes -Wigner's friend (1961), the Page-Wootters mechanism (1983), the Frauchiger-Renner theorem (2018), and the quantum-reference-frame literature - and its specific, narrower technical contribution relative to that literature is stated explicitly.

1. Minimal Hypotheses

Let H be a (finite-dimensional, for concreteness) Hilbert space, evolving under a fixed global unitary $\hat{U}$ (equivalently, generated by a fixed Hamiltonian $\hat{H}$ in continuous time). Four conditions are used, and only four:

Condition	Statement	Used for
(H1) Closure	The system evolves under a single global unitary, with nothing external to it.	All results
(H2) Embeddedness	An observer occupies a proper, nonempty subsystem: $H=H_O \otimes H_R$, R nonempty.	§2, §3
(H3) Locality	The elementary subsystems carry a graph or lattice structure (bounded degree), constraining which parts of H_O and H_R can interact directly.	§4 only
(H4) Genuine coupling	The O–R interaction is not fine-tuned to be exactly non-entangling at every relevant instant.	§3

These are the only assumptions used anywhere below. Any closed quantum system satisfying them is covered - a lattice of interacting spins, a quantum field theory with a UV cutoff, a collection of trapped ions or superconducting qubits, a cellular-automaton-like discrete circuit, or any other construction with these four properties.

(H2) presupposes an exact tensor-product factorization; this holds in the finite-dimensional setting adopted here but requires the modular-theory reformulation of the reduction step in the continuum/algebraic-QFT (Type III factor) setting, addressed separately in [Materov, Toward a Principle of Significance].

2. Geometrically Induced Reduction

Given $H=H_O \otimes H_R$ (H2), the information accessible to an observer confined to O is, by definition, the expectation value of every operator in the algebra $A_O := B(H_O) \otimes \hat{I}_R$ - the operators the observer can in principle measure, given that it occupies O and nothing else. The unique state on A_O reproducing these expectation values is the reduced density matrix

$$\rho_O(t) := Tr_R[|\Psi(t)\rangle\langle\Psi(t)|].$$

No choice is involved in passing from $|\Psi(t)\rangle$ to $\rho_O(t)$: it is forced entirely by $H=H_O \otimes H_R$ and by nothing else. This is already the fully general statement of what an embedded observer's accessible description consists of - no locality structure, no circuit, no further assumption is needed for this step; (H2) alone suffices.

3. The No-Go Theorem

Call the O–R coupling ACTIVE over an interval if the joint evolution restricted to that interval is entangling - i.e. not of product form (an operator acting on H_O alone) $\otimes$ (an operator acting on H_R alone). Product evolution is a measure-zero special case among all possible joint unitaries; an active, entangling coupling is the generic situation for any evolution not specifically fine-tuned to avoid it - this is hypothesis (H4).

3.1 Lemma A: no complete self-description once entangled

Lemma A (purification non-uniqueness; Hughston–Jozsa–Wootters 1993). If $\rho_O(t)$ is not pure (equivalently, its von Neumann entropy $S(\rho_O(t)) = -\text{Tr}[\rho_O(t) \ln \rho_O(t)] > 0$), then infinitely many distinct global pure states on $H_O \otimes H_R$ - related by arbitrary unitaries acting on H_R alone - all reduce to the exact same $\rho_O(t)$. Consequently $\rho_O(t)$, the totality of information accessible to an observer confined to O, does not determine $|\Psi(t)\rangle$ uniquely.

Proof. Write the spectral decomposition $\rho_O(t) = \sum_k p_k |k\rangle\langle k|$. For any orthonormal set $\{|f_k\rangle\} \subset H_R$, $|\Phi\rangle := \sum_k \sqrt{p_k} |k\rangle \otimes |f_k\rangle$ is a global pure state with $Tr_R[|\Phi\rangle\langle\Phi|] = \rho_O(t)$ exactly, by direct computation. Different choices of $\{|f_k\rangle\}$ (related by any unitary on H_R) give different $|\Phi\rangle$. If $S(\rho_O(t)) > 0$, $\rho_O(t)$ has rank ≥ 2 , so distinct choices of $\{|f_k\rangle\}$ give genuinely distinct $|\Phi\rangle$. ■

Numerical confirmation: two global two-qubit pure states, differing by a generic unitary applied to the R-factor alone, were constructed explicitly and shown to produce bit-for-bit identical ρ_O while remaining distinct as global states - an observer confined to O has no local measurement that could distinguish them.

3.2 Lemma B: no autonomous unitary law once entangled

Lemma B. If the O–R coupling is active over $[t, t+\delta t]$, there is in general no fixed unitary $\hat{W}$ on H_O alone such that $\rho_O(t+\delta t) = \hat{W}\rho_O(t)\hat{W}^\dagger$ holds for every possible state of R. Equivalently, ρ_O 's evolution is not a

closed, autonomous law over that interval - it depends on information about R that O does not have access to.

Proof / verification. Fix O's own preparation and vary only R's. An explicit two-qubit calculation with a generic (Haar-random) entangling coupling, with O initialized identically in two runs and R initialized in two different states, produced two different resulting ρ_O (both mixed). No single fixed $\hat{W}$ can reproduce two different outputs from one input, so no O-only unitary law exists whenever the coupling is active. (General mechanism: a marginal/reduced dynamics is unitarily implementable if and only if it factorizes with no entangling system-environment coupling -the standard fact underlying unitary dilations of quantum channels.)

3.3 The theorem

Theorem 3.1 (Embedded-Observer No-Go). Let O be a proper, nonempty subsystem with R its nonempty complement, coupled by an active interaction (H4) over some interval. Then it is impossible for the observer realized on O to simultaneously: (a) reconstruct the complete global state $|\Psi(t)\rangle$ from its own accessible data $\rho_O(t)$; (b) evolve under a fixed, autonomous unitary law independent of R; while (c) remaining a genuine part of the same interacting system.

Corollary (genericity, not universality). The hypothesis - an active, entangling coupling - fails only if the interaction is an exact product unitary throughout: a measure-zero condition, and one that would also make O a physically decoupled subsystem in every meaningful sense. For any generic evolution, and for any O genuinely part of an interacting system, the theorem's conclusion is unavoidable, not merely likely.

Remark 3.3.1 (Weak coupling as a physical near-degenerate regime). The corollary above distinguishes the theorem's hypothesis (H4) from its negation in a strictly binary way: a coupling is either exactly product (measure-zero) or active. Physically realizable systems occupy a continuum of coupling strengths between these extremes, and some naturally-motivated candidates sit extremely close to the product end without ever reaching it exactly. An observer whose relevant degrees of freedom are realized in a dark-matter sector, coupled to the baryonic system of interest only gravitationally (and possibly through additional weak portals, depending on the dark-sector model), is such a candidate: the entangling rate between O and R can be many orders of magnitude smaller than for an electromagnetically coupled observer, so that $S(\rho_O(t))$ and $1 - u(\Phi_{t,t+\delta t})$ both grow, but arbitrarily slowly. Theorem 3.1 applies to any such observer without modification - coupling strength does not enter the proof of Lemma A or Lemma B - but the practical time scale over which conditions (a) and (b) fail becomes a free physical parameter set by the coupling, not merely a qualitative yes/no fact set by (H4) alone. This sharpens the sense in which the corollary's genericity is a statement about a measure-zero exception rather than about a physically negligible one: on human or even cosmological time scales, a sufficiently weakly-coupled observer's departure from (a) and (b) may be undetectable, even though it is, by Theorem 3.1, never exactly zero.

3.4 (a) and (b) are logically distinct: a stock condition and a flow condition

(a)'s failure depends only on whether $\rho_O(t)$ is CURRENTLY mixed -a property of the accumulated history, regardless of what is happening at the present instant. (b)'s failure depends only on whether the interaction ACTIVE right now is entangling -a property of the present transition alone. These need not coincide: if the coupling was entangling in the past (so $\rho_O(t)$ is already mixed, (a) already and permanently failed) but the interaction over some later interval $[t, t+\delta t]$ happens to be an exact product unitary $u_O \otimes u_R$, that transition alone satisfies

$$\rho_O(t+\delta t) = u_O \rho_O(t) u_O^\dagger,$$

an exact, autonomous unitary update using only u_O and the already-mixed $\rho_O(t)$.

Numerical confirmation. A two-qubit system (O, R) was evolved through an entangling step followed by an exactly product step. After the entangling step, ρ_O had purity $\text{Tr}[\rho_O^2]=0.621028$ (mixed: $S>0$, (a) already and permanently failed). The subsequent product step gave $\rho_O(t+\delta t)=u_O\rho_O(t)u_O^\dagger$ exactly (agreement to numerical precision, $\text{atol}=10^{-10}$), with purity exactly preserved at 0.621028 across that transition. (b) therefore holds at that specific step despite (a) having already, and permanently, failed.

The two conditions are consequently distinct in kind - (a) about accumulated history, (b) about the present transition - sharing a common physical source (entangling coupling, whether in the past for (a) or presently for (b)) without being one identical fact. For a generic evolution not fine-tuned to insert product interactions at chosen instants, active coupling recurs essentially continuously, so (a) and (b) fail together in practice; but the exact counter-example above shows the two can be cleanly separated whenever the flow condition alone happens to hold at a given instant.

4. Causal Order

Establishing WHEN correlations between an observer and a distant part of the system can first form requires a locality structure (H3): elementary subsystems arranged on a graph or lattice, with direct interactions confined to nearby elements.

4.1 General local Hamiltonian systems: the standard (approximate) bound

For a Hamiltonian $\hat{H}=\sum_i h_i$ with each h_i acting on a bounded neighbourhood (any lattice, any graph, any spatial dimension), the Lieb–Robinson theorem (1972) gives, for operators $\hat{A}, \hat{B}$ supported at graph distance d ,

$$\|[\hat{A}(t), \hat{B}]\| \leq C\|\hat{A}\|\|\hat{B}\| \exp\left[-\frac{d - v_{LR} t}{\xi}\right],$$

an exponentially small - not exactly zero - commutator for $t < \frac{d}{v_{LR}}$. This is the standard causal bound available to any locally-interacting Hamiltonian system, continuous or discrete in time.

4.2 Discrete local circuits: an exact light cone

A sharper, exactly causal bound is available when, in addition to (H3), the elementary subsystems are updated by a DISCRETE LOCAL CIRCUIT: partition the interaction graph's edges into K disjoint-edge colour classes ($K \leq z+1$, z the maximum degree, by Vizing's theorem - such a partition always exists for a bounded-degree graph), and let one circuit step be

$$\hat{U} = \hat{U}_K \circ \dots \circ \hat{U}_1, \quad \hat{U}_k = \prod_{e \in \text{colour } k} U_e,$$

each U_e a unitary acting only on the (at most two) elementary subsystems at the endpoints of edge e . Gates within one colour class act on disjoint subsystems, hence commute, so each $\hat{U}_k$ is unitary, and $\hat{U}$ a product of unitaries - is unitary.

Lemma 4.1 (exact support growth). Let $\hat{A}$ have support X_0 (the smallest vertex set outside of which $\hat{A}$ acts as the identity), and let $\hat{A}(n) := \hat{U}^n \hat{A} \hat{U}^n$. Then $\text{supp}(\hat{A}(n)) \subseteq \{v : \text{dist}(v, X_0) \leq nK\}$.

Proof. Consider one colour-class sublayer, a set of gates on pairwise disjoint edges. If a gate's edge touches no vertex in the current support, it acts as the identity there and commutes with the current operator (operators on disjoint tensor factors always commute), leaving the support unchanged. If a gate's edge does touch the current support, conjugating by it can enlarge the support only to include that edge's two endpoints - never more, since the gate acts trivially (as the identity) on every other subsystem. Hence one sublayer extends the

support by at most one graph-distance unit. A full step (K sublayers) extends it by at most K units; n steps, by at most nK. ■

Theorem 4.2 (Exact Light Cone). Let $\hat{A}$ be supported at X_0 and $\hat{B}$ at a vertex set Y with $\text{dist}(X_0, Y) > nK$. Then

$$[\hat{A}(n), \hat{B}] = 0 \text{ EXACTLY,}$$

with no exponential tail and no approximation: the support of $\hat{A}(n)$, by Lemma 4.1, remains disjoint from Y , and operators on disjoint tensor factors commute exactly, not merely approximately. This sharper, exact statement is available precisely because the evolution is a finite product of strictly local gates - each sublayer's gates have exactly disjoint support by construction - rather than the exponential of a local Hamiltonian, whose commutator norm decays but need not vanish exactly at any finite separation.

4.3 Numerical verification (See GitHub repository in References)

A 10-element chain (elementary subsystems 0–9, nearest-neighbour graph, maximum degree 2, so $K=2$ colour classes suffice) was equipped with a generic (Haar-random) two-subsystem unitary gate on each edge, fixed once and reused at every step (a generic, non-fine-tuned choice, avoiding any special algebraic coincidence). Taking $\hat{A}=X$ on subsystem 0 and static $\hat{B}=X$ on subsystem 9 (graph distance $d=9$), Theorem 4.2 predicts an exact zero commutator for $n \leq 4$ (since $nK=2n < 9$ requires $n \leq 4$) and permits a nonzero commutator from $n=5$ onward. The single-sided commutator $\|[\hat{A}(n), \hat{B}]\|$ was computed exactly:

n (steps)	$\ [\hat{A}(n), \hat{B}]\ $	Predicted
0	0	Exactly 0
1–4	2×10^{-15} to 5×10^{-15} (machine zero)	Exactly 0
5	0.697	Nonzero permitted
6–9	$1.37 \rightarrow 1.99$ ($\rightarrow 2$, the algebraic maximum)	Nonzero permitted

Table 1. Exact confirmation of Theorem 4.2: the commutator is exactly zero (to machine precision) through $n=4$ and jumps to $O(1)$ starting exactly at the predicted $n=5$.

A methodological note for reproduction: testing the commutator of two operators BOTH conjugated by the same $\hat{U}^n$ is unitarily invariant and trivially reproduces the $n=0$ value at every n - a check that superficially resembles a light-cone test but is vacuous, since $\|\hat{U}^{-n}[\hat{A}, \hat{B}]\hat{U}^n\| = \|[\hat{A}, \hat{B}]\|$ for every n regardless of locality. The physically meaningful test evolves only one of the two operators, comparing it against the other held fixed.

5. Formalizing Stock vs. Flow

§3.4's distinction between the reduction failure (a) and the autonomy failure (b) can be stated using two independently established, already-named quantities from quantum information theory, rather than left as a verbal distinction.

Stock: the von Neumann entropy $S(\rho_O(t))$, the standard measure of how mixed $\rho_O(t)$ currently is, and hence (via Lemma A) of how badly the global state fails to be reconstructible from O 's data at time t . This is a STATE property: it depends on the entire history that produced $\rho_O(t)$, and once positive, cannot be reduced to zero again by any operation confined to O alone.

Flow: the CHANNEL property governing the elementary map $\Phi_{t, t+\delta t}$ taking $\rho_O(t) \mapsto \rho_O(t+\delta t)$. Wallman, Granade, Harper & Flammia (2015) define exactly this quantity for a general quantum channel Φ as its unitarity, $u(\Phi)$ - the average change in purity a channel induces over input pure states, a standard tool from

randomized-benchmarking practice for separating coherent (unitary-like) from incoherent (entangling/dissipative) noise. $u(\Phi_{t,t+\delta t})=1$ exactly when the elementary step is unitary (Lemma B's condition); $u<1$ signals genuine entangling coupling at that specific step, independent of $S(\rho_O(t))$.

With this vocabulary, the exact counter-example of §3.4 restates as: $S(\rho_O)$ can be strictly positive (stock already exhausted) while $u(\Phi_{t,t+\delta t})=1$ (flow momentarily perfect) at a specific step - two independent axes, not one. This connects the present result to the repeated-interaction (“collision model”) literature in open quantum systems (Rau 1963; for a comprehensive modern treatment, Ciccarello, Lorenzo, Giovannetti & Palma 2022), where a system undergoes a sequence of elementary interactions with fresh environment ancillas: the standard exponential approach to a stationary state in that literature is a flow statement (each collision's unitarity sets the rate), while the system's own entropy growth is the corresponding stock statement - the same two axes, in a setting where they are usually not given separate names, because in the simplest collision models with i.i.d. fresh ancillas stock and flow degrade together by construction and so are not normally worth distinguishing.

A gravitationally-coupled dark-sector observer illustrates a further point in this vocabulary: $u(\Phi_{t,t+\delta t})$ can remain extremely close to, but not exactly equal to, 1 over cosmologically long intervals, so that the flow condition is nearly but not exactly satisfied indefinitely, while $S(\rho_O)$ still accumulates monotonically, however slowly, so the stock condition still fails in the limit $t \rightarrow \infty$. The two axes of §3.4 therefore separate not only in the exact, instant-by-instant sense already demonstrated numerically, but also in an asymptotic, rate-based sense: near-vanishing flow-failure over a long window is compatible with eventual, unavoidable stock-failure.

6. Relation to Prior Work

The claim that an observer cannot occupy a fully external, God's-eye vantage point on a system unitarily including itself is not new. Wigner's 1961 friend argument already showed that treating a friend's measurement as an ordinary unitary interaction, from the outside, conflicts with the friend's own account of having obtained a definite outcome. Brukner (2018; building on an earlier 2016 discussion) derives a no-go theorem for observer-independent facts along related lines, showing that a shared, observer-independent account of measurement outcomes cannot be maintained once locality, freedom of choice, and the universal validity of quantum theory are assumed jointly, in a two-laboratory extension of the Wigner's-friend scenario. The Page–Wootters mechanism (1983) develops the same embeddedness into a positive account of time itself as relational, emerging from correlations between a clock subsystem and the rest - closely related to, and part of the inspiration for, the reduction argument of §2. Frauchiger & Renner (2018) sharpen the Wigner's-friend structure into an explicit inconsistency theorem: assuming quantum theory is universally valid, together with natural consistency requirements between nested observers, leads to contradictory conclusions among the observers themselves - a different, and in some respects stronger, kind of no-go result (an inconsistency, rather than an impossibility-of-simultaneous-properties), obtained through an explicit multi-agent Gedankenexperiment rather than through the entanglement/purification route used here. The present theorem is a complementary, more minimal result: rather than deriving a contradiction from a specific multi-agent protocol, it derives an unconditional limitation - no complete self-description, no autonomous unitarity - directly from entanglement generation across an observer's own physical boundary, for any embedded observer whatsoever, without needing to construct a Frauchiger–Renner-style scenario at all.

A precise word is needed on quantum reference frames (QRF), since the present setting could be mistaken for that literature. QRF work (Giacomini, Castro-Ruiz & Brukner 2019, and subsequent work) studies observers that can be placed in coherent superposition of reference-frame configurations (e.g. position or momentum) - a superposition over WHICH FRAME is used to describe the same physics, not over measurement outcomes. Vanrietvelde (2026) clarifies this distinction explicitly, defending the operational

meaning of position-superposed QRFs while explicitly distinguishing them from Wigner's-friend-type outcome-superposed observers, which raise separate (and in some proposed resolutions, disputed) issues. The present theorem belongs entirely to the latter (Wigner's-friend) category - O is a fixed subsystem with a fixed physical role as observer, and the question is what O can know and how O evolves, not whether O 's own reference frame is superposed. No claim about QRF-type frame superpositions is made or needed here, and the theorem should not be read as a QRF result.

A note on experimental tests of related no-go theorems, and why this one is not among them (yet). The extended-Wigner's-friend "Local Friendliness" (LF) programme (Brukner 2018; Bong et al. 2020, "A strong no-go theorem on the Wigner's friend paradox," *Nature Physics*; Proietti et al. 2019, "Experimental test of local observer independence," *Science Advances*) has produced actual experimental violations of LF inequalities, using single photons as the "Friend." This has been directly and specifically criticized on the grounds that a photon path lacks the persistent, recoverable memory a genuine observer would have (Baumann & Brukner 2024 show that any awareness by the friend of Wigner's operation on her own memory conflicts with no-signaling in these scenarios; related concerns are raised throughout the extended-Wigner's-friend literature). It is tempting to read the present theorem as bearing directly on that debate, but this temptation should be resisted without further work: the LF theorems are Bell-type inequality statements about a specific multi-observer, multi-setting protocol, while Theorem 3.1 is an unconditional statement about any single embedded observer once genuinely entangled, and no rigorous bridge between the two has been constructed here. If anything, one plausible (but unproven) implication runs opposite to the intuitive reading: since Theorem 3.1 places the same limitation on any observer with genuine persistent memory, not only on a memoryless photon, the memory critique's implicit assumption - that a "real" observer would retain complete, recoverable self-knowledge and thereby behave differently - may itself not survive scrutiny under the present framework. Establishing this rigorously is left for future work; no experimental claim, for or against the LF experiments, is made here.

The monotonic, operation-on- O -irreversible growth of $S(\rho_O(t))$ documented here (via Lemma A) gives an exact, general-purpose formalization of the irreversibility that decoherence-based accounts of branching (Zurek; the quantum-Darwinism literature) require: once an observer's subsystem has become entangled with the rest of a closed unitary system, no operation confined to that subsystem can restore access to the pre-entanglement global description. This is consistent with, and lends direct support to, the branching mechanism underlying the Many-Worlds Interpretation, rather than counting against it.

What should still be avoided is any claim that the paper "annuls," "refutes," or "rules out" MWI, or any competing interpretation. The theorem is interpretation-neutral by construction - it follows from (H1)-(H4) alone, without adjudicating between unitary-only (MWI-type) and collapse-based readings of measurement, and the one place it does connect substantively to an interpretational debate (§5's stock quantity, as above) supports the unitary/branching side rather than opposing it

7. Corollary: No Escape by Extension

A natural objection to Theorem 3.1 is that it only concerns observers *already* fixed as a proper subsystem of a *given* closed system H . One might hope to escape the theorem by introducing a further observer W , external to H , who watches the O - R system from outside and is thereby exempt from (H2). This section shows that this hope cannot be realized: any W capable of *physically learning* about H 's evolution is, by the theorem's own hypotheses, forced into exactly the role of an embedded observer of a strictly larger closed system - and Theorem 3.1 reapplies to W directly, without modification.

Definition 7.1 (Informative coupling)

Let H_W be an auxiliary Hilbert space, and let $H' := H \otimes H_W$, equipped with a joint unitary $\hat{U}$ generating the evolution of the extended system (H' itself satisfying (H1), i.e., H' is taken to be closed). Say W is informative about H over an interval $[t, t+\delta t]$ if the restriction of $\hat{U}$ to that interval is entangling across the bipartition $H | H_W$ - i.e. not of product form (an operator on H alone) $\otimes$ (an operator on H_W alone). This is precisely condition (H4), applied to the H - W split rather than the O - R split.

Corollary 7.2 (No Escape by Extension)

Let $(H, \hat{U})$ be a closed unitary system satisfying (H1), and let W be any system coupled to H via a joint unitary $\hat{U}$ on $H' = H \otimes H_W$, with H' itself satisfying (H1). Exactly one of the following holds.

(i) Uninformative case. If W is never informative about H (the H - W coupling is an exact product throughout), then no correlation between W 's data and H 's subsequent evolution is dynamically established beyond whatever was fixed at preparation. W learns nothing about H 's dynamics that it did not already encode initially. W is "external" only in the vacuous sense of being causally irrelevant to - and informationally cut off from - the evolution of H .

(ii) Informative case. If W is informative about H over some interval (Definition 3.5.1 holds), then $H' = H \otimes H_W$ satisfies (H1) by construction, W is a proper, nonempty subsystem of H' (so (H2) holds for H'), and the W - H coupling satisfies (H4) for the bipartition (W, H) of H' . Theorem 3.1 therefore applies **verbatim**, with the substitution $O := W$, $R := H$: W cannot simultaneously (a) reconstruct the complete global state of H' from its own accessible data, and (b) evolve under a fixed, autonomous unitary law - for exactly as long as it remains informative about H .

Proof.

(i) A product joint unitary $u_H \otimes u_W$ acting throughout $[t, t+\delta t]$ acts within each tensor factor separately; conjugation by a product unitary cannot increase the mutual information $I(H : W)$ of any state (this is the standard fact that local operations cannot generate correlation - entangling operations are precisely the non-product ones, by definition of (H4)/Definition 7.1). Hence any correlation present in $\rho_{HW}(t+\delta t)$ already existed in $\rho_{HW}(t)$, and was not established by this transition. Iterating over any interval throughout which the coupling remains product, W 's marginal knowledge of H is frozen at its initial value.

(ii) Immediate from the hypotheses: (H1) holds for H' by construction; (H2) holds for H' because W is by hypothesis a proper nonempty subsystem ($H_W \neq H'$, $H_W \neq \text{trivial}$); (H4) holds for the (W, H) bipartition by Definition 7.1. All hypotheses of Theorem 3.1 are satisfied with the roles of O and R played by W and H respectively, and the theorem's conclusion follows without any new argument. ■

Corollary 7.3 (Regress)

Iterating the construction - introducing a further observer W' of the once-extended system $H' = HH_W$ - reproduces the dichotomy of Corollary 7.2 one level up: at every stage, W' is either (i) causally irrelevant to H' , or (ii) itself a proper embedded subsystem of a strictly larger closed system $H'' = H' \otimes H_{W'}$, subject to Theorem 3.1 in turn. No finite chain of extensions terminates in an observer that is simultaneously (a) physically informed about the system it observes and (b) exempt from Theorem 3.1's conclusion.

Remark 7.4 (Relation to the Wigner's-friend regress)

Corollary 7.2-7.3 recovers, as a direct structural consequence of (H1)-(H4) alone, the regress familiar from the Wigner's-friend literature: promoting the friend's status by adding a further external observer (Wigner) merely relocates the same limitation to Wigner, once Wigner is treated as a physical system genuinely coupled to the friend. The present derivation shows this is not a contingent feature of the specific two-observer Wigner's-friend protocol - it requires no measurement postulate, no later interference experiment, no assumption about the friend's report - but a structural fact about *any* closed unitary extension whatsoever, following from the same four minimal hypotheses used throughout this paper.

Remark 7.5 (The only escape lies outside the formalism).

The dichotomy of Corollary 7.2 is exhaustive within (H1)-(H4). An observer that is to remain both physically informed about a closed unitary system's evolution and exempt from Theorem 3.1 cannot be introduced within this formalism: case (i) sacrifices informedness, case (ii) sacrifices exemption. The only way to introduce such an observer is to step outside the hypotheses altogether - e.g. by postulating a channel from H to W that is not unitary (violating (H1) for any joint closed description), or by placing W outside the scope of quantum-mechanical description as formalized here. Such a postulate is neither licensed nor excluded by the present framework.

7.6 Rank/Spectrum Refinement of the Informative Case

7.6.A.1 Rank-theoretic form of Theorem B

Recall the setup: $S \otimes A_1 \otimes \dots \otimes A_n$, redundant chain of copying links (Definition 2), cut $O_j = S \otimes A_1 \otimes \dots \otimes A_j$. Lemma A already gives $\text{rank}(\rho_j) = |\{i : c_i \neq 0\}|$, constant across the redundant segment. Theorem B showed the spectrum of ρ_j can change across a generic (non-copying) link. We now show the same is true, more dramatically, of the *rank* itself - directly in the $r(O, n)$ currency of Definition 8.6.

Corollary 7.6.B.1 (Rank inflation across a generic link). There exist a redundant prefix $S \rightarrow A_1$ (Definition 1) and a generic unitary W on $A_1 \otimes A_2$ such that $r(O_1, \text{before } W) = 1$, $r(O_1, \text{after } W) = \dim(O_1)$, i.e. the same subsystem $O_1 = S \otimes A_1$ goes from pure (fully self-contained, zero entanglement with anything) to maximally mixed relative to its own dimension (the largest rank Definition 8.6 allows for that subsystem), purely as a function of which side of a single generic interaction the cut is evaluated on.

Proof (explicit, verified construction).

Take $\dim S = \dim A_1 = 2$, $\dim A_2 = 4$. Let $c = (0.6, 0.8)$, and form the redundant state $|\psi^0\rangle = \sum_i c_i |i\rangle_S |i\rangle_{A_1} |0\rangle_{A_2}$ via a single copying link $S \rightarrow A_1$ (Definition 1). At this stage, $O_1 = S \otimes A_1$ is in a pure state $(\sum_i c_i |i\rangle, i)$, tensored with A_2 's blank state - so $r(O_1) = 1$ trivially: $\dim(O_1) = 4$, but the state is pure, hence rank 1.

Now apply a Haar-random unitary W on $A_1 \otimes A_2$ ($\dim 2 \times 4 = 8$; W acts identically on every S -branch, since S is a spectator). Numerically (verified, seed fixed for reproducibility):

$r(O_1, \text{before } W) = 1$, eigenvalues: $\{0, 0, 0, 1\}$

$r(O_1, \text{after } W) = 4$, eigenvalues: $\{0.0164, 0.0872, 0.3189, 0.5774\}$

The jump from 1 to $4 = \dim(O_1)$ is the maximum possible increase in rank Definition 8.6 permits for a subsystem of this dimension. Since W was Haar-random (not fine-tuned), this is the typical, not exceptional, outcome once $\dim A_2 \geq \dim(O_1)$ -matching the genericity language of Corollary 8.5a directly, now expressed as a rank statement rather than only a spectral one. ■

Corollary 7.6.B.2 (Consequence for $N(O)$). By Theorem 8.6/Corollary 8.6a and the §4.6 machinery (Theorem 4.6.4) of *Toward a Principle of Significance*, both $r(O, n)$ and any ε -significant ensemble

cardinality $N_e(O)$ computed from ρ_O are functions of ρ_O 's spectrum alone. Corollary B.1 therefore transfers immediately: **$N(O_1)$ computed just before a generic interaction and $N(O_1)$ computed just after it need not agree, and in the generic case will not.** Cut Invariance (§A of the earlier lemma) is consequently not a property of the system, but of the specific link crossed - precisely the same distinction Corollary 8.5a already draws for the No-Go theorem's own hypothesis, now shown to propagate downstream into the *Principle of Significance's* $N(O)$ as well.

7.6.A.2 Worked example: the photon ping-pong as a minimal quantized model

Scope and an explicit caveat. The classical ping-pong recursion (arbitrary real p, E, x, t) has no Hilbert space at all, so none of §8 or the Principle of Significance applies to it directly - this was established earlier in our conversation. What follows is *not* a derivation from the actual relativistic Compton S-matrix; it is a minimal, schematic qubit quantization, meant only to show *that* cut-dependence already appears at the level of a single generic scattering vertex, without invoking decoherence, dissipation, or any large-N argument. A genuine treatment would need continuous energy-momentum registers and the real QED vertex; nothing here should be read as a quantitative prediction for the real process.

Setup. Three qubits: S = electron A's "has it recoiled" register, A_1 = the photon's register, A_2 = electron B's register. Each vertex is modeled as the two-qubit partial-swap ("beam-splitter") gate $V(\theta)|00\rangle=|00\rangle$, $V(\theta)|01\rangle=\cos\theta|01\rangle+\sin\theta|10\rangle$, $V(\theta)|10\rangle=-\sin\theta|01\rangle+\cos\theta|10\rangle$, $V(\theta)|11\rangle=|11\rangle$, parametrized by a coupling angle θ ($\theta=0$: no interaction; $\theta=\frac{\pi}{2}$: full exchange) - a standard schematic for "partial transfer of one quantum of excitation," structurally analogous to a beamsplitter or Jaynes–Cummings-type exchange, standing in for the unmodeled real vertex. $V(\theta)$ is generic (non-copying) for every $0<\theta<\frac{\pi}{2}$: it does not satisfy the characterization proposition of Theorem B (its output on $|1\rangle\otimes|0\rangle$ is a genuine superposition, not a product state $|1\rangle\otimes|\phi\rangle$), so Theorem B applies directly. Initial state $|\Psi^0\rangle = |1\rangle_S|0\rangle_{A_1}|0\rangle_{A_2}$ (electron A "about to be struck," photon and B blank - the qubit analogue of $p_A=p_B=0$, ε_0 fixed). Apply $V(\theta_1)$ on (S, A_1) [vertex 1: $A\leftrightarrow\text{photon}$], then $V(\theta_2)$ on (A_1, A_2) [vertex 2: $\text{photon}\leftrightarrow B$] - exactly mirroring the causal order of the real process.

Verified results (three coupling-angle choices; all computed directly, not estimated):

θ_1, θ_2	spec ρ_S ($O = \{\text{electron A}\}$)	spec ρ_{S,A_1} ($O = \{A, \text{photon}\}$)	spec ρ_{A_2} ($O = \{\text{electron B}\}$)
$\frac{\pi}{4}, \frac{\pi}{4}$	{0.5, 0.5}	{0.25, 0.75}	{0.25, 0.75}
$\frac{\pi}{3}, \frac{\pi}{6}$	{0.25, 0.75}	{0.1875, 0.8125}	{0.1875, 0.8125}
1.0, 1.3	{0.2919, 0.7081}	{0.3426, 0.6574}	{0.3426, 0.6574}

Rank stays at 2 throughout (an artifact of using qubits - Corollary B.1 shows rank inflation requires the traced-out side to have enough room; a more realistic higher-dimensional register would show it too), but **the spectrum genuinely depends on which cut is evaluated**: $O=\{\text{electron A alone}\}$ and $O=\{\text{electron A} + \text{photon}\}$ give different weight distributions at every (θ_1, θ_2) tested, and hence -via §4.6: different $N(O)$. Note also that spec $\rho_{S,A_1} = \text{spec} \rho_{A_2}$ exactly in every row: this is not a coincidence but the standard

complementary-Schmidt-spectrum fact for a pure global state, and is itself a useful cross-check that the computation is correct.

7.7 Four Candidate Exits, and Why Each Either Narrows the Theorem's Scope or Fails to Escape It.

7.7.A. Algebraic structure: (H2) may fail outside finite dimensions

(H2) presupposes $H = H_O \otimes H_R$ for any admissible bipartition. This holds in finite dimensions and non-relativistic QM, but in continuum algebraic QFT the local algebra of observables in a subregion is generically a Type III₁ von Neumann factor: no trace, no density matrices, no well-defined partial trace, and no tensor-product decomposition $H = H_O \otimes H_R$ at all. $\rho_O(t)$ as defined in §2 is then simply not defined.

Caveat (genuine scope limitation, not a refutation): this does not falsify Theorem 3.1 for the finite-dimensional/factorizable setting it is proved in - the paper already says "for concreteness" about finite dimensions. It does mean the theorem's extension to continuum QFT is not automatic and requires the same machinery as §9 of the companion "Principle of Significance" note (Witten's crossed-product construction, Type II algebras, Araki relative entropy in place of von Neumann entropy). Recommended framing: state this explicitly as an open extension, analogous to how the companion paper treats it, rather than leaving (H2) to silently assume factorizability.

7.7.B. Hidden variables (Bohmian mechanics): an ontological, not operational, exit

In de Broglie–Bohm theory, a definite particle configuration exists at all times, independent of entanglement - an additional layer of ontology beyond ρ_O . One could argue a Bohmian agent's actual configuration is never truly 'unknown', only epistemically inaccessible due to quantum equilibrium (the ψ^2 -distribution constraint on knowable initial conditions).

Caveat: this does not touch Lemma A as an operational claim - $\rho_O(t)$ still fails to determine $|\Psi(t)\rangle$ from any measurement protocol confined to O, which is exactly what Definition 3.2's operational sense (echoed in the companion paper) requires. What changes is only the ontological status of (a)'s failure: in a hidden-variable ontology, there is a further fact of the matter (the actual configuration) that Theorem 3.1 does not speak to, but that fact remains permanently inaccessible to O under the equilibrium hypothesis. This is best flagged as a distinction between the ontological and operational readings of (a) - using exactly the vocabulary already introduced in Definition 8.1 - rather than as a counterexample.

7.7.C. QBism: dissolving the question rather than answering it

QBism (Fuchs, Caves, Schack) denies that p is an objective property of a system at all - it is an agent's personal credence function, not a physical fact that could be 'possessed' or 'inaccessible.' Under this reading, asking whether O can reconstruct $|\Psi(t)\rangle$ is a category error: $|\Psi\rangle$ was never a claim about the world available to be reconstructed.

Caveat: this is the cleanest of the four in the sense that it does not contest any proof step - it rejects the realist framing the question is posed in. It is a genuine instance of Remark 7.5's "step outside the hypotheses altogether" (specifically, outside the implicit realism about $|\Psi\rangle$ that (H1)–(H2) are usually read with), and should be listed alongside the two non-unitary-channel and outside-QM-description examples Remark 7.5 already gives, as a third named category of exit - not a refutation of the theorem within its own terms.

7.7.D. Retrocausality/the two-state vector formalism - full treatment

This is the most technically involved of the four, and the one requiring a careful mathematical argument rather than a one-paragraph caveat, since a superficial reading suggests postselection might let an observer access data effectively 'from the future.' It does not, and the reason is worth proving rather than asserting.

7.7.D.1 Setup (Aharonov–Bergmann–Lebowitz 1964; Aharonov–Vaidman)

A system pre-selected in $|\psi\rangle$ at t_1 evolves forward under the same global $\hat{U}$ as elsewhere in this paper: $|\psi(t)\rangle = \hat{U}(t, t_1)|\psi\rangle$. A system post-selected in $|\phi\rangle$ at t_2 (the outcome of a measurement performed at t_2) evolves backward under the same $\hat{U}$: $\langle\phi(t)| = \langle\phi|\hat{U}(t_2, t)$. For an intermediate measurement with projectors $\{\Pi_i\}$ at $t_1 < t < t_2$, the ABL rule gives

$$P(a_i | \psi, \phi) = \frac{|\langle\phi(t)|\Pi_i|\psi(t)\rangle|^2}{\sum_j |\langle\phi(t)|\Pi_j|\psi(t)\rangle|^2}$$

and the weak value for a weakly-coupled pointer is $A_w = \frac{\langle\phi(t)|\hat{A}|\psi(t)\rangle}{\langle\phi(t)|\psi(t)\rangle}$ (generically complex, generically outside the spectrum of $\hat{A}$).

7.7.D.2 Equivalence lemma: TSVF is standard QM's own retrodiction, not new physics

A direct computation (already present in ABL 1964) shows the ABL rule is mathematically identical to ordinary forward-evolving QM's Bayesian conditioning on the t_2 outcome: computing $P(a_i|\psi)$ by forward propagation from $|\psi\rangle$ alone, then conditioning via Bayes' rule on having subsequently observed ϕ at t_2 , reproduces the ABL probability exactly. TSVF is a symmetric-in-time restatement of the same predictions unitary QM already makes, not a distinct empirical theory and not a modification of the Schrödinger/ $\hat{U}$ dynamics.

7.7.D.3 Consequence 1: (H1) is not touched

$\hat{U}$ remains the sole dynamical law in both directions; $\langle\phi(t)|$ is obtained by applying the same $\hat{U}$ backward, not by any non-unitary or externally-sourced process. Nothing about TSVF, as an interpretation of standard unitary QM, violates global closure.

7.7.D.4 Consequence 2: no operational escape from Lemma A or Lemma B

Define O 's operationally accessible algebra at time t , $A_O(t)$, as containing only the outcomes of operations O has actually performed up to and including t - the same operational standard Definition 3.2 already uses. ABL probabilities and weak values are computed using $\langle\phi|$, the outcome of a measurement performed at t_2 . For $t < t_2$, $\langle\phi|$ is not an element of, and cannot be derived from, $A_O(t)$: it requires data from an event that has not yet occurred relative to O . TSVF quantities are therefore ensemble statistics computable only by an agent with access to $A_O(t_2)$, i.e. retrodictively, after t_2 . This is structurally identical to Corollary 7.2's treatment of an external observer W - informative only once actually causally coupled, and contributing nothing beforehand. $\rho_O(t)$, as already defined in §2 by a partial trace over data available up to t alone, remains the complete operational description available to O regardless of whether a time-symmetric two-vector description exists as a mathematical background structure.

7.7.D.5 The stronger move: a final boundary condition as ontological postulate

A genuinely different (and stronger) claim - not TSVF-as-interpretation but a modification of what counts as "the state" - posits a fixed final boundary condition as a physical fact about the system, not a retrodictive bookkeeping device. This is the structure of the Horowitz–Maldacena (2004) black-hole final-state

proposal and of Gell-Mann-Hartle's generalized quantum mechanics with both initial and final density matrices. If cited as a route around Theorem 3.1 in a quantum-gravity context, this distinction - postulate, not interpretation - should be stated explicitly, since only the postulate route is a substantive departure from the paper's premises at all.

7.7.D.6 A necessary caveat for this stronger move: known unitarity obstruction

Gottesman & Preskill (2004) showed that the Horowitz–Maldacena final-state proposal, as originally stated, does not generically preserve exact unitarity of black-hole evaporation for arbitrarily entangled infalling matter - cloning-type problems arise outside a special (maximally entangled) final-state projection. Any citation of a final-boundary-condition postulate as a physical (not merely interpretive) escape route should cite this objection alongside the original proposal; citing Horowitz-Maldacena without it substantially overstates how settled this route is.

7.7.D.7 Even under the strong postulate, only the ontological reading of (a) is affected

Suppose a fixed final state does, for some non-degenerate dynamics, select a unique global purification $|\Phi\rangle$ consistent with both boundaries - this is not automatic and depends on the specific dynamics, but grant it for the sake of argument. This gives, at most, an ontological fact in the sense of Definition 8.1: a well-defined $|\Psi(t)\rangle$ exists. It does not give O an operational route to (a), because determining which purification is the correct one still requires access to ϕ , unavailable to O before t_2 by the same argument as D.4. The strong postulate, even if fully accepted as physics, alters only Theorem 3.1(a)'s ontological status, not its operational content - precisely the distinction Definition 8.1 already draws for a different purpose in §8.

7.7.D.8 Theorem (Two-State Purification Non-Uniqueness)

Claim. If $\rho_O(t)$ has rank ≥ 2 (equivalently $S(\rho_O(t)) > 0$), or $\sigma_O(t)$ has rank ≥ 2 (equivalently $S(\sigma_O(t)) > 0$), then there exist infinitely many distinct pairs $(|\Psi_i\rangle, |\Phi_f\rangle)$ of global pure boundary conditions - related to a given pair by unitaries acting on H_R alone, chosen independently for the forward and backward branches - that reproduce exactly the same operationally accessible pair $(\rho_O(t), \sigma_O(t))$.

Proof. $|\Psi(t)\rangle$ and $|\Phi(t)\rangle$ are two distinct vectors in H , with no algebraic relation to one another beyond both being generated by the same fixed global unitary $\hat{U}$ acting on their own, independently chosen boundary data - this independence is not an added assumption but the defining feature of the two-state formalism itself: pre-selection $|\Psi_i\rangle$ and post-selection $\langle\Phi_f|$ are, by construction, freely and separately specifiable physical inputs (this is exactly what distinguishes the ABL rule, a conditional probability computed across two independent boundary conditions, from ordinary single-boundary quantum mechanics). Apply the Lemma A / HJW construction (§3.1) to each vector separately. For $\rho_O(t) = \sum_k p_k |k\rangle\langle k|$ (rank ≥ 2 by hypothesis), any orthonormal $|f_k\rangle \subset H_R$ gives $|\Psi(t)\rangle = \sum_k \sqrt{p_k} |k\rangle \otimes |f_k\rangle$ with $\text{Tr}_R[|\Psi(t)\rangle\langle\Psi(t)|] = \rho_O(t)$ exactly; running $\hat{U}$ backward to t_i recovers a corresponding $|\Psi_i\rangle$. Independently, for $\sigma_O(t) = \sum_l q_l |l\rangle\langle l|$ (if also rank ≥ 2), any orthonormal $|g_l\rangle \subset H_R$ gives $|\Phi(t)\rangle = \sum_l \sqrt{q_l} |l\rangle \otimes |g_l\rangle$ with $\text{Tr}_R[|\Phi(t)\rangle\langle\Phi(t)|] = \sigma_O(t)$ exactly; running $\hat{U}$ forward to t_f recovers a corresponding $|\Phi_f\rangle$. Since $\{|f_k\rangle\}$ and $\{|g_l\rangle\}$ parametrize two separate purifications of two separate vectors on the same factor H_R , with no constraint of the formalism linking them, they vary independently; the (rank($\rho_O(t)$)-dimensional) unitary freedom in $\{|f_k\rangle\}$ and the (rank($\sigma_O(t)$)-dimensional) unitary freedom in $\{|g_l\rangle\}$ therefore compose into a strictly larger, product family of boundary-condition pairs $(|\Psi_i\rangle, |\Phi_f\rangle)$, all distinct as global data, all consistent with the identical accessible pair $(\rho_O(t), \sigma_O(t))$. If either marginal has rank ≥ 2 , this family is already infinite. ■

7.7.D.9 Recommended citation-safe statement about two-state/retrocausal

Two-state/retrocausal reformulations of quantum mechanics are mathematically equivalent to standard unitary QM conditioned via Bayes' rule on a later measurement outcome (Aharonov-Bergmann-Lebowitz, 1964); they neither violate (H1) nor grant an embedded observer any operationally accessible information beyond $\rho_O(t)$ prior to the post-selection event, so Theorem 3.1 is unaffected by adopting the two-state vector formalism as an interpretation of standard QM. A genuine departure would instead require positing a final boundary condition as an ontological postulate (e.g. Horowitz & Maldacena, 2004) rather than a retrodictive re-description of existing statistics - a move independently known to face unitarity obstructions (Gottesman & Preskill, 2004) - and even this stronger move alters only the ontological reading of Theorem 3.1(a) in the sense of Definition 8.1, not its operational content, which is unaffected in either case.

8. Corollary: No Escape by Identification with the Whole

§7 addressed the hope of escaping Theorem 3.1 by placing the observer *outside* H. A complementary hope is to place the observer *at* H itself - identifying the observer not with a proper subsystem but with the whole closed system, so that some designated part $R' \subset H$ is "known" simply because it lies within the same total state the observer is identified with. The hand/body intuition is a natural way to state this hope: the whole organism, so the thought goes, already "contains" full knowledge of its own hand, entanglement notwithstanding. This section shows the hope trades on an ambiguity in what "the observer knows" means, and that resolving the ambiguity either empties the claim or reintroduces Theorem 3.1 unchanged.

Definition 8.1 (Locus of knowledge)

Given $H = H_M \otimes H_{R'}$ for some designated candidate "reader" subsystem M and its complement R' , say a claim of the form "**M knows the state of R**" is:

ontological, if it means only that the global pure state $|\Psi(t)\rangle$ - of which R 's reduced state is a marginal - exists and is well-defined; or

operational, if it means that some algebra of observables $A_M := B(H_M) \otimes \hat{I}_{R'}$, confined to M, reproduces (via measurement outcomes accessible to M alone) information about R 's state - i.e. the sense of "accessible data" already used throughout §2–§3.

Only the operational sense is a physical claim about *knowledge*, in the sense relevant to Theorem 3.1; the ontological sense is a claim about the existence of a state, true trivially for every closed system regardless of any observer.

Corollary 8.2 (No Escape by Identification with the Whole)

Let $H = H_M \otimes H_{R'}$, with $R' \subseteq H$ nonempty and proper. Then:

(i) The degenerate identification $M = H$, $R' = \emptyset$. If the "super-observer" is identified with the whole of H (H2's requirement that R be nonempty is violated by construction), no operational claim of knowledge about anything is being made at all: there is no designated complement whose state could be *inaccessible*, since nothing lies outside M. Any claim that the whole "knows" its own state in this configuration is the ontological claim of Definition 8.1 - true, but empty of the operational content the word "knows" ordinarily carries. This is not a case Theorem 3.1 fails to cover; it is the case its hypothesis (H2) explicitly excludes as vacuous, and no physical work is done by pointing to it.

(ii) The non-degenerate identification: a reader must be named. Any operational claim that "the whole knows about its part R " that is not vacuous in the sense of (i) requires specifying *which* physical degrees of freedom within H actually carry the correlation with R' - a memory, a sensory pathway, a neural correlate, or any other physically instantiated reader $M \subsetneq H$, $R' \subseteq H \setminus M$. But once M is named, M is by construction a proper, nonempty subsystem of H (H2 holds for the pair $(M, H \setminus M)$), and if M 's coupling to R' is genuinely correlating (H4 holds), Theorem 3.1 applies verbatim to M with the substitution $O := M$, $R := H \setminus M$. The "super-observer" has not escaped the theorem; it has been resolved, on inspection, into an ordinary embedded observer, and the theorem was never evaded - only its subject was, prior to specification, left unnamed.

Proof. (i) is immediate from (H2): R nonempty is a stated hypothesis of Theorem 3.1, so $M=H$, $R'=\emptyset$ simply falls outside its domain, and - since $A_M = B(H) \otimes \hat{I}_\emptyset$ is the full algebra of observables on H - no restriction of accessible data occurs, hence no operational knowledge-claim distinct from the trivial one is available to evaluate. (ii) is immediate from Theorem 3.1 itself, once O, R are read as $M, H \setminus M$ respectively; no new argument is required beyond confirming (H2) and (H4) hold for this bipartition, which holds by the construction of M as a genuinely correlating reader. ■

Corollary 8.3 (The cost of acquired correlation) The hand/body intuition contains a second, independent claim worth making precise: that a designated reader M can *come to know* R' through a process of interaction, where beforehand it did not. This is correct, and Lemma A gives it an exact quantitative form.

Let $H = H_M \otimes H_{R'}$ be closed, with global state pure at all times (H1). Since $S(\rho_M(t)) = S(\rho_{R'}(t))$ for a globally pure bipartite state, the mutual information between M and R' satisfies

$$I(M:R';t) = S(\rho_M(t)) + S(\rho_{R'}(t)) - S(\rho_{MR'}(t)) = 2S(\rho_M(t)),$$

since $S(\rho_{MR'}(t)) = 0$ (global purity). Consequently:

Any acquisition of correlational knowledge by M about R' — any increase of $I(M:R')$ from an initial value of zero - is identically, and only, an increase in $S(\rho_M(t))$: the entanglement entropy of M relative to its own complement. This is exactly the quantity whose positivity is the hypothesis of Lemma A. M cannot acquire nonzero operational knowledge of R' without simultaneously, and to the same exact degree, satisfying Lemma A's condition for the non-reconstructibility of the global state from M 's own data. The two are not merely correlated in practice; for a closed bipartite system they are two readings of one and the same number.

Proof. Direct computation from the definition of mutual information and the identity $S(\rho_M) = S(\rho_{R'})$ for a pure $|\Psi\rangle$ on $H_M \otimes H_{R'}$ (Araki–Lieb / standard entanglement-entropy symmetry). At $t=0$, if $\rho_M(0)$ is taken pure (M initially uncorrelated with R' , as in the "hand not yet felt" starting condition of the motivating example), $I(M:R';0)=0$; any subsequent active, non-fine-tuned coupling (H4) generates $S(\rho_M(t))>0$ by Lemma B's mechanism, and the identity above converts this directly into $I(M:R';t)>0$. ■

Remark 8.4 (What the hand example actually shows)

The intuition behind the motivating example - that a prior lack of felt correlation between a reader and a bodily part is dissolved by a sequence of interactions and observations - is not a counterexample to Theorem 3.1; by Corollary 8.3 it is a direct instance of Lemma A and Lemma B operating exactly as stated, with the specific numerical identification $I = 2S$ making the trade explicit: the knowledge gained is bought at the precise price of the reader's own entanglement entropy, not obtained for free by expanding the

reader's identity to include the whole. This is the same phenomenon documented, in a different vocabulary, by the decoherence and quantum-Darwinism literature (proliferation of redundant records of a system's state into an environment, read out by localized observers) - the present corollary states its exact accounting relation for a bipartite closed system, without needing the further apparatus (many independent environment fragments, redundancy) of that literature.

Remark 8.5 (Symmetry with §7)

§7 and §8 are mirror images of a single fact about (H1)–(H4): the boundary of "the observer" can be pushed outward (§7, adding W external to H) or inward (§8, seeking a reader M internal to H distinct from H itself), and in neither direction does relocating the boundary produce an observer exempt from Theorem 3.1 - it only ever produces a new, and equally embedded, instance of the same bipartition the theorem already covers. The only genuinely exempt case in either direction is degenerate: an external W with no informative coupling (§7.2(i)), or an internal M identified with the whole and hence with no designated complement (§8.2(i)) - and both degenerate cases purchase their exemption by giving up the very thing ("knowing something about a part it is not") that made the observer interesting to consider in the first place.

9. Toward an Experimental Test

Conditions (a) and (b) of Theorem 3.1, taken alone, are not new experimental territory: that an entangled subsystem's reduced state is mixed and does not evolve autonomously is routine, extensively verified quantum mechanics (entanglement generation, state tomography, and decoherence are observed constantly on existing hardware). Two more specific, less commonly isolated claims of this paper are directly testable on present-day gate-based quantum processors.

9.1 Testing the stock/flow separation (§3.4)

Protocol: on a two-qubit register (O, R), apply a calibrated entangling gate (e.g. CNOT or CZ, >99.5% fidelity on current superconducting or trapped-ion hardware) to reach a known mixed ρ_O with measured purity $\text{Tr}[\rho_O^2] < 1$. Perform full state tomography on O. Then apply ONLY single-qubit operations to O and R separately - an exactly product interaction by construction, with no two-qubit gate active. Perform tomography on O again. Prediction: purity(ρ_O) is unchanged, and $\rho_O(t+\delta t) = u_O \rho_O(t) u_O^\dagger$ for the applied single-qubit rotation u_O , to within tomography error ($\sim 10^{-2}$ – 10^{-3} on current hardware) - confirming condition (b) holds exactly for that transition despite (a) having already failed. A deviation beyond the tomography error budget would falsify the exact-unitarity prediction for a nominally product step, and would itself be diagnostic of residual entangling crosstalk in the "product" gate as implemented.

9.2 Testing the exact light cone (§4.2–§4.3)

Protocol: implement a genuine brickwork (colour-sublayered) local circuit on a many-qubit register, of the kind already used in existing digital quantum simulation and information-scrambling experiments (cf. Landsman et al. 2019, "Verified quantum information scrambling," Nature, for a directly relevant experimental platform and measurement technique). Measure an out-of-light-cone correlator between a fixed operator at one end of the register and a Heisenberg-evolved operator at the other, as a function of circuit depth n , exactly as in §4.3. The distinguishing signature: Theorem 4.2 predicts the correlator should be consistent with EXACTLY zero (bounded only by gate-error rates, not by any intrinsic exponential tail) before the predicted causal threshold $n = \lceil \frac{d}{K} \rceil$, in contrast to the residual exponential tail a genuinely continuous-time (analog Hamiltonian) evolution would produce at the same nominal depth. Distinguishing these two hypotheses requires an honest gate-error noise-floor characterization - the observed residual signal below threshold should scale with the calibrated two-qubit gate error rate, not with any additional model-

dependent parameter, for the exact-zero prediction to be considered confirmed rather than merely consistent with noise.

Statement of contribution. What is specifically new in the present treatment, relative to this literature, is narrower and more technical than the underlying phenomenon, and is stated plainly rather than implied: (i) an exact, not exponentially-suppressed, causal bound available specifically for discrete local circuits (§4.2–§4.3), sharper than the generic Lieb–Robinson bound because of the strict edge-disjointness of circuit sublayers; and (ii) the explicit logical separation of the reduction failure (stock, $S(\rho_O)$) from the autonomy failure (flow, channel unitarity $u(\Phi)$), together with a constructive counter-example demonstrating the two are not one fact restated - a distinction the collision-model and open-quantum-systems literature has the tools to make (§5) but does not typically draw out explicitly, since in the standard i.i.d.-ancilla setting the two axes happen to coincide. Neither point overturns or supersedes Wigner, Page–Wootters, or Frauchiger–Renner; both sharpen one specific technical corner of an already well-established family of results, using already-established quantities (von Neumann entropy, channel unitarity) rather than new formalism.

References

- Wigner, E. P. (1961). Remarks on the mind-body question. In I. J. Good (ed.), *The Scientist Speculates*. Heinemann, London.
- Page, D. N., & Wootters, W. K. (1983). Evolution without evolution: Dynamics described by stationary observables. *Phys. Rev. D*, 27, 2885.
- Lieb, E. H., & Robinson, D. W. (1972). The finite group velocity of quantum spin systems. *Commun. Math. Phys.*, 28(3), 251–257.
- Vizing, V. G. (1964). On an estimate of the chromatic class of a p-graph. *Diskret. Analiz*, 3, 25–30.
- Hughston, L. P., Jozsa, R., & Wootters, W. K. (1993). A complete classification of quantum ensembles having a given density matrix. *Phys. Lett. A*, 183(1), 14–18.
- Frauchiger, D., & Renner, R. (2018). Quantum theory cannot consistently describe the use of itself. *Nature Communications*, 9, 3711.
- Giacomini, F., Castro-Ruiz, E., & Brukner, Č. (2019). Quantum mechanics and the covariance of physical laws in quantum reference frames. *Nature Communications*, 10, 494.
- Vanrietvelde, A. (2026). Specifying the operational meaning of quantum reference frame transformations. *arXiv:2607.03417*.
- Rau, J. (1963). Relaxation phenomena in spin and harmonic oscillator systems. *Phys. Rev.*, 129(4), 1880–1888.
- Ciccarello, F., Lorenzo, S., Giovannetti, V., & Palma, G. M. (2022). Quantum collision models: Open system dynamics from repeated interactions. *Physics Reports*, 954, 1–70.
- Wallman, J., Granade, C., Harper, R., & Flammia, S. T. (2015). Estimating the coherence of noise. *New Journal of Physics*, 17, 113020.
- Nielsen, M. A., & Chuang, I. L. (2000). *Quantum Computation and Quantum Information*. Cambridge University Press.
- Brukner, Č. (2016). On the quantum measurement problem. In R. Bertlmann & A. Zeilinger (eds.), *Quantum [Un]Speakables II*, Springer. *arXiv:1507.05255*.
- Brukner, Č. (2018). A no-go theorem for observer-independent facts. *Entropy*, 20(5), 350. *arXiv:1804.00749*.
- Baumann, V., & Brukner, Č. (2024). Wigner's friend's memory and the no-signaling principle. *Quantum*, 8, 1481.
- Proietti, M., Pickston, A., Graffitti, F., Barrow, P., Kundys, D., Branciard, C., Ringbauer, M., & Fedrizzi, A. (2019). Experimental test of local observer independence. *Science Advances*, 5(9), eaaw9832.
- Bong, K.-W., Utreras-Alarcón, A., Ghafari, F., Liang, Y.-C., Tischler, N., Cavalcanti, E. G., Pryde, G. J., & Wiseman, H. M. (2020). A strong no-go theorem on the Wigner's friend paradox. *Nature Physics*, 16, 1199–1205.
- Landsman, K. A., Figgatt, C., Schuster, T., Linke, N. M., Yoshida, B., Yao, N. Y., & Monroe, C. (2019). Verified quantum information scrambling. *Nature*, 567, 61–65.

- Aharonov, Y., Bergmann, P. G., & Lebowitz, J. L. (1964). Time symmetry in the quantum process of measurement. *Physical Review*, 134(6B), B1410–B1416.
- Aharonov, Y., & Vaidman, L. (1990). Properties of a quantum system during the time interval between two measurements. *Physical Review A*, 41(1), 11–20.
- Aharonov, Y., Popescu, S., & Tollaksen, J. (2010). A time-symmetric formulation of quantum mechanics. *Physics Today*, 63(11), 27–32.
- Gell-Mann, M., & Hartle, J. B. (1990). Quantum mechanics in the light of quantum cosmology. In W. H. Zurek (ed.), *Complexity, Entropy, and the Physics of Information*. Addison-Wesley.
- Reeh, H., & Schlieder, S. (1961). Bemerkungen zur Unitäräquivalenz von Lorentzinvarianten Feldern. *Nuovo Cimento*, 22(5), 1051–1068.
- Bisognano, J. J., & Wichmann, E. H. (1976). On the duality condition for quantum fields. *Journal of Mathematical Physics*, 17(3), 303–321.
- Takesaki, M. (1970). *Tomita's Theory of Modular Hilbert Algebras and Its Applications*. Lecture Notes in Mathematics 128. Springer.
- Witten, E. (2018). Notes on some entanglement properties of quantum field theory. *Reviews of Modern Physics*, 90, 045003.
- Zurek, W. H. (2003). Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75, 715–775.
- Zurek, W. H. (2009). Quantum Darwinism. *Nature Physics*, 5, 181–188.
- Everett, H. (1957). "Relative state" formulation of quantum mechanics. *Reviews of Modern Physics*, 29(3), 454–462.
- Materov, S. (2026) "Toward a Principle of Significance", DOI 10.5281/zenodo.22207800
- Materov, S. (2026) "Computational Appendix - The Embedded-Observer No-Go Theorem".
https://github.com/SergejMaterov/Embedded_Observer_NoGo