Causality Is Not a Causal Chain

From Causal Relations to the Organization of Living Systems

Wolfgang Stegemann

Abstract

Causality in science and philosophy is often analyzed in terms of isolable cause effect relations. One variable changes another, one event brings about another, and a mechanism is decomposed into individual causal steps. This procedure is methodologically indispensable and extraordinarily successful in physics, chemistry, biology, and neuroscience. Yet the possibility of experimentally isolating local causal relations does not imply that the system under investigation is itself organized in a causally factorized way. At precisely this point, a systematic category error arises in the analysis of living systems. Organisms do not consist merely of interconnected causal chains. In recurrent, autocatalytic processes, they generate and maintain the conditions under which their own components can be causally effective. This paper therefore proposes a distinction between **causal relation** and **causal organization**. Local causal relations may be factorable while the overall system is not sufficiently factorable at the organizational level. Life thus introduces no new kind of causality. It changes the organizational form of physical causal relations. This position is distinguished from autopoiesis, Rosen's closure to efficient causation, Kauffman's work constraint cycles, Moreno and Mossio's closure of constraints, Noble's biological relativity, and strong forms of emergent and downward causation. In the final part, the concept of **causal collapse**, developed in earlier work, is reformulated. What collapses is not causality itself, but the possibility of decomposing the relevant neural dynamics, under an appropriate partition, into causally independent subprocesses. Causal collapse is thereby specified as a limiting case of causal non-factorizability within a living, plastic, and recurrently closed neural system.

Keywords: causality, organization, autocatalysis, autopoiesis, organismic closure, factorization, emergence, causal collapse, consciousness

1. The Problem Lies Not in Causality, but in Its Tacit Form

When causality is discussed in science and philosophy, what is usually meant is a relation between distinguishable quantities. In its elementary form, it can be represented as

$$A \rightarrow B.$$

A change in A leads, under certain conditions, to a change in B. This form is indispensable to scientific analysis. Without the possibility of analytically distinguishing variables, manipulating them, and comparing their effects, there would be neither experimental physics nor molecular biology, pharmacology, or neuroscience.

The problem therefore does not begin with this concept of causality. It begins where a methodological operation is transformed into an ontological claim. Because a relation can be analytically isolated, it is assumed that the system itself consists of isolable relations. Successful

decomposition of an object of investigation is taken to imply that the object is internally decomposable in the same way.

This inference is especially consequential in biology. A gene can be knocked out. A receptor can be blocked. A neural population can be stimulated or inhibited. Such interventions can reveal genuine causal dependencies. If manipulating A produces a change in B, then the statement

$$A \rightarrow B$$

is justified under the relevant experimental conditions. What is not justified without further argument is the stronger conclusion that A possesses its causal role independently of the system in which it occurs, or that B is sufficiently explained by A as an isolated cause.

This is why a second category is required: **causal organization**. The question concerning a causal relation is: What does A do to B? The question concerning causal organization is: Which system conditions make it possible for A to possess this causal role in this system at all?

The distinction can initially be written as follows:

$$A \rightarrow B$$

denotes a local causal relation, whereas

$$(A \rightarrow B) \mid \mathcal{O}$$

indicates that this relation is realized within an organization $\mathcal{O}$. Here, $\mathcal{O}$ is not an additional cause alongside A. The notation is intended only to express that the concrete causal role of A depends on conditions that are provided, stabilized, or themselves generated by the system.

The central thesis of this paper can therefore be stated as follows:

$$\boxed{\text{Causal relation} \neq \text{causal organization}}$$

and, consequently,

$$\boxed{\text{Local factorization} \not\Rightarrow \text{factorization of the system.}}$$

The thesis is not directed against causal analysis. It is directed against confusing a method of analysis with the organizational form of the object being analyzed.

2. Physical Causality Is Not Simply Linear

A clarification is necessary, because otherwise the criticism would target the wrong issue. Physics by no means recognizes only linear causal chains. Nonlinear dynamics, feedback, coupled oscillators, phase transitions, chaotic systems, and networks have long been part of the standard repertoire of physical description. It would therefore be incorrect to characterize life simply by claiming that circular or nonlinear causality appears in place of linear causality.

Even a structure such as

$$A \rightarrow B \rightarrow C \rightarrow A$$

is not biological as such. Feedback loops occur in technical control systems just as they do in physical systems. A thermostat changes heating output, heating output changes temperature, and

temperature feeds back onto the thermostat. Such a control loop can be fully recurrent without being alive.

The specifically biological step therefore does not lie in circularity alone. It lies in the fact that the circularly coupled processes participate in producing and maintaining their own conditions of realization. The components of a living system do not merely act on one another. They are part of a process in which the conditions of their own efficacy are continually reproduced.

This distinction is crucial because it separates the present approach from a mere theory of feedback. Recurrence is a dynamical property. Self-production is an organizational property. Only their combination generates the form of causal dependence characteristic of living systems.

3. Autopoiesis: The Right Intuition, but Not Yet a Developed Theory of Causality

With the concept of autopoiesis, Maturana and Varela took a decisive step. In the classical formulation, Varela, Maturana, and Uribe describe living systems as networks of production processes whose components in turn produce and maintain the network that produces those components (Varela, Maturana & Uribe 1974). Maturana soon afterward specifies that the autonomy of living systems consists in self-production. A living system is a network of production processes that is continuously and recursively realized through the interactions of components generated by the system itself (Maturana 1975).

This conception shifts attention from substance to organization. An organism is not alive because it consists of particular molecules. What matters is the way processes and components are connected into a self-producing whole.

For the question pursued here, however, this determination is not yet sufficient. Autopoiesis tells us how living organization is structured. It does not systematically distinguish between the causality of individual processes and the organization of that causality. The concept of recursive production contains this insight implicitly, but it does not elevate it into a separate conceptual distinction.

This is precisely where a later ambiguity arises. If all relevant processes within an autopoietic network are recursively connected, the question emerges whether this entails a special biological causality or merely a special organization of physical and chemical processes. For the present position, the latter interpretation is sufficient and theoretically more economical.

No second kind of causality need be introduced. The reaction of an enzyme remains chemical. Ion transport remains electrochemical. The effect of a membrane remains physically and chemically describable. What is new is not the type of the individual causal relation, but the fact that its conditions of realization are themselves produced and maintained within the system.

Autopoiesis can therefore be conceptually sharpened as follows:

$$\text{Autopoiesis} = \text{recurrent self-production of the conditions of causal efficacy.}$$

This formulation does not claim to replace classical autopoiesis theory. It isolates precisely the aspect relevant to a theory of causal organization.

4. Autocatalysis as an Elementary Form of Causal Self-Reference

The idea can be illustrated even more basically with autocatalytic processes. In an ordinary reaction chain, a product may emerge from a preceding reaction:

$$A \rightarrow B \rightarrow C.$$

In an autocatalytic context, a product contributes directly or indirectly to its own further production. In simplified form:

$$A \rightarrow B \rightarrow C \rightarrow A.$$

The significance of this structure does not lie in causality becoming circular and therefore mysterious. At every point in time, local reaction steps remain directed. A acts on B, B on C, and C contributes to the renewed production or stabilization of A. The circle arises at the level of organization, not because individual cause effect relations lose their temporal direction.

This is exactly why the concept of causal organization is analytically useful. The individual relations can still be investigated in factorized form. Reaction rates can be measured, catalytic effects isolated, and individual components removed. But the role of each component is partly conditioned by the context that maintains that component itself.

The causal role of A is therefore not identical with an intrinsic property of A. A molecule can produce entirely different effects under different conditions. In an organism, these conditions are not merely externally given. A substantial part of them is generated by the system itself: concentrations, membrane potentials, compartments, pH ranges, transport pathways, enzymes, regulatory molecules, and spatial proximities.

The organism therefore does not produce the laws of nature according to which its components act. But to a considerable extent it produces the **boundary conditions** under which particular physical and chemical possibilities are actually realized.

This is the minimal sense in which organization is causally relevant.

5. Rosen: Closure to Efficient Causation and the Question of the Conditions of Efficacy

Robert Rosen formulated this problem in especially radical terms. In *Life Itself*, he characterizes living systems in terms of “closure to efficient causation” (Rosen 1991). In the later reconstruction by Cárdenas et al., the core idea is explained as follows: a system is closed to efficient causation in Rosen’s sense when the components that function as efficient causes within the system are materially generated by the system itself (Cárdenas et al. 2010).

Rosen’s achievement lies in making the question of cause itself reflexive. It is not enough to ask what produces B. One must also ask what produces the structure or function that produces B. The investigation thereby shifts from individual reaction sequences to the generation of the system’s causal conditions.

This step is central to the present approach. Nevertheless, Rosen’s Aristotelian terminology is not adopted here. The distinction between material and efficient cause is historically and systematically burdened and only partially compatible with empirical biology. Montévil and Mossio

themselves note that Rosen's model, although theoretically pioneering, remains too abstract for the concrete identification of biological causes (Montévil & Mossio 2015).

The formulation proposed here is therefore deliberately weaker and closer to operational analysis:

$$A \rightarrow B \mid \mathcal{O}.$$

It does not ask for different Aristotelian kinds of causes. It asks which organization $\mathcal{O}$ provides the conditions under which a local relation $A \rightarrow B$ is realized.

The distance from Rosen therefore consists in two points. First, no separate ontological category of efficient causation is required. Second, closure is not understood as an absolute property of the system. Living systems are thermodynamically, materially, and causally open in many respects. What is closed is only the recurrent context of certain conditions required for the continuation of the organization.

6. Kauffman: Work Constraint Cycles and the Physical Realization of Autonomy

Stuart Kauffman has sought to connect the abstract idea of organismic autonomy more closely to physical processes. In his work on autonomous agents, the so-called work constraint cycle plays a central role. Work generates or maintains constraints, while constraints in turn channel processes in such a way that further work can be performed. In the conception developed jointly with Philip Clayton, autocatalytic reproduction, work cycles, boundaries, the construction of constraints, and evolved agency belong to the conditions of minimal biological agency (Kauffman & Clayton 2006).

This approach is very close to the idea developed here. It shows that biological organization cannot be understood solely as a flow of matter. What matters are structures that constrain energy flows and thereby enable particular processes, while those structures are themselves generated or maintained by such processes.

Here too, however, a conceptual shift is useful. The expression *constraint* can easily suggest that an additional entity acts on a process. In fact, what is involved is often a spatial, structural, or dynamical condition of the process itself. A membrane, for example, is not an abstract constraint alongside a transport process. Its physical structure is part of the conditions under which certain forms of transport occur and others are prevented.

For the present theory, therefore, it is not decisive whether such conditions are called constraints. What matters is their recurrent dependence: conditions of causal efficacy are generated by processes whose course, in turn, depends on precisely those conditions.

Schematically:

$$P_1 \rightarrow C_1,$$

$$C_1 \Rightarrow P_2,$$

$$P_2 \rightarrow C_2,$$

$$C_2 \Rightarrow P_1.$$

The symbol ⇒ is not meant here to denote a different kind of causality. It merely marks that a condition structures the space of possibilities of a process. Physically, this structuring remains realized through concrete interactions.

7. Montévil and Mossio: Closure of Constraints and the Crucial Disagreement

Montévil and Mossio have presented what is probably the most precise formal account of organizational closure to date. They distinguish between processes and constraints. Processes are changes occurring under open nonequilibrium conditions. Constraints are relatively stable structures that influence such processes and display a certain persistence over relevant timescales. Biological organization arises when such constraints stand in relations of mutual dependence and thereby form a closure of constraints (Montévil & Mossio 2015).

This theory captures an essential point: the stability of biological organization cannot be explained by individual reaction chains alone. The structures that enable and channel processes themselves depend on other structures and processes within the system. Organizational closure consists in this mutual dependence.

At this point there is an important proximity to the present position. Nevertheless, a fundamental difference remains in the ontological interpretation.

Mossio, Bich, and Moreno explicitly argue for closure as an **emergent causal regime**. Configurations, they maintain, possess ontologically irreducible properties by virtue of their relational organization and therefore have causal powers of their own. On this basis, biological systems realize a higher-level form of causation (Mossio, Bich & Moreno 2013). The later book *Biological Autonomy* likewise discusses biological emergence and interlevel causation explicitly as part of its theoretical framework (Moreno & Mossio 2015).

From the perspective defended here, this ontological upgrading is unnecessary.

If a membrane prevents certain molecules from passing through, no emergent causal power of membrane organization is required. The effect arises from the physical properties of the membrane structure. If that membrane is in turn generated and maintained by metabolic processes, a recurrent organizational relation emerges. Organization is indispensable for explanation, but it need not be treated as an additional causal entity.

The disagreement can be stated succinctly:

$$\text{Montévil/Mossio: organizational closure} \Rightarrow \text{distinctive causal regime,}$$

whereas the present view holds:

$$\text{organizational closure} = \text{distinctive organization of ordinary causal relations.}$$

The explanatory autonomy of biological organization is thereby acknowledged without attributing ontologically autonomous causal powers to it.

8. No Downward Causation: Levels of Description Are Not Levels of Causation

Debates about biological organization frequently invoke upward and downward causation. Denis Noble's theory of biological relativity is a prominent example. Noble argues that biological systems contain no a priori privileged level of causation. Processes at different scales influence one another, so causal explanation exclusively in terms of genes or molecules is insufficient (Noble 2012).

The empirical core of this criticism is convincing. Organisms are multiscale systems. Membrane potentials influence protein states, and protein states influence membrane potentials. Tissue structures alter local conditions, while local processes maintain tissue structures. Organismal states alter hormonal, neural, and metabolic processes, which in turn alter the organismal state.

The formulation becomes problematic only when these interactions are interpreted as causation between ontologically distinct levels. A level of description is not a spatially or causally autonomous region of the world. "Organism," "organ," "cell," "protein," and "ion" refer to different resolutions of the same material process.

In the mechanistic debate, Craver and Bechtel have shown that supposed top down causes can often be reconstructed as mechanistically mediated effects without invoking mysterious interlevel causes (Craver & Bechtel 2007). This objection is important, but from the perspective developed here it does not go quite far enough. The alternative must not consist simply in dissolving the problem into ever finer mechanistic decomposition.

The organism need not be conceived as an additional cause standing above its parts. But its organization must not be reduced to a mere list of partial causes either. Organization denotes the structured dependence of the conditions under which these partial causes are effective.

Thus:

no causation between levels of description,

but rather

one causally organized reality described at different levels.

In this way, Noble's justified anti-reductionism can be preserved without adopting the metaphor of downward causation.

9. Mechanistic Explanation and Its Limit

Contemporary philosophy of science has developed mechanistic explanation as an alternative to simple reduction models. For Craver, a mechanism consists of entities and activities organized spatially and temporally so that together they produce a phenomenon (Craver 2007; Craver 2025). Organization is therefore explicitly part of the explanation.

This is a substantial advance over a conception according to which explanation merely consists in decomposing a whole into parts. Nevertheless, a difference remains between mechanistic explanation and the theory proposed here.

Mechanistic explanation asks primarily which components and activities constitute a particular phenomenon. Organization arranges these components such that the explanandum comes about. For many biological questions, that is exactly right.

In a self-producing system, however, an additional question arises: Where do the components and conditions presupposed by the mechanism come from? How are they maintained? How are the conditions of their functioning continuously reproduced?

A technical machine can be fully explained mechanistically without the mechanism manufacturing its own parts. In an organism, precisely this self-production is constitutive.

An organismic explanation must therefore connect two dimensions:

$$\text{mechanistic relation} + \text{organizational reproduction}.$$

The first shows **how** a local process works. The second shows **how the conditions under which it works persist within the system**.

This is not intended as a rival to mechanistic explanation. It is a supplement required precisely where the object does not merely function but produces the conditions of its own functioning.

10. Factorization: From a Methodological Principle to a Philosophical Problem

The concept of factorization is the actual key to the position developed here.

In an analytical model, a complex system is decomposed into variables or subsystems:

$$S = \{X_1, X_2, \dots, X_n\}.$$

One then investigates relations such as

$$X_i \to X_j.$$

Such decomposition is methodologically legitimate as long as it remains clear that the isolated relation is a relation **under particular system conditions**.

The problem arises when it is assumed that the total causality of S can be represented completely as the sum of independent partial relations:

$$K(S) = \sum_{i,j} K\left(X_i, X_j\right).$$

For self-producing systems, this is generally not the case. The causal properties of the components depend partly on conditions that are themselves generated by other components of the system.

Instead:

$$K\left(X_i, X_j\right) = f\left(X_i, X_j, \mathcal{O}\right),$$

where $\mathcal{O}$ is itself realized through relations among the X_i.

This produces a recursive structure:

$$\mathcal{O} = F(X_1, X_2, \ldots, X_n),$$

while at the same time

$$K(X_i, X_j) = f(\mathcal{O}).$$

This recursion does not mean that the whole mystically causes its parts. It means that the local efficacy of the parts depends on an organization jointly realized by those very parts.

The analytical error therefore consists in inferring from

$$\exists\, K(X_i, X_j)$$

to

$$K(S) = K(X_i, X_j).$$

In words: evidence for a local causal dependence does not amount to a complete explanation of a system property.

11. Three Typical Errors in Biological Causal Explanation

The distinction between relation and organization makes it possible to specify several recurring fallacies more precisely.

11.1 Genetic Reductionism

If a genetic variation influences a phenotypic trait, a genuine causal relation exists. It does not follow, however, that the trait is "contained in the gene." A gene acts only within a cellular organization that provides transcription, translation, regulation, metabolism, membrane structures, and developmental conditions.

The error is:

$$G \rightarrow P$$

is reinterpreted as

$$G = \text{bearer of } P.$$

Causal participation is turned into ontological localization.

11.2 Neuroscientific Localizationism

The fact that lesioning or stimulating a brain region changes a capacity demonstrates causal relevance. It does not show that the capacity is localized in that region as an isolated property. The region has its function within recurrent neural, vascular, metabolic, and organismic contexts.

Here too,

$$N_i \rightarrow E$$

is too quickly turned into

$$N_i = \text{location of } E.$$

11.3 Functionalist Transfer

The error becomes especially consequential in functionalism. If a biological system realizes a functional relation and a machine produces the same input output relation, this is often taken to establish functional equivalence at the level of system properties.

From

$$I \to O$$

in two systems, however, only equivalence of this relation follows. It does not follow that

$$\mathcal{O}_{bio} = \mathcal{O}_{tech}.$$

A technically reproduced relation may be identical while the conditions of its production are fundamentally different. Precisely this difference disappears when function is completely detached from genesis and self-production.

12. Why a Technical Control Loop Is Not an Autocatalytic Organism

The distinction from technical systems is important because modern machines can be highly recurrent, adaptive, and self-regulating. An artificial neural network contains feedback. A robot can monitor its battery and seek out a charging station. A software agent can diagnose its own states and restart processes.

All of this shows that feedback, regulation, and even functional self-maintenance are insufficient to characterize biological organization.

The difference lies in the origin of the conditions of realization. In a technical system, the causal architecture is constructively presupposed as a functioning organization. Its components need not be generated through the system's own operational context. Even if a robot could replace individual components, the production system, material cycle, energy infrastructure, and design logic would still be presupposed unless they too were integrated into the same self-producing context.

In an organism, metabolism, membrane formation, repair, synthesis, regulation, and reproduction are intertwined both historically and in the present. Organization is not merely an arrangement of parts, but a process in which the conditions of that arrangement are continuously generated.

It is therefore necessary to distinguish between two senses of "for the system." A technical state can be functionally relevant to the continued existence of a machine because design and control logic make it so. In an organism, relevance arises within a system whose persistence is itself the result of recurrent self-production.

The difference is therefore not complexity. It is organizational genesis.

13. The Historical Dimension: Organization Has Become What It Is

Living organization is not only synchronically recurrent; it is historically evolved. Every present causal relation in an organism is the result of a phylogenetic and ontogenetic history during which the conditions of its efficacy have changed.

This matters for a theory of causation. A technical mechanism can often be explained from its current structure because its design is externally documentable and its function can largely be separated from its history of origin. In an organism, by contrast, present organization is a condensation of historical change.

The causal role of a structure therefore cannot always be derived from its present form alone. Developmental processes alter the spaces of possibility available to later processes. Earlier states generate boundary conditions that channel later development. Evolution changes not only components but the organization in which components become causally effective.

The formula

$$A \rightarrow B \mid \mathcal{O}$$

thus acquires a temporal extension:

$$A_t \rightarrow B_t \mid \mathcal{O}_t,$$

where

$$\mathcal{O}_t = F(\mathcal{O}_{t-1}, E_{t-1}, \Delta_t).$$

E stands for environmental conditions and Δ for developmental change. The concrete organization at a given time is therefore the result of its own history and its coupling to the environment.

The approach thereby also differs from purely synchronic closure models. Organization is not only closed. It has become what it is.

14. From Biological Closure to Causal Collapse

The argument so far concerns living systems in general. An autocatalytically organized organism generates and maintains conditions under which its components can become causally effective. Yet this does not by itself entail consciousness. Self-production, organizational closure, and recurrence are necessary points of reference for the theory developed here, but taken alone they do not constitute a theory of experience.

The transition to consciousness therefore requires a further step. This step was developed in a series of earlier papers under the concept of **causal collapse** (Stegemann 2025a, 2025b, 2025c). Its integration into a broader developmental line from autopoiesis through causal poiesis to causal collapse was subsequently elaborated further (Stegemann 2025d). The present paper retains this concept but clarifies the notion of causality presupposed by it. The expression is not intended to mean that causality disappears. It denotes a collapse of **sufficient causal factorization**.

14.1 Why Nervous Systems Create a New Organizational Situation

With the emergence of nervous systems, chemical, metabolic, and endocrine processes are supplemented by a much faster form of spatially distributed coupling. Electrical and electrochemical signals make it possible for distant functional domains to enter into recurrent relations within short time windows. Varela, Lachaux, Rodriguez, and Martinerie (2001) described this problem in terms of large-scale integration of distributed neural activity and discussed phase synchronization as a possible mechanism of dynamic binding. Such findings initially show only that neural systems possess temporally coordinated long-range relations. They do not yet explain why consciousness should arise from them.

For the present theory, what matters is that the nervous system is not an externally added information network. It is part of a living system whose material, energetic, and regulatory conditions are continuously generated and maintained by the organism. Rapid neural recurrence therefore builds upon an already existing autocatalytic organization.

Schematically, a developmental line can be represented as follows:

Autocatalysis → self-production → neural recurrence → increasing non-separability.

This does not yet determine a threshold for consciousness. It merely identifies the structural precondition under which such a threshold can meaningfully be formulated.

14.2 The Original Thesis of Causal Collapse

In the earlier papers, causal collapse was described as a state in which previously functionally distinguishable neural processes lose their independent causal separability through recursive coupling (Stegemann 2025a, 2025b, 2025c). From the outset, the decisive point was not the sheer number of connections but a change in the state of the system. The system is not conscious because many neurons are active or because information is processed simultaneously at many locations. What matters is the organizational form of their mutual dependence.

As long as functional domains can be described with sufficient independence, a system can be approximated by decomposing it into subprocesses. For a partition

$$\mathcal{P} = \{S_1, S_2, \dots, S_k\}$$

such a description would be adequate if the global dynamics F could be reconstructed from the dynamics of the subsystems with only limited loss:

$$F \approx F_{S_1} \otimes F_{S_2} \otimes \cdots \otimes F_{S_k}.$$

The symbol $\otimes$ is not intended here to posit a particular physical operation. It serves only as formal shorthand for a composite description in which subprocesses can be modeled separately and subsequently combined.

Causal collapse denotes the limiting case in which such decomposition is no longer sufficient for the relevant dynamics. For every relevant partition $\mathcal{P}$, a non-negligible remainder of the global dynamics persists:

$$D(F, F_{\mathcal{P}}) > \varepsilon.$$

Here, D does not designate an already fixed empirical measure. It stands generally for the deviation between the actual system dynamics and a description reconstructed from separate subsystem

dynamics. Nor is ε a universal natural constant. It marks a theoretically or empirically determinable boundary of sufficient explanation. The formula therefore does not yet operationalize causal collapse. It merely makes explicit what is meant by factorization.

The central claim is:

$$\boxed{\text{Causal collapse} = \text{collapse of sufficient causal factorization.}}$$

14.3 What Does Not Collapse in Causal Collapse

This clarification is necessary because the expression "collapse of causality" can be misleading. It might be read as implying that the physical causal order breaks down or that an indeterministic event occurs. Neither is intended.

Every individual neural interaction remains physically describable. Membrane potentials, synaptic transmission, neuromodulatory influences, and metabolic conditions do not lose their causal character. Nor does "collapse" refer to a quantum physical collapse. The concept denotes only a change in the **possibility of describing the overall causal structure**.

Locally, the following still holds:

$$A \to B.$$

What becomes problematic is the stronger inference:

$$A \to B \quad \Rightarrow \quad A \text{ and } B \text{ can be explained independently of the global state.}$$

It is precisely this inference that no longer holds under causal collapse. The local relation remains real, but its concrete efficacy depends so strongly on recurrent counter-effects, modulations, and state conditions that an isolated description no longer reconstructs the system state sufficiently.

The concept therefore concerns not the existence of causes but their **separability**.

14.4 Non-Factorizability Is More Than Correlation

This distinction is equally important. Two neural variables can be strongly correlated without being causally non-separable. A common cause may alter both at the same time. Conversely, processes that are highly causally dependent on one another may, under certain conditions, display only weak linear correlations.

Causal collapse therefore cannot be defined by functional connectivity or synchrony alone. Such measures can indicate coupled dynamics, but they are not direct measures of causal non-factorizability.

Perturbational methods are methodologically closer to the issue. Using the Perturbational Complexity Index, Casali and colleagues (2013) showed that conscious states are associated with complex, spatially distributed cortical responses to a targeted perturbation. Koch, Massimini, Boly, and Tononi (2016) likewise discuss differentiation and integration of neural activity as relevant quantitative markers of conscious states. These findings do not prove causal collapse. They do, however, show that the response of the entire system to a local perturbation can be investigated empirically and that states of consciousness differ precisely in the range and differentiation of such system responses.

For the theory proposed here, the decisive question would therefore not be how strongly two regions correlate, but how strongly a local perturbation distributes its effects across the system through recurrent dependencies and to what extent a partition destroys the explanatory account of those effects.

The empirical target thus shifts from

correlation

to

perturbationally determinable causal non-separability.

14.5 A Threshold, Not Merely More Integration

The term collapse implies a transition. It is therefore not claimed that every additional recurrent connection makes a system slightly more conscious. Rather, the theory claims that the quality of organization changes once functionally segregated processes can no longer be treated with sufficient independence for the relevant system state.

Before the threshold, a partition may still be explanatorily sufficient:

$$D(F, F_{\mathcal{P}}) \leq \varepsilon.$$

Beyond the threshold, for the relevant partitions:

$$D(F, F_{\mathcal{P}}) > \varepsilon.$$

This threshold formulation is initially theoretical. It does not claim that a universal measure is already known from which consciousness could be read off unambiguously. This is precisely where the approach differs from premature operationalization. Before a measure can be fixed, one must first clarify what kind of non-separability is to be measured.

Causal collapse is therefore not an equation of consciousness with network density, synchrony, recurrence, or mathematical integration. It denotes a specific system state in which the causal role of the processes involved can be adequately determined only within their shared recurrent dynamics.

14.6 What the Present Clarification Adds to the Earlier Papers

The 2025 papers on causal collapse initially developed the model from the problem of consciousness. Consciousness was described as a state of far-reaching causal inseparability in a recurrent neural system and distinguished from models based merely on information processing or functional accessibility (Stegemann 2025a, 2025b, 2025c). In *Autopoiesis, causal poiesis, causal collapse*, this idea was placed more explicitly within an evolutionary and organizational developmental line of living systems (Stegemann 2025d).

The present paper adds a conceptual clarification that was not yet sufficiently explicit in those works. If one speaks of causal collapse, one must first clarify the concept of causality being presupposed. The answer now is: the collapse concerns neither physical causality nor the existence of local cause effect relations. It concerns the **factorizability of their organization**.

This also changes the theoretical status of the concept. Causal collapse no longer appears as an isolated additional principle for explaining consciousness. It becomes the limiting case of a more general development:

causal relation → recurrent causal organization → autocatalytic self-production.

Building on this:

autocatalytic self-production → rapid neural recurrence → collapse of sufficient factorization.

This sequence does not claim a linear historical ladder. It denotes a conceptual reconstruction of increasing organizational dependence. Consciousness is not deduced from autocatalysis. Rather, the additional organizational form required for causal collapse within an already living system is specified.

15. Distinguishing the Approach from Theories of Integration

Talk of non-factorizability must not be confused with the mere claim that a system is "integrated." Integration is initially a structural, dynamical, or statistical property. Many non-conscious biological and technical systems can be highly integrated.

The difference becomes especially clear in relation to Integrated Information Theory. IIT characterizes consciousness in terms of a system's intrinsic cause effect power and attempts to capture irreducibility formally by comparing the whole system with appropriate partitions (Oizumi, Albantakis & Tononi 2014). At first glance, the proximity to the concept of non-factorizability used here is therefore considerable. Both approaches ask what is lost when the system is partitioned.

The theoretical status of this question, however, differs. In IIT, irreducible causal structure itself becomes the constitutive core of consciousness. The approach defended here does not begin with a formal property of arbitrary systems. Non-factorizability acquires its significance only within an already living, autocatalytically self-producing, plastic, and historically evolved neural system.

Accordingly:

Integration ≠ consciousness

and likewise:

Irreducibility under a partition ≠ causal collapse.

Only the combination of biological genesis, organismic self-production, rapid recurrent neural coupling, and the loss of sufficient causal factorization forms the theoretical framework of causal collapse.

The difference is therefore not merely one of degree. IIT seeks a formal property that can in principle be determined independently of the type of system. Causal collapse, by contrast, is tied to a particular genesis and organizational form. A formally identical network structure in a technically constructed system would therefore not be phenomenally equivalent merely because of its mathematical irreducibility.

The advantage of this position is that a structural feature is not immediately ontologized. Integration, complexity, and irreducibility can be markers of a particular organization. They are not yet its explanation.

16. Distinguishing the Approach from Emergence

The concept of emergence is often used when systems display properties that do not belong to their components in isolation. In a weak sense, this statement is unproblematic. A heartbeat is not a property of an individual cardiac muscle cell, and a vortex is not a property of a single water molecule.

The difficulty begins when relational novelty is taken to imply new causal powers. Mossio, Bich, and Moreno take precisely this step when they describe organizational closure as an emergent causal regime with ontologically irreducible properties (Mossio, Bich & Moreno 2013).

The approach defended here does not require this step.

A system property can be explanatorily irreducible without introducing ontologically additional causality. Explanatory irreducibility then means only that the property can be understood only with reference to the organization. It does not mean that alongside physical interactions there exists a second class of biological forces.

One can therefore distinguish:

$$\text{ontological reduction} \neq \text{explanatory reduction.}$$

The fact that an organism is fully physically realized does not mean that a description of its molecules explains its organization. Conversely, the necessity of organizational concepts does not imply that organization is an additional ontological entity.

This position avoids both eliminative reductionism and causal emergentism.

17. One Ontology, Multiple Descriptions

The concept of causality is often unnecessarily complicated by ontologizing levels of description. Physiological, psychological, and phenomenological descriptions then appear as separate domains between which effects would somehow have to be transmitted.

An organismic approach requires no such duplication. There is one organism. The same organism can be described molecularly, cellularly, physiologically, behaviorally, or phenomenologically. These descriptions capture different relations and regularities. They do not generate additional objects.

The same follows for causality: causal relations exist in the real system, not between descriptions. A psychological description does not cause a neural change, just as a neural description does not cause an action. Both may capture different aspects of the same organismic process.

This view is central to the theory of causal organization because it prevents organization itself from being turned into a higher-level ontological agent.

The organism as a whole does not “act downward” on its cells. Rather, local cellular processes are embedded in conditions realized by the ongoing state of that same organism.

18. Scientific Consequence: Intervention Establishes Participation, Not Automatically Organization

The distinction has direct methodological consequences.

Experimental interventions are especially effective at making dependencies visible. If A is manipulated and B changes, causal participation by A can be justified. The cleaner the intervention, the stronger the claim.

Yet even a perfect intervention initially answers only a local question. It shows that A is relevant to B under the given conditions. It does not automatically show how those conditions arose, how they are maintained, or what role the same relation plays in other organizational states.

Logically, an intervention should therefore justify only the following inference:

$$I(A) \Rightarrow \Delta B \Rightarrow A \text{ is causally relevant to } B \text{ under } \mathcal{O}.$$

Without an additional argument, the following inference is not warranted:

$$I(A) \Rightarrow \Delta B \Rightarrow A \text{ fully explains } B.$$

This difference is particularly important in complex biological systems because interventions often trigger secondary system adaptations. An organism responds to perturbations. Thus, not only one variable but often the organization itself changes.

19. The Actual Theoretical Shift

The literature on autopoiesis, closure, biological autonomy, work constraint cycles, and biological relativity has already described different parts of the problem with considerable clarity. The theory proposed here therefore does not claim to have discovered organismic recurrence for the first time, nor does it introduce self-production, closure, or multiscalarity as new phenomena.

Its independent contribution lies in a different conceptual ordering.

First, it distinguishes between **causal relation** and **causal organization**. A local cause effect relation and the organizational form that enables its concrete efficacy are not identical.

Second, local experimental factorization is explicitly separated from systemic factorization. The fact that an intervention isolates a causal dependence does not show that the organism consists of independently recombinable causal modules.

Third, organizational closure is not interpreted as a new kind of causality. Life adds no specifically biological force to physics. It organizes physical causal relations in such a way that conditions of their efficacy are recurrently generated and maintained within the system.

Fourth, levels of description are not treated as levels of causation. This removes the need to connect psychological, biological, or physical descriptions through an additional top down or interlevel causation.

Fifth, historical genesis is incorporated into causal organization. The present role of a component depends not only on current structure. It is part of a phylogenetic and ontogenetic development that has generated the relevant space of possible causal efficacy.

Sixth, the concept of causal collapse developed in earlier work is embedded within this more general theory of causality (Stegemann 2025a, 2025b, 2025c, 2025d). Its specific content can now be stated more precisely: it is not causality that collapses, but the sufficient factorization of the relevant recurrent neural dynamics.

The theory can therefore be compressed into the following statement:

$$\boxed{\text{Life generates no new causality, but a new organization of causality.}}$$

For complex neural systems, the further claim is:

$$\boxed{\text{Causal collapse} = \text{collapse of sufficient factorization within recurrent neural organization.}}$$

The concept of causal collapse is thereby not replaced but more narrowly specified. The earlier theory of consciousness acquires an explicit theory of causality as its foundation.

20. Conclusion

Living systems pose a special challenge to the concept of causality not because they suspend the laws of nature. They pose a challenge because the conditions of causal efficacy themselves become part of the organization to be explained.

A molecule acts according to the same physical and chemical principles inside and outside an organism. Within an organism, however, it is embedded in a context that continuously generates concentrations, spatial boundaries, reaction partners, energy gradients, and regulatory conditions. Its concrete causal role is therefore neither merely intrinsic nor independent of the system.

Classical causal analysis remains fully valid. But it must be supplemented by an analysis of organization.

The decisive distinction is:

$$\boxed{\text{Causal relation} \neq \text{causal organization.}}$$

And it follows that:

$$\boxed{\text{Local factorization} \not\Rightarrow \text{factorization of the organism.}}$$

This distinction makes it possible to avoid reductionism without introducing special emergent forces. It permits a unified ontology without denying the explanatory autonomy of biological organization. At the same time, it creates a conceptual transition from autocatalysis through organismic closure to neural non-factorizability.

Causal collapse consequently acquires a more precise meaning. The concept denotes neither a rupture of physical causality nor an additional causal power. It denotes the limiting case in which the neural processes relevant to the system state can no longer be factorized with sufficient independence without losing the organizing dynamics itself. What collapses is explanation in terms of separate causal modules.

This also makes clear what the present paper adds to the earlier work on causal collapse. Those papers developed a model of consciousness (Stegemann 2025a, 2025b, 2025c) and later situated it

within a broader development from autopoiesis and causal poiesis (Stegemann 2025d). Here, the concept of causality presupposed by that model is made explicit.

The organism is therefore not causally mysterious. It is causally organized in a different way. And causal collapse is not the end of causality, but the point at which its isolated factorization ceases to provide a sufficient explanation.

References

Cárdenas, M. L., Letelier, J. C., Gutierrez, C., Cornish-Bowden, A. & Soto-Andrade, J. (2010). Closure to efficient causation, computability and artificial life. *Journal of Theoretical Biology, 263*(1), 79–92. https://doi.org/10.1016/j.jtbi.2009.11.010

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M. A., Laureys, S., Tononi, G. & Massimini, M. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Science Translational Medicine, 5*(198), 198ra105. https://doi.org/10.1126/scitranslmed.3006294

Craver, C. F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.

Craver, C. F. (2025). Mechanistic Explanation. *Open Encyclopedia of Cognitive Science*. https://doi.org/10.21428/e2759450

Craver, C. F. & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology & Philosophy, 22*, 547–563. https://doi.org/10.1007/s10539-006-9028-8

Kauffman, S. A. (2000). *Investigations*. Oxford: Oxford University Press.

Kauffman, S. A. & Clayton, P. (2006). On emergence, agency, and organization. *Biology & Philosophy, 21*, 501–521. https://doi.org/10.1007/s10539-005-9003-9

Koch, C., Massimini, M., Boly, M. & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience, 17*(5), 307–321. https://doi.org/10.1038/nrn.2016.22

Maturana, H. R. (1975). The organization of the living: A theory of the living organization. *International Journal of Man-Machine Studies, 7*(3), 313–332. https://doi.org/10.1016/S0020-7373(75)80015-0

Montévil, M. & Mossio, M. (2015). Biological organisation as closure of constraints. *Journal of Theoretical Biology, 372*, 179–191. https://doi.org/10.1016/j.jtbi.2015.02.029

Moreno, A. & Mossio, M. (2015). *Biological Autonomy: A Philosophical and Theoretical Enquiry*. Dordrecht: Springer. https://doi.org/10.1007/978-94-017-9837-2

Mossio, M., Bich, L. & Moreno, A. (2013). Emergence, closure and inter-level causation in biological systems. *Erkenntnis, 78*(Suppl. 2), 153–178. https://doi.org/10.1007/s10670-013-9507-7

Noble, D. (2012). A theory of biological relativity: no privileged level of causation. *Interface Focus, 2*(1), 55–64. https://doi.org/10.1098/rsfs.2011.0067

Oizumi, M., Albantakis, L. & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS Computational Biology, 10*(5), e1003588. https://doi.org/10.1371/journal.pcbi.1003588

Rosen, R. (1991). *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*. New York: Columbia University Press.

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

Stegemann, W. (2025b). *Consciousness: Expanding the Model of Causal Collapse*. Zenodo. https://doi.org/10.5281/zenodo.15351381

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

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

Varela, F. J., Lachaux, J. P., Rodriguez, E. & Martinerie, J. (2001). The brainweb: Phase synchronization and large scale integration. *Nature Reviews Neuroscience, 2*(4), 229–239. https://doi.org/10.1038/35067550

Varela, F. G., Maturana, H. R. & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems, 5*(4), 187–196. https://doi.org/10.1016/0303-2647(74)90031-8