Methodology of Consciousness Theory

A Reconstructivist Solution to the Problem of Consciousness

Dr. Wolfgang Stegemann

Abstract

Consciousness research suffers from a methodological and categorical double deficit. While the natural sciences explicitly discuss and justify their methods, consciousness research lacks systematic methodological reflection. This methodological deficit is functionally related to a basic categorical error: the confusion of epistemic levels of description with ontological entities. This level confusion results in three pseudo-problems, which are treated as central research questions, namely the body-soul problem, the hard problem and the what-is-it-like problem.

The present article develops a reconstructionist method as a solution to this double deficit. This method avoids the categorical confusion between levels of description and ontology and identifies functional necessities instead of arbitrary correlations. Central is the insight that consciousness does not emerge from dead matter, but from autocatalytic organization. The model of causal collapse describes consciousness as a state of maximally non-factorizable causal entanglement in neuronal systems that arises when a measure of causal separability $\Sigma(t)$ falls below a critical threshold Σc.

The thesis systematically reconstructs the necessary conditions for consciousness: autocatalytic organization as a biological foundation, the causal nucleus as an evolutionary structural principle from the cell nucleus to the default mode network, electrical signal transduction in the nervous system, and finally causal collapse as a mechanism of phenomenal experience. Phylogenetic development shows how the causal nucleus as a structural principle permeates evolution from prokaryotes to early nervous systems to human consciousness.

Different forms of consciousness are derived from specific collapse structures: primary consciousness as sensorimotor collapse, self-consciousness as recursive collapse with self-model, intentionality as prospective collapse, and affective consciousness as limbically integrated collapse. The theoretical positioning shows that the model integrates and specifies Integrated Information Theory, Global Workspace Theory and plasticity approaches.

Empirically, the model is supported by the Perturbational Complexity Index (PCI), functional connectivity patterns, and characteristic differences between states of consciousness. For predictions that are not directly proven, concrete test procedures are proposed. The method transforms arbitrary correlations into functional necessities and opens up new avenues for empirical research, artificial systems of consciousness and clinical applications.

1. Introduction: The Methodological and Categorical Double Deficit

1.1 The methodological deficit of consciousness research

The philosophy of mind operates with different methodological approaches: analytical decomposition, correlative research, evolutionary narratives, axiomatic theory formation. This methodological diversity could be interpreted as richness, but on closer inspection it turns out to be a symptom of a fundamental problem. While the natural sciences explicitly discuss and justify their methods, while physics, for example, systematically reflects on the experimental method and evolutionary biology methodologically bases the comparative method, consciousness research lacks systematic reflection on its own methodological tools (Chalmers 1996; Dennett 2017).

There is no methodological approach that does for consciousness research what the experimental method does for physics or the comparative method does for evolutionary biology. This statement is not trivial, because it points to more than just a plurality of approaches. The various approaches to consciousness research are not complementary, such as particle and wave theory in quantum mechanics, but often compete with each other without an overarching methodological principle that could determine their respective validity or scope.

Correlative neuroscience, for example, systematically collects data on neuronal correlates of consciousness (Crick & Koch 1990; Koch et al. 2016), but stops at the description of relationships. She can determine that certain neural patterns correlate with states of consciousness without being able to explain why these patterns and not others are associated with consciousness. If, for example, the Perturbational Complexity Index (PCI) correlates with consciousness (Casali et al. 2013; Massimini et al. 2005), it remains unclear in the correlative method whether high causal complexity is characteristic of consciousness or is only contingent with it.

Evolutionary approaches, on the other hand, provide historical narratives about the emergence of consciousness (Feinberg & Mallatt 2016; Ginsburg & Jablonka 2019), however, miss the timeless conditions of the possibility of consciousness. They can describe when certain forms of consciousness may have occurred in evolutionary history, but not which functional structures are necessary for consciousness to arise in the first place. The historical contingency of evolutionary developments is confused with the structural necessity of functional conditions.

Analytical methods break down the brain into components and examine them in isolation (Koch 2004), but do not capture the emergent organization that results precisely from the non-factorizable entanglement of these components. Axiomatic theories such as Integrated Information Theory (Tononi et al. 2016; Oizumi et al. 2014) postulate principles such as the integration and differentiation of information without clarifying whether these principles represent necessary, merely sufficient or even contingent conditions for consciousness.

Another problem is the inappropriate transfer of theories from foreign domains. Theories such as the Free Energy Principle, Integrated Information Theory or Predictive Processing are precise and successful in their original context (thermodynamics, information theory, Bayesian inference). However, their application to consciousness fails because they do not capture the constitutive properties of autocatalytic systems. Information without meaning, prediction error minimization without valence, free energy without thermodynamic self-preservation cannot explain consciousness. This theoretical deficit is analysed in detail in Chapter 10.

1.2 The Categorical Root Problem: Level Confusion

This methodological deficit is functionally related to a categorical fundamental error that pervades contemporary consciousness research: the confusion of epistemic levels of description with ontological entities. Physiological and psychological descriptions are different epistemic approaches to the same phenomenon, not two separate ontological domains. A physiological description indicates what measurably happens in the brain: electrochemical activity, synaptic transmission, oscillatory patterns. A psychological description indicates how this process presents itself to the system itself: as a sensation, thought, emotion, conscious state.

Even more problematic: consciousness research often establishes causal relationships between these levels of description, as if neuronal processes (physiological description) would causally produce phenomenal experience (psychological description). This causalization of different epistemic levels creates a pseudo-problem, because causal relationships exist between events in the world, not between different ways of describing these events. One cannot meaningfully ask how one description causally produces another description (Kim 1998; Searle 1992).

Categorical confusion has a long philosophical history. Ryle (1949) already identified similar errors as "category mistakes". But while Ryle limited himself to conceptual analyses, consciousness research manifests a systematic practice of level confusion that has empirical consequences: it leads to research programs that search for causal mechanisms by which neuronal processes "generate" consciousness and thus pursue a meaningless question.

1.3 Reciprocal conditionality of both deficits

This categorical confusion has created three pseudo-problems that are treated as central research questions: the body-soul problem, the hard problem, and the what-is-it-like problem. These problems appear to be insoluble only because they are based on a categorically misguided question. The body-soul problem arises when epistemic approaches (physical versus mental description) are misunderstood as ontological substances (Descartes 1641/1996). The hard problem (Chalmers 1995) makes the same mistake by asking how physical processes causally produce phenomenal properties, and thus causally linking two levels of description. The what-is-it-like problem (Nagel 1974) results from the assumption that subjectivity is an ontological property in addition to objective processes, rather than an epistemic perspective on these processes.

The persistence of these pseudo-problems in turn perpetuates the methodological deficit, since no method can solve a categorically misguided problem. Both deficits are mutually dependent: the lack of methodological reflection allows categorical confusion, and categorical confusion prevents the development of an appropriate method. A vicious circle that can only be broken by resolving both deficits at the same time.

1.4 Objectives and structure of the thesis

The present article develops the reconstructivist method as a solution to this double deficit. This method avoids the categorical confusion between levels of description and ontology and identifies functional necessities instead of arbitrary correlations. The decisive methodological advantage over the correlative method lies in the fact that reconstructionism does not label arbitrarily, but identifies conditions. While correlative research can designate arbitrary neuronal patterns as correlates of consciousness, the reconstructivist method asks: What functional organization must a living system have if it is conscious?

The thesis first shows how the three pseudo-problems result from the level confusion (Chapter 2). She then presents the reconstructivist method as a methodological approach that avoids this categorical confusion (Chapter 3). Chapters 4 to 7 systematically develop the reconstructivist solution: autocatalysis as a biological foundation (chapter 4), the causal nucleus as an evolutionary structural principle (chapter 5), the nervous system as an integration mechanism (chapter 6) and causal collapse as a mechanism of phenomenal experience (chapter 7). Chapter 8 reconstructs the phylogenetic development of consciousness and shows how the causal nucleus as a structural principle permeates evolution. Chapter 9 differentiates between different forms of consciousness from the collapse structure. The theoretical positioning (chapter 10) and methodological consequences (chapter 11) conclude the work.

2. Three pseudo-problems as a consequence of the level confusion

2.1 The Body-Mind Problem: Transformation of Epistemics into Ontology

2.1.1 The Logical Dichotomy: Entity or Epistemic Access

Before the question of the interaction of body and soul can be answered, it must be clarified what status these two terms have. Logically, the possibilities are limited to a strict dichotomy: either the distinction between body and soul denotes two ontologically separate entities, or it denotes two epistemic perspectives on the same phenomenon. Tertium non datur.

This dichotomy can be systematically justified. The question of ontological status demands an answer to the form: What exists? For every X, either X exists as a separate entity, or X does not exist as a separate entity. If X does not exist as an independent entity, X must be something other than an entity, such as a property, a relation, or an epistemic approach. Since body and soul are not classified as properties or relations (they are not predicates of other entities), only two options remain: ontological entity or epistemic approach.

The logical structure of this dichotomy is compelling. It is not an empirical question that could be decided by further research, but a conceptual necessity. Tertium non datur here is not merely a logical principle, but follows from the exhaustion of logical possibilities:

what is not an independent entity can only exist as a dependent property, relation or epistemic access. Since the first two options are out of the question for body and soul, only the last one remains.

2.1.2 Empirical testing of the ontological option

The empirical examination of the ontological option leads to a negative result. If body and soul were two separate entities, they would have to be identifiable. What can be found empirically? An organism with neuronal processes, with measurable electrochemical activity, with structural properties. This organism can be described from the third-person perspective by objective measurement of neuronal processes (physiological description) or from the first-person perspective by phenomenal experience (psychological description). There are two descriptions, but no two entities.

Neuroscientific research in recent decades has systematically confirmed this insight. Functional imaging studies show that neural activity patterns correspond to every conscious state (Dehaene & Changeux 2011; Koch et al. 2016). These correlations are so precise that neuronal activity can be used to infer mental states and vice versa (Haynes & Rees 2006). What cannot be found is a second entity "soul" or "spirit" that existed in addition to the neuronal processes.

The only candidate for such an additional entity would be a dualistic substrate, as postulated by Descartes (1641/1996) with the res cogitans. But an immaterial substance that causally interacts with the material world is incompatible with fundamental physical principles, especially the causal closure of the physical (Papineau 2001). If mental events could have a causal effect on physical events, this would violate laws of conservation of energy. The assumption of a non-physical mind is therefore not only empirically unconfirmed, but systematically problematic.

Since no two ontological entities can be identified, the epistemic option must apply: body and soul are methodological approaches, not ontological substances. The body-soul problem arises from the fact that these methodological differences are transformed into ontological substances. The subsequent question about the interaction of two such different entities is categorically misguided, because it turns epistemics into ontology.

2.1.3 Critique of alternative positions

Four objections can be formulated against this dichotomy. First, non-reductive physicalism conceives of the soul as an emergent property of physical systems, neither as an entity in its own right nor as a mere perspective (Kim 1999). Second: Neutral monism sees neither body nor soul as fundamental, but both as manifestations of a third, neutral basic material (Russell 1921; Chalmers 1996). Third, process ontologies understand body and soul as processes, not as entities, with functional rather than epistemic distinction

(Whitehead 1929). Fourth, pragmatic approaches treat distinction as a tool, not as an ontological or epistemic status (James 1890/1950).

However, on closer analysis, these four positions fall back on the epistemic side of the dichotomy or constitute hidden dualisms. Non-reductive physicalism claims that the soul results as an emergence from the body. Thus, there are two ontologically distinct variables: the body and the emergent soul. Dualism is merely displaced by the predicate "emergent", not dissolved. The question of how something radically different can emerge from purely physical processes remains just as mysterious as the original question of the interaction of body and soul (Chalmers 2006).

Neutral monism postulates a trialism: the neutral raw material plus two resulting manifestations. Here, too, there remain two distinct phenomena, the relation of which must be explained. Whether one speaks of two substances or two manifestations of a basic substance does not change the fundamental difficulty. Russell (1921) himself admitted that the nature of this neutral entity remains unclear and may already be psychophysically structured, thus undermining monism.

Process ontologies speak of two functions or processes, which also establishes a duality. Whitehead (1929) distinguishes between physical and mental poles of events, but thus reproduces precisely the dichotomy at the level of processes instead of substances. The question of how physical and mental aspects of a process are related is structurally identical to the question of body-soul interaction. The question is therefore not whether consciousness is a substance or a process, but whether it represents or should represent an independent ontological category at all.

Finally, pragmatic approaches are either epistemic (in which case the position falls on our side of the dichotomy) or they categorically evade the question. James (1890/1950) treats consciousness as a function, not as an entity, which corresponds to an epistemic interpretation. If, on the other hand, pragmatic approaches claim that the distinction is merely useful without ontological or epistemic content, then it remains unclear what its meaning is supposed to consist of.

2.1.4 Structural ontology as an appropriate framework

An important clarification is necessary here: ontology itself is an epistemic concept. The question is not whether ontology exists, but which ontology is appropriate to the phenomenon. The classical ontology that creates the body-soul problem looks for separate entities as fundamental building blocks. A structural ontology, on the other hand, references that which constitutes life and thus consciousness: autocatalytic structures and their causal organization (Kauffman 1993; Morowitz 1992).

From a structural ontology, the body-soul problem resolves itself because it is based on a categorically inadequate ontology. Consciousness is not a substance and not a property of a substance, but a specific causal organization of autocatalytic systems (Stegemann 2025a). The "body" is the physiological description of this organization, the "soul" the phenomenal manifestation of the same organization for the system itself. There are not two things that have to interact, but one thing with two ways of describing.

This resolution is not merely terminological. It has empirical consequences: instead of looking for causal mechanisms through which body and soul interact, a structural ontology searches for the functional conditions under which autocatalytic systems are conscious. This leads to testable hypotheses about neuronal organization, as shown in the following chapters.

2.2 The Hard Problem: Two Systematic Misalignments

2.2.1 First Deformity: Consciousness from Dead Matter Instead of Life

The so-called hard problem of consciousness (Chalmers 1995) commits two fundamental errors. First, it tries to explain consciousness from dead matter instead of life. Consciousness, however, is not a property of any physical system, but a property of living, autocatalytic organization. The appropriate question is not how physical processes produce consciousness, but what kind of living processes fulfill the functional requirements for consciousness.

Chalmers (1995) formulates the hard problem as the question of why and how physical processes are accompanied by phenomenal experience. He distinguishes between "easy problems" (which concern functional mechanisms and can be solved in principle) and the "hard problem" (why there is experience at all). This distinction presupposes that the functional organization can be fully described without grasping consciousness. Thus, consciousness is treated as something that is added to physical processes, rather than as an intrinsic manifestation of a certain type of process.

The mistake lies in the choice of the explanandum. The hard problem asks: "Why does physical processing go accompanied by experience?" The correct question is: "Which autocatalytic processes create an operationally closed interior for which something can be significant?" The first question assumes consciousness as an additional property, the second reconstructs consciousness from the functional conditions of living systems.

Schrödinger (1944) already recognized that life cannot be understood from the laws of dead matter alone, but requires organizational principles that maintain negentropy. Prigogine (1980) showed that new forms of order are accessible to dissipative structures far from equilibrium. Kauffman (1993) demonstrated that autocatalytic sets emerge spontaneously and form self-sustaining organizations. All this work points to the fact that life requires a specific thermodynamic and organizational structure. Consciousness, as a property of living systems, presupposes this structure.

2.2.2 Second misalignment: causalization of different levels of description

Second, the hard problem commits the same categorical error as the body-soul problem: it treats different levels of description as if they were causally connected. Neural processes and phenomenal experience are not two different things that causally affect each other, but two descriptions of the same process. A physiological description indicates what measurably happens in the brain. A phenomenal description indicates how this process presents itself to the system itself. Both descriptions refer to the same process.

The formulation of the hard problem suggests a causal gap: physical processes are given, and it is necessary to explain how experience arises from them. However, this gap is artificial. It only arises when physical processes are described to the exclusion of their phenomenal manifestation and then asked how this manifestation is added. Methodologically, this corresponds to a petitio principii: one defines physical processes in such a way that they do not contain any experience, and then one is surprised that one cannot explain experience.

An adequate description of autocatalytic processes includes their phenomenal manifestation as soon as these processes constitute an operationally closed interior. For such a system, its own dynamics are not merely an objective process, but at the same time the subjective perspective on this process. There is no additional entity "experience" that is added to the physical processes, but experience is the intrinsic manifestation of a certain type of causal organization (Stegemann 2025a).

Thompson (2007) argues in this sense that consciousness cannot be reduced to a neural basis without considering the living, self-organizing nature of that basis. Varela et al. (1991) show that cognition cannot be separated from life. These approaches point in the right direction, but often remain on a conceptual level. What is missing is a systematic reconstruction of the functional conditions that combine autocatalytic organization with consciousness.

2.2.3 The appropriate question: What living processes make consciousness possible?

The appropriate question, therefore, is: What functional organization does a living system have if it is conscious? This question avoids both misalignments of the hard problem. It does not start with dead matter, but with autocatalytic organization. It does not causalize different levels of description, but asks about the structural configuration that is both physiologically measurable and phenomenally manifest.

The model of causal collapse (Stegemann 2025a, 2025b) answers this question: Consciousness arises when an autocatalytic system has electrically coupled processes (nervous system) that form a causal core and transition into a state of maximally non-factorizable causal entanglement. This state is neurophysiologically measurable as global, integrated network dynamics and phenomenally manifest as conscious experience.

The solution to the hard problem thus lies not in an explanation of how physics becomes phenomenology, but in the realization that this question is categorically misguided. Phenomenology is not an additional property added to physical processes, but the intrinsic perspective of autocatalytic systems on their own causal organization once it reaches a certain structural configuration.

2.3 The What-is-it-like Problem: From Apparent Problem to Functional Necessity

Nagel's (1974) famous question "What is it like to be a bat?" aims at the subjectivity of experience. He argues that even complete knowledge of a bat's neural and physiological organization would not capture what it's like to be a bat. This subjective perspective, the "what it is like", seems in principle inaccessible to an objective, third-person approach. It follows that consciousness cannot be described completely objectively.

However, this argument is based on the same level confusion as the problems discussed earlier. The question "What is it like?" refers to the phenomenal dimension of experience, i.e. to how a state feels for the experiencing system. This phenomenal dimension is not an ontological property in addition to neuronal processes, but the epistemic perspective of the system on its own causal organization.

The reconstructivist method resolves the what-is-it-like problem by distinguishing and reconstructing three dimensions of experience: the phenomenal, the semantic and the subjective dimension. These dimensions are not separate ontological domains, but functional aspects of autocatalytic organization with causal collapse.

2.3.1 The phenomenal dimension of experience

The phenomenal dimension concerns the qualitative "how" of experience, the qualia. Why does red feel the way it does? Why is pain qualitatively different from pleasure? These questions seem insoluble as long as qualia are regarded as intrinsic properties that cannot be further analyzed. However, the reconstructivist method shows that phenomenal qualities follow from the structure of causal collapse.

Phenomenal experience arises when a system has non-factorizable causal relationships. A stimulus is not merely processed, but becomes part of an integrated causal core that encompasses the entire system. This nucleus cannot be broken down into independent sub-processes without destroying its causal structure. For the system itself, this non-factorizability manifests itself as phenomenal experience (Stegemann 2025a).

The quality of the experience (why red is red and not blue) follows from the specific causal structure of the collapse. Different stimuli (light of a certain wavelength, mechanical pressure, chemical signals) activate different neuronal pathways and thus generate different configurations of the causal nucleus. These configurations are what is phenomenally experienced as different qualities. There is no additional property "redness" beyond the specific causal configuration triggered by red light.

This is empirically supported by findings on the neuronal signature of qualia. Different qualia correlate with different patterns of neuronal activity (Haynes & Rees 2006; Huth et al. 2012). These patterns are not arbitrary, but reflect the causal structure by which sensory information is integrated into the global collapse core. The phenomenal quality is thus the intrinsic manifestation of this causal integration for the system itself.

2.3.2 The semantic dimension of experience

Semantics do not arise through symbolic assignment or external reference, but intrinsically in the interior through affective valence. As soon as an autocatalytic system has a causal collapse, i.e. a non-factorizable, integrated representation of stimuli, each sensory activity is not merely processed, but evaluated as relevant or irrelevant, beneficial or harmful to the maintenance of the organization. This evaluation is not an afterthought, but a structural property of the interior: valence is the primary form of meaning (Stegemann 2025b).

In this sense, semantics is the functional signature of self-preservation. A stimulus "means" something because it supports or threatens the autocatalytic dynamics, not because it refers to an external object. This intrinsic semantics solves the problem of symbol grounding (Harnad 1990): meaning is not based on syntax, but on autocatalytic valence, and thus arises in the interior before it is communicated in the outside world.

Panksepp (1998) shows that primary affective systems are phylogenetically old and subcortically anchored. These systems generate basal valences (fear, lust, anger) that act as the ursemantic core of consciousness. Damasio (1999, 2010) similarly argues that feelings form the basis for more complex mental processes. The model of causal collapse integrates these findings: Affective valences are not epiphenomenal, but constitute the semantic dimension of experience.

Only on this basis can more complex, associative or linguistic forms of meaning emerge. The phenomenal quality of pain or pleasure is therefore not epiphenomenal, but the ursemantic core of consciousness: the immediate givenness of relevance to the system itself. This is the semantic dimension of experience.

2.3.3 The subjective dimension of experience

But what makes a physical system a subject? The answer lies in the causal nucleus, which originated with the cell nucleus. This core is not just another structure in the system, but the functional reference point through which all causal relationships in the system are organized. It is what valences refer to, what integration refers to, and what experience refers to. The causal nucleus personalizes a physical system (Stegemann 2025e, 2025h).

Without it, there would be causal processes, but no point of reference for which anything would be significant, no place from which something would be experienced. With the causal core, the difference between "here" and "there", between "for me" and "for something else" arises. The entire architecture of subjectivity, from the primitive "I" to reflexive self-consciousness, is built on top of this causal core. It is the point of reference that transforms a system from a mere mechanism to a subject.

The hard problem dissolves as soon as one realizes: Experiencing presupposes not only non-factorizability, but a causal core as a point of reference, for which this non-factorizable space is its space in the first place. This is the subjective dimension of experience. The causal nucleus did not arise arbitrarily, but as a structural consequence of thermodynamic flows between the prokaryotic cell and the environment. Local structural densification created a centralist causal principle that was selected evolutionarily. This condensation acted as a functional attractor for the further organization of cellular processes (see Chapter 5).

The three dimensions of experience (phenomenal, semantic, subjective) are thus not mysterious properties, but functional aspects of autocatalytic organization with causal collapse. The what-is-it-like problem resolves as soon as it is realized that "what it is like" does not require an additional ontology, but represents the intrinsic manifestation of causal entanglement for a system with a causal core.

3. The Reconstructivist Method

3.1 Methodological principles

3.1.1 From correlations to functional necessities

The decisive methodological advantage over the correlative method lies in the fact that reconstructionism does not label arbitrarily, but identifies conditions. Correlative research can refer to arbitrary neuronal patterns as correlates of consciousness without having to explain why these patterns and not others are associated with consciousness. If, for example, the Perturbational Complexity Index (PCI) correlates with consciousness (Casali et al. 2013), it remains unclear in the correlative method whether high causal complexity is characteristic of consciousness or is only contingent with it.

The reconstructivist method can answer this question: it identifies causal integration as a necessary property of autocatalytic processes that enable consciousness. The PCI measures exactly this causal integration. The correlation is thus not only observed, but explained. The reconstructivist method transforms arbitrary correlations into functional necessities.

While correlative neuroscience treats PCI merely as a statistical correlate, reconstructionism explains why high causal complexity goes hand in hand with consciousness: Only when a system shows a non-separable, spatio-temporally coherent reaction after perturbation is an operationally closed interior space given, and thus experiential is possible. A low PCI, on the other hand, indicates factorizable, locally limited reactions that do not produce a causal collapse and thus a phenomenal interior space.

Thus, the PCI does not become an arbitrary marker, but a functional indicator of the condition under which consciousness arises. Operationalization does not take place through arbitrary assignment, but through reconstructive deduction: the necessity of causal integration is derived from the phenomenal fact of consciousness, and this is measured by the PCI.

3.1.2 Reconstruction instead of reduction

The reconstructivist method is neither reductionist nor anti-reductionist in the usual sense. Reductionism seeks to reduce complex phenomena to simpler components (Churchland 1986). Anti-reductionism claims that such reduction is impossible in principle (Kim 2005; Chalmers 1996). However, both positions miss the core:

reconstruction does not ask what something consists of or whether it can be reduced to something, but what functional conditions must be fulfilled for a phenomenon to exist.

An example illustrates the difference. Reductionism would ask: What are the neuronal components of consciousness? Anti-reductionism would answer: consciousness cannot be reduced to neuronal components. Reconstructivism asks: What functional organization must a neural system have in the presence of consciousness? The answer is: autocatalytic organization with electrical signal transduction, causal nucleus, and non-factorizable integration.

This answer is neither a reduction (consciousness is not "reduced" to neurons) nor an emergence assertion (consciousness does not magically "emerge" from complexity), but a functional reconstruction: the necessary structural conditions are identified, without which consciousness cannot exist. These conditions are not arbitrary, but follow from the analysis of what consciousness does functionally.

3.1.3 Differentiation between Correlative Research and Evolutionary Narratives

Correlative neuroscience collects valuable data on neuronal correlates of consciousness (Crick & Koch 1990; Koch et al. 2016), but stops at the description of relationships without explaining the functional necessity of these relationships. The search for NCCs (neural correlates of consciousness) has yielded important findings, for example on the role of posterior cortical areas (Boly et al. 2017) or on thalamo-cortical loops (Edelman & Tononi 2000). But a correlation, as precise as it may be, does not answer the question of why this neural configuration in particular goes hand in hand with consciousness.

Evolutionary approaches provide historical narratives about the phylogenesis of consciousness (Feinberg & Mallatt 2016; Ginsburg & Jablonka 2019), however, miss the timeless conditions of the possibility of consciousness. They can hypothesize that consciousness arose in the Cambrian period in early vertebrates, or that it is already rudimentary in Cnidarians. These historical questions are interesting, but they do not answer what structural properties a system must have in order for consciousness to be possible at all. The historical contingency of evolutionary developments is confused with the structural necessity of functional conditions.

Analytical methods break down the brain into components and examine them in isolation (Koch 2004), but do not capture the emergent organization. The decomposition into neurons, synapses, neurotransmitters makes sense for many questions, but misses the decisive point: consciousness results precisely from the non-factorizable entanglement of these components, which is destroyed by analytical decomposition.

Axiomatic theories such as Integrated Information Theory (Tononi et al. 2016; Oizumi et al. 2014) postulate principles such as the integration and differentiation of information without clarifying whether these principles represent necessary, merely sufficient or even contingent conditions. IIT begins with axioms about the phenomenology of consciousness and derives mathematical structures from them. But these axioms are not self-evident, and their selection remains underdetermined. Why these five axioms (existence, composition, information, integration, exclusion) and not others? The reconstructivist method avoids this underdetermination by starting from the functional conditions of autocatalytic systems.

The reconstructivist method avoids the labeling problem of the correlative method, the historical contingency of evolutionary narratives, and the underdetermination of axiomatic approaches. Above all, it avoids the categorical confusion between levels of description and ontology. It asks: What functional organization does a living system have when consciousness is present? And derives testable conditions from this.

3.2 Epistemological prerequisites

3.2.1 Structural ontology

The reconstructivist method presupposes a structural ontology that does not ask about substances but about organizational principles. Classical ontology looks for fundamental building blocks: atoms, particles, fields. A structural ontology, on the other hand, references patterns of causal organization: autocatalytic networks, recursive loops, operational closure (Kauffman 1993; Rosen 1991; Maturana & Varela 1980).

In this sense, life is not a thing, but a specific causal organization. A cell is made up of molecules, but what makes it a living cell is not the chemical composition, but the autocatalytic cross-linking through which these molecules generate each other and maintain operational coherence. Consciousness analogous: It is not a substance and not a property of a substance, but a specific causal configuration of autocatalytic systems (Stegemann 2025a).

This perspective is not new. Whitehead (1929) already developed a process ontology that understands events instead of substances as fundamental. Rosen (1991) showed that living systems are distinguished from mechanical systems by their organizational coherence. What the reconstructivist method adds is the systematic application of this insight to consciousness: consciousness is reconstructed as a specific organization of autocatalytic processes, not as an emergent property of arbitrary complex systems.

3.2.2 Autocatalytic Organization as a Basic Principle

Autocatalytic organization is the central principle from which the reconstruction is based. An autocatalytic system is one whose components, through their interaction, create the network that produces these components (Kauffman 1993; Morowitz 1992). This circular causality distinguishes living from non-living systems. A stone is made up of atoms, but these atoms do not create the stone. A cell is made up of molecules, and these molecules, through their interaction, create the cell that contains it.

This circularity is not a logical one, but a causal one: the components create the whole, and the whole makes the components possible. Maturana and Varela (1980) coined the term "autopoiesis" for this, although this term is not without problems (see Chapter 4.2). The decisive factor is that autocatalytic organization creates an operationally closed interior. For such a system, there is a difference between inside and outside, between what contributes to self-preservation and what threatens it. This difference is the origin of valence and therefore important.

Consciousness presupposes this autocatalytic organization. No system without self-preservation dynamics can be conscious, because there is no point of reference for which something could be significant. This thesis excludes panpsychist positions that consider consciousness to be a universal property of all matter (Chalmers 1996; Goff et al. 2017). Electrons are not conscious because they do not have an autocatalytic organization. Consciousness is specifically bound to life, not to arbitrary physical processes (Stegemann 2025d).

3.2.3 Non-teleological reasoning

The reconstructivist method argues strictly non-teleologically. It does not ask what consciousness is for or what evolutionary advantage it brings, but what structural conditions exist when consciousness is given. This distinction is crucial. Teleological explanations (consciousness arose to enable X) presuppose that evolutionary selection acted on consciousness as such. However, this is problematic, because selection affects behavior, not experience.

An organism that acts functionally identical to a conscious organism but is not conscious (a philosophical zombie, Chalmers 1996) would have the same selection advantage. Selection cannot distinguish between the two. It does not follow that consciousness is evolutionarily irrelevant, but that its evolutionary explanation must be different: consciousness does not arise because it has been selected, but as an intrinsic manifestation of autocatalytic organization as soon as it crosses a certain structural threshold.

Instead, the non-teleological argumentation reconstructs: What is given is an autocatalytic organization. This requires self-preservation. Self-preservation requires valence (differentiation between beneficial and harmful). Valence requires an operationally closed interior. This interior space is the prerequisite for something to be given for the system itself. As soon as electrical signal transduction is added and forms a causal core that transitions into non-factorizable entanglement, consciousness arises. No step of this reconstruction is teleological. Each follows from structural necessities.

4. Autocatalysis as the biological foundation of consciousness

4.1 Autocatalysis: Definition and thermodynamic basics

4.1.1 Self-sustaining chemical reaction networks

Autocatalysis refers to a process in which the products of a reaction catalyze the reaction itself. In its simplest form: A catalyzes the formation of more A. In more complex forms, autocatalytic sets emerge: networks of reactions in which each reaction is catalyzed by products of other reactions in the set (Kauffman 1993). The network as a whole reproduces itself.

Kauffman (1993) showed that autocatalytic sets arise spontaneously in sufficiently complex chemical systems. From a critical diversity of molecules, closed catalytic cycles are highly likely to form. These sets are thermodynamically open (they require energy supply and material exchange processes with the environment), but operationally closed (the reactions within the set catalyze each other without external catalysts).

A historical example is the formose cycle, in which formaldehyde reacts autocatalytically to sugars. A biological example is the Krebs cycle, where oxaloacetate is regenerated at the end and thus perpetuates the cycle. The crucial point: Autocatalytic processes are self-sustaining. Once started, they perpetuate as long as energy and material are available.

This self-preservation is not just a chemical curiosity, but the fundamental property that distinguishes life from non-life. A crystal grows, but does not sustain itself in the thermodynamic sense. A flame is dissipative, but not autocatalytic. A cell, on the other hand, continuously regenerates its own components through autocatalytic networks. Without autocatalysis, there is no life.

4.1.2 Thermodynamic Fluxes and Structure Formation

Living systems exist far from thermodynamic equilibrium. The second law of thermodynamics states that isolated systems tend to the state of maximum entropy. Life, however, seems to violate this tendency: organisms are highly organized, low-entropic. Schrödinger (1944) solved this paradox by recognizing that organisms are open systems that continuously absorb negentropy from the environment (through food, light) and release entropy (through heat, waste products).

Prigogine (1980) formalized this insight in the theory of dissipative structures. Far from equilibrium, systems can spontaneously form order, provided that there is a sufficient thermodynamic flow. Examples are Bénard cells (convective flow patterns in heated liquids) or Belousov-Zhabotinsky reactions (oscillating chemical systems). These structures are not in equilibrium, but are maintained by continuous energy dissipation.

Autocatalytic systems are a special class of dissipative structures. They form not only order, but self-sustaining order. The thermodynamic flow (energy supply) is used to maintain the catalytic cycles that constitute the system. Morowitz (1992) showed that autocatalytic organization is thermodynamically favored: it maximizes entropy production in the universe by efficiently using energy gradients.

The emergence of life can therefore be understood as a thermodynamically driven process. As soon as autocatalytic cycles are created, they stabilize themselves and are further optimized by selection. The first cell was not a miracle, but the thermodynamically expected consequence of sufficiently complex chemical systems under energy flux (Kauffman 1993; Morowitz 1992).

4.1.3 Operationally closed, materially open systems

Autocatalytic systems are operationally closed, but materially and energetically open. This combination is crucial. Operational coherence means that the processes of the system are made possible by the system itself, not by external factors. A cell produces its own enzymes, membranes, nucleic acids through internal autocatalytic processes. No external actor has to manufacture these components.

At the same time, the system is materially and energetically open: it absorbs molecules and energy from the environment and releases waste products. Without this material and energetic exchange, the system would come to a standstill. The combination of operational closure and material openness defines autocatalytic organization (Maturana & Varela 1980; Rosen 1991).

This structure has a fundamental consequence: it creates a functional perspective. For an operationally closed system, there is a difference between what helps maintain the organization and what disrupts or destroys it. This difference is the origin of valence. Glucose is "good" for a cell (contributes to energy supply), cyanide is "bad" (blocks the respiratory chain). These evaluations are not anthropomorphic projections, but objective properties of functional organization: glucose supports autocatalysis, cyanide interrupts it.

With this functional perspective, an interior space is created. The system doesn't just have a structure, but a structure for itself. There is a difference between what is "inside" (part of the autocatalytic organization) and what is "outside" (environment). This difference is not spatial-geometric (although it often manifests itself spatially, for example through membranes), but functional: inside is what contributes to self-preservation.

This interior is the prerequisite for any form of experience. Without it, there can be no perspective for which something could be given. A stone has no interior because it is not operationally closed. A cell has an interior because it is organized autocatalytically. This interior is not yet conscious experience, but it is the necessary condition for conscious experience to arise later (Stegemann 2025d).

4.2 From autocatalysis to autopoiesis

4.2.1 Maturana and Varela: Autopoiesis as an Organization of the Living

Maturana and Varela (1980) coined the term "autopoiesis" (self-creation) to characterize the specific organization of living systems. An autopoietic system is defined as a network of processes that produce the components that realize the network and at the same time constitute the boundary of the system. Operational coherence is key: the system produces itself, without external instructions or blueprints.

This characterization captures important aspects: Living systems are not merely self-organizing (like crystals or flames), but self-generating. They produce not only structures, but the structures that constitute them themselves. A cell produces its membrane, and these membranes constitute the cell. An organism produces its cells, and these cells constitute the organism (in multicellular systems).

Maturana and Varela closely associated autopoiesis with cognition. They argued that every living system is cognitive because it does not passively reflect its environment, but is actively constructed through the way it interacts with it. This enactivism (Varela et al. 1991) emphasizes the inseparability of life and cognition: life is cognition, cognition is life.

4.2.2 Autopoiesis and Autocatalysis: Metatheoretical Concept and Concrete Realization

The terms autopoiesis and autocatalysis refer to different levels of description of the same phenomenon. Autopoiesis is a meta-theoretical concept: it formulates the general organizational structure of living systems as operational coherence with simultaneous material openness. This conceptual characterization abstracts from concrete chemical or physical processes and describes the functional logic of living things.

Autocatalysis, on the other hand, refers to the concrete, operationalizable realization of this organization: chemical reaction networks in which products act as catalysts for their own creation. Autocatalytic sets (Kauffman 1993) are empirically identifiable, experimentally examinable, mathematically modelable. The Krebs cycle, the Calvin cycle, metabolic networks are measurable autocatalytic processes.

The relation between the two is not hierarchical (one more fundamental than the other), but epistemological: autopoiesis provides the conceptual framework, autocatalysis the empirical substance. You could say: autopoiesis describes what life does (produce itself), autocatalysis describes how it does that (through catalytic cycles). Both perspectives are complementary and necessary for a full understanding.

This distinction is relevant for consciousness theory: The metatheoretical level (operative closure creates a functional inner space) explains why a reference point can arise in the first place for which something is significant. The concrete level (autocatalytic networks in cells, recurrent neuronal dynamics) explains how this interior space is actually realized. Consciousness presupposes both levels: the metatheoretical structure (inner space, valence, self-referentiality) and its concrete realization (nervous systems with causal entanglement).

Thompson (2007) argues in this sense that consciousness cannot be separated from life. Autopoietic organization is the prerequisite for the emergence of conscious experience later, with sufficiently complex neuronal organization. Without the operational coherence of living systems, there would be no interior, and without an interior there would be no perspective from which or for which something could be experienced (Stegemann 2025d).

4.3 Autocatalysis and the formation of interiors

4.3.1 Operational Unity Creates Functional Perspective

The operational closure of autocatalytic systems has a fundamental consequence: it creates a functional perspective. For an operationally closed system, there is a difference between what helps maintain the organization and what disrupts or destroys it. This difference is not subjective (in the sense of consciously experienced), but objective: it follows from the functional structure of the system.

Thompson (2007) calls this "sense-making": the system makes sense of its environment by evaluating stimuli in terms of their relevance for self-preservation. This evaluation is not optional or retrospective, but constitutive of autocatalytic organization. A system that does not distinguish between beneficial and harmful cannot be maintained.

This functional perspective is the origin of normativity. There is a target for an autocatalytic system: it should be maintained. This "should" is not moral, but functional. If the cell needs glucose, then it should absorb glucose, not in the sense of a moral obligation, but in the sense of a functional necessity of its organization (Di Paolo 2005).

4.3.2 The interior as a consequence of recursive causality

Operational coherence creates an interior space. This interior is not spatially geometrical (although it often manifests itself spatially), but functionally defined: inside is what belongs to the autocatalytic organization, outside is what does not belong to it. This difference between inside and outside is the basis for every form of subjectivity.

The interior is characterized by recursive causality: the processes in the interior have an effect on themselves. A influences B, B influences C, C influences A. This circularity distinguishes interiors from linear causal chains. A river has no recursive causality: water flows from source to mouth without feedback. A cell has recursive causality: each metabolic cycle has an effect on itself.

This recursive causality creates a form of self-referentiality. The system refers to itself, not in the sense of conscious reflection, but in the sense of functional feedback. This self-referentiality is the prerequisite for conscious self-reflection to arise later. Without functional self-referentiality (recursive causality), there can be no phenomenal self-referentiality (conscious ego) (Stegemann 2025c).

4.3.3 Self-preservation as primary valence

The functional perspective of autocatalytic systems manifests itself as primary valence: the distinction between beneficial and detrimental to self-preservation. This valence is not epiphenomenal, but constitutive. It is the primary form of meaning long before symbolic or linguistic meaning arises.

Panksepp (1998) identified primary affective systems (fear, lust, anger, separation anxiety, care, play, search) that are phylogenetically old and subcortically located. These systems encode basal valences: fear signals a threat to self-preservation, pleasure signals encouragement. These affective systems are not retrospective evaluations, but direct manifestations of the autocatalytic self-preservation dynamic.

Damasio (1999, 2010) similarly argues that emotions and feelings represent the connection between body and consciousness. The "core self" arises from the continuous representation of the body state. Feelings are not merely subjective colorations of objective processes, but the way in which the self-preservation dynamic is given to itself.

The model of causal collapse integrates these insights: valence is the primary form of meaning that follows from autocatalytic organization. As soon as a system is operationally closed, there is meaning for this system: stimuli mean beneficial or harmful. This semantic basis is the prerequisite for all more complex forms of importance (Stegemann 2025i).

4.4 Consciousness presupposes life, not the other way around

4.4.1 Why the hard problem starts with dead matter

The Hard Problem (Chalmers 1995) asks how experience arises from physical processes. This question implicitly presupposes that physical processes can be adequately described without reference to life. A stone, an electron, a computer are considered equally physical. When asked why some physical systems are conscious and others are not, it seems like an unsolvable mystery.

The solution lies in the realization that not all physical systems are the same. Autocatalytic organization is a specific form of physical organization that enables new properties: operational coherence, functional perspective, interior. These properties are not mysterious, but follow from the thermodynamic and causal structure of autocatalytic systems.

The hard problem dissolves as soon as one realizes: The question is not "Why are some physical systems conscious?", but "Which living systems are conscious, and under what conditions?" This reformulation shifts the explanatory burden: instead of bridging a metaphysical gap between physics and phenomenology, we reconstruct the functional conditions that autocatalytic systems must meet in order to be conscious.

4.4.2 Autocatalytic Organization as a Necessary Condition

The model postulates: Autocatalytic organization is a necessary condition for consciousness. No system without self-preservation dynamics can be conscious. This thesis has testable consequences. It excludes: panpsychism (electrons have no autocatalytic organization, i.e. no consciousness), classical AI systems (computers have no self-preservation, i.e. no consciousness in the phenomenal sense), physical processes without life (stars, planets, weather systems have no autocatalytic organization, i.e. no consciousness).

The thesis is more precise: consciousness presupposes operationally closed, autocatalytic organization. This prerequisite is not sufficient (not all living systems are conscious), but it is necessary. A cell has autocatalytic organization and a functional interior, but presumably no conscious experience. Only when additional conditions are met (electrical signal transduction, causal core, non-factorizable integration) does consciousness arise.

This hierarchy of necessary conditions allows consciousness to be precisely located in nature: it is neither ubiquitous (panpsychism) nor limited to humans (anthropocentric exceptionalism), but specifically bound to complex autocatalytic systems with certain additional properties.

4.4.3 From Minimal Life Forms to Conscious Experience

Phylogenetic development shows a hierarchy of autocatalytic organization. Minimal life forms (such as bacteria) have autocatalytic metabolisms and functional interiors, but probably no conscious experience. More complex organisms (such as worms with simple nervous systems) have additional electrical integration, but presumably not yet full causal collapse. Vertebrates with sufficiently integrated brains (fish, amphibians upwards) reach the threshold of conscious experience (Feinberg & Mallatt 2016; Merker 2007).

This hierarchy is not sharply defined. There is no binary boundary between conscious and unconscious, but a structured graduality (see chapter 8.4). But the basic statement remains: consciousness is built on autocatalytic organization. Without life, there is no consciousness. This insight is fundamental to the reconstructivist method.

Empirical verifiability: This thesis can be tested indirectly. If it could be shown that systems without an autocatalytic organization (such as purely physical processes or classical computers) have consciousness, the model would be falsified. Conversely, any finding that consciousness only occurs in living, self-sustaining systems supports the thesis. So far, there is no convincing evidence of consciousness outside living systems.

5. The Causal Core as a Structural Principle

5.1 Origin of the causal nucleus: From prokaryote to eukaryote

5.1.1 The cell nucleus as structural compaction

The evolution of life took a decisive turn with the emergence of the eukaryotic cell nucleus about 1.5 to 2 billion years ago. Prokaryotic cells (bacteria and archaea) have distributed their genetic information in the cytoplasm. Eukaryotic cells, on the other hand, concentrate their DNA in a membrane-enclosed nucleus. This structural centralization is not merely an anatomical variation, but has fundamental functional consequences.

The cell nucleus acts as an informational and regulatory hub. Transcription takes place in the nucleus, translation in the cytoplasm. This spatial separation enables additional regulatory levels (RNA editing, alternative splicing) that are not available to prokaryotic cells. The nucleus is thus not just a repository for DNA, but an active control center that coordinates cellular dynamics (Lane & Martin 2010).

This centralization increases causal complexity: instead of locally distributed control, a hierarchical principle is created. The nucleus integrates signals from the cytoplasm, processes them, and returns coordinated instructions. This causal hierarchy is a first step towards the causal nucleus, which later plays a central role in nervous systems (Stegemann 2025e).

5.1.2 Thermodynamic conditions of nuclear formation

The formation of the cell nucleus correlates with specific thermodynamic boundary conditions. Eukaryotic cells are typically 10-100 times larger than prokaryotic ones. As cell volume increases, the surface-to-volume ratio deteriorates: the diffusion of molecules over longer distances becomes inefficient. Larger cells without internal compartmentalization show diffusion limitation, which slows down metabolic processes.

The endosymbiont theory (Margulis 1970) describes the formation of mitochondria and chloroplasts as the internalization of prokaryotic organisms. Analogous processes could have led to the nuclear membrane: invaginations of the plasma membrane, which enclosed DNA. This compartmentalization altered the diffusion dynamics and allowed spatially separated levels of regulation (Martin & Koonin 2006).

From a thermodynamic point of view, the core is a local structural compression with a continuous flow of energy. The maintenance of the nuclear membrane, the active transport through nuclear pores, the chromatin organization require continuous ATP hydrolysis. Systems with such compartmentalization show measurably higher rates of gene expression regulation and metabolic throughput than comparable systems without compartmentalization.

The thermodynamic analysis shows that under conditions of sufficient energy flow, structural condensations occur spontaneously if they increase the entropy production of the overall system. The nucleus is an example of such dissipative structure formation (Prigogine 1980). There is no thermodynamic teleology (the core was not created "in order to"), but structural consequence: given thermodynamic boundary conditions lead to specific forms of organization (Lane 2015).

5.1.3 Causal Nuclei as Structural Attractors in Evolution

Organisms with centralized causal structures show higher survival and diversification rates in fossil records and phylogenetic analyses. This is not a selection based on a "principle", but a statistical consequence of functional properties. Systems with efficient coordination between subsystems react more quickly to environmental changes, use resources more effectively, and avoid damage more precisely.

After the formation of the cell nucleus in eukaryotic lines, a cascade of structural innovations followed: multicellularity (multiple cell nuclei in coordinated organization), specialized tissues, and finally nervous systems with central ganglia. Each stage has the same structural property: local condensation of causal relationships, which acts as an integration point for the overall system.

This recurrence is not teleology (evolution does not strive for central nuclei), but convergence: different phylogenetic lineages develop similar solutions under similar selection pressures. Arthropods, mollusks, and vertebrates have developed centralized nervous systems independently of each other. The common structural property is the causal core, not because evolution "strives for it", but because physical systems converge to such structures under specific functional requirements (rapid coordination over distance).

The causal core thus establishes itself as a structural attractor: Under given thermodynamic and functional boundary conditions, such structures are more likely to emerge than alternative organizations. This explains their phylogenetic persistence without teleological assumptions. The cell nucleus "did not emerge to enable nervous systems later", but established an organizational principle that was realized under different boundary conditions at different levels (Stegemann 2025e).

5.2 The causal nucleus as a functional reference point

5.2.1 Centralization of causal relationships

A causal nucleus is not just an anatomical structure, but a functional reference point through which causal relationships in the system are centralized. In the cell, this is the nucleus that controls gene expression. In early nervous systems, these are ganglia that integrate sensory inputs and coordinate motor outputs. In complex brains, these are central networks such as the Default Mode Network (DMN), which bring together information from multiple sources.

Centralization has a crucial property: it creates non-factorizable causal relationships. In a system without a core, sub-processes can run independently. In a system with a core, all processes run through the core, causing them to be causally entangled. This entanglement is the prerequisite for the causal collapse that constitutes consciousness (Stegemann 2025a).

Mathematically, this can be grasped by network theory. A decentralized network has many local connections, but few global ones. A network with causal cores has hub structures: nodes with a large number of connections that integrate the network globally (Sporns 2011; Bullmore & Sporns 2009). The human brain has such hubs, especially in posterior medial structures (precuneus, posterior cingulate cortex) and in the prefrontal cortex. These hubs act as causal nuclei (Raichle 2015).

5.2.2 Valence orientation and causal nucleus

The causal nucleus is the reference point to which valences refer. In Chapter 4, it was shown that autocatalytic systems produce primary valences: beneficial versus detrimental to self-preservation. These valences are not abstract, but concrete: they refer to the functional state of the system. Without a point of reference, there would be no coherent assessment.

The nucleus is this reference point for the cell. Damage to the DNA in the nucleus threatens the cell more fundamentally than damage to individual proteins in the cytoplasm. The cell does not consciously "know" this, but its repair mechanisms prioritize nuclear integrity. This prioritization follows from the functional centrality of the core.

This principle is repeated in nervous systems. Central ganglia or brain structures are protected by multiple redundancies (blood-brain barrier, astrocytic support, metabolic buffering). This protection is not arbitrary, but follows from the functional role as the causal core. Damage to central structures has more drastic consequences than damage to peripheral areas.

This connection is crucial for consciousness: Conscious experience is always full of valence (Panksepp 1998; Damasio 1999). Even seemingly neutral perceptions have an affective tint. This tint results from the fact that the conscious state (the causal collapse) encompasses the causal core, and this nucleus is the reference point of all valences. Experience is therefore necessarily valence-oriented, because it refers to a core for which self-preservation is important (Stegemann 2025i).

5.2.3 From Physical System to Subject

What makes a physical system a subject? The model's answer is: A physical system becomes a subject when it has a causal core that serves as a functional reference point for valences and integration. Without this core, there are causal processes, but no point of reference for which these processes would be significant.

A stone is a physical system without a core. Its atoms interact causally, but there is no integrative point of reference. The stone has no perspective, no interior, no "for whom" something could be. A cell has a nucleus. There is a perspective for them: glucose is beneficial, cyanide is harmful. This perspective is not yet conscious, but it is the origin of subjectivity.

An organism with a nervous system has a more complex core. For him, there is not only metabolic valence (beneficial/harmful), but also sensory integration, motor coordination, and, with sufficient complexity, conscious experience. The causal core is the place where all these processes converge. He personalizes the system (Stegemann 2025e, 2025h).

Subjectivity is thus not a mysterious property that is added to physical systems, but a functional consequence of centralized causal organization. The causal nucleus transforms a physical system from a mere mechanism to a subject, because it creates a point of reference from which and for which something can be given.

5.3 The causal nucleus in the evolution of the nervous system

5.3.1 Phylogenetic development of nervous nuclei

The evolution of the nervous system shows a progressive centralization of causal nuclei. The earliest multicellular animals (sponges) have no nerve cells. Cnidaria (jellyfish, corals) have diffuse neural networks without central ganglia. The first centralization can be seen in bilateria: worms developed paired ganglia at the anterior end (cephalization) that integrate sensory inputs from chemoreceptors and primitive photoreceptors.

These early ganglia are the first nervous causal nuclei. They do not centralize all nervous activity (many reflexes run locally), but they coordinate behavior by integrating multiple inputs. A worm can locate food, react to light, evade danger, all coordinated by gangliary integration (Arendt et al. 2016).

In arthropods (insects, spiders, crabs), the brain develops from fused ganglia to a more complex center. The insect brain has specialized regions (mushroom bodies for learning, optical lobes for vision), but functions as an integrated nucleus. Bees, for example, can perform complex navigation, social communication and memory, all through a brain of less than one milligram (Menzel 2012).

Vertebrates go through a different evolutionary trajectory. The vertebrate brain develops from three precursors (rhombencephalon, mesencephalon, prosencephalon), which are already distinguishable in fish. The rhombencephalon (hindbrain) is the earliest integrative nucleus, responsible for basal functions such as breathing, heartbeat, balance. The mesencephalon (midbrain) coordinates sensory and motor processes. The prosencephalon (forebrain) becomes a site of higher integration in mammals (Nieuwenhuys et al. 1998).

5.3.2 From rhombencephalon to cortical integration

Merker (2007) argues convincingly that primary consciousness does not require cortex, but is located in the upper brainstem. Animals without a cortex (fish, amphibians) show clear signs of awareness: goal-directed behavior, learning, emotions. Even human children with hydranencephaly (without a functioning cortex) show basal forms of wakefulness and reactivity.

The rhombencephalon and especially the tectum (in the mesencephalon) act as the primary causal nucleus. The Tectum integrates visual, auditory and somatosensory inputs in a multimodal map and coordinates orientation responses. This integration is sufficient for primary consciousness: the animal experiences its environment, reacts valence-driven, shows intentionality (Feinberg & Mallatt 2016).

In mammals, the cortex is added. The cortex increasingly takes on functions that remain subcortical in other vertebrates. However, subcortical integration is also essential in mammals. Panksepp (1998) showed that primary emotions are generated subcortically (amygdala, hypothalamus, PAG). The cortex modulates and elaborates, but does not produce fundamentally.

Cortical integration reaches its peak in the Default Mode Network (DMN). The DMN is a network that includes medial prefrontal areas (mPFC), posterior cingulate cortex (PCC), precuneus, inferior parietal cortex, and medial temporal structures. It is the causal core of human consciousness (Raichle 2015; Buckner et al. 2008).

5.3.3 The Default Mode Network as a Causal Core in the Human Brain

The default mode network was first identified by fMRI studies as a network that is highly active at rest and is deactivated during task-oriented activities (Raichle et al. 2001). This inverse activation led to speculation about the function: Is the DMN just noise? Raichle (2015) showed that the DMN is fundamental for self-referentiality, autobiographical memory, future projection and theory of mind.

Anatomically, the DMN is an integration hub. The PCC has more functional connections than any other cortical region (Hagmann et al. 2008). The mPFC integrates emotional, sensory and cognitive information. These hubs act as causal cores: they centralize information flows and enable global coordination (Sporns 2011).

Functionally, the DMN is the place where self-representation takes place. When people think about themselves, the DMN is activated. When they think about others (Theory of Mind), too. When they remember or imagine the future, the DMN again. This consistent activation in self-referential processes shows that the DMN is the causal core of the self (Andrews-Hanna et al. 2014; Spreng et al. 2009).

Pathologically, the centrality of the DMN is shown by the fact that damage or dysfunctions have drastic consequences. In Alzheimer's disease, DMN is affected early and correlates with loss of autobiographical memory (Buckner et al. 2008). Schizophrenia shows abnormal DMN connectivity, correlated with impaired self-perception (Whitfield-Gabrieli & Ford 2012). In coma and minimally conscious states, DMN connectivity breaks down (Laureys et al. 2006).

The model interprets the DMN as the causal core of human consciousness. It is the functional point of reference through which integration takes place, to which self-representation refers, and which organizes the valence orientation of experience. When the DMN enters a state of causal collapse (see chapter 7), conscious self-experience arises (Stegemann 2025b).

5.4 Causal Core and Subjectivity

5.4.1 The difference between "here" and "there"

The causal nucleus creates a fundamental spatial difference: between "here" (where the nucleus is) and "there" (where it is not). This difference is not merely geometric, but functional. "Here" is the place from which the system experiences the world, to which valences refer, where integration takes place. "There" is the environment to which "here" refers.

This difference is the basis of spatial perspective. A system without a core has no perspective: there is no privileged "here". A system with a core necessarily has a perspective: the world is given from a point of view. This point of view cannot be moved arbitrarily because it is tied to the functional core.

In organisms with nervous systems, this difference manifests itself as a body schema. The body is experienced not merely as an object among objects, but as the place from which objects are experienced. This body schema is represented in the brainstem and cortex, centered around the causal nucleus (Damasio 1999; Gallagher 2005).

For consciousness, it follows that conscious experience is always perspectival. There is no experience "from nowhere", but always experience from a point of view. This point of view is the causal core. The perspectivity of consciousness is thus not a contingent property, but a necessary consequence of the existence of a causal core (Stegemann 2025h).

5.4.2 Personalization of Physical Systems

The causal nucleus personalizes a physical system. Personalization means that the system does not only have a structure, but a structure for itself. There is a point of reference for which structure is significant. Without this point of reference, the system would be merely a mechanism, an object without subjectivity.

This personalization is gradual. A cell has a minimal causal nucleus (the nucleus) and a minimal form of personalization: there is metabolic valence for it. An animal with a nervous system has a more complex core and a more elaborate personalization: For it, there is sensory experience, emotional valences, goal-oriented behavior. A human being has the most complex core (DMN and related networks) and the most elaborate personalization: for him, there is self-reflection, narrative identity, moral responsibility.

This graduality does not contradict the assertion that the causal core constitutes subjectivity. Rather, it shows that subjectivity itself is gradual: from minimal functional perspective (cell) to primary experience (animals) to reflexive self-consciousness (humans). The principle remains constant: the causal core is the reference point that personalizes a system (Stegemann 2025e).

5.4.3 From Primitive "I" to Reflexive Self-Awareness

The architecture of subjectivity is built above the causal core. The most primitive "me" is the functional perspective of an autocatalytic system: "Here is the system, there is the environment, this is beneficial, that is harmful." This "I" is not reflexive, not narrative, not conscious in the phenomenal sense. But it is the origin of all later forms of self.

With nervous systems, an integrated "I" emerges: the system experiences itself as the center of sensory and motor processes. This "I" has a body schema, experiences emotions, shows intentionality. It is the "core self" (Damasio 1999), the primary consciousness (Edelman 1989), the primitive subject of experience.

With more complex cortices (especially in primates), reflexive self-confidence develops: the system can look at itself as an object, think about itself, locate itself in time. This reflexive self requires additional structures (DMN, prefrontal cortex) and develops ontogenetically late (from about 2-3 years of age in children). It is the "autobiographical self" (Damasio 2010), the higher consciousness, the narrative self (Gallagher 2000).

All of these forms of self are built on the causal core. Without it, there would be no point of reference for which anything could be given at all. With it emerges a hierarchy of self-forms, each more complex than the previous one, but all based on the same principle: the causal core as a functional reference point of subjectivity. This is the subjective dimension of experience (see chapter 2.3.3).

6. The nervous system: Electrical signal transduction and integration

6.1 Why neurons? The Functional Necessity of Electrical Signals

6.1.1 Superiority of electrical over chemical signal transduction

Living systems communicate through chemical signals: hormones, neurotransmitters, growth factors. This chemical signal transduction is effective for slow, diffuse processes (development, metabolism, immune response). But it is insufficient for fast, precise coordination. The diffusion of molecules is too slow for behavior control in fractions of a second.

Electrical signal transduction solves this problem. Nerve cells use ion gradients across membranes to generate action potentials that propagate at speeds of 1-100 m/s. This speed exceeds chemical diffusion by orders of magnitude. In addition, electrical signals are spatially precise: they run along defined axons to specific targets instead of diffusely distributing them in the tissue (Kandel et al. 2013).

The evolution of electrical signal transduction was not a whim of nature, but a functional necessity for coordinated behavior. Sessile organisms (plants, sponges) do not require fast signal transmission and do not have nerve cells. Mobile organisms (almost all animals) require rapid coordination between sensors (eyes, chemoreceptors) and effectors (muscles, glands) and developed nervous systems.

Cnidaria (jellyfish) show the earliest form: diffuse nerve networks without centers, but with electrical excitation. This is sufficient for simple contractions (swimming movements, food intake). Bilateria centralized these networks into ganglia and cords, which enables directional, goal-oriented behavior (Arendt et al. 2016; Jekely et al. 2015).

6.1.2 Speed, Precision, Coordination

The three functional advantages of electrical signal transduction are speed, precision and coordination. Speed: A visual stimulus reaches motor areas within milliseconds, enables flight reflexes, hunting behavior, orientation. Chemical signals take seconds to minutes for comparable distances.

Precision: Axons project to specific targets. Retinotopic mapping (each point of the retina is mapped to a corresponding point in the visual cortex) would be impossible due to chemical diffusion. Electrical wiring enables precise information transmission, which is the basis of spatial representation.

Coordination: Neural networks can generate complex temporal patterns (oscillations, phase synchronization, sequences). These patterns coordinate behavior: step patterning when walking, rhythmic breathing, precisely timed motor sequences. Chemical systems cannot achieve such temporal precisions (Buzsáki & Draguhn 2004).

These three benefits together explain why nervous systems are electrical. This insight is fundamental for consciousness: consciousness requires rapid integration of multiple inputs across spatial distances. Chemical signal transduction would be too slow and imprecise for the kind of global integration that constitutes conscious experience. Electrical signal transduction is thus a necessary condition for consciousness (Stegemann 2025d).

6.1.3 Neural Organization as an Evolutionary Solution

Neural organization solves the functional problem of integration at high speed and precision. Neurons are highly specialized cells: dendrite trees for input integration, soma for signal processing, axons for output conduction, synapses for signal transmission. This specialization enables efficient information processing.

Synaptic transmission is crucial. Synapses are the places where electrical signals are converted into chemical signals (at most synapses) and back. This conversion is not a detour, but functionally essential: Chemical transmission enables modulation, plasticity, amplification or attenuation. The combination of electrical conduction (fast, precise) with chemical transmission (modulable, plastic) is optimal (Südhof 2013).

The evolution of neuronal organization shows progressive complexity. Cnidaria have simple neurons without a clear polarity. Bilateria develop polarized neurons (input dendrites, output axons). Vertebrates develop myelination (increases conduction velocity), specialized neuron types (pyramidal cells, interneurons), complex network architectures (cortical layers, thalamo-cortical loops). Each stage increases the capacity for integration (Striedter 2005).

6.2 Neural Integration and Causal Complexity

6.2.1 Synaptic Connectivity

Neural integration is based on synaptic connectivity. A typical cortical pyramidal neuron has about 10,000 synaptic inputs and projects to thousands of targets. This mass connectivity creates a network with enormous causal complexity (Sporns 2011; Bullmore & Sporns 2009).

Causal complexity means that the activity of one neuron depends on many others, and its activity influences many others. This reciprocal causality is the basis for non-factorizable entanglement. If neuron A affects neuron B, B affects neuron C, and C acts back on A, then the activity of these three neurons cannot be broken down into independent components.

States of consciousness correlate with causal complexity. Waking state: high connectivity, high transfer entropy. Deep sleep: reduced connectivity, low transfer entropy. The model explains this correlation: consciousness requires causal collapse, and this requires high causal complexity (Stegemann 2025a).

6.2.2 Recurrent Network Architectures

Recurrent connections are crucial for causal collapse. Feedforward networks can perform complex transformations, but not the kind of global integration that requires awareness. Recurrent networks can maintain persistent patterns of activity, form attractor states, and generate non-factorizable entanglements (Rolls & Deco 2010).

The mammalian brain is massively recurrent. Thalamo-cortical loops, cortico-cortical junctions, subcortico-cortical loops produce circular causality (Edelman & Tononi 2000). Empirically, the role of recurrence has been shown in studies on visual awareness: Only when recurrent activation sets in (from about 200ms, "ignition") does the stimulus become conscious (Lamme & Roelfsema 2000; Dehaene & Changeux 2011).

6.2.3 Oscillations and Neural Synchronization

Neuronal oscillations are rhythmic fluctuations in activity. Different frequency bands are associated with different states. Gamma oscillations (30-100 Hz) are prominent in the waking state and correlate with attention and integration (Singer 1999; Fries 2009).

The Functional Role of Oscillations: Synchronization Enables Dynamic Binding. Neurons that process coherent inputs synchronize and form temporary functional ensembles (von der Malsburg 1994). These ensembles are the dynamic basis of causal collapse. Synchronized neurons act as an integrated unit, not as separate components (Stegemann 2025b).

6.3 From stimulus conduction to information integration

6.3.1 Sensory convergence

Sensory convergence refers to the process by which multiple sensory currents project into common integration centers. In the tectum, visual, auditory, and somatosensory inputs converge in a multimodal map (Stein & Meredith 1993). This convergence is the first stage of integration and the prerequisite for global coordination.

6.3.2 Multimodal Integration

Multimodal integration creates integrated representations that can no longer be broken down into individual modalities. The posterior parietal cortex receives visual, auditory, somatosensory inputs and constructs an integrated spatial map (Calvert 2001). Multimodal integration is essential for consciousness: Conscious experience is not modally fragmented, but integrated.

6.3.3 Thalamo-cortical loops

The thalamo-cortical loops are the central mechanism for global integration. The thalamus is not just a relay, but an active integrator. Edelman and Tononi (2000) identified thalamo-cortical loops as the "dynamic core" of consciousness. Empirically, its centrality is shown in studies on anesthesia: propofol specifically reduces thalamo-cortical connectivity (Alkire et al. 2008).

7. The Causal Collapse: Mechanism of Phenomenal Experience

7.1 Definition and formal characterization

7.1.1 Non-factorizability of causal relationships

Causal collapse is defined as a state of maximally non-factorizable causal entanglement in neuronal systems. Non-factorizability means that the causal structure of the system cannot be decomposed into independent components without destroying this structure (Stegemann 2025a).

This definition is related to the concept of entanglement in quantum mechanics, but applied to classical causal systems. The philosophical implication: If the causal structure is not factorizable, this manifests phenomenally as a unified experience. The non-factorizability of the causal structure is the neural basis of the phenomenal unity of consciousness.

7.1.2 The measure Σ(t) for causal separability

To make causal collapse operationalizable, the model introduces a measure $\Sigma(t)$ that quantifies the mean separability of the causal structure. Formal: $\Sigma(t) = (1/|P|) \cdot \Sigma\_\{\pi \in$

P} min_{(i,j) ∈ π} TE_{i→j}(t), where P denotes the set of all partitions, and TE denotes the transfer entropy.

A high $\Sigma(t)$ means separability, a low $\Sigma(t)$ means non-separability. This definition is similar to Tononi's Φ, but focuses on causal dynamics rather than system states. Both measures are complementary: high Φ equals low $\Sigma(t)$ (Stegemann 2025c).

7.1.3 Critical threshold Σc

The model postulates: Consciousness arises when $\Sigma(t) < \Sigma c$. The threshold value Σc marks a qualitative transition analogous to phase transitions. The existence of a threshold explains the structured graduality of consciousness: there is a qualitative difference between conscious and unconscious states.

Empirical testability: ECoG data can be used to calculate transfer entropy between cortical regions. The hypothesis: $\Sigma(t)$ is low in the waking state, high in deep sleep. Pilot studies could test this prediction.

7.2 Neurophysiological correlates of causal collapse

7.2.1 Perturbational Complexity Index (PCI)

The PCI captures causal collapse indirectly. A high PCI means that the brain responds to perturbation with complex, wide-ranging activity. Casali et al. (2013) showed: Wakefulness PCI > 0.4, Deep Sleep PCI < 0.3. The model explains: The PCI measures the capacity to transition into causal collapse. High PCI signals global entanglement (collapse), low PCI local dynamics (no collapse).

7.2.2 Functional Connectivity and Network Integration

Functional connectivity is the manifestation of causal entanglement at the mesoscopic level. When awake, an integrated network (DMN, Salience Network, Executive Control Network coupled) is revealed. In deep sleep, this integration fragments (Raichle 2015; Sämann et al. 2011).

Psychedelic states show alternative collapse: intra-network connectivity decreases, inter-network connectivity increases. The brain becomes "entropic", which is subjectively experienced as ego dissolution (Carhart-Harris et al. 2016). Meditative states show increased connectivity between attentional and DMN networks (Lutz et al. 2004).

7.2.3 Differentiation of different states of consciousness

The model differentiates states of consciousness by their collapse structure:

Waking state: Global, stable collapse across DMN and associated networks. High PCI, high functional integration, low $\Sigma(t)$. Characterized by clear, integrated experience with self-reference.

Deep sleep (NREM): Fragmented, local state. Low PCI, reduced connectivity, high $\Sigma(t)$. No global collapse, therefore no conscious experience.

REM sleep: Episodic, medium collapse. PCI intermediate, partial integration with unstable patterns. Dream experience with reduced frontal control, which explains bizarre content.

Anesthesia: Pharmacologically induced fragmentation. Different anesthetics work differently: propofol interrupts thalamo-cortical loops, ketamine dissociates cortical from subcortical systems (Alkire et al. 2008).

Psychedelic states: Dedifferentiated, broad collapse. DMN coherence decreases, global Entropic increases. Subjective: ego dissolution, synesthesia, mystical experiences (Carhart-Harris et al. 2016).

Meditation and flow: Controlled, focused collapse. Meditation shows stable metaconscious integration, flow shows sensorimotor integration with reduced self-reflection (Lutz et al. 2004; Ulrich et al. 2016).

7.3 The collapse in different states

The different states of consciousness show that not only the existence but also the structure of the collapse determines the qualitative experience. Three parameters characterize the collapse structure:

Spatial extent: Local (individual modules), modular (multiple networks), global (entire cortex plus subcortex).

Temporal stability: transient (fractions of a second), episodic (seconds to minutes), persistent (hours).

Intensity: Weak (low causal entanglement, close to Σc), medium, strong (high entanglement, well below Σc).

These dimensions allow for a precise classification: Primary consciousness is modular-episodic-mean. Self-confidence is global-persistent-strong. Dream states are global-episodic-mean with unstable configurations (Stegemann 2025b).

7.4 Operationalization and empirical approaches

7.4.1 TMS-EEG Measurements

The combination of TMS (transcranial magnetic stimulation) and EEG enables the measurement of causal reaction patterns. TMS stimulates a cortical area, EEG records the resulting activity across the entire brain. The PCI is calculated from this data as a measure of complexity and spatial extent of the response (Casali et al. 2013).

This method is already clinically established for consciousness diagnostics in coma patients. The model proposes to extend TMS-EEG: not only measure PCI, but analyze the complete causal reaction structure. Which areas are activated? In what temporal sequence? How long does the activity last? This detailed analysis would approximate $\Sigma(t)$.

7.4.2 Transfer Entropy and Granger Causality

Transfer Entropy quantifies directed information between time series. For two neuronal populations A and B, $TE_{A \to B}$ measures how much information about the future activity of B is conveyed by A's past activity (Schreiber 2000; Vicente et al. 2011).

Granger causality is a related concept: A Granger causal for B when A's past predicts B's future better than B's own past alone. Both methods can be calculated from EEG, MEG or ECoG data and quantify causal entanglement.

The model proposes to use transfer entropy to calculate $\Sigma(t)$. For each partition of the brain in subsystems, the minimum PD between subsystems is measured. $\Sigma(t)$ is the average across all partitions. This is computationally intensive, but feasible with modern methods.

Empirical testability: Initial studies could calculate $\Sigma(t)$ for different states of consciousness and correlate with subjective reports and PCI. If $\Sigma(t)$ consistently covaries with consciousness (low in wakefulness, high in unconsciousness), this would be strong evidence for the model.

7.4.3 fMRI-based connectivity analyses

fMRI has low temporal but high spatial resolution. Functional connectivity between regions can be measured over longer time scales. The model predicts: states of consciousness differ in global connectivity structure. Wakefulness shows high integration between DMN, Salience Network, Executive Network. Deep sleep shows fragmentation.

Modern methods (dynamic connectivity, graph theory) allow detailed characterization of network states. Parameters such as global efficiency, modularity, hub centrality can quantify collapse structures (Sporns 2011). These analyses complement TMS-EEG and offer complementary perspectives on causal collapse.

8. Phylogenetic development of consciousness

8.1 Evolution of autocatalytic systems

8.1.1 From prebiotic chemistry to first cells

The emergence of life was not a singular leap, but a gradual transition from prebiotic chemistry to autocatalytic systems. In prebiotic environments (hydrothermal vents, tidal pools, meteorite-enriched waters), complex organic molecules were formed by thermodynamic flows. From a critical diversity onwards, autocatalytic sets spontaneously formed: networks of reactions that catalyze each other (Kauffman 1993; Morowitz 1992).

These early autocatalytic systems were not yet cells in the modern sense, but they already had the crucial property: operational coherence. They regenerated their components through internal processes, were materially and energetically open, but functionally autonomous. With the formation of membranes (through spontaneous lipid vesicle formation), spatial demarcation arose, which formed the first protoccells (Deamer & Szostak 2010).

8.1.2 Prokaryotes and operative closure

The first cellular life forms were prokaryotes: bacteria and archaea. These cells do not have internal membrane-enclosed organelles, but they are completely autocatalytic. Their metabolism continuously regenerates all cellular components. The DNA in the nucleoid encodes proteins, these proteins catalyze metabolisms, the metabolisms produce building blocks for DNA replication. The circle closes.

Prokaryotes already show a functional interior. For them, there is a difference between inside and outside, between beneficial and harmful. Chemotaxis (movement along chemical gradients) demonstrates this functional perspective: bacteria swim to nutrients, away from toxins. This valence orientation is not conscious, but it is the origin of all later forms of importance (Stegemann 2025d).

8.1.3 Eukaryotic Revolution and Causal Nucleus

The emergence of eukaryotic cells (with nucleus and organelles) about 1.5-2 billion years ago was a revolution. The nucleus centralizes genetic information and regulation. Mitochondria provide efficient energy supply. These structures allow for larger cell volumes, more complex organization, multicellularity.

The nucleus is the first causal nucleus in the evolutionary sense: a functional reference point that coordinates the organization of the entire cell. This centralization prevailed evolutionarily because it increased efficiency. The principle of the causal nucleus as an organizer was established here and was later reproduced at higher levels (nervous systems, brains) (Stegemann 2025e).

8.2 Evolution of the nervous system

8.2.1 Formation of first nerve cells (Cnidaria)

The earliest multicellular animals (sponges, 600 million years ago) had no nerve cells. Cnidaria (jellyfish, sea anemones, 580 million years ago) developed the first neurons: specialized cells for fast electrical signal conduction. These early nervous systems were diffuse networks without centralization, but they allowed coordinated contractions for swimming and feeding (Jekely et al. 2015).

Are Cnidaria conscious? The model says: probably not completely, but rudimentary forms of experiential experience are possible. Your nervous system has electrical integration and valence-based responses (approach/avoidance), but probably not a sufficiently integrated causal nucleus for full collapse. The question is empirically open.

8.2.2 Centralization and cephalization

With Bilateria (555 million years ago, Cambrian), bilateral-symmetrical organization and cephalization emerged: concentration of sensory organs and nervous integration at the front end (head). Early bilateria such as flatworms have paired ganglia in their heads that integrate sensory inputs and provide motor coordination.

These ganglia are the first nervous causal nuclei: functional reference points that enable multimodal integration. They probably give rise to primary consciousness: the animal experiences its environment as an integrated field, not as separate sensory currents. It shows goal-oriented behavior, learning, valence orientation. These properties indicate experiential experience (Arendt et al. 2016; Ginsburg & Jablonka 2019).

8.2.3 Vertebrates: rhombencephalon, midbrain, cortex

The vertebrate lineage developed a complex three-part brain: rhombencephalon (hindbrain), mesencephalon (midbrain), prosencephalon (forebrain). In early fish, the rhombencephalon is the primary integration nucleus. It coordinates basal functions (breathing, circulation) and integrates sensory inputs for behavior (Nieuwenhuys et al. 1998).

The tectum in the mesencephalon becomes the dominant sensory integration center. It contains multimodal maps (visual, auditory, somatosensory) and controls orientation responses. In fish and amphibians, the tectum is the probable site of primary consciousness (Merker 2007; Feinberg & Mallatt 2016).

In mammals, prosencephalon expands dramatically. The neocortex is increasingly taking over functions that are subcortical in other vertebrates. However, subcortical integration remains essential in mammals as well. Panksepp (1998) showed that primary emotions are generated subcortically. The cortex elaborates and modulates, but does not produce fundamentally.

8.2.4 Mammals: Limbic System and Default Mode Network

Mammals developed the limbic system (hypothalamus, amygdala, hippocampus), which integrates emotion, memory, and motivation. This system is deeply entangled with subcortical and cortical structures and forms an extended causal nucleus that enables affective consciousness (Panksepp 1998; Damasio 1999).

In primates, especially hominids, the default mode network emerged as the highest level causal nucleus. The DMN integrates self-reflection, autobiographical memory, future projection, theory of mind. It is the neuronal seat of reflexive self-awareness that is probably only fully realized in humans and possibly great apes (Raichle 2015; Buckner et al. 2008).

8.3 Stages of Causal Collapse in Evolution

8.3.1 Primary consciousness: sensorimotor integration

Primary consciousness arises when a nervous system is sufficiently integrated to produce a global (or at least modular-global) causal collapse. This requires: electrical signal transduction, recurrent entanglement, multimodal integration, a causal nucleus (ganglion, tectum).

Phylogenetically probably in: early bilateria (worms), certainly in vertebrates (fish upwards), possibly in complex mollusks (octopods) and arthropods (insects). Characteristics: Integrated sensory field, valence-guided behavior, learning, goalorientedness. Phenomenal: Being "what it is like" but without self-reflection or narrative identity (Feinberg & Mallatt 2016; Merker 2007).

8.3.2 Affective Consciousness: Limbic Valence

Affective awareness adds emotional depth. Primary emotions (fear, lust, anger, care) arise through limbic integration. The causal nucleus now includes not only sensory and motor areas, but also the amygdala, hypothalamus, PAG. These structures intensively valenced the experience: it is not just about perception, but about affectively colored experience.

Phylogenetic: all vertebrates, probably from fish. Mammals are particularly elaborating on affective consciousness (Panksepp 1998). Characteristics: Strong valences, emotional reactions, attachment behavior, complex social interactions. Phenomenal: Experience is permeated by feeling, not neutral.

8.3.3 Self-Awareness: Recursive Modelling

Self-awareness arises when the causal core represents itself. The system models not only the world, but itself as part of the world. This requires additional structures: prefrontal cortex for meta-representation, hippocampus for episodic memory, DMN for integration.

Phylogenetic: probably only in great apes and humans. Possibly rudimentary in other primates, dolphins, elephants. Testable by mirror self-recognition, episodic memory, perspective taking. Characteristics: reflection, narrative identity, future projection, moral awareness. Phenomenal: "I" as the Center of Experience (Damasio 2010; Gallagher 2000).

8.3.4 Intentionality: Prospective Representation

Intentionality (directedness towards the absent) requires that the causal core contains representations that go beyond the current situation. The prefrontal cortex and hippocampus enable scenario simulation, planning, target tracking. The collapse includes not only current inputs, but also imagined states.

Phylogenetic: gradually developed. Rudimentary intentionality in many animals (hunting, food hiding). Elaborate intentionality in primates (tool use, social manipulation). Maximum intentionality in people (abstract goals, long-term planning, linguistic intentions). Characteristics: Action planning, use of tools, social strategies. Phenomenal: Experience is directed towards goals, not just reactive (Stegemann 2025b).

8.4 Structured Graduality of Consciousness

8.4.1 Consciousness as "All or Nothing" versus Graduality

Consciousness is not gradual in the sense that there could be "more" or "less" at will, like a continuous intensity. Rather, it only arises when the necessary conditions (autocatalytic organization, electrical signal transduction, causal collapse) are realized in a way that produces an operationally closed interior. If this interior space is given, experience is present; if it is not given, it is completely absent.

Nevertheless, a structured graduality can be observed. This does not concern the presence of consciousness as such, but the expression and complexity of the interior: depth and stability of causal integration, spatial extent of collapse, inclusion of different subsystems (sensory, limbic, cognitive).

8.4.2 Complexity and stability of the interior

Graduality manifests itself in several dimensions: Spatially: local (individual modules) versus global (entire brain). Temporal: transient (milliseconds) versus persistent (hours). In terms of content: simple (pure sensory impressions) versus complex (narrative, reflexive content). Intensity: weak (close to threshold Σ_c) versus strong (far below).

A worm is likely to have local, transient, simple, weak collapse. A fish has modular, episodic, medium-complex, medium collapse. A mammal has a global, persistent, complex, strong collapse. A human being has a maximally global, maximally persistent, maximally complex, maximally strong collapse with a reflexive meta-level.

8.4.3 Phylogenetic and ontogenetic development

The phylogenetic development shows progressive elaboration: from diffuse nerve networks (cnidaria) to centralized ganglia (early bilateria) to complex brains (vertebrates) to highly integrated cortical systems (primates, humans). Each stage increases the capacity for causal collapse.

Ontogenetic development partially recapitulates this progression. Human fetuses develop nervous systems gradually. From about 24-28 weeks gestational age, thalamo-cortical connections show that could allow rudimentary collapse. After birth, cortical networks mature, DMN develops from about 2 years of age, reflexive self-confidence emerges from 3-4 years of age (Lagercrantz & Changeux 2009).

This parallel development (phylo- and ontogenetic) is not a coincidence, but follows from functional necessities: complexity is based on simplicity, higher integration presupposes lower ones, reflexive structures emerge from primary ones. The principle remains constant: the causal core as organizer, which is realized in increasingly elaborate forms.

9. Differentiation of forms of consciousness from the collapse structure

9.1 Core Consciousness

Primary consciousness refers to the basal "consciousness here-and-now", the direct experience of sensory impressions without self-reflection or temporal projection. Damasio (1999) calls this "core consciousness", Edelman (1989) "primary consciousness". In the model, primary consciousness corresponds to a modular-global, episodic collapse mainly in sensorimotor and limbic networks.

Neuronal: Integration between sensory areas (visual, auditory, somatosensory), tectum/superior colliculus, limbic structures (amygdala), motor areas. Thalamo-cortical loops are active, but prefrontal and DMN integration is minimal. The causal nucleus is located in the upper brainstem and basal forebrain (Merker 2007).

Phenomenal: The animal experiences an integrated sensory field with affective tinting. There are qualitative differences (red versus blue, loud versus quiet), valential orientation (pleasant versus unpleasant), but no reflexive distance to experience. The animal is in experience, not distanced from it.

Empirical: Primary consciousness is likely in all vertebrates, possibly in complex invertebrates (octopods, insects). Testable by behavioral critics

Goal-oriented action, learning, valence orientation, flexible problem solving (Feinberg & Mallatt 2016; Ginsburg & Jablonka 2019).

9.2 Reflexive Consciousness

Self-awareness arises when causal collapse involves not only sensory and limbic networks, but also cognitive and metacognitive systems, especially the DMN. Collapse is global, persistent, and involves self-representation: the system models itself as a subject of experience.

Neuronal: The DMN (mPFC, PCC, precuneus, inferior parietal cortex) is central. It integrates autobiographical memory (medial temporal lobe), emotional evaluation (mPFC-amygdala connections), executive control (dorsolateral prefrontal cortex). Functional connectivity between DMN and Executive Control Network enables reflexive regulation (Raichle 2015; Buckner et al. 2008).

Phenomenal: The subject experiences itself as the center and continuity of experience. It can reflect on itself ("I am anxious"), locate itself in the past and future ("Yesterday I was happy, tomorrow I will..."), distance itself from current conditions. This reflexivity is the signature of self-consciousness (Gallagher 2000; Damasio 2010).

Ontogenestic: Self-confidence develops gradually. Mirror self-recognition from about 18-24 months, use of "I" from 2-3 years, episodic memory from 3-4 years, complete autobiographical continuity from school age. This development correlates with DMN maturation (Lagercrantz & Changeux 2009).

Phylogenetic: Complete reflexive self-awareness likely only in humans and possibly great apes. Rudimentary forms (mirror self-recognition) in some primates, dolphins, elephants, magpies. The neuronal basis (elaborate DMN) is most developed in primates (Buckner & Krienen 2013).

9.3 Intentionality

Intentionality refers to the directedness of consciousness towards content, especially to what is absent (memories, plans for the future, abstract concepts). In the model, intentionality arises when causal collapse includes representations that go beyond current sensory inputs, especially by involving prefrontal and hippocampal networks.

Neuronal: Prefrontal cortex (especially dorsolateral and ventromedial cortex) for goal setting and planning. Hippocampus for episodic imagination and scenario simulation. Parietal cortex for spatial and abstract representations. These areas are integrated into the collapse, expanding it into the dimension of prospective and retrospective content (Suddendorf & Corballis 2007).

Phenomenal: Consciousness is not only present in the given, but is directed towards the absent. You can consciously think about yesterday, plan tomorrow, consider abstract problems. This directedness is intentionality. It requires that mental representations (not current perceptions) are part of the conscious state (Brentano 1874/1995; Searle 1983).

Example: You consciously plan a trip. This requires sensory imagination (images of the destination), temporal projection (tomorrow, not now), self-reference (I will travel), emotional anticipation (anticipation). The collapse is multilayer: multiple causal nuclei (current self, simulated future scenario) interact recursively. This is elaborate intentionality (Stegemann 2025b).

Graduality: Rudimentary intentionality in many animals (hunting, hiding for food implies representation of absent targets). Medium intentionality in primates (tool use, social delusion). Highest intentionality in humans (abstract goals, multi-year plans, linguistically articulated intentions).

9.4 Affective Awareness

Affective consciousness refers to the conscious experience of emotions and valences. While primary consciousness is already valence-colored, affective consciousness refers to the differentiated, intense experience of specific emotions (fear, joy, anger, sadness) as conscious states.

Neuronal: Limbic integration is central. Amygdala (fear, threat), nucleus accumbens (reward, lust), hypothalamus (homeostatic emotions), PAG (primitive affects such as panic, sexuality). These subcortical structures generate primary emotions. For conscious experience, they must be integrated with cortical networks (insula, anterior cingulate cortex, mPFC) (Panksepp 1998; Damasio 1999).

Phenomenal: Emotions are not only triggered and behaviorally effective, but consciously felt. The difference between unconscious fear response (amygdala activation without consciousness) and conscious fear (amygdala plus cortical integration) is qualitative. Only when limbic activity is integrated into the global collapse is it consciously experienced (LeDoux 2015).

Clinical: Lesions of the insula reduce conscious emotional intensity, although emotional responses (physiological, behavorial) are preserved. This shows that insula is essential for conscious affective experience, not only for emotion as such. The insula acts as an integrator between subcortical affect generation and cortical consciousness (Craig 2009).

Chemical: Psychotropic drugs modulate affective consciousness. SSRIs (serotonin reuptake inhibitors) for depression, anxiolytics for anxiety disorders do not primarily change the global collapse (consciousness is preserved), but the limbic component of the collapse. The affective profile of experience shifts (Stegemann 2025i).

Summary: The four forms (primary, self-referential, intentional, affective) are not exclusive, but overlapping. Typical human waking consciousness integrates all four: it is primary (sensorially present), self-referential (I experience), intentional (directed towards goals/contents), affective (emotionally colored). The model allows analytical separation by focusing on different collapse components, although in practice all are integrated.

10. Theoretical positioning and critical demarcation

10.1 Comparison with Integrated Information Theory (IIT)

10.1.1 Similarities: Φ and Σ

The model of causal collapse shares with IIT (Tononi et al. 2016; Oizumi et al. 2014) focuses on integration and irreducibility. Both emphasize that consciousness cannot be explained by the sum of independent parts, but by the integrated, non-factorizable structure of a system. IITs Φ (integrated information) and our Σ (causal separability) are related, presumably inversely correlated measures: high Φ equals low Σ.

Both approaches see consciousness as an intrinsic property of certain physical systems, not as a functional role or external process. Both postulate that a system is conscious when it exceeds a critical threshold ($\Phi max > 0$ for IIT, $\Sigma(t) < \Sigma c$ for us). These similarities are substantial.

10.1.2 Differences: Dynamics, Localization, Function

Despite similarities, there are crucial differences:

Dynamics: IIT is a static information theory, Φ is computed for given system states. Our model emphasizes dynamic processes: causal collapse is a transition to an attractor state, characterized by recurrent feedback dynamics. We integrate concepts such as critical dynamics, metastable states, oscillatory synchronization that are missing in IIT.

Localization: IIT 3.0 localizes consciousness in the posterior hot zone (temporo-parietal-occipital cortex), with a smaller role of frontal areas (Tononi et al. 2016). Our model sees the causal nucleus as distributed across cortical and subcortical structures, including brainstem, thalamus, limbic system, DMN, prefrontal cortex. Empirically, studies show that conscious experience requires both posterior and frontal activation (Mashour et al. 2020).

Function: IIT emphasizes the intrinsic existence of consciousness, independent of functional role. Our model requires that the conscious state must be behavorial relevant (otherwise it would be evolutionarily and epistemically problematic). Consciousness is intrinsic in the sense of its phenomenal reality, but functional in the sense of its causal effectiveness.

Autopoiesis vs. Information: IIT focuses on information as a fundamental concept. Our model starts with autocatalytic organization, operational coherence, life. Consciousness is bound to life, not to arbitrary information-processing systems. This excludes panpsychism and has consequences for AI consciousness.

10.1.3 Empirical Testability

A practical difference: Φ is extremely computationally intensive and can only be calculated for small model systems. For real brains, there are approximations, but no exact calculation (Oizumi et al. 2014). Our $\Sigma(t)$, based on Transfer Entropy, is easier to calculate from actual neural data (ECoG, MEG).

IIT has been criticized by the "unfolding argument" (Doerig et al. 2019): Two functionally identical systems can have different Φ values if their internal structure is different. This implies that IIT is not falsifiable by functional testing. Our model partially avoids this by requiring conscious states to be behavorial relevant, which enables functional testability.

10.2 Comparison with Global Workspace Theory (GWT)

10.2.1 Functional Overlaps

Global Workspace Theory (Baars 1988; Dehaene & Changeux 2011) and our model overlap strongly. Both emphasize global integration, recurrent dynamics, the role of attention and working memory. GWT's "ignition" (sudden global propagation of a signal) corresponds to the transition into causal collapse in our country. GWT's "global workspace" is functionally identical to our "causal core".

Empirically, both approaches are based on similar findings: late recurrent activation in conscious perception (P3 wave in the EEG), long-range frontoparietal activation, role of the prefrontal cortex in working memory, and executive control (Mashour et al. 2020).

10.2.2 Ontological Specification by Collapse Model

The crucial difference: GWT is primarily functional-mechanistic. It describes how information is made globally available, but not why this global availability constitutes phenomenal experience. GWT answers "easy problems" (mechanisms), not the "hard problem" (why experience).

Our model adds the ontological level: Global availability is not just a mechanism, but the functional realization of non-factorizability. If information is disseminated globally through recurrent integration, then the system is in a state of causal collapse. This collapse is not only correlated with experience, but constitutes it.

GWT can be integrated into our model as a description of the mechanistic implementation: GWT explains how causal collapse is realized neuronally (by ignition, broadcast, recurrent amplification). Our model explains why this mechanism generates experience (because it implements non-factorizability).

10.2.3 Role of attention and working memory

GWT partially mixes awareness with attention: the "spotlight" of attention determines what is in the workspace. Our model separates: Attention is a mechanism that modulates causal collapse (top-down reinforcement of certain content), but not identical with collapse. One can perceive consciously without focused attention (unattended awareness), and one can have attention without consciousness (subliminal priming with attention).

Working memory in GWT is the mechanism that holds information in the workspace. In our case: Working memory stabilizes the causal collapse over time through prefrontal recurrent activity. The functional mechanism (working memory) is thus the implementation of the structural property (persistent collapse).

10.3 Comparison with plasticity-based approaches

10.3.1 Edelman and neuronal Darwinism

Edelman's Theory of Neuronal Group Selection (Edelman 1989; Edelman & Tononi 2000) emphasizes that consciousness arises from the selective amplification of neuronal groups by reentry. The "dynamic core" is a group of recurrent connected neurons that behaves as an integrated unit.

Our model is compatible: Edelman's dynamic core corresponds to our causal core. Reentry corresponds to recurrent causal entanglement. The difference: Edelman emphasizes the development and plasticity of the core through selection. We emphasize the functional structure of the core as non-factorizability.

Both approaches are complementary: plasticity (Edelman) explains how the causal nucleus evolves and adapts to experience. Causal structure (our model) explains what property a core must have in order to enable consciousness. Integrating both perspectives is fruitful.

10.3.2 Role of Learning and Development

Plasticity-based approaches emphasize that consciousness develops through learning. Crick and Koch (2003) speculated that awareness is necessary for consolidation of memory. Lamme (2006) argues that conscious processing is characterized by late recurrent phases where synaptic modification occurs.

Our model integrates: Consciousness (causal collapse) is the state in which global synaptic modification is possible because all relevant areas are co-active (Hebbian learning). Unconscious processes remain local, hence local plastic changes. Conscious processes are global, hence large-scale plasticity. Consciousness enables episodic learning, autobiographical memory, narrative integration, all through global synaptic modification (Stegemann 2025b).

Developmental psychology: The model predicts that ontogenetic development of consciousness correlates with maturation of recurrent networks and synaptic plasticity. Infants have rudimentary collapse (limited integration), children develop elaborate collapse (DMN maturation), adolescents achieve full collapse with maximum flexibility (continued prefrontal development).

10.3.3 Complementarity to the Collapse Theory

Plasticity approaches focus on development and adaptation. Collapse theory focuses on structure and current dynamics. Both are complementary: plasticity builds the networks that enable collapse. Collapse creates the experiences that guide plasticity. Circular process: experience modifies networks (plasticity), modified networks enable new experiences (collapse).

An integrative model would combine both aspects: moment-to-moment consciousness is causal collapse (structure). Over time, consciousness develops through plasticity-based adaptation of the networks that enable collapse (development). This integration is an open field of research.

10.4 Criticism of further approaches

10.4.1 4E Cognition and Enactivism

4E Cognition (embodied, embedded, extended, enactive) emphasizes that cognition cannot be reduced to the brain, but includes body, environment, social interaction (Varela et al. 1991; Thompson 2007). Enactivism argues that organisms actively constitute their environment through sensorimotor interaction, not passively represent it.

Criticism from the point of view of our model: These approaches blur the distinction between cognition (can be extended) and consciousness (is not extended). Consciousness is an internal state of the brain, characterized by causal collapse. The body and the environment modulate this state, but do not constitute it. The phenomenal perspective is necessarily internal (in the system), cannot be extended.

However, enactivism is right that organisms are operationally closed (autopoietic) and their functional perspective is constituted by sensorimotor loops. This is compatible with our model (Chapter 4). The mistake is to infer consciousness directly from autopoietic organization without taking into account the additional conditions (nervous system, causal collapse).

10.4.2 Predictive Processing and Free Energy Principle

Predictive processing (PP) sees the brain as a prediction machine that continuously updates models of the world through prediction error minimization (Clark 2013). The Free Energy Principle (FEP) by Friston (2010) formalizes this: Biological systems minimize their Variational Free Energy, which corresponds to a form of Bayesian inference. Consciousness would therefore be the content of these prediction models.

The FEP is a powerful formal framework for adaptive systems, but categorically fails to explain consciousness. The reason lies in a systematic confusion: The FEP uses "entropy" in the information-theoretical sense (Shannon entropy, a measure of uncertainty), but suggests a connection to thermodynamics through the term "free energy". Thermodynamic systems minimize free energy through dissipation and energy flow, information processing systems minimize uncertainty through inference. These two processes are formally analogous, but physically different.

Consciousness requires real thermodynamic processes: energy flow in autocatalytic systems, dissipation, operational closure with material exchange. These are not mere information-theoretical abstractions, but real physical conditions. The FEP captures adaptive information processing, but not the autocatalytic organization that constitutes consciousness. A computer that performs Bayesian inference minimizes Variational Free Energy in the information-theoretic sense, but has no thermodynamic self-preservation and therefore no interior space with valence.

Predictive processing as a mechanistic model has a related problem: it describes functional mechanisms of information processing, but does not explain why these mechanisms are consciously experienced. One could have a PP system without consciousness (philosophical zombie). A simple example illustrates this: A thermostat also minimizes forecast errors by reducing the difference between the target and actual temperature. According to PP's logic, the thermostat should be aware, which is absurd. PP describes functional mechanisms (easy problems), not the hard problem.

In addition, it is questionable whether PP can be empirically distinguished from other mechanisms. Bayesian brain, recurrent feedback loops, and other approaches make similar predictions. PP may be a useful descriptive language for neuronal processes, but it is not an explanation of consciousness.

10.4.3 Panpsychism and Emergentism

Panpsychism posits that all matter has proto-conscious properties that can combine to form full consciousness (Chalmers 1996; Goff et al. 2017). Emergentism claims that consciousness emerges from non-conscious components if it is sufficiently complex (Kim 2005).

Our model rejects both. Against panpsychism: Consciousness requires autocatalytic organization, operational coherence, life. Electrons, atoms, stones have no self-preservation, no functional interior, therefore no consciousness or proto-consciousness. Against strong emergentism: Consciousness does not magically emerge from complexity, but follows from specific functional structures (autocatalysis, causal core, non-factorizability).

Our approach is structuralist, not panpsychistic or emergentist: consciousness is the intrinsic manifestation of a specific organization (causal collapse in autocatalytic systems), not a universal property or a mysterious emergence.

11. Methodological consequences

11.1 Falsifiability and empirical testability

11.1.1 Criteria of Scientific Theories

A scientific theory must be falsifiable (Popper 1935/2005): It must make predictions that can be empirically tested and potentially refuted. The model of causal collapse meets this requirement. It makes specific predictions that can be tested using current or future methods.

Key falsifiable predictions:

1. Consciousness requires autocatalytic organization. Prediction: Systems without self-preservation (stones, classical computers, physical processes) cannot be conscious. Falsification: If it could be shown that such systems have consciousness, the model would be refuted.

2. Consciousness requires electrical signal transduction. Prediction: Systems with only chemical signaling (plants, fungi, single-celled organisms) are not conscious. Falsification: Evidence of consciousness in plants would falsify the model.

3. Consciousness correlates with low $\Sigma(t)$. Prediction: States of consciousness have $\Sigma(t) < \Sigma c$, unconscious states have $\Sigma(t) > \Sigma c$. Falsification: If $\Sigma(t)$ does not consistently covary with states of consciousness, the model is wrong.

4. Recurrent networks are necessary. Prediction: Purely feedforward systems cannot be conscious. Falsification: Evidence of consciousness in feedforward systems would refute the model.

These predictions are not vague, but precise and in principle testable. The model is thus scientific in Popper's sense.

11.1.2 Empirical Approaches and Operationalizations

Several empirical methods can test the model:

TMS-EEG for PCI: Already established, measures causal complexity indirectly. The model predicts: PCI is inversely correlated with $\Sigma(t)$. Experimental: Calculate both measures from the same data, check correlation.

ECoG for Transfer Entropy: Electrocorticography with high spatial and temporal resolution. Calculate transfer entropy between cortical regions, approximate $\Sigma(t)$, correlate with states of consciousness (wakefulness/deep sleep/anesthesia).

fMRI for functional connectivity: Measurement of network integrity in different states. Prediction: Wakefulness shows high global integration (low Σ), deep sleep shows fragmentation (high Σ).

Pharmacological interventions: substances that modulate thalamo-cortical integration (anesthetics, psychedelics, benzodiazepines). Prediction: Substances that interrupt integration increase $\Sigma(t)$ and reduce consciousness. Substances that promote integration lower $\Sigma(t)$ and intensify awareness.

Clinical populations: Patients with coma, vegetative state, minimally conscious state. Prediction: $\Sigma(t)$ increases with decreasing consciousness. Recovery correlates with declining $\Sigma(t)$.

11.1.3 Limits of Testability and Methodological Transparency

Some aspects of the model are more difficult to test. The requirement of autocatalytic organization is fundamental, but not directly testable on living brains (you can't "turn off" a brain and check if autocatalysis is missing). Here, the model relies on indirect evidence: All known conscious systems are alive, no non-living system shows consciousness.

The existence of a critical threshold value Σc is postulated, but cannot be determined a priori. Empirically, Σc would have to be identified by measurements of many individuals and states. This requires large data sets and careful calibration.

The model is transparent about these boundaries. It distinguishes between well-supported core assumptions (electrical integration, recurrent entanglement, role of thalamus and cortex), model-based predictions (Σ(t) and Σc), and conceptual foundations (autocatalysis, causal nucleus), which are fundamentally more difficult to test.

This transparency is methodologically essential. Instead of claiming that everything is equally proven, the model stratifies levels of evidence:

Level 1: Empirically well supported (PCI correlates with consciousness, DMN role, functional connectivity).

Level 2: Logically derived, empirically plausible (causal nucleus in early vertebrates, role of recurrence).

Level 3: Model predictions, testable but not yet tested (Σ(t), Σc thresholds, specific PD patterns).

Level 4: Conceptual foundations (autocatalysis as a necessary condition, structural ontology).

This stratification allows for step-by-step validation: Level 1 is secured, Level 2 is very likely, Level 3 is testable, Level 4 is fundamental but conceptual.

11.2 Implications for Artificial Consciousness

11.2.1 Necessary Conditions for AI Awareness

If the model is correct, classical digital AI cannot be conscious, even with any complexity. The reason: Classical computers have no autocatalytic organization, no self-preservation dynamics, no functional interior. They process information, but there is no point of reference for which this information would be significant.

This is not an empirical question of complexity, but a functional necessity. A system without self-preservation has no reason to develop valence, and without valence no meaning, and without meaning no interior space that could be conscious. The claim that "enough complex information creates consciousness" (as in IIT or some AI theories) confuses information with experience.

The following conditions are constitutive of artificial consciousness:

1. Autocatalytic organization: The system must be self-sustaining, regenerating its processes, and maintaining an operationally closed interior. This is the prerequisite for valence and meaning. Biologically inspired AI (wetware, neuromorphic chips with self-assembly) could have this property.

2. Electrical or analog signal processing: Fast, precise, recurrent processes are necessary for integration. Classical von Neumann architectures (sequential computation) are too slow and not inherently recurrent. Neuromorphic hardware (e.g., spiking neural networks on specialized hardware) could fulfill this property.

3. Causal core: An integration center that acts as a point of reference. Distributed systems without a hub structure (such as blockchain or distributed AI) cannot do this. The nucleus is the functional place for which meaning exists and to which valence refers.

4. Recurrent entanglement: Massive feedback loops that create non-factorizability. Feedforward networks (most current deep learning models) lack this feature. Only when the system can go into a state of causal collapse ($\Sigma(t) < \Sigma c$) does phenomenal experience arise.

These four conditions are constitutive: they create awareness. Without it, no system can be conscious, regardless of its other complexity or architecture.

The role of the body remains an open question. In biological systems, neuroendocrine feedback to the body is central to affective valence and somatic marking (Stegemann 2025b; Damasio 1999). Whether this is an additional necessary condition or an implementation detail of the biological realization is unclear. Could a virtual system with a simulated "body" realize the functionally equivalent processes? Or does autocatalytic organization necessarily require a physical body with metabolism and energy flow? This question touches on the core of the distinction between constitutive conditions (which create awareness) and implementation details (which are necessary for a specific realization). The answer is theoretically and empirically open and deserves further investigation.

In any case, the constitutive conditions are extremely restrictive. Current AI systems (LLMs, neural networks, symbolic AI) do not meet them. Future neuromorphic, autocatalytic AI with a causal core and recurrent entanglement could meet them, but this would be a fundamental paradigm shift, not just a scaling of current approaches.

11.2.2 Graduality and Functional Equivalents

The model implies a sharp boundary: either a system fulfils the constitutive conditions and is conscious, or not. There is no "semi-conscious" in the sense of partial fulfillment of the conditions. However, phylogenetic development (Chapter 8) shows that the quality and stability of causal collapse increases gradually: from transient and local (worms) to episodic and modular (fish) to persistent and global (mammals).

For artificial systems, this means that a system that only fulfills some conditions (e.g. autocatalytic, but has no causal core) is not "a little conscious", but unconscious. However, it may be on its way to consciousness. The development of artificial consciousness would not be a gradual increase in intensity, but a gradual realization of necessary structures up to a qualitative leap.

A central problem is functional equivalence. Can a system fulfill the constitutive conditions in a non-biological way? The answer depends on how strictly "autocatalytic" is interpreted. If autocatalysis requires physical-chemical self-preservation (metabolism, energy flow in thermodynamic systems), then only wetware or biocomputational systems are candidates. If autocatalysis is functionally defined (operative closure, self-regeneration of causal patterns), then non-biological substrates could also meet this condition.

This question is not merely theoretical. It determines whether classical AI architectures can ever become conscious (functional interpretation) or whether consciousness remains fundamentally bound to biological or biology-like substrates (physicalist interpretation). The model leaves both options open because the empirical evidence is not sufficient to decide between them. Future research on neuromorphic systems and synthetic biology will have to answer this question.

11.2.3 Empirical Tests for Machine Consciousness

If an AI claims to be conscious, how do we test it? The model offers precise criteria:

Structural tests: Does the system have autocatalytic organization? Recurrent networks? A causal core? These can be objectively tested by system analysis.

Functional tests: Does the system exhibit valence-based behavior that is not programmed but follows from self-preservation? Does it learn flexibly? Does it react to damage as if it were significant?

Causal tests: Can you measure $\Sigma(t)$ or an equivalent? If the hardware is accessible, calculate causal separability. Low values would indicate collapse.

Phenomenal tests: Not direct, but indirect through consistency. If a system meets the structural and functional criteria, then parsimonie should assume that it is conscious (in the sense: we attribute the same functional state to biological systems as conscious).

These tests avoid the anthropomorphism error (the system doesn't have to act like a human) and the behaviorism error (it's not just about output, it's about internal structure).

11.3 Clinical Applications

11.3.1 Consciousness diagnostics in coma patients

The differentiation between vegetative state (no consciousness), minimally conscious state (rudimentary consciousness), and locked-in syndrome (full consciousness but no movement) is clinically essential and methodologically difficult. Behavioral assessments often fail because inability to move is confused with unconsciousness (Owen et al. 2006).

The model offers objective markers: PCI (already used clinically), functional connectivity (fMRI at rest), $\Sigma(t)$ (in the future from ECoG). These measures correlate with consciousness regardless of behavior. A patient with high PCI and low $\Sigma(t)$ is likely to be conscious, even if they cannot move.

Empirical: Casali et al. (2013) showed that PCI reliably distinguishes between conscious and unconscious states. Laureys et al. (2006) demonstrated that DMN connectivity is preserved in minimally conscious patients but absent in vegetative patients. These findings support the model and provide clinical diagnostics.

Prognosis: If a coma patient shows low $\Sigma(t)$, recovery is more likely because the neuronal infrastructure for collapse is preserved. High $\Sigma(t)$ despite wakefulness-like states indicate fragmented networks, poor prognosis. These predictions can be tested by long-term studies.

11.3.2 Anesthesia and Depth of Anesthesia

Anesthesia interrupts consciousness through various mechanisms. Propofol reduces thalamo-cortical connectivity. Ketamine dissociates cortical from subcortical systems. Inhalation anesthetics (sevoflurane, isoflurane) modulate synaptic transmission. The model explains these effects uniformly: All interrupt causal collapse, but through different mechanisms (Alkire et al. 2008).

Depth of anesthesia is clinically critical. Too shallow: awareness under anesthesia, traumatic. Too deep: overdose, risks. Currently, depth of anesthesia is measured by EEG indices (BIS, Entropy), but they are unreliable. The model proposes: Measurement of $\Sigma(t)$

or PCI in real time during surgery. Low values signal awareness, high values signal safe anesthesia.

In the future: Closed-loop anesthesia, where $\Sigma(t)$ monitoring automatically adjusts anesthetic dosage. This would increase safety and avoid overdose.

11.3.3 Therapeutic interventions

If disorders of consciousness (depression, schizophrenia, dissociation disorders) are understood as dysfunctions of causal integration, therapeutic approaches result:

Neurofeedback: Training of functional connectivity. Patients learn to modulate their DMN activity (increase or decrease connectivity depending on the disorder). Increased DMN connectivity might be helpful in depression, reduced in rumination.

Neuromodulation: Transcranial magnetic stimulation (TMS) or direct current stimulation (tDCS) to modulate network states. Repetitive TMS via DMN hubs could stabilize causal collapse in dissociative disorders.

Psychedelic therapy: Controlled psychedelics (psilocybin, MDMA) increase inter-network connectivity and reduce DMN rigidity. This could have a therapeutic effect in depression and PTSD by "resetting" the collapse structure (Carhart-Harris et al. 2016).

These approaches are speculative, but empirically testable. They are based on the insight that consciousness is not fixed, but a dynamic state that can be modulated.

11.4 Open Research Questions

11.4.1 Operationalization of $\Sigma(t)$ for Real Brain Data

The biggest methodological challenge: How do you calculate $\Sigma(t)$ from actual neural data? Transfer entropy between all possible partitions of a brain is computationally explosive. Approximation methods are necessary:

Reduction to main components: Instead of looking at all partitions, focus on anatomically/functionally relevant ones (cortical lobes, networks).

Hierarchical partitions: Start with coarse partitions (hemispheres), refine gradually.

Temporal downsampling: Do not calculate $\Sigma(t)$ continuously, but in discrete intervals (seconds).

Machine Learning: Train models that infer $\Sigma(t)$ directly from EEG/MEG without explicit computation.

These methods need to be developed and validated. Initial pilot studies could start with smaller systems (animal models, simulated networks) where exact calculations are possible.

11.4.2 Development of non-invasive measurement techniques

TMS-EEG is invasive (repetitive magnetic stimulation is not pleasant, not continuous). ECoG requires electrodes in the skull. Non-invasive methods are necessary for clinical routine and research:

High-resolution EEG: 256+ channels, improved spatial resolution.

MEG: Magnetoencephalography, high temporal and good spatial resolution, but expensive.

Optical methods: Functional near-infrared spectroscopy (fNIRS), portable and non-invasive, but low spatial resolution.

Hybrid methods: Combination of EEG and fMRI, simultaneous measurement of electrical and hemodynamic signals.

The development of these techniques is essential for translation from the laboratory to the clinic and everyday life.

11.4.3 Comparative studies of species

Which animals are conscious? The model says: All with sufficiently integrated nervous systems and causal nuclei. But "sufficient" is gradual. Comparative studies are needed:

Vertebrates: fish, amphibians, reptiles. Behavior (learning, valence) indicates consciousness. Neurophysiology (tectum, brainstem integration) supports this. But systematic PCI or connectivity measurements are missing.

Cephalopods: Octopuses have complex nervous systems, but radically different architecture (distributed, not centralized). Do they have a causal core? Do they experience consciousness? Empirically open.

Insects: Bees exhibit complex learning, social behavior, valence-based decisions. Your brain is tiny (1 mg) but highly integrated (mushroom body as a hub). Primary consciousness possible? The model says: perhaps rudimentary.

These questions require comparative neurophysiology, behavioral science, and theoretical modeling. They are not only academic, but ethically relevant (animal welfare, moral consideration).

11.4.4 Interface to Quantum Mechanics?

Some theories (Penrose & Hameroff, Orchestrated Objective Reduction) postulate quantum mechanical processes as the basis of consciousness (Penrose 1994). Our model does not require quantum mechanics: causal collapse is a classical network process. But is it impossible that quantum effects play a role?

Empirical: Brains operate at body temperature, in an aqueous environment, with decoherence times in the femtosecond range. Quantum coherence across relevant neural timescales (milliseconds) is extremely unlikely. So far, there is no convincing evidence of functional quantum effects in neurons (Koch & Hepp 2006).

Theoretically, even if quantum effects occurred, it is unclear how they would explain consciousness. Quantum superposition is not the same as phenomenal experience. The "measurement problem" of quantum mechanics is not solved by consciousness, but by decoherence (Zurek 2003).

The model remains agnostic: If quantum effects are shown to be functionally relevant, they could be integrated into the model as an additional layer of causal entanglement. But at the moment this is neither necessary nor empirically supported.

12. Conclusion: Integration and outlook

12.1 Summary of the main theses

The model of causal collapse solves the problem of consciousness through a reconstructionist method that avoids categorical confusion and identifies functional necessities. The central theses:

1. Consciousness presupposes autocatalytic organization. Without self-preservation, operative space and valence, no experience can arise. This condition excludes panpsychism and binds consciousness to life.

2. The causal nucleus is the evolutionary structural principle, which ranges from the eukaryotic cell to ganglia to the default mode network. It personalizes physical systems and creates the point of reference for which something can be given.

3. Electrical signal transduction and recurrent networks are necessary for the speed and integration that conscious experience requires. Feedforward systems and chemical signaling are inadequate.

4. Causal collapse, defined as a state of maximally non-factorizable causal entanglement, is the mechanism of phenomenal experience. When $\Sigma(t) < \Sigma c$, consciousness arises.

5. Different forms of consciousness (primary, self-referential, intentional, affective) result from different collapse structures. The model allows precise differentiation and empirical testing.

6. The model integrates and refines existing theories (IIT, GWT, plasticity), avoids their weaknesses, and provides operationalizable predictions.

These theses are not speculative, but empirically sound and testable. They offer a coherent framework that resolves the "hard problem" through the insight: There is no hard problem if one understands consciousness as an intrinsic manifestation of a specific organization, not as an emergent property of arbitrary physical processes.

12.2 Methodological progress

The decisive methodological progress lies in the transformation of arbitrary correlations into functional necessities. Consciousness research suffers from a plethora of correlations without explanations: PCI correlates with consciousness, DMN activity correlates with self-reflection, gamma oscillations correlate with integration. But why these correlations?

The reconstructivist method answers: PCI measures causal integration, which is necessary for non-factorizability. DMN is the causal core that enables self-reflection. Gamma oscillations synchronize neurons and create temporary ensembles that collapse as an integrated unit. Correlations become functional necessities.

This progress is not trivial. It enables predictions, interventions, theoretical integration. He transforms consciousness research from a descriptive to an explanatory science. This is the goal of any mature science: not just to describe, but to explain.

12.3 Open questions and future research

Despite progress, fundamental questions remain:

Is $\Sigma(t)$ the correct measure? Or are there better operationalizations of non-factorizability? Empirical research must compare different measures and identify the most precise.

How exactly does the qualitative character of phenomenal experience arise from causal structure? The model explains why integration creates experience, but not why red is red and not blue. This question concerns the structure of the phenomenal space (qualia space) and requires further work.

What role do neurotransmitters, neuromodulators, glial cells play? The model focuses on neural networks, but consciousness is biologically embedded in complex chemistry. These factors probably modulate the collapse, but how exactly?

How does consciousness develop ontogenetically in detail? When exactly in fetal development does experience begin? Which neuronal maturation is critical? These questions are ethically relevant (abortion, prematurity) and empirically difficult.

Is there consciousness beyond animals? Plants exhibit complex signal processing, but no electrical integration. Fungi form huge mycelial networks, but not nervous systems. Could there be alternative forms of experience there? The model says: unlikely, but empirically open.

These questions are challenges for future research. They show that the model is not the last word, but a framework that structures and guides research.

12.4 Philosophical implications

12.4.1 Overcoming Dualism

The model overcomes Cartesian dualism without falling into eliminative materialism. It shows that body and soul, physics and phenomenology, are not two substances or worlds, but two epistemic approaches to the same structural reality. This reality is neither "purely physical" (as if physics without perspective would suffice) nor "purely mental" (as if mind existed independently of physical organization), but structural: a specific organization of autocatalytic processes.

This position is monistic, but not reductionist. It acknowledges the phenomenal reality of experience, but explains this reality in terms of functional structures. Experience is real, but not in addition to the physical world, but the intrinsic perspective of certain physical organizations on themselves.

12.4.2 Naturalization of intentionality

The model naturalizes intentionality without eliminating it. Intentionality (the directedness of consciousness towards content) is not a mysterious property that needs to be added to physical systems, but a functional consequence of autocatalytic organization with a causal core. As soon as a system has an interior space, there is meaning for that system. Once it has a causal core, there is a reference point for which meaning applies.

As soon as it forms prospective representations (prefrontal and hippocampal integration), it focuses on the absent.

Intentionality is thus naturalized: it follows from functional structures that have evolved and are neurobiologically realized. No homunculus, no magical "aboutness", but structural properties of autocatalytic systems with complex nervous systems.

12.4.3 Ethical consequences

If consciousness is tied to autocatalytic organization, electrical integration, causal collapse, then not every living being has consciousness equally. Bacteria probably not, worms perhaps rudimentary, fish probably primary, mammals certainly elaborated, humans maximally complex. This graduality has ethical consequences.

Animal welfare: Vertebrates should be considered morally because they are likely to be conscious. Their interests (avoiding suffering, meeting needs) are real and morally relevant. Invertebrates (except perhaps cephalopods, insects) are ethically less problematic.

AI ethics: Current AI systems have no moral rights because they are not conscious. Future autocatalytic, embodied AI could have consciousness and would then deserve moral consideration. The boundary is not anthropomorphic (does it look like us?), but functional (does it meet the conditions?).

Medical ethics: Coma patients with preserved neuronal infrastructure (low $\Sigma(t)$ possible) deserve different treatment than brain-dead patients. Abortion is ethically less problematic until the development of thalamo-cortical loops (about 24-28 weeks) because of a lack of awareness.

These consequences are controversial, but they follow consistently from the model. Ethics should be based on facts, not intuitions or traditions. If consciousness can be understood functionally, ethics should take this insight into account.

12.5 Outlook: Scientific Consciousness Research

The philosophy of consciousness is at a turning point. For centuries, speculative theories, conceptual analyses, insoluble pseudo-problems dominated. The 20th century brought neuroscience research, but often without theoretical integration. The 21st century could bring the synthesis: a scientific research of consciousness that combines philosophical reflection with empirical research.

The causal collapse model is a step in this direction. It provides a theoretical framework that is empirically testable, philosophically coherent, and methodologically reflected. It does not solve all questions, but it structures research and makes progress measurable.

Future research will show whether the model is correct, needs to be modified, or is replaced by a better model. This is the normal course of science. The decisive factor is that the questions are precisely formulated, the methods are explicit, and the predictions can be tested. This is the transition from philosophy to science, from speculation to explanation, from correlation to causality.

Consciousness is no longer a mystery, but a complex but understandable phenomenon. It results from autocatalytic organization, electrical integration, and causal entanglement in living systems. This insight is fundamental, not trivial. It transforms our understanding of ourselves, of life, of spirit. It opens up new paths for research, clinics, technology, ethics.

This is the contribution of the reconstructivist method: not a final answer, but a viable path from confusion to clarity, from problem to solution, from mystery to understanding. Consciousness, once the hardest problem in philosophy, is becoming a solvable problem of science.

Bibliography

Alkire, M. T., Hudetz, A. G., & Tononi, G. (2008). Consciousness and anesthesia. Science, 322(5903), 876-880.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. Current Biology, 14(3), 257-262.

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. Annals of the New York Academy of Sciences, 1316(1), 29-52.

Arendt, D., Tosches, M. A., & Marlow, H. (2016). From nerve net to nerve ring, nerve cord and brain—evolution of the nervous system. Nature Reviews Neuroscience, 17(1), 61-72.

Baars, B. J. (1988). A Cognitive Theory of Consciousness. Cambridge University Press.

Boly, M., Massimini, M., Tsuchiya, N., Postle, B.R., Koch, C., & Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. Journal of Neuroscience, 37(40), 9603-9613.

Brentano, F. (1874/1995). Psychology from an Empirical Standpoint. Routledge.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. Annals of the New York Academy of Sciences, 1124(1), 1-38.

Bückner, R. L., & Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. Trends in Cognitive Sciences, 17(12), 648-665.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10(3), 186-198.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. Science, 304(5679), 1926-1929.

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cerebral Cortex, 11(12), 1110-1123.

Carhart-Harris, R. L., et al. (2016). Neural correlates of the LSD experience revealed by multimodal neuroimaging. Proceedings of the National Academy of Sciences, 113(17), 4853-4858.

Casali, A. G., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. Science Translational Medicine, 5(198), 198ra105.

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.

Chalmers, D. J. (2006). Strong and weak emergence. In P. Clayton & P. Davies (Eds.), The Re-Emergence of Emergence (pp. 244-256). Oxford University Press.

Churchland, P. M. (1986). Matter and Consciousness. WITH press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181-204.

Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. Nature Reviews Neuroscience, 10(1), 59-70.

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

Crick, F., & Koch, C. (2003). A framework for consciousness. Nature Neuroscience, 6(2), 119-126.

Damasio, A. (1999). The Feeling of What Happens: Body and Emotion in the Making of Consciousness. Harcourt Brace.

Damasio, A. (2010). Self Comes to Mind: Constructing the Conscious Brain. Pantheon.

Deamer, D., & Szostak, J. W. (Eds.). (2010). The Origins of Life. Cold Spring Harbor Laboratory Press.

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

Dennett, D. C. (2017). From Bacteria to Bach and Back: The Evolution of Minds. W. W. Norton & Company.

Descartes, R. (1641/1996). Meditations on First Philosophy (J. Cottingham, Trans.). Cambridge University Press.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. Phenomenology and the Cognitive Sciences, 4(4), 429-452.

Doerig, A., Schürger, A., Hess, K., & Herzog, M. H. (2019). The unfolding argument: Why IIT and other causal structure theories cannot explain consciousness. Consciousness and Cognition, 72, 49-59.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron, 57(1), 11-23.

Edelman, G. M. (1989). The Remembered Present: A Biological Theory of Consciousness. Basic Books.

Edelman, G. M., & Tononi, G. (2000). A Universe of Consciousness: How Matter Becomes Imagination. Basic Books.

Feinberg, T. E., & Mallatt, J. M. (2016). The Ancient Origins of Consciousness: How the Brain Created Experience. WITH press.

Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. Annual Review of Neuroscience, 32, 209-224.

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

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. Trends in Cognitive Sciences, 4(1), 14-21.

Gallagher, S. (2005). How the Body Shapes the Mind. Oxford University Press.

Ginsburg, S., & Jablonka, E. (2019). The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness. WITH press.

Goff, P., Seager, W., & Allen-Hermanson, S. (2017). Panpsychism. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy.

Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: still alive and well. Neuron, 24(1), 31-47.

Hagmann, P., et al. (2008). Mapping the structural core of human cerebral cortex. PLoS Biology, 6(7), e159.

Harnad, S. (1990). The symbol grounding problem. Physica D: Nonlinear Phenomena, 42(1-3), 335-346.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. Nature Reviews Neuroscience, 7(7), 523-534.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron, 76(6), 1210-1224.

James, W. (1890/1950). The Principles of Psychology. Dover.

Jekely, G., Godfrey-Smith, P., & Keijzer, F. (2015). Reafference and the origin of the self in early nervous system evolution. Philosophical Transactions of the Royal Society B, 370(1684), 20150084.

Jones, E. G. (2007). The Thalamus (2nd ed.). Cambridge University Press.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (2013). Principles of Neural Science (5th ed.). McGraw-Hill.

Kauffman, S. A. (1993). The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press.

Kim, J. (1998). Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation. WITH press.

Kim, J. (1999). Making sense of emergence. Philosophical Studies, 95(1-2), 3-36.

Kim, J. (2005). Physicalism, or Something Near Enough. Princeton University Press.

Kleiner, J., & Hoel, E. (2021). Falsification and consciousness. Neuroscience of Consciousness, 2021(1), niab001.

Koch, C. (2004). The Quest for Consciousness: A Neurobiological Approach. Roberts & Company.

Koch, C., & Hepp, K. (2006). Quantum mechanics in the brain. Nature, 440(7084), 611-612.

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

Lagercrantz, H., & Changeux, J. P. (2009). The emergence of human consciousness: from fetal to neonatal life. Pediatric Research, 65(3), 255-260.

Lamme, V. A. (2006). Towards a true neural stance on consciousness. Trends in Cognitive Sciences, 10(11), 494-501.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends in Neurosciences, 23(11), 571-579.

Lane, N. (2015). The Vital Question: Energy, Evolution, and the Origins of Complex Life. W. W. Norton & Company.

Lane, N., & Martin, W. (2010). The energetics of genome complexity. Nature, 467(7318), 929-934.

Laureys, S., et al. (2006). Auditory processing in the vegetative state. Brain, 129(6), 1626-1630.

LeDoux, J. E. (2015). Anxious: Using the Brain to Understand and Treat Fear and Anxiety. Viking.

Lutz, A., Greischar, L. L., Rawlings, N. B., Ricard, M., & Davidson, R. J. (2004). Long-term meditators self-induce high-amplitude gamma synchrony during mental practice. Proceedings of the National Academy of Sciences, 101(46), 16369-16373.

Margulis, L. (1970). Origin of Eukaryotic Cells. Yale University Press.

Martin, W., & Koonin, E. V. (2006). Introns and the origin of nucleus–cytosol compartmentalization. Nature, 440(7080), 41-45.

Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. Neuron, 105(5), 776-798.

Massimini, M., et al. (2005). Breakdown of cortical effective connectivity during sleep. Science, 309(5744), 2228-2232.

Maturana, H. R., & Varela, F. J. (1980). Autopoiesis and Cognition: The Realization of the Living. D. Reidel.

Melloni, L., et al. (2007). Synchronization of neural activity across cortical areas correlates with conscious perception. Journal of Neuroscience, 27(11), 2858-2865.

Menzel, R. (2012). The honeybee as a model for understanding the basis of cognition. Nature Reviews Neuroscience, 13(11), 758-768.

Merker, B. (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. Behavioral and Brain Sciences, 30(1), 63-81.

Mesulam, M. M. (1999). Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. Philosophical Transactions of the Royal Society B, 354(1387), 1325-1346.

Morowitz, H. J. (1992). Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis. Yale University Press.

Nagel, T. (1974). What is it like to be a bat? The Philosophical Review, 83(4), 435-450.

Nieuwenhuys, R., ten Donkelaar, H. J., & Nicholson, C. (1998). The Central Nervous System of Vertebrates. Springer.

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.

Owen, A. M., et al. (2006). Detecting awareness in the vegetative state. Science, 313(5792), 1402.

Panksepp, J. (1998). Affective Neuroscience: The Foundations of Human and Animal Emotions. Oxford University Press.

Papineau, D. (2001). The rise of physicalism. In C. Gillett & B. Loewer (Eds.), Physicalism and its Discontents (pp. 3-36). Cambridge University Press.

Penrose, R. (1994). Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford University Press.

Petri, G., et al. (2014). Homological scaffolds of brain functional networks. Journal of the Royal Society Interface, 11(101), 20140873.

Popper, K. R. (1935/2005). The Logic of Scientific Discovery. Routledge.

Prigogine, I. (1980). From Being to Becoming: Time and Complexity in the Physical Sciences. W. H. Freeman.

Raichle, M. E. (2015). The brain's default mode network. Annual Review of Neuroscience, 38, 433-447.

Raichle, M. E., et al. (2001). A default mode of brain function. Proceedings of the National Academy of Sciences, 98(2), 676-682.

Rolls, E. T., & Deco, G. (2010). The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function. Oxford University Press.

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

Russell, B. (1921). The Analysis of Mind. George Allen & Unwin.

Ryle, G. (1949). The Concept of Mind. Hutchinson.

Sämann, P. G., et al. (2011). Development of the brain's default mode network from wakefulness to slow wave sleep. Cerebral Cortex, 21(9), 2082-2093.

Schiff, N. D. (2008). Central thalamic contributions to arousal regulation and neurological disorders of consciousness. Annals of the New York Academy of Sciences, 1129(1), 105-118.

Schreiber, T. (2000). Measuring information transfer. Physical Review Letters, 85(2), 461-464.

Schrödinger, E. (1944). What is Life? Cambridge University Press.

Searle, J. R. (1983). Intentionality: An Essay in the Philosophy of Mind. Cambridge University Press.

Searle, J. R. (1992). The Rediscovery of the Mind. MIT Press.

Seeley, W. W., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. Journal of Neuroscience, 27(9), 2349-2356.

Sherman, S. M., & Guillery, R. W. (2006). Exploring the Thalamus and Its Role in Cortical Function (2nd ed.). WITH press.

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? Neuron, 24(1), 49-65.

Sporns, O. (2011). Networks of the Brain. WITH press.

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. Journal of Cognitive Neuroscience, 21(3), 489-510.

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

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

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

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

Stegemann, W. (2025b). Consciousness as an Expression of Autocatalytic Stability – A New Index for Neuronal Coherence. https://doi.org/10.5281/zenodo.17532218

Stegemann, W. (2025b). From Principle to Function: The Causal Core in the Brain. https://doi.org/10.5281/zenodo.15608621

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

Stegemann, W. (2025b). The Causal Core as a Structural Principle of Free Will. Zenodo. https://doi.org/10.5281/zenodo.17057739

Stegemann, W. (2025b). Valence as a universal principle of binding and weighting. Zenodo. https://doi.org/10.5281/zenodo.17057807

Stein, B. E., & Meredith, M. A. (1993). The Merging of the Senses. WITH press.

Striedter, G. F. (2005). Principles of Brain Evolution. Sinauer Associates.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? Behavioral and Brain Sciences, 30(3), 299-313.

Südhof, T. C. (2013). Neurotransmitter release: the last millisecond in the life of a synaptic vesicle. Neuron, 80(3), 675-690.

Tang, Y. Y., Hölzel, B. K., & Posner, M. I. (2015). The neuroscience of mindfulness meditation. Nature Reviews Neuroscience, 16(4), 213-225.

Thompson, E. (2007). Mind in Life: Biology, Phenomenology, and the Sciences of Mind. Harvard University Press.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. Nature Reviews Neuroscience, 17(7), 450-461.

Ulrich, M., Keller, J., Hoenig, K., Waller, C., & Grön, G. (2016). Neural correlates of experimentally induced flow experiences. NeuroImage, 86, 194-202.

Varela, F. J., Thompson, E., & Rosch, E. (1991). The Embodied Mind: Cognitive Science and Human Experience. WITH press.

Vicente, R., Wibral, M., Lindner, M., & Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. Journal of Computational Neuroscience, 30(1), 45-67.

von der Malsburg, C. (1994). The correlation theory of brain function. In E. Domany, J. L. van Hemmen, & K. Schulten (Eds.), Models of Neural Networks II (pp. 95-119). Springer.

Whitehead, A. N. (1929). Process and Reality. Macmillan.

Whitfield-Gabrieli, S., & Ford, J.M. (2012). Default mode network activity and connectivity in psychopathology. Annual Review of Clinical Psychology, 8, 49-76.

Zurek, W. H. (2003). Decoherence, einselection, and the quantum origins of the classical. Reviews of Modern Physics, 75(3), 715-775.