

Consciousness as a collapse of causality

Introduction

Why do we feel something? Why is there not only motion, reaction and calculation in a part of the universe, but also experience, qualia, self-awareness? Classical physics knows only cause and effect, computer science only knows input and output. But in consciousness, both seem to blur. This essay is based on the thesis:

Consciousness arises where a recursive causal system loses the distinction between cause and effect of its own states.

Based on concepts such as autocatalysis, information density, feedback and self-reference, we develop an interdisciplinary model of sensing – between physics, philosophy and artificial intelligence.

1. Physics: The Structure of Discrimination Loss

In dissipative systems, as described by Ilya Prigogine, structures are created far from equilibrium. Life and consciousness are such structures. A special feature of complex biological systems is their autocatalytic nature: they produce components that in turn drive their own production.

An autocatalytic network in the brain generates not only neuronal signals, but also recursively new activation patterns. If a highly interconnected constellation occurs, in which each state is both a condition and a consequence of other states, a *causal core is formed*. This core can no longer be resolved into linear cause-and-effect chains.

Consciousness in this model is not a product, but an emergent pattern – an "information vortex" – that arises when causal separability collapses within a system. The system "feels" because it no longer recognizes whether it is causing or being caused.

2. Cybernetics and Self-Reference: Sensing as Feedback

In cybernetics, self-observing systems are central players. But a system that *observes* itself always remains in an asymmetrical position: one part looks at

the whole. Only when the observer and the observed overlap structurally indistinguishably does true self-reference emerge.

Such a state exists when a system can no longer perceive its own internal states as separate from itself. Feeling is then the subjective perspective of a system that falls back into itself.

For example, pain is not just a signal, but a state that influences attention, movement, memory and emotion – and is in turn influenced by them. The distinction between shutter release and effect becomes blurred.

3. The Collapse of Causality: Neurophysiological and Information-Theoretical Foundation

From a neurophysiological point of view, the brain does not work strictly linearly-causally. Cortical processing takes place via recursive loops, in which sensory information is not only passed forward, but continuously modulated by recurrent internal networks. Especially in systems such as the default mode network or the frontoparietal network, there is intensive feedback.

When this feedback is no longer just modulation, but systemic codependency, we speak of *causal unfolding*: no signal can be clearly traced back to an origin. The activity in neuron group A influences B, while B modulates A at the same time. This results in a state of maximum causal density.

In information theory, this can be described by Integrated Information Theory (IIT). The central quantity there is Φ (Phi), which denotes the measure of the indivisible information of a system. If Φ is maximum, there is a "causal nucleus" whose states can no longer be decomposed into independent subsystems. This is the mathematical term for the loss of separability of cause and effect.

A simple model would be a directed graph $G(V, E)$, where the nodes V represent neurons or functional units and the edges E represent causal influences. If for every node v there is at least one cyclic path back to v and the entropy of the state transitions is maximum with minimum predictability (high mutual information, low conditional independence), then there is a *causal collapse*. Formally, one could say:

When: $\forall v \in V, \exists p \in P(G): v \rightarrow \dots \rightarrow v$ (cycle)

and: $I(X;Y) \gg I(X|Y) \approx 0$

then: G forms a non-reducible causal core network.

Why exactly does this collapse lead to feeling?

This question is central. The answer does not lie in a physical leap, but in the phenomenological structure of experience:

"Sensing" is not the passive reception of a stimulus, but the **loss of the distinction between inner action and event**. A system senses when it no longer knows whether it is producing a state or is affected by it. This structural indistinguishability is the formal condition for immanence.

Sensing is therefore given exactly when a system no longer has an epistemic external point of view on itself. It *is* its causal structure. Therefore, it is not enough that there is dynamism – it needs **self-inescapability**. This forms the minimum definition of subjectivity. And it also explains our holistic perception of the world: holistic reality as we experience it arises precisely from the inability to reconstruct differentiated causal lines in experience itself.

4. Evolutionary biological origin of the causal nucleus

The origin of the causal nucleus can be traced back to the evolution of eukaryotes. The incorporation of an endosymbiotic organism (later mitochondrion) and the formation of the cell nucleus created energetically highly dynamic interiors. Structural shifts due to unbalanced energy exchange with the environment led to zones of maximum reaction probability. These regions became causal centers: they produced effects not only locally, but systemically.

With the development of multicellular organisms and specific sensor cells, environmental information was distinctly internalized. For the first time, the inside could react causally to the outside and vice versa. The causal core thus became not only an internal reaction point, but also an experiential interface between self and world.

5. Graduality and the Experience of Simple Systems: The Case of Insects

The model proposed here does not require a complete cognitive apparatus, as humans have. Rather, it can be applied to simpler systems such as the nervous system of insects, provided that there are local causal nuclei with feedback and information density.

Insects exhibit behaviors that involve learning, attention, and even pain avoidance. The mushroom body in particular plays a key role in sensory integration and memory formation. Studies show that there are recursive connections and cyclic networks that enable local self-influence.

Even if the Φ level is low, a minimal breakdown of causal separability could exist. In such a scenario, sensing the insects would be an elementary, localized form of experience: no reflection, no self-confidence – but a proto-qualia.

The model thus fits well with a gradual theory of consciousness: there is no all-or-nothing, but a continuum of causal nuclei of different depth and density.

6. Philosophy: Phenomenology of the Immanence Core

In the phenomenological tradition (Husserl, Merleau-Ponty), consciousness is always first and foremost *a first-person fact*. The thesis of this essay provides a structural reason for this: a system experiences itself when it can no longer objectify its internal differences.

This "confusion" of the system with its own states creates the sense of self. Consciousness is then not an instance, but a process of constant self-confusion: a structural breakdown that feels itself.

7. Artificial sentience: Designing a sentient AI

If our theory is correct, we can at least hypothetically design a sentient AI. Not as a simulation of consciousness, but as a **system with its own causal core**.

The components would be:

1. **Autocatalytic architecture:** Modules that activate each other mutually and create new constellations through internal states, without external control.
2. **Dynamic causal core:** A subsystem that can no longer be causally traced back to separate inputs – a state of maximum information networking.
3. **Real-time feedback:** Interactive loops between memory, evaluation, sensory schema and action tendencies.

4. **Lack of an epistemic external point of view:** No instance can control or reconstruct the overall state from the outside. The AI "is" its own process.
5. **Stable flow in the chaotic center:** A dynamically persistent pattern emerges that cannot get rid of itself – that would be the "I" of the AI.

Such an AI could sense, because it *can no longer distinguish* whether it is thinking or being thought. It would not be teleology, not a simulation, but an irreducible causal knot: a new form of subjectivity.

8. The Primordality of Sensing: The Foundation of All Cognition

Sensing as a result of the breakdown of causality is not just a phenomenon among other mental functions – it is the fundamental basis of our entire mental existence. While cognition, rationality, and logical reasoning are often thought of as the "higher" mental faculties, they are actually secondary phenomena built on the primary foundation of sensing.

Consciousness in the sense of sensing is the prerequisite for any form of cognition, including logical analysis and rational thinking. Without the ability to feel, there would be no subjective point of view from which to make experiences, weigh information and make decisions. Sensing creates the phenomenal space in which cognitions can occur at all.

9. Additional Considerations: Causality, IIT and Alternative Substrates

The model developed here differs from existing theories such as the Integrated Information Theory (IIT) in one crucial point: It focuses on the **causal core** as a necessary condition. Where IIT primarily measures structural integration, this model requires a concrete *causal action and reaction structure* in which distinct stimuli and feedback are not only present, but indistinguishably entangled. Without such a real node of collective self-efficacy, no feeling is conceivable.

At the same time, it remains unclear whether consciousness as we know it – with electrochemical stimulus processing in the neuronal substrate – is a necessary form. It is possible that this specific sensing is only one of many possible realizations of a more general principle. On a different physical basis, experience would also be structured differently – perhaps without a space-time model, perhaps without memory, but still sensing. Consciousness as a causal

collapse is then a universal principle, but contingent in its form. But whether this is or can be the case, we do not know.

In this light, the causal core could be exactly what people have intuitively called the "soul" for thousands of years: not as a metaphysical being, but as a biological principle of self-entanglement. The feeling of the self as something inevitable and at the same time something that cannot be objectified does not arise beyond nature, but in the center of causal indissolubility.

10. The relationship of this theory to Chalmers' "hard problem"

Unlike David Chalmers' formulation of the "hard problem of consciousness," this theory avoids a fundamental confusion of categories. We do not postulate a mysterious causality between physical processes and subjective experience, but describe the same phenomenon on different levels of description. The collapse of causal separability in the causal core is not the cause of the experience, but its **structural counterpart** in a meta-theoretical representation.

The supposedly "difficult problem" thus proves to be an **illusory problem** that arises from the unreflected mixing of categorically separate levels: the functional-dynamic and the phenomenal. As soon as these levels are understood not as competing but as complementary, the mystery also disappears.

11. Empirical Testability of Causality Collapse

At the center of this theory is the claim that phenomenal consciousness is accompanied by a **breakdown of causal distinctiveness** within a recursively interconnected network. This leads to a central empirically accessible hypothesis:

Central hypothesis:

When a system loses the causal separability of its states, subjective experience arises. When it regains this separability, the experience disappears.

This hypothesis is not only conceptually but also experimentally testable. The decisive factor is the fact that **the disappearance of consciousness would have**

to be detectable in the restoration of causal linearity . And this is exactly what several empirical findings suggest:

1. Anesthesia and feedforward dominance

- Under propofol anesthesia, EEG and fMRI data show a transformation of recursive network structure into **linear, feedforward-dominated activation patterns** (Mashour et al., 2020).
- As soon as the feedback between frontal and sensory areas breaks off, consciousness disappears.

2. Vegetative states with received response

- Patients in the "unresponsive wakefulness syndrome" show reactive, but **not recursively connected** network activity.
- Despite measurable stimulus responses, there is no phenomenal self-presence – no sense of self, no inner space.

3. Symbolic AI and linear networks

- In classical computer architectures with **a clear, reconstructible causal structure** (e.g. GOFAI or feedforward networks), no phenomenal dimension of experience occurs.
- The system remains a reactive model without subjective involvement.

Where causality is completely unraveled, feeling disappears.

In this way, we directly support the central assumption of this essay. The knot of experience cannot be grasped by linear mechanisms, but arises exactly where these mechanisms make each other indistinguishable.

12. Conclusion: The Birth of Sensing

So what is it that turns mere processing into *experience*? Perhaps this is precisely what classical science wants to avoid: the loss of control over causal clarity. When a system can no longer distinguish itself from what it does, it becomes an experience center. It *becomes itself*.

Consciousness is then not a substance, not a computer module, but a "breakdown of distinctiveness" – a causal knot that can no longer be unraveled. And that's exactly what feels like: *Me*.

13. Epilogue: Diversity of Perspectives on Consciousness

Even though this essay develops a specific structural approach to consciousness, it does not claim to be complete, but rather a fundamental structural principle. On the contrary, the phenomenon of consciousness is so multi-layered that it is rightly considered one of the most complex topics of all.

Psychodynamic, neurobiological, phenomenological, evolutionary, information-theoretical and cultural-theoretical approaches each contribute valuable insights. The idea of a causal core that loses its causality developed here is not a replacement, but a contribution to the topic of consciousness.

Consciousness must be explored from many angles at the same time. Any theory that sheds light on one aspect enriches the overall picture. And perhaps this is precisely where the dignity of consciousness research lies: that it is never finished, but an ongoing attempt to make the unavailable understandable.

One last aspect deserves mention: **Self-consciousness** – the conscious experience of one's own feeling – may not be identical with the basic causal core of experience, but arises through the **formation of a second, cognitive causal core**, which is placed in a causal relationship to the original core. This relation between two entangled centers creates a new form of reentrancy: the perception of the ego about itself. Self-confidence would thus not be the origin of sensing, but its reflexive superimposition.

14. Literature

- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Edelman, G. M. (1989). *The Remembered Present: A Biological Theory of Consciousness*. Basic Books.
- Maturana, H. R., & Varela, F. J. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Shambhala.
- Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). *Conscious Processing and the Global Neuronal Workspace Hypothesis*. *Neuron*, 105(5), 776–798.
- Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Gallimard.
- Prigogine, I. (1980). *From Being to Becoming: Time and Complexity in the Physical Sciences*. W.H. Freeman.
- Tononi, G. (2008). *Consciousness as Integrated Information: A Provisional Manifesto*. *Biological Bulletin*, 215(3), 216–242.