The Physical Infrastructure of Experience:

Criticality, Non-Factorizability, and the Causal Core as Interpreter

Abstract

Contemporary theories of consciousness share a common methodological deficit: they describe functional correlates of experience without determining the structural conditions under which experience as such becomes possible at all. The present article develops a framework for the physical infrastructure of experience, the set of necessary structural conditions, not sufficient causal mechanisms, that a physical system must satisfy in order to make phenomenal experience possible. Drawing on the theory of causal collapse and the concept of the causal core as interpreter (Stegemann 2026a, 2025a, 2025b), the article argues that three mutually conditioning structural features constitute this infrastructure: (1) criticality, understood as the dynamic regime of phase transitions in which characteristic length scales diverge; (2) non-factorizability, understood as the ensuing impossibility of decomposing the causal structure of the system into independent partial causes; and (3) the causal core as a dynamic attractor in the system's phase space, providing the reflexive dimension without which integrated causal unity would remain inert. Mathematically, three independent formal strands converge on the same structural situation: the Hammersley-Clifford theorem, the non-analyticity of the partition function, and the violation of the entanglement entropy area law, the last as a formal characterization tool, not a mechanistic hypothesis about neural quantum processes. An explicit demarcation from Integrated Information Theory shows that non-factorizability in the present framework is a categorical structural property, not a scalar measure. Comparison with superconducting coherence sharpens the anti-panpsychist argument, and concrete proposals for empirically distinguishing active from passive criticality are developed. The framework does not claim to solve the hard problem; it claims to determine the physical preconditions without which the hard problem cannot arise at all.

Keywords: *consciousness; causal collapse; non-factorizability; criticality; causal core; interpreter; physical infrastructure of experience; autocatalytic closure; attractor; phase transitions; Integrated Information Theory*

1. Introduction

The question of how physical processes give rise to subjective experience has shaped the philosophy of mind for at least three decades, since Chalmers (1995) gave it the canonical formulation as the *hard problem of consciousness*. Contemporary theoretical responses have been numerous and varied: Integrated Information Theory (Tononi 2004, 2014) attempts to quantify consciousness through a measure of integrated information; Global Workspace Theory (Baars 1988; Dehaene & Changeux 2011) identifies consciousness with the broadcast of information across a global neural workspace; Predictive Processing (Clark 2016; Friston 2010) models the brain as a hierarchical prediction machine that minimizes free energy. Each approach has generated substantial empirical research programmes and theoretical elaborations. Yet they share a structural deficit that, from the perspective developed here, is not accidental but systematic: they operate within a correlative methodology that identifies neural or functional correlates of experience without asking the deeper question of which structural conditions a physical system must satisfy in order to be capable of experience at all (Stegemann 2026b).

The present article does not attempt to solve the hard problem. It undertakes something more modest but, in a certain sense, more fundamental: determining the *physical infrastructure* of experience. By infrastructure is meant the set of necessary structural conditions, not sufficient mechanisms, that a physical system must satisfy in order for the question of experience to be applicable to it at all. A system that fails to satisfy these conditions is not one to which the question of whether it has experience can be meaningfully posed; a system that satisfies them is one for which the explanatory gap that constitutes the hard problem genuinely opens up.

The framework rests on three mutually conditioning structural concepts developed in a series of earlier publications (Stegemann 2025a, 2025b, 2025c, 2025d, 2025e, 2025f, 2025g, 2026a, 2026c): causal collapse, non-factorizability, and the causal core as interpreter. Criticality, the dynamic regime of phase transitions, provides the physical realization condition of non-factorizability, and the causal core as dynamic attractor in the system's phase space provides the reflexive dimension that transforms integrated causal unity into the structural precondition of perspectivity. These three features are not independent theoretical postulates but convergent characterizations of a single structural situation described from dynamic, causal, and reflexive perspectives.

Section 2 introduces the conceptual background of causal factorizability and its breakdown. Section 3 develops the connection between criticality and non-factorizability on the basis of mathematical results from statistical mechanics, probability theory, and quantum information. Section 4 introduces the causal core and the concept of the interpreter, and develops the argument for why reflexive causal structure constitutes the minimal structure of perspectivity. Section 5 develops the positive concept of physical infrastructure, demarcates it from standard approaches to neural correlates of consciousness, and draws an explicit boundary with Integrated Information Theory. Section 6 addresses three objections. Section 7 concludes.

2. Causal Factorizability and Its Breakdown

2.1 The Standard Picture

In the standard treatment of causal systems, whether in Woodward's (2003) interventionist framework, Pearl's (2000) Structural Causal Models, or the Humean tradition, a system is causally characterizable if its behavior can be represented as a function of the independent values of its component variables. The joint probability distribution $P(X_1, \ldots, X_n)$ factorizes along a directed acyclic graph (DAG) as a product of conditional distributions:

$$P(X_1, \ldots, X_n) \ = \sum_i P(X_i \mid pa(X_i))$$

where $pa(X_i)$ denotes the set of parent variables of X_i in the DAG. This factorizability encodes the causal assumption that each variable is determined by its direct causes without residual global dependencies. Pearl (2000) formalizes this as the *causal Markov condition*: conditional on its parent variables, each variable is independent of all non-descendants. The implicit presupposition is the acyclicity of the causal graph. As argued in Stegemann (2025b), this presupposition is structurally violated by living, autocatalytic systems, generating a series of category errors in the application of causal modelling to consciousness research.

2.2 Autocatalytic Systems and the Breakdown of DAG Representation

An autocatalytic system is one in which the products of certain reactions serve as catalysts for other reactions in the same set, so that the system as a whole reproduces its own catalytic closure (Kauffman 1993; Varela, Maturana & Uribe 1974). In the terminology of Stegemann (2025b, 2026c), autocatalytic closure is the minimal structural condition for *causal poiesis*: the capacity of a system to be the cause of its own causal structure. The decisive mathematical consequence is the presence of cycles in the causal structure: A causes B, which causes C, which in turn modifies the catalytic activity of A. The causal graph is not a DAG but a fixed-point system of simultaneous equations. The causal Markov condition fails not as an exception but as a structural feature, and with it, the factorizability of the joint distribution fails accordingly.

The Hammersley-Clifford theorem (Hammersley & Clifford 1971; Lauritzen 1996) formalizes the connection: a strictly positive probability distribution factorizes with respect to a graph G if and only if it satisfies the Markov properties with respect to G. Non-factorizability in the technical sense means that no finite undirected graph can represent the dependency structure of the system, dependencies permeate all scales and cannot be exhaustively captured by any local structure of conditional independence. This mathematical result provides the formal counterpart to the phenomenological observation that conscious systems cannot be adequately described by decomposing their behavior into independent component processes. The indivisibility of conscious experience, what Husserl (1900/1975) analyzed as the horizon structure of intentionality and what Merleau-Ponty (1945/1966) described as the motor unity of the bodily subject, has a precise structural correlate in the mathematical non-factorizability of autocatalytic causal systems.

3. Criticality as the Physical Realization Condition of Non-Factorizability

3.1 The Physics of Critical Points

Phase transitions are characterized by the behavior of the correlation length ξ, which measures the spatial extent over which fluctuations in one part of the system are correlated with fluctuations in another. Away from the critical point, ξ is finite: correlations decay exponentially as $C(r) \sim exp\left(-\frac{r}{\xi}\right)$. At the critical point T_c, the correlation length diverges: $\xi \rightarrow \infty$. Correlations follow a power law $C(r) \sim r^{-\alpha}$ without a characteristic scale. In thermodynamic terms, the partition function

$$Z = \sum exp\left(-\frac{H}{kT}\right)$$

develops a non-analyticity at T_c: the free energy $F = -kT \ln Z$ cannot be written as a sum of independent subsystem contributions. The system is non-decomposable in a rigorous thermodynamic sense. Criticality is the physical realization condition of non-factorizability: the dynamic regime in which non-factorizability becomes a structural necessity rather than an approximation error.

3.2 Self-Organized Criticality and Biological Systems

The relevance of criticality to neural systems is extensively documented. Beggs and Plenz (2003) showed that spontaneous neural activity in cortical networks follows a power-law distribution of avalanche sizes, a hallmark of criticality. Subsequent work by Chialvo (2010), Shew and Plenz (2013), and Haimovici et al. (2013) established that the brain operates near a critical point, associated with maximal information transmission, dynamic range, and perturbation sensitivity. What is less often noted is that biological systems do not merely happen to lie near the critical point: they actively maintain proximity to it through self-organized regulatory mechanisms. Autocatalytic closure in the sense of Stegemann (2025b, 2026c), drawing on Maturana and Varela (1980), is precisely such a mechanism. The autocatalytic system does not passively occupy the critical point; it *stabilizes* it as its normal mode of operation. This is what Stegemann (2025a, 2025d) has called *causal collapse*: the systematic displacement of component-level causation by a global causal mode that admits no decomposition into independent partial causes.

3.3 Quantum Information and Entanglement Entropy

A further mathematical characterization of non-factorizability at criticality comes from quantum information theory. A bipartite quantum state

$$|\psi\rangle \in H_A$$

$\otimes H_B$ is non-factorizable (entangled) if and only if its Schmidt rank exceeds 1, that is, if it cannot be written as $|\psi\rangle = |\varphi\rangle_A \otimes |\chi\rangle_B$. The entanglement entropy $S(\rho_A) = -Tr(\rho_A \, log \, \rho_A)$ quantifies the degree of non-factorizability.

Away from criticality, an area law holds: $S \sim L^{d-1}$, meaning that entanglement entropy scales with the boundary area of a region, not its volume. At the critical point this is violated and $S \sim log \, L$, a logarithmic scaling calculated exactly for conformal field theories (Calabrese & Cardy 2004; Vidal et al. 2003). The violation is a precise formal indicator that non-factorizability has become system-wide.

The epistemic status of this argument must be kept clear. Entanglement entropy is used here as a mathematical tool for the formal characterization of non-factorizability, not as a claim about the mechanism of neural processing. This distinguishes the present framework fundamentally from hypotheses such as Orch-OR (Penrose 1994; Hameroff 1998), which postulate a mechanistic role for quantum processes in neural microtubules. Whether biological systems exploit quantum entanglement in the technical sense is an open empirical question of no relevance to the structural theory developed here. What matters is that the same structural situation, non-factorizability at criticality, is independently characterized by three formalisms: the Hammersley-Clifford theorem, the non-analyticity of the partition function, and the violation of the area law.

4. The Causal Core as Interpreter

4.1 From Unity to Reference: The Causal Core as Attractor and the Constitution of Perspectivity

Non-factorizability establishes that the system constitutes an irreducible causal unity. Causal unity alone, however, does not yet constitute the infrastructure of experience. A coherent laser pulse, a ferromagnetic domain at the Curie point, or the global synchronization of a reaction-diffusion system all represent forms of non-factorizable causal unity that we do not typically regard as candidates for experience. The additional structural feature required is reflexivity: the system must not merely constitute a causal unity but be related to itself *as* a unity.

What exactly is the causal core (CC)? It is neither a localized substructure of the system nor a specific node in the neural network. It is a *dynamic attractor in the phase space of the system*, localized not in physical space but in the space of possible system states. In dynamical systems theory, an attractor is a subset of phase space to which a system converges under its own dynamics and near which it returns after perturbations. The CC as attractor is that region of phase space in which the non-factorizable global state of the system assumes its most stable operative form, the form in which the global state is not merely instantiated but used by the system itself as the starting point for processing incoming signals.

This characterization resolves a potential conceptual tension. One might ask: if the CC is not a local substructure but the 'functional centre' of self-referential closure, is that not a covert homunculus? The attractor characterization shows that it is not. An attractor is not an additional instance alongside the system dynamics; it is a structural property of the dynamics themselves, the system's tendency to return to a specific region of phase space. Equally, the CC is not an instance that 'reads' or 'evaluates' the global state; it is the region of phase space in which the global state assumes an internally stable, self-referential form. Purely functional and structural, without any explanatory extra-instance.

A further clarification is required. One might ask: why does an internal reference point constitute perspectivity rather than merely another physical mechanism? The answer lies in the formal minimal structure of perspectivity. Perspectivity does not presuppose consciousness, it presupposes an internal difference: the difference between the system as a whole and the state in which the system is accessible to itself as a whole. This difference is not itself a state of the system among others; it is the formal condition for the system to *have* states at all, for something to count as the system's state rather than merely being a physical state of the system.

Husserl (1928/1966) showed in his analysis of inner time-consciousness that self-affection, the way in which an experience grasps itself in the act of occurring, cannot be understood as an additional act alongside primary experience, but is its immanent structure. The causal core as interpreter is the physical locus of this structure: not as a mechanism that generates consciousness, but as the causal realization condition of the minimal difference without which no state can hold for a system. This step is modest but precise: it claims not that reflexive causal structure explains experience, but that a system lacking this structure is not one to which the question of experience is applicable, since it lacks the formal minimal structure without which no state can count as *its own* state. The reflexive layer is therefore also infrastructure, infrastructure of perspectivity, not explanation of experience.

This is the function of what Stegemann (2026a) has called the *causal core as interpreter*. The CC is not an additional component of the system, nor a localized substructure within it. It is the functional centre of the system's self-referential causal closure: the point at which the global non-factorizable state of the system is not merely instantiated but *operated* by the system itself. In the terminology of Stegemann (2026a), the CC does not process incoming neural patterns representationally but interprets them through nonlinear interference with the system's own coherent global state.

4.2 The Hologram Analogy and Its Formal Content

In optical holography, information is not stored locally but as a distributed interference pattern across the entire holographic medium. This information is accessible only through the interaction of the stored pattern with a coherent reference beam. The reference beam stores no additional information; it provides the phase reference that resolves the distributed pattern into a coherent image. In a non-factorizable causal system, the global state is distributed in a structurally analogous way. The causal core plays the role of the reference beam: it is the internal phase reference that makes the distributed global state operatively accessible within the system. Without such a reference, the global state is real but inert, causally efficacious at the physical level but not available to the system as its own state.

Formally, this role is analogous to the global phase in a coherent quantum state. The phase factor $e^{i\varphi}$ in $|\psi\rangle \; = \; e^{i\varphi}|\varphi\rangle$ is not a subsystem, has no spatial localization, and is not detectable by any local measurement, yet without a phase reference the system cannot exhibit interference, the global coherence that defines its non-factorizable character. The causal core plays an analogous role: not a part among parts, but the internal reference that makes the non-factorizable causal unity of the system operatively reflexive.

4.3 Interpretation as Nonlinear Interference

Stegemann (2026a) has formalized the interpretive function of the causal core through phase-amplitude coupling. In neural systems exhibiting criticality, the global state can be characterized by the phase of slow oscillatory components (theta, alpha band), which organizes the amplitude of faster, local oscillations (gamma band), a case in which the global non-local state modulates local subsystem activity without being localizable in any subsystem. The interpretive function consists in this: incoming neural patterns are not processed by linear mapping onto stored representations, but are evaluated through their *interference* with the current global coherent state. The same incoming pattern receives different evaluations depending on the current global state, and the global state is itself modified by the incoming pattern in a way that affects all subsequent interactions. This is not representation in the classical cognitivist sense but the ongoing modulation of a non-factorizable system state by local inputs, where the relevant causal factor is the global configuration of relations.

A dynamical-systems-theoretic clarification suggests understanding interpretation as *trajectory in the attractor basin*. If the causal core is an attractor in the system's phase space, then interpretation is the specific form the system dynamics assumes under the influence of an external stimulus: interpretation is the trajectory of the system in the attractor basin, modified by the current stimulus. Different external inputs generate different trajectories; the current global state, the position in the attractor basin, determines how these trajectories unfold. The same incoming signal generates different trajectories depending on the starting point in the attractor basin, and different signals can generate convergent trajectories when the global state dominates. This is more precise than the hologram analogy and closes the gap between structural characterization and dynamic description. The hologram analogy describes the structural aspect of this process; the trajectory in the attractor basin describes the dynamic aspect.

5. The Physical Infrastructure of Experience

5.1 Three Structural Layers

The framework developed in the preceding sections can be summarized as a three-layered account of the physical infrastructure of experience. The layers are not independent: each is the condition of the next, and together they constitute a single, internally differentiated structural situation.

The first layer is the *dynamic layer*: the autocatalytic system that actively maintains its proximity to the critical point. Without autocatalytic closure, without the capacity of the system to be the cause of its own causal organization, the subsequent structural features cannot arise. A system that merely happens to lie near the critical point for extrinsic reasons does not qualify; the system must actively stabilize its own critical regime through the feedback mechanisms of its autocatalytic structure (Stegemann 2025e, 2025b).

The second layer is the *causal layer*: non-factorizability as the necessary consequence of the system's critical dynamic regime. At this layer the system constitutes an irreducible causal unity: no decomposition into independent partial causes can represent its causal structure without loss of essential causal content. The Hammersley-Clifford theorem establishes that this unity cannot be captured by any finite graph of conditional independencies; the non-analyticity of the partition function establishes its thermodynamic irreducibility; and the violation of the entanglement entropy area law establishes its quantum-information-theoretic analogue (Stegemann 2025a, 2025c, 2025d).

The third layer is the *reflexive layer*: the causal core as internal reference point, which transforms the system's non-factorizable causal unity into a reflexive structure, one in which the global integrated state is operatively accessible to the system itself as an integrated state. A system possessing the first two layers but lacking the third has non-factorizable causal unity but no internal difference between system state and the holding of that state for the system; it has causal integration without the structural precondition of experiential integration (Stegemann 2026a, 2025h).

A question arises concerning the relationship among the three layers: is reflexivity genuinely an independent third layer, or a specific form of non-factorizability already sufficiently determined by the first and second? The answer lies in the character of the first layer. Autocatalysis does not merely produce non-factorizability as such, but a specific *kind* of non-factorizability: a dynamic, self-sustaining one. The laser pulse or the ferromagnetic domain at the Curie point also exhibit non-factorizability, but a static, externally imposed one that collapses when external conditions are removed. The autocatalytically generated non-factorizability is one the system actively produces and whose producer it simultaneously is. This self-referentiality of the productive dynamics, the system generates the conditions of its own non-factorizability, is not added from outside to non-factorizability but contained within it. The causal core as attractor is the dynamic expression of this self-referential productive dynamics: the point in phase space at which self-sustaining non-factorizability passes into an internally accessible form. Reflexivity is therefore not an independent third layer alongside the second, but the complete expression of a non-factorizability generated by autocatalysis, and only by autocatalysis. The three layers are analytically distinguishable but ontologically a single, internally differentiated structural situation.

5.2 Infrastructure versus Correlates: Demarcation from Integrated Information Theory

The distinction between the present framework and the prevailing search for neural correlates of consciousness (NCC) is categorical, not merely terminological. The NCC research programme (Crick & Koch 1990; Koch et al. 2016) seeks the minimal neural mechanisms sufficient for a given conscious content. The present framework does not operate at this level. It does not seek the neural mechanisms sufficient for specific conscious contents, but the structural conditions necessary for consciousness as a form, the conditions under which the question of experience becomes applicable to a physical system at all. A system lacking the three infrastructure layers is not one for which the question of which neural events correlate with which conscious contents can be meaningfully posed; it is one for which the form of that question is itself inapplicable.

A separate demarcation from Integrated Information Theory (IIT) is required because a structural kinship might superficially appear to obtain. Tononi's Φ is also a measure of non-factorizability: it quantifies the extent to which the whole generates more information than the sum of its parts. The Exclusion Postulate of IIT identifies the subsystem with maximum Φ as the bearer of consciousness, which might superficially recall the causal core. The difference is categorical in nature and concerns three levels. First, Φ is a *scalar measure*, assigning a number to every system and permitting gradual gradations. Non-factorizability in the present framework is a *categorical structural property*: a system is either factorizable or not, tertium non datur. Second, IIT despite its complexity remains within the correlative paradigm, it still seeks a neural substrate that *has* consciousness. The present framework asks what structural form a system must have for the question of consciousness to be applicable to it. Third, the Exclusion Postulate, localization of consciousness in the subsystem with maximum Φ, is precisely what the present framework structurally excludes: non-factorizability is by definition not localizable in a subsystem. The causal core is not a subsystem with maximum integration but the internal reference point of a system-wide, irreducible causal unity.

5.3 The Hard Problem Reconsidered

The framework developed here does not dissolve the hard problem but clarifies its structure. Chalmers (1995, 1996) formulates the hard problem as the question why physical processes give rise to subjective experience, why there is something it is like to be in certain physical states. The framework of physical infrastructure does not answer this question; it identifies the structural conditions under which the question first becomes applicable. A system lacking the three infrastructure layers is not one for which the hard problem arises. There is no question why the causal processes of a thermostat give rise to experience, not because the question has a trivial answer, but because the thermostat does not satisfy the structural conditions under which experience is a possibility. Whether the presence of these structural conditions is sufficient to explain why there is experience, why, as Chalmers puts it, the lights are on, remains open.

At this point the boundary of the framework with respect to the hard problem can be determined still more precisely. The hard problem asks: why is there for this system a what-it-is-like, a *being-for-me*? The infrastructure framework does not answer this question; but it can specify more precisely why the three-layered infrastructure is the necessary condition for this question to be meaningfully posed at all. Without autocatalytically generated, self-sustaining non-factorizability, there is no global state instantiated as a unity. Without an attractor as internal reference point, there is no internal difference between system state and the holding of that state. And without this internal difference, the concept of the for-me is categorially inapplicable, not because the system could not have the for-me, but because the formal minimal structure is absent at which the question of the for-me can dock at all. The hard problem arises only for systems that satisfy the three infrastructure layers, and its specific difficulty, the unbridgeability of the explanatory gap, depends on the reflexive structure described here being present on the physical side. The framework does not close the gap; but it shows that the gap has a determinate form and that this form is determined from the physical side.

6. Objections and Replies

6.1 The Panpsychist Objection

An objection to the framework is that it is too permissive: if non-factorizability at criticality is the central structural feature, then every physical system at a phase transition would satisfy the infrastructure conditions and would have to be credited with some form of experience. The reply is that the framework is more restrictive than this objection suggests. The first layer, autocatalytic closure actively stabilizing its own critical regime, excludes simple physical systems at phase transitions. A ferromagnetic material at the Curie point is at a critical point but does not actively maintain this regime through self-organized autocatalytic feedback; it is placed there by external temperature control. The self-organization of criticality through internal feedback is an additional structural requirement distinguishing biological systems with autocatalytic closure from physical systems that are merely coincidentally critical (Stegemann 2025e).

One might object that this criterion is resistant to operationalization: how is it empirically decided whether a system actively self-maintains criticality or merely happens to be near it? The answer is that the demarcation criterion operates at the level of structural conditions, not at the level of empirical measurement procedures, analogously to how the theoretical definition of Turing-completeness is not touched by the question of whether a concrete physical system instantiates it. The development of suitable empirical operationalizations of autocatalytic closure for neural systems is an independent research task that presupposes but must be distinguished from the conceptual framework. Nevertheless, concrete empirical candidates for the distinction can be named.

First, the *perturbation response kinetics* of active and passive criticality differ systematically. A system that actively stabilizes criticality returns to the critical regime after perturbation on a characteristic time scale, exhibiting a signature distinct from the passive relaxation of a physical system in a temperature gradient: the return is driven by internal feedback dynamics, which manifest in the structure of the avalanche statistics following the perturbation. Shew and Plenz (2013) have documented corresponding differences between pharmacologically manipulated and unmanipulated cortical preparations.

Second, *metabolic signatures* provide a further candidate. Autocatalytic systems must expend energy to maintain their critical organization against thermal fluctuations and external perturbations. The resting metabolic rate of the brain, drawing on approximately 20 per cent of total body metabolism at only two per cent of body mass, is a rough indicator. More precisely measurable would be the difference in metabolism between resting state and perturbed state: a system actively stabilizing criticality should show a characteristic additional metabolic cost during return from the sub- or supercritical regime, which does not occur in passive physical systems.

Third, active and passive criticality differ in their *response to chronic disturbances*. Physical systems externally placed at the critical point leave it as soon as boundary conditions change. Biological systems with autocatalytic closure show a tendency to actively maintain their critical regime even under altered boundary conditions, within limits set by the capacity of the autocatalytic feedback. Pathological states such as epileptic activity (Meisel et al. 2012) or criticality changes under anaesthesia (Tagliazucchi et al. 2016) could be interpreted as cases in which this capacity is temporarily exhausted or suppressed.

The superconductor example also permits an important sharpening of the operationalization problem. One might object that a superconductor in a cryostat also returns to its coherent state after perturbation. The answer lies in the *endogeneity of the restoring force*. In the superconductor, the drive to return is thermodynamic relaxation to the ground state, entirely driven by the external temperature bath: remove the cryostat, and the superconducting state collapses immediately and irreversibly, because no internal dynamics exist to maintain it. The restoring force is exogenous, it lies outside the system. In the living neural system, the restoring force is generated by the system's own metabolic feedback loops. This difference is empirically measurable: one tests whether return dynamics after perturbation are preserved even when external environmental conditions are varied while internal metabolism remains intact. A system with an endogenous restoring force shows robustness to such external variations; a system with an exogenous restoring force does not.

Furthermore, the third layer, the causal core as internal reference point that constitutes the minimal difference between system state and the holding of that state for the system, introduces a further condition that most non-biological critical systems fail to meet. The panpsychist conclusion does not follow from the infrastructure approach.

6.2 The Epiphenomenalist Objection

A second objection is that even if the framework correctly identifies the physical infrastructure of experience, it leaves open the possibility that experience itself is epiphenomenal, that the three-layered structure is causally efficacious while experience rides along without causal contribution. This objection is well-formulated as a metaphysical concern but misidentifies the claim of the present framework. The framework does not claim that the three-layered infrastructure is identical with experience or that it solves the hard problem. It claims that the infrastructure is the physical side of the structural situation in which the hard problem arises. If the physical infrastructure includes a genuinely reflexive causal structure, one in which the global state modulates its own processing through the interpreter function, then the causal closure argument against mental causation may need to be reconsidered within the framework of non-factorizable causal structure rather than the standard linear causal framework (Stegemann 2025h).

6.3 The Empirical Objection

A third objection is empirical: the claim that neural systems actively maintain criticality is not universally accepted. There are models of neural cognition that do not require criticality (Destexhe & Contreras 2006), and the empirical evidence for brain criticality remains contested (Dehghani et al. 2012; Touboul & Destexhe 2010). The reply operates on two levels. At the empirical level, the evidence for criticality in neural systems has increased substantially since Beggs and Plenz (2003), and while debates about specific implementations continue, the broad conclusion that neural systems operate near critical points has significant empirical support (Plenz et al. 2021). At the conceptual level, the framework does not require criticality in the precise statistical-mechanical sense applicable to equilibrium systems; it requires the weaker claim that neural systems exhibit the structural feature of which criticality is the paradigm case, the divergence of effective correlation lengths and the ensuing impossibility of factorized description.

7. Conclusion

The present article has developed a framework for the physical infrastructure of experience, understood as the set of necessary structural conditions that a physical system must satisfy in order for the question of experience to be applicable to it. The framework identifies three mutually conditioning structural layers: the dynamic layer of autocatalytic closure, which maintains criticality; the causal layer of non-factorizability as the structural consequence of criticality; and the reflexive layer of the causal core as dynamic attractor in the system's phase space, which transforms non-factorizable causal unity into the structural precondition of perspectivity by constituting the minimal difference between system state and the holding of that state for the system.

The framework is supported by converging mathematical characterizations from statistical mechanics, probability theory, and quantum information. Quantum information theory is used as a mathematical characterization tool, not as a mechanistic hypothesis about neural processes, a distinction that fundamentally separates the present approach from Orch-OR-type theories. Comparison with superconducting quantum coherence makes clear that the infrastructure conditions are not satisfied by all forms of physical non-factorizability, but only by those in which the system actively maintains its own critical organization through autocatalytic closure.

The explicit demarcation from Integrated Information Theory shows that the present framework is not merely a reformulation of IIT: non-factorizability is a categorical structural property, not a scalar measure; the framework operates at the level of infrastructure conditions rather than correlates; and the Exclusion Postulate of IIT, the localization of consciousness in the subsystem with maximum Φ, is conceptually incompatible with a non-factorizable framework.

The framework also permits a structural answer to a question that in consciousness research often remains implicit: why is experience a phenomenon of life rather than of inanimate matter? The answer is not a biologistic thesis, it follows from the logic of the framework itself. Autocatalytic closure is not one among several possible realizations of the infrastructure conditions; it is the only known physical structure that actively and endogenously stabilizes criticality, rather than having it externally imposed. Thus the first layer is not merely the causal substrate of the second and third, it is the point at which life and mind structurally coincide. A system that lives in the sense of autocatalytic closure generates, in the very act of living, the physical infrastructure under which the question of experience becomes applicable to it. Conversely, a system lacking this closure cannot instantiate non-factorizability of the right kind, not self-referential and endogenously stabilized, but at most passively imposed. Autocatalytic closure is therefore the bridge between life and mind: not in the sense of an identity thesis, life is not experience, but in the sense of an enabling relation: experience is possible only where life in the sense of autocatalytic self-maintenance obtains.

The framework does not solve the hard problem of consciousness. But it identifies with greater precision the physical side of the explanatory gap and provides the framework in which this gap can be addressed with the structural precision it requires. The question shifts from *Which mechanisms generate experience?* to *What must a physical system be for the question of experience to be applicable to it?* That is not a more modest claim, it is a more fundamental one.

References

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.

Beggs, J. M. & Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *Journal of Neuroscience*, 23(35), 11167–11177.

Calabrese, P. & Cardy, J. (2004). Entanglement entropy and quantum field theory. *Journal of Statistical Mechanics: Theory and Experiment*, P06002.

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

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Chialvo, D. R. (2010). Emergent complex neural dynamics. *Nature Physics*, 6(10), 744–750.

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Crick, F. & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2, 263–275.

Dehaene, S. & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.

Dehghani, N. et al. (2012). Avalanche analysis from multielectrode ensemble recordings in cat, monkey, and human cerebral cortex during wakefulness and sleep. *PLOS ONE*, 7(6), e40110.

Destexhe, A. & Contreras, D. (2006). Neuronal computations with stochastic network states. *Science*, 314(5796), 85–90.

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

Haimovici, A., Tagliazucchi, E., Balenzuela, P. & Chialvo, D. R. (2013). Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Physical Review Letters*, 110(17), 178101.

Hameroff, S. (1998). Quantum computation in brain microtubules? The Penrose-Hameroff 'Orch OR' model of consciousness. *Philosophical Transactions of the Royal Society A*, 356(1743), 1869–1896.

Hammersley, J. M. & Clifford, P. (1971). *Markov fields on finite graphs and lattices*. Unpublished manuscript.

Husserl, E. (1900/1975). *Logical Investigations*. Max Niemeyer Verlag.

Husserl, E. (1928/1966). *On the Phenomenology of the Consciousness of Internal Time (1893–1917)* (R. Boehm, Ed.). Martinus Nijhoff.

Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford University Press.

Koch, C., Massimini, M., Boly, M. & Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience*, 17(5), 307–321.

Lauritzen, S. L. (1996). *Graphical models*. Oxford University Press.

Maturana, H. R. & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. D. Reidel.

Meisel, C., Storch, A., Hallmeyer-Elgner, S., Bullmore, E. & Gross, T. (2012). Failure of adaptive self-organized criticality during epileptic seizure attacks. *PLOS Computational Biology*, 8(1), e1002312.

Merleau-Ponty, M. (1945/1966). *Phenomenology of Perception* (C. Smith, Trans.). Routledge.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Penrose, R. (1994). *Shadows of the mind: A search for the missing science of consciousness*. Oxford University Press.

Plenz, D. et al. (2021). Self-organized criticality in the brain. *Frontiers in Physics*, 9, 639389.

Shew, W. L. & Plenz, D. (2013). The functional benefits of criticality in the cortex. *The Neuroscientist*, 19(1), 88–100.

Stegemann, W. (2025a). Consciousness as a collapse of causality. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.17563267

Stegemann, W. (2025b). Autopoiesis, causal poiesis, causal collapse. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.15599500

Stegemann, W. (2025c). Consciousness, Expanding the model of causal collapse. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.15351381

Stegemann, W. (2025d). Integrated model of causal collapse. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.15350050

Stegemann, W. (2025e). Consciousness as an expression of autocatalytic stability, A new index for neuronal coherence. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.17532218

Stegemann, W. (2025f). From principle to function: The causal core in the brain. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.15608621

Stegemann, W. (2025g). The mechanism of phenomenal experience. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.17698505

Stegemann, W. (2025h). The causal core as a structural principle of free will. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.17057739

Stegemann, W. (2026a). The causal core as interpreter. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.18340973

Stegemann, W. (2026b). Methodology of consciousness theory. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.18131340

Stegemann, W. (2026c). A framework for consciousness: Autopoiesis, causal poiesis, causal collapse. *Zenodo Preprint*. https://doi.org/10.5281/zenodo.18328704

Tagliazucchi, E., von Wegner, F., Morzelewski, A., Brodbeck, V., Jahnke, K. & Laufs, H. (2016). Breakdown of long-range temporal dependence in default mode and attention networks during deep sleep. *Proceedings of the National Academy of Sciences*, 110(38), 15387–15392.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5, 42.

Tononi, G. (2014). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin*, 215(3), 216–242.

Touboul, J. & Destexhe, A. (2010). Can power-law scaling and neuronal avalanches arise from stochastic dynamics? *PLOS ONE*, 5(2), e8982.

Varela, F. J., Maturana, H. R. & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, 5(4), 187–196.

Vidal, G., Latorre, J. I., Rico, E. & Kitaev, A. (2003). Entanglement in quantum critical phenomena. *Physical Review Letters*, 90(22), 227902.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.