

The Embedded-Observer No-Go Theorem

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

DOI: 10.5281/zenodo.21730119

Sergej Materov

© 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.

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

$$\rho_O(t) := \text{Tr}_R[|\Psi(t)\rangle\langle\Psi(t)|].$$

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 $\text{Tr}_R[|\Phi\rangle\langle\Phi|] = \rho_O(t)$ 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 $\text{rank} \geq 2$, 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 (Nielsen, M. A., & Chuang, I. L. 2000; Stinespring, W. F. 1955). 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.

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 is 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

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

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

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.

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. 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 is 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.

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.
- Stinespring, W. F. (1955). "Positive functions on C*-algebras." *Proc. Amer. Math. Soc.*, 6(2), 211–216
- Materov, S. (2026) "Computational Appendix — The Embedded-Observer No-Go Theorem".
https://github.com/SergejMaterov/Embedded_Observer_NoGo