Consciousness - Expanding the Model of Causal Collapse

Abstract:
The model of *causal collapse* describes consciousness as a state of maximally inseparable causal entanglement in neuronal systems (Stegemann 2025). In this article, the original model is expanded in an interdisciplinary way. First, neurophysiological correlates and measurable indicators of causal collapse are explained, such as the **Perturbational Complexity Index (PCI)** from *TMS-EEG* experiments, functional connectivity patterns, and characteristic differences between states of consciousness (wakefulness, deep sleep, REM sleep, psychedelic states). Secondly, there is a systematic differentiation of different **types** of consciousness – from *primary consciousness* to *self-consciousness* and *intentionality* to *affective consciousness* – each reconstructed from the structure of causal collapse. Third, the extended model is theoretically positioned in comparison to competing approaches such as *Integrated Information Theory (IIT), Global Workspace Theory (GWT)* and plasticity-based models. Despite different perspectives, it turns out that the **causal core** of consciousness can be understood as a decentralized, dynamic network state that offers both consistent theoretical description and empirical testability. Finally, possibilities for connection to empirical research and implications for artificial consciousness systems are discussed.

1. Introduction

The question of how **consciousness** arises from biological brain dynamics is the subject of intensive research. The *model of causal collapse* provides an integrative explanatory approach by defining consciousness as a special dynamic state in which neuronal cause-and-effect relationships **intertwine so densely** that they can no longer be broken down into separately causal sub-processes. This phenomenon – **causal collapse** – occurs when a measure of causal separability (hereafter Σ) falls below a critical threshold value Σ_c, which corresponds to maximum inseparability. Formally, the conscious state of collapse can be described by the **inequality** $\Sigma < \Sigma_c$. In such moments, the brain forms a *causal nucleus* in which neuronal processes are highly integrated and simultaneously differentiated in a way rich in information. Theoretical considerations from computer science and neuroscience support the assumption that consciousness arises precisely when a system can simultaneously produce **integrated** and **differentiated** activity patterns. In other words, a "sweet spot" must be reached between total fragmentation and mere uniformity – a state **of maximum complexity** in which the whole is greater than the sum of its parts. The model of causal collapse ties in here and

offers a framework in which neurobiological findings, phenomenological descriptions and systems theoretical concepts can be united.

In the original *Integrated Model of Causal Collapse*, a multidimensional classification scheme was proposed to describe manifestations of this collapse in space, time, and intensity. For example, a distinction was made between **local**, **modular** and **global** collapse events (spatial dimension), between **transient** (short-term) and **chronic** (long-lasting) collapse states in the temporal dimension, as well as between **weak** and **strong collapse** Collapse depending on the degree of inseparability (intensity dimension). In the state of consciousness, there is typically a *global*, metastable and strong collapse, i.e. large parts of the brain form an integrative causal core that neither immediately decays nor is completely rigid. It is important to emphasize that this core is **not limited to a specific location**, but results from the *coordinated activity of distributed networks*. Modern brain research findings show that aspects of **the self and experience** are tied to the dynamically coupled activity of a large-scale network, in particular the **default mode network (DMN)** and related nodes. The DMN primarily comprises medial prefrontal cortex (mPFC) and posterior singular cortex (PCC) as well as other cortex areas and acts as an integrative hub. It has a unique ability to bring together information from different areas of the brain. Such a distributed *core dynamic* thus fulfills the role of a causal core: it integrates sensory inputs, memory content and internal states into a coherent overall experience that is indivisible to the system.

The aim of this paper is to substantially expand the model of causal collapse along three axes: **(1)** by a closer link with empirical neurophysiology, **(2)** by reconstructing different forms of consciousness from the collapse structures, and **(3)** by comparing and differentiating it with other leading theories of consciousness. The aim is to make the model more theoretically consistent and empirically compatible. In the following, the neurophysiological foundation of the causal collapse nucleus is first dealt with in order to create measurable points of contact (Chapter 2). Subsequently, the systematic differentiation of primary consciousness, self-consciousness, intentionality and affective-emotional consciousness is carried out on the basis of the model (chapter 3). In Chapter 4, the extended model is positioned against Integrated Information Theory (IIT), Global Workspace Theory (GWT) and plasticity-oriented approaches. Finally, Chapter 5 summarizes the most important points (conclusion) and outlines future research questions.

2. Neurophysiological foundation of causal collapse

A central concern in the extension of the model is the link with **empirically measurable variables** of brain activity. If the occurrence of consciousness is indeed accompanied by a causal collapse in the brain, it should be possible to find corresponding markers in neurophysiological data. In recent years, important indicators have been developed that are considered a *fingerprint* of conscious states. These include, in particular, measures of **integrative complexity** such as the Perturbational Complexity Index (PCI), neuronal **connectivity patterns** in resting EEG/fMRI, and characteristic changes in these variables when changing the state of consciousness (e.g., during the transition from wakefulness to deep sleep, under anesthesia, or altered states of consciousness induced by psychedelics).

2.1 Perturbational Complexity Index (PCI) and critical causal entanglement: PCI is a measure that quantifies the combined **integration and differentiation** of cortical reactions. It is determined experimentally by using *transcranial magnetic stimulation (TMS)* to specifically stimulate the cerebral cortex and the resulting EEG response patterns are recorded and compressed. The idea behind this is that a conscious brain will respond to a local impulse with a complex, globally distributed activity pattern, whereas an unconscious (e.g., deeply asleep or anesthetized) brain will only show simple, localized responses. In fact, it has been shown that PCI consistently reaches high levels in awake subjects, while it drops drastically during deep sleep or under anesthesia. Casali et al. (2013) report that the PCI was able to reliably assign individuals to their state of consciousness – high scores in wakefulness and REM dream phases, significantly lower scores in non-REM sleep and vegetative states. These findings fit seamlessly with the concept of causal collapse: High PCI values indicate a state in which widely distributed brain areas cooperatively form a common complex pattern – which corresponds to a strong global collapse in the model. Conversely, a low PCI signals a loss of global integration (local reactions dominate) and thus the absence of a causal core. **TMS-EEG studies** by Massimini and colleagues provide impressive evidence in this regard: In deep sleep, a TMS impulse evokes only local slow waves in the immediate region, while in the waking state the same impulse triggers a widely propagating cascade of differentiated EEG activity. The brain in the conscious state thus behaves like a *single, inseparable system* that reacts to a disorder as a whole – a characteristic feature of causal collapse.

In addition to the TMS-supported PCI measurements**, effective connection analyses** also underpin the idea of a critical causal entanglement. Thus, in the pathological case of a coma or minimally conscious state, the effective

connectivity between brain areas breaks down: perturbations no longer propagate over a large area, which corresponds to a decayed causal nucleus. Current work attempts to calculate the concept of *Integrated Information* (Φ) – closely related to our Σ measure – directly from resting fMRI data. Initial results indicate that healthy awake brains have higher integrated information than unconscious states, which is consistent with the collapse model.

2.2 Functional connectivity in different states of consciousness:
Another approach to empirical testing of the model is the analysis of the *functional connectivity* of the brain, i.e. the statistical interactions between regions under different states. Here there are clear **differences in profile** between fully conscious and altered states of consciousness, which can be interpreted in the light of the causal collapse.

- **Waking state vs. deep sleep:** In the waking resting state, there is typically a pronounced **network integration** – in particular, networks such as the default mode network (DMN) and fronto-parietal networks are closely functionally coupled and form "hubs" of global communication. In deep sleep (non-REM), on the other hand, this network fragments: Studies show that *default mode* connectivity is drastically reduced and communication tends to be limited to local clusters. Specifically, it was observed that in deep sleep the frontal cortex is functionally decoupled from the posterior cortex, which fits the lack of holistic consciousness content. The model would interpret this to mean that no global collapse occurs during deep sleep (at best local "mini-collapses"), so that no uniform causal core exists – accordingly, consciousness is missing. Conversely, the fully formed DMN hub in the waking state implies an *integrated collapse across cognitive and sensory domains*, i.e. a persistent causal core that connects self-perception and environmental context. Interestingly, the degree of DMN integration correlates with the intensity of consciousness: In patients with impaired consciousness (e.g., minimally conscious state), connectivity in the DMN is reduced in a graduated manner according to the level of consciousness.
- **REM sleep (dreaming):** REM sleep is a special case, as there is a high level of internal awareness (dreams) despite sleep. Neurophysiologically, an intermediate pattern emerges in the REM: Certain cortical networks (associative visual areas, limbic structures) are highly active and sometimes well connected, while frontal control networks are relatively dampened. The collapse model logically describes the dream state as an *episodic, middle collapse with fluctuating global expansion*. This means that at times the neural dynamics reach collapse conditions (which

explains the intense dream experiences), but fall apart again due to a lack of prefrontal stabilization. Empirically, this is in line with the fact that PCI levels in REM sleep are lower than in wakefulness, but significantly higher than in deep sleep and sometimes reach into the range of minimal consciousness. Global integration is thus there, but less stable and purposeful (which is subjectively reflected in the bizarre, changeable nature of dreams).

- **Psychedelic states:** Altered states of consciousness induced by psychedelic substances (such as LSD, psilocybin) offer an insightful test field for the model, as the conscious experience does not disappear, but changes qualitatively. Functional imaging under psychedelics shows characteristic network changes: In particular, **the internal connectivity** of established networks such as the DMN **is reduced**, while at the same time **the inter-network connectivity** increases. Simply put, psychedelics dissolve rigid functional modules and promote more mixed global communication. For example, Carhart-Harris et al. (2016) report a decrease in coherence within the DMN in LSD (especially between medial prefrontal and posterional cingulate cortex) with a simultaneous increase in connections between DMN and normally separated networks. The brain is in a more entropic, flexible state. Within the framework of our model, this can be interpreted as an *alternative collapse*: instead of a clearly defined causal core that keeps certain contents stable (e.g. the sense of self in the normal waking state), a broad-based collapse arises under psychedelics, which leads to unusual couplings. The collapse is **global**, but less **focused**, which could subjectively explain phenomena such as **ego dissolution** and **synesthesia**. In fact, the disintegration of the DMN under psychedelics correlates with the degree of ego dissolution: the more the normal DMN core activity dissolves, the more the person reports the feeling that the self is dissolving. In the model term, this means that the usual *self-referential causal core* is replaced by an atypical collapse that intertwines sensory, emotional and memory contents differently. Notably, studies also measure an increase in **the entropy** of brain activity (broader frequency spectrum, more variable connectivity), which fits well with the picture of a less constrained, exploratory collapse state.
- **Meditation and flow states:** There are also gradual variations within normal wakefulness. For example, flow states **(e.g. being completely absorbed in an activity) are neurophysiologically characterized by** high functional integration with simultaneous weakening of **self-referential** activity. According to the model, flow leads to a *global, strong collapse* with merged attention and action, but temporarily reduced self-

monitoring – which is subjectively experienced as a "loss of sense of time and self" with maximum action efficiency. **Meditative states**, on the other hand, can take on different patterns depending on the meditation technique. Experienced meditators often show **increased connectivity between attention networks and DMNs** in EEG/fMRI, i.e. an unusual coupling of systems that are normally active in opposite directions. For example, in long-time resting practitioners, there is an increased functional connection between the dorsal attention network (frontoparietal) and medial DMN regions. This can be interpreted as an integration of meta-attention and self-awareness – in the model, a *controlled collapse* that allows for both deep immersion (low external distraction) and persistent meta-awareness. Meditation thus seems to be able to specifically influence certain parameters of causal collapse: for example, **temporal stability** (through practice, an observer consciousness can be maintained over a longer period of time) and integration **via domains** (sensory, interoception, self-modeling merge into a holistic awareness). Neurophysiologically, this is accompanied by specific synchronization patterns (e.g., increased gamma coherence fronto-parietal) and greater resilience of the DMN against interruption. The collapse model can integrate such findings by understanding meditation as training to keep a *causal nucleus in a desired attractor state* – active navigation in the phase space of the brain, so to speak (see Chapter 4.1 on attractor dynamics).

Recent studies show, for example, a significant reduction in DMN coherence (ReHo) with a simultaneous increase in intermodular connectivity in psilocybin (Carhart-Harris et al., 2014), which the model interprets as a dedifferentiated, broadly networked collapse. Similarly, EEG studies on meditation (Lutz et al., 2004) show that experienced practitioners maintain high-frequency gamma synchrony fronto-parietal – a stable, coherent collapse nucleus across metacognitive networks. Both cases show that the qualitative structure of collapse syndrome varies in a neurophysiologically measurable way.

In summary, various objective measures – PCI, functional connectivity, signal entropy – support the idea that **conscious states are associated with globally integrated, complex brain** dynamics, while unconscious or less conscious states exhibit fragmentation and simplicity. The extended model of causal collapse makes these correspondences explicit: The **causal core** of a conscious state manifests itself neurophysiologically in a dynamic network that is (a) distributed over a large area, (b) highly feedback, (c) information-rich, and (d)

relatively stable against disturbances. This can be measured, among other things, by high complexity measures (PCI, Φ) and robust, far-reaching causal interactions. In the next chapter, it will be examined how *different forms of consciousness* – from raw feeling to reflected self-perception – can be interpreted as specific collapse configurations.

2.3 Operationalization of the Σ(t) measure

The model of causal collapse uses Σ(t) as a measure of the mean separability of the causal structure of a recursive system at a time t. To make this measure empirically usable, we propose to operationalize Σ(t) by causally directed coupling measures such as transfer entropy (TE) or Granger causality (GC) between functional subsystems. Formally, Σ(t) can be defined approximately as:

$$\Sigma(t) = \min_{P \in \mathrm{P}} \left[\frac{1}{|P|} \sum_{(i,j) \in P} \mathrm{TE}_{i \leftrightarrow j}(t) \right]$$

Where P denotes the set of all partitions of the system into disjoint subsystems. Transfer entropy $\mathrm{TE}_{i \leftrightarrow i}$ measures the bidirectional causal coupling between subsystem i and j. The higher Σ(t), the more causally decoupled the parts are; the lower Σ(t), the more the system collapses into a common causal core.

In practice, TMS-EEG, fMRI or ECoG data can be used to record the interactions between brain areas in real time. The calculation of Σ(t) is done by partitioning relevant areas (e.g. frontal vs. parietal), determining the bilateral causal interactions and mean aggregation. This makes it possible to check whether the global causal collapse occurs in certain states (e.g. wakefulness, REM, meditation). This operationalization complements existing measures such as the PCI with an explicitly causally structured perspective.

3. Differentiation of forms of consciousness from the collapse structure

Consciousness is not monolithic; it shows different **levels of quality and aspects** that are relevant from a philosophical and neurocognitive point of view. Four frequently discussed facets are: **(i)** *primary consciousness* (also called core or basal consciousness), **(ii)** *self-consciousness* (reflexive or higher consciousness), **(iii)** *intentionality* (directedness of consciousness towards contents/goals) and **(iv)** *affective consciousness* (the feeling of emotions/valences). The extended model of causal collapse provides a framework for systematically understanding all these forms. Each form of

consciousness can be reconstructed as a specific *expression of the collapse dynamics in the* brain – with characteristic neural *collapse profiles*, each of which includes or emphasizes certain components. In the following, these four types of consciousness will be considered individually and differentiated according to the collapse structure.

3.1 Primary consciousness (core consciousness):
By primary consciousness we mean the basic *"consciousness here-and-now"* that non-human animals also have – the direct experience of sensory impressions and sensations without demanding self-reflection or temporal projection. Damasio refers to this as *core consciousness*, while Edelman speaks of *primary consciousness*, which belongs to many animals. In our model, primary consciousness corresponds to a **modular to global, but episodic collapse** mainly in sensory-emotional networks. Specifically, when sensory areas (e.g. visual, auditory cortex areas) together with basal affective structures (limbic system, thalamus) temporarily form an integrated causal core, a *feeling experience* arises – which includes taste qualia, colours, pain, etc. This nucleus is *modular/global* in the sense that it spans several modules (such as sensory input + evaluation by amygdala), but does not yet keep the whole brain in permanent collapse. It is *episodic*, i.e. it appears situationally and can disintegrate again (for example, when attention wanders or consciousness sinks into sleep). The intensity is usually *medium collapse*: sufficiently inseparable to create a uniform experience, but not maximum in terms of depth of integration. Neuroanatomically, structures such as **thalamo-cortical loops** and integrative nodules in the upper brainstem (reticular formation, tectum) correspond to this, which generate a basic level of awareness. Animal studies suggest that even mammals and birds have a rather complex primary consciousness. For example, despite abnormal neuroanatomy (no neocortex), birds have a highly integrated tectum-basal ganglia-thalamus system, which presumably enables sustained wakefulness – evidenced by complex behavior and perceptual performance. This fits with the observation that *core consciousness* is not exclusively human, but is also present in many animals. In the collapse model, this can be explained by the fact that even those brains achieve a sufficiently strong causal collapse via sensory and affective domains: As soon as the neuronal interactions form an integrated pattern, subjective sensing arises. Primary consciousness is thus the **basic stage in** the model: a collapse that *generates phenomenal experience* (the *"something-is-like" feeling, something it is like*), but still without an elaborate self or narrative content.

3.2 Self-consciousness (reflexive consciousness):
Self-consciousness refers to the ability to be aware of one's own self as subject

and object – i.e. *consciousness that has itself as its object*. This is often equated with the term *higher consciousness* or *reflected consciousness*. In the model, self-confidence arises when the causal collapse not only includes sensory-affective networks, but also **extends to cognitive and metacognitive systems** and **chronically stabilizes**. In practice, this means that the brain forms a sustained global collapse that integrates memory, language and a self-model in addition to momentary perceptions . This *global, strong interdependence* creates the phenomenon of the **self-experience** – a center of experience that carries through time and distinguishes between self and world. Neurally, this is reflected in the **coupling between the default mode network (DMN)** – which is important for self-reference and autobiographical memory – **and fronto-parietal executive networks** – responsible for attention and cognitive control. The DMN has a unique position because it integrates information from a wide variety of sources (interoception, episodic memory, scenario design) as a hub. Studies show that DMN areas (mPFC, PCC) are among the most interconnected nodes in the brain and mediate global connectivity. This connectivity allows me to generate a **higher-order self-model** that summarizes sensory and mental content as *my experience*. In the collapse model, self-confidence thus corresponds to a *chronic, global collapse of high complexity* – the causal nucleus permanently encompasses almost the entire brain and includes self-modeling loops in particular.

Subjectively, this is expressed in the ability to think about oneself (reflection), to locate oneself in the past and future and to introduce a **first-person perspective** into experiences. These properties arise because in the collapsed state, the **prefrontal cortex**, among other things, is closely connected to the hippocampal memory system and the sensory association areas. In this way, current perceptions can be placed in the context of one's own life story – for example, one recognizes a place and knows "I've been there before" or one experiences pain and thinks "something hurts me". Neuropsychologically, this correlates with specific indicators: The **DMN activates during self-centered thought processes** and **deactivates itself in** the event of loss of consciousness (e.g., sleep, anesthesia) in line with the decrease in self-awareness. People with damage to DMN nodules (e.g. due to dementia) often show impairments in self-orientation. All of this supports the view that self-awareness is at its core a product of a *meta-collapse structure*: the collapse virtually collapses *on itself* in that the brain not only integrates data, but also carries along a model of its own integration performance (a kind of self-reference in the network). The model speaks of **recursive collapse** here – the causal dynamics contain a self-representation. In development, self-confidence occurs only at a late stage: Toddlers develop their first constant sense of self (mirror recognition, use of

"I") at about 2–3 years of age, which is related to the maturation of fronto-parietal networks and narrative memory structures. The model reflects this in that a lasting self-awareness only emerges with the differentiation of nested collapse structures and *the stabilization of the self-model through narrative integration*. From a phylogenetic point of view, full self-consciousness seems to be limited to a few species (humans, possibly great apes, dolphins, etc.), which correlates with the presence of sufficiently complex brains. It is important to emphasize that **self-awareness gradually grows out of primary consciousness** as the collapse state increases in depth and scope. The model thus allows the two to be seen not as separate phenomena, but as a continuum: from basal experience (a simple core, e.g. "there is something there") to reflected experience ("**I** experience something, and I know it"). Neuroscientifically, this means the transition from mainly thalamo-cortical loops to an inclusion of prefrontal and parietal areas in the nucleus, culminating in the DMN/executive network node.

3.3 Intentionality (directedness and goal-relatedness):
In philosophy, the term *intentionality* means the "attitude towards something" – that mental states can refer to content or objects (e.g. a thought is about something). In everyday life, we also understand it to be goal-oriented: wishes, intentions, beliefs refer to future actions or states of the world. Brentano regarded intentionality as a characteristic of the psychic: *every mental act is directed towards an object*. How can intentionality be interpreted in the context of causal collapse? In the model, intentionality arises when the collapsed core **encloses** representational structures *that point beyond the here-and-now* – in particular, repositories for concepts, **modelling of the external world** and **planning networks**. In other words, the causal core must not only integrate the current sensory event, but also contain internal *models of something else*, so that consciousness has a *content beyond* itself. Neuroscientifically, two components could be essential here: the **hippocampal-entorhinal system** for imagination and memory content (spatial maps, episodic scenarios) as well as **prefrontal and parietal cortex areas** for action planning and goal representation. When these merge with the conscious core, an *intentional consciousness results*.

An example: Let's consciously imagine going on a trip tomorrow. At this moment, we are not only aware of ourselves, but our conscious content *is about* a possible future. This requires the brain to create a **mental representation** ("trip tomorrow") in a state of collapse and integrate it with the self-model and evaluation instances (one may feel anticipation). This state is causally complex because it combines sensory imagination (images of the trip), temporal projection (tomorrow), and self-reference (I will do it). The

model describes this as *a multilayer collapse with prospective recursion*: multiple causal nuclei (e.g., a core of current self-perception and a simulated core of the anticipated situation) interact recursively. Intentionality thus requires an **expansion of the collapse structure into the dimension of the modeling plane** – the conscious system has "models in it" to which it is directed. The special role of the **prefrontal cortex**, for example, becomes empirically visible here: It is necessary to initiate goal-oriented actions and to direct thoughts to what is absent. Studies show that only consciously perceived stimuli lead to robust activations in prefrontal areas and subsequent action areas. In the Global Neural Workspace, it is postulated that conscious content is needed to be widely available in the brain – for example, to be represented as an intention in the frontal cortex. Our model agrees, but goes further: the **causal collapse connects the sensory representations to these frontal intentions so closely** that an inseparable *intentional field emerges* (one can then no longer just "see the landscape" without at the same time "wanting to go somewhere" – perception itself is imbued with intentionality).

Philosophically, the connection between consciousness and intentionality has led to debates as to whether all consciousness is intentional. Searle argued that many conscious states are intentional, but not all (e.g., pain as a qualitative state without a clear "object"). In our framework, this can be made more precise: a *purely phenomenal consciousness* (like pure pain) would correspond to a collapse without the inclusion of representational networks – i.e. primary consciousness without intentionality. But as soon as the brain relates a painful stimulus to something ("My foot hurts because I stumbled" or "I have to stop the pain"), it becomes intentional content. The model allows for both constellations, so that consciousness and intentionality are closely related, but not strictly identical. Intentionality occurs when the causal core contains **symbolism** and **directionality**, which is common in human brains through language and concepts. In animals without language, consciousness may often be less intentional in the abstract sense (although it can of course also be directed at objects, such as a predator consciously pursuing prey – this is rudimentary intentionality in the space of action). The *theory of mind* – understanding that others have mental states and intentions – would be an even more complex step, in which a model of an alien collapse (of another consciousness) is simulated in one's own collapse. This requires additional nesting and occurs phylogenetically and ontogenetically late (in the child from about ~4 years of age, parallel to the second "social" causal nucleus for the perspective of others).

In summary, **intentionality in the collapse model is the dimension of directed *content***: the conscious causal core includes states that have a meaning for the

system beyond immediate experience – be it the representation of an external object, a remembered episode or a future intention. Technically speaking, the collapse is functionally *expanded here*: in addition to the phenomenal integration node (which makes everything tangible), there are *semantic and pragmatic nodes* that are integrated into the collapse dynamics (e.g. language networks that generate meaning and frontal lobes that maintain action goals). The resulting consciousness is *intentionally interwoven* – it always has a "theme" or "goal". This aspect is examined again in chapter 4.2 when comparing it with global workspace theories, where the focus is on the availability for *reportable action* (a clearly intentional criterion).

3.4 Affective consciousness (emotional feeling):

While intentionality emphasizes *cognitive directionality*, affective awareness focuses on the *quality of feeling*, especially emotions and moods. Affective consciousness is the conscious experience of emotions – e.g. feeling fear, feeling joy, feeling grief – including the associated body sensations and valences (pleasant/unpleasant). Many researchers, such as Jaak Panksepp, see emotions as a fundamental form of consciousness that is phylogenetically very old and deeply anchored subcortically. Panksepp speaks of *primal affects*, which arise from the activation of primordial brain regions (hypothalamus, amygdala, PAG, etc.) and provide conscious emotional tonality. In our model, affective consciousness already played a role in *the primary consciousness* – described there as basic *sensing with valence*. In fact, the earliest conscious states are probably evolutionarily simple positive or negative sensations (pleasure/displeasure), coupled with homeostatic needs. The causal collapse in simple nervous systems is likely to have initially taken place mainly between sensory and affective core structures (e.g. in early vertebrates between the rhomboid brain and the simple cortex).

For a fully developed affective consciousness – as humans know it, where emotions are experienced in a nuanced and reflected way – both **limbic networks** (which generate emotion) **and cortical evaluation and context networks** (which interpret emotion) must be integrated in the collapse. An emotion such as fear initially arises unconsciously as a reaction (amygdala trigger), but becomes a conscious feeling of fear when this activity synchronizes with the cortical core of consciousness. The model postulates that **emotional collapse** occurs when limbic structures (amygdala, hypothalamus, ventral striatum) and cortex transiently form an inseparable unit. As a result, the emotion no longer becomes just a stimulus-reaction scheme, but a qualitatively experienced feeling. Neurophysiologically, this is supported by findings that the **insula and the anterior cingulate cortex** – nodes between the limbic system

and DMN – are activated when feelings become conscious (e.g., conscious pain perception correlates with insula activity).

Affective awareness also has a **motivational dimension**: feelings drive behavior. Thus, the *affective collapse nucleus* also strongly influences the output side of the system (willingness to act via basal ganglio-thalamo-cortical loops). Our model requires that a conscious core should be **causally effective** in behavior – feelings fulfill this par excellence: they modulate decisions, attention and memory. For example, a strong fear takes over the causal core (you can hardly think of anything else, sensory impressions are subordinate to fear) and influences the entire system dynamics (flight reflexes, vigilance, etc.). This corresponds to an *overarching emotional collapse*. On the other hand, consciousness can also regulate emotions when cognitive control networks are involved in the collapse (e.g., by prefrontal inhibition of the amygdala in cognitive reevaluation). A fully affectively conscious system results when the collapse unites the **core areas that generate feelings** and the **areas that** interpret feelings. As a highly developed being, humans do not experience emotions purely "instinctively", but usually embedded in self-reflection ("I am angry *because*...") – that is, affective consciousness is often intertwined with self-confidence and intentionality. Nevertheless, one can pick it out analytically: In the collapse model, for example, one could construct a state *of purely affective consciousness*, where an emotion fills the consciousness without much cognitive content (for example, moments of overwhelming joy or deep sadness in which one is completely absorbed in the feeling). Here you would have a *functional-emotional collapse*: sensory and cognitive modules are carried away and synchronized by the limbic dynamics.

Interestingly, neurochemical changes – e.g. in **antidepressants or drugs** – show strong effects on affective consciousness, which also influences collapse. MDMA ("ecstasy"), for example, increases prosocial feelings and decreases anxiety, which can be interpreted as a shift in the emotional collapse core: The fear network (amygdala) is functionally decoupled, instead reward and social attachment systems (oxytocin, serotonin effects) couple more strongly into the nucleus. The person then feels connected and fear-free – subjectively a changed consciousness. This demonstrates that **chemical modulation** of the nodes involved in the collapse qualitatively alters consciousness. The model can serve as a framework to describe such effects (e.g. "Emotional collapse shifts from fear domain to attachment domain").

In summary, *affective consciousness* in the model is the aspect that arises from the inclusion of evolutionarily old, subcortical **affect circuits** in the causal nucleus. Depending on the involvement of these circuits, one speaks of a

predominantly **emotional collapse** (emotional tone dominates the experience) or, if they are missing, of a purely *noetic* (cognitive) consciousness, which would be rather cold-rational. In fact, it is speculated that there may be forms of consciousness (such as in AI or hypothetical aliens) that have cognitive integration but no feelings – our model would see this as a collapse without a limbic component, which may be unstable or incomplete. In any case, Panksepp's finding that **primary emotions are homologous organized in subcortical networks of all mammals** implies that affective consciousness is one of the earliest manifestations and forms the foundation on which higher forms (self, intentionality) saddle up. Damasio also argues that the *core self* arises from the linking of body signals (feeling of being alive) with sensory content – i.e. is dependent on feelings as a supporting pillar. The **model of causal collapse integrates these perspectives** by considering *affective feeling* as an integral part of the collapsed nucleus, which is present from the beginning and is not disappeared but modulated in more complex forms of consciousness.

Interim conclusion: The four analyzed forms of consciousness can be characterized by different **profiles of causal collapse** . **Primary consciousness**: modular/moderate global collapse, especially sensory-limbic, **self-awareness**: global, persistent collapse incl. self-model networks, **intentionality**: collapse with frontal and memory-based representations directed towards the absent, **affective consciousness**: Collapse with intense limbic involvement, which shapes the quality of feelings. In reality, these facets usually occur in combination – for example, our everyday consciousness has elements of all of them (we experience sensory impressions with emotional coloring, know about ourselves and are mentally related to the world and the future). But analytically, separation helps to understand which *subsystems* must contribute to the collapse in order to produce a certain quality. The model thus creates a bridge between phenomenological **diversity of consciousness** and specific **neural constellations**: For example, self-consciousness is tied to DMN coupling, intentionality to frontohippocampal involvement, affect to limbic wiring, etc. In the next section, we compare how this extended model is positioned against other theories and where there are overlaps or differences.

4. Theoretical positioning: comparison with IIT, GWT and plasticity-based approaches

Consciousness research has produced a number of different theoretical models, each of which emphasizes certain aspects. The extended model of causal collapse can now be considered in the context of these competing approaches. Here we focus on three influential currents: **(a)** *Integrated*

Information Theory (IIT) according to Tononi and colleagues, **(b)** Global *Workspace Theory (GWT)* according to Baars/Dehaene, and **(c)** approaches that focus primarily on *neuronal plasticity and learning* as the basis of consciousness. We discuss similarities and differences in order to work out the unique selling points of our model.

4.1 Comparison with Integrated Information Theory (IIT):
IIT is one of the most formal theories of consciousness and defines consciousness via the quantity Φ (Phi), which quantifies the irreducible **integrated amount of information** of a system. According to IIT, a system has consciousness exactly if (and insofar) as there is a *state with maximum irreducible integration* – the so-called *Main Complex*. At first glance, our collapse model has a lot of similarities with IIT: Both emphasize integration and indivisibility as core criteria. In fact, Σ, the measure of causal separability, could be considered inverse to Φ – the smaller Σ (more inseparable), the higher Φ (more integrated). The KK criterion (causal collapse criterion) of our model also explicitly requires a minimum complexity, "measured by Φ or similar measures". That is, we recognize Φ as a meaningful quantification of what we qualitatively describe as collapse. Consequently, the collapse model shares a central content with IIT: **consciousness is an intrinsic, emergent property of a sufficiently complex and integrated networked system**. Both theories postulate a singular core (IIT: "complex", we: "causal core"), each of which forms the substrate of a single conscious entity.

Nevertheless, there are important differences:

- **Localization vs. distribution:** Classical IIT (version 3.0) often assumes that the location of the highest Φ in the brain – Tononi suspects the *posterior hot zone* (temporo-parietal-occipital cortex) – is decisive. It is argued that posterior associative areas in particular carry consciousness, while frontal areas are less involved (from an IIT point of view, they could rather function as observers). Our model, on the other hand, emphasizes that although the causal core is decentralized, it can still *encompass the entire system* and, especially in the case of self-awareness, also includes prefrontal networks. Empirically, there is discussion here: A recent adversarial study between IIT and GNW found that conscious content is decodeable in both posterior and frontal regions, but IIT predicts that the actual integration occurs posteriorly. Our model would position itself more flexibly here: In simple forms of consciousness (primarily sensory), the collapse nucleus may indeed be posterior, but in self-reflexive states it must be encompassed frontoposteriorly. So we don't see a hard contradiction, but rather a **contextual extension**: IIT's criterion can be

applied to subnetworks, but the full collapse includes more, especially for complex content.

- **Dynamics and Critical Edge:** IIT is a static information theory – Φ is computed for a given system state. The collapse model, on the other hand, emphasizes the *dynamic property* that consciousness is a state on the **edge of chaos**, metastable between order and disorder. We attach importance to the idea of consciousness as an **attractor state** in phase space. In this way, we integrate system-theoretical concepts (e.g. mixture of positive and negative **Lyapunov exponents** – a sign of critical dynamics) into the description. IIT is largely silent on dynamics, except that it implicitly assumes that a system must form a maximum of integration in a stationary setting. Rather, our model sees consciousness as *a process-based collapse* that is maintained by constant feedback. This fits in with observations such as the TMS "ignition" phenomenon mentioned above: Conscious content *ignites* a feedback network. In IIT terminology, one could say: Φ is not simply high, but the system is moving at a critical point where it reacts to perturbations with sustained but structurally changeable activity.
- **External function vs. intrinsic existence:** IIT emphasizes that consciousness exists *for itself* (intrinsic ownership) and does not have to have a special information-processing purpose. The collapse model does not contradict this – we define consciousness just as intrinsically via causal indivisibleness. However, we also emphasize that the collapse core must be *behaviorally relevant*. In other words, in order for us to speak of a "real" conscious core, it should have an influence on the system output (otherwise it would be epistemically difficult to determine and biologically unusual). This is more reminiscent of considerations from the Global Workspace/Prefrontal Cortex models. However, it is more of an additional pragmatic demand than a definitional must. A highly integrated subsystem, which would have no effect in complete isolation, would be unlikely in the biological organism – so the model practically excludes such a thing, which IIT would logically allow.
- **Mathematical formalization:** IIT has an elaborated formal basis (state spaces, concept structures, etc.), while the collapse model has so far argued more qualitatively and conceptually. We are aiming for formalization in the future, for example via an explicit term for Σ that is related to Φ. In 4.2 below, we discuss some formulas of attractor dynamics. However, it is important to note that the **predictions** of both theories are similar in many respects – e.g. both say that with loss of consciousness (sleep, anesthesia) the integrative complexity collapses. The PCI experiments were even derived from IIT in a theory-driven

manner and also fit the collapse model. Where there are differences would be, among other things, the **focus on cortex vs. subcortical contributions**: IIT representatives sometimes have the view that subcortical activities (e.g. brainstem) do not contribute significantly to the conscious content because their Φ is lower in isolation. The collapse model would counter that consciousness is not fully functional without the integration with brainstem/limbic – we assume that the thalamus, for example, as a clock and mediator, is essential to the collapsed nucleus (which is evidence-based, because thalamic stimulation can influence consciousness, and thalamus is active in almost all consciousness networks). Here, then, our approach extends the *physical base* of the nucleus to more than just the cortex. Overall, one can say that the **model of causal collapse is related to the spirit of IIT**, but expands it with a *system-dynamic perspective* and locates it clearly in the **neuroanatomical structure** (including the brainstem and frontal areas). This should make it compatible to be tested more empirically (e.g. by measuring network criticality).

A central difference to IIT is its empirical and technical accessibility: While the calculation of Φ is very complex and can only be applied to simplified model networks, the collapse measure Σ(t) can be estimated relatively efficiently in real data using existing methods of causality analysis (e.g., transfer entropy, Granger). This makes the model more experimentally compatible – for example, for characterizing altered states of consciousness or evaluating artificial systems. Σ(t) can also be transferred to AI systems by measuring internal network couplings and checking whether the collapse of a recursive architecture is a given. IIT remains predominantly theoretical here so far. The collapse model thus allows not only an alternative point of view, but also a concretely testable dynamic of the transition to conscious states.

4.2 Comparison with the Global Workspace Theory (GWT):

The Global Workspace Theory and related models (Global Neuronal Workspace, GNW) understand consciousness primarily as a function of the *global availability of information in the brain*. According to Baars (1988), consciousness is like a spotlight in the theater that shines selected content on the stage (global working memory) so that many specialized agents (unconscious modules) can access it. Dehaene and Changeux transfer this to the brain: A stimulus is consciously experienced when it creates the threshold for global spread – in particular, triggers a late "ignition" mediated by recurrent activity in fronto-parietal networks, which lifts the stimulus into the working

memory and thus makes it accessible for reporting/decision-making. The difference between conscious and unconscious lies in **attention and dissemination**: unconscious processes remain local and short, conscious processes become globally broadcasted and long-lasting (approx. >0.3s).

Our collapse model shares GWT's emphasis on **globality**: we also define consciousness through global integration. The fact that **recurrent loops** are crucial is also emphasized in both approaches. The "late ignition" described in GWT corresponds exactly to the point at which Σ falls below the critical value in our model and a stable collapse occurs – namely then the feedback loops have extended the local processing so much that an indivisible whole is created. GWT thus provides a *functional description*, so to speak, of how the collapse occurs: Through sufficiently strong, balanced feedforward and feedback signals, a **self-reinforcing activity loop** (attractor) is created, which represents the corresponding consciousness content event. In fact, some GNW models define consciousness as precisely this permanently self-amplified state ("neuronal thunderstorm"). This allows GWT to be embedded as a **functional sub-model** in our structural theory model. We would say: The *global workspace* is the operational side of the causal core – what GWT calls "in the workspace" we call "in the collapse core". It's basically the same amount of neurons, just described from a different perspective (GWT: information flow function, collapse model: cause-effect unit).

Nevertheless, there are accent differences:

- **Consciousness as a medium of communication vs. as a property:** GWT emphasizes the **usefulness** of consciousness – why it evolved: It facilitates the flexible coupling and coordination of many brain functions through global exchange. Consciousness therefore has the function of a *broadcast transmitter* or blackboard system. Our model is more neutral in terms of teleology; it describes consciousness rather as **the emergence of a holistic state**. In a way, we provide a more microscopic explanation *of why* global broadcast phenomenally generates consciousness: because this broadcast situation corresponds to a causal state of indivisibleness. GWT itself avoids statements about subjective experience (it only outlines mechanisms that accompany it). This is exactly where the collapse model comes in: When GWT says "a piece of content X is in the global workspace", our model says "the brain is in a state of collapse that has X as an integral part – therefore X is experienced". In other words: GWT describes **conditions**, collapse model describes **essences** of consciousness. Both approaches are therefore **more than rival in** terms of complementarity.

- **Role of the self and deeper structures:** Classical GWT versions (Baars) did not have a distinct concept of the self; Dehaene's GNW has it implicitly via frontal systems, but focus is on *content*. Our model explicitly addresses the self-nucleus (see 3.2). So we integrate GWT with a self-model (similar to some newer extensions of GWT, e.g. Graziano's Attention Schema, which introduce a model of itself as part of the workspace). In addition, as mentioned, our model includes subcortical circuits. The Global Neural Workspace Theory focuses primarily on the cortex (the "ignition loop" between sensory areas and the prefrontal cortex). However, more recent data suggest that the thalamus and basal ganglia are also part of the global network of consciousness (e.g. intralaminar thalamus as a clock generator). Our model sees these as an integral part of the collapse, while GWT often remains somewhat fuzzy here.
- **Consciousness vs. attention:** GWT mixes consciousness with attention mechanisms in places (spotlight metaphor). The collapse model makes a distinction in principle: attention can be understood as a mechanism that promotes collapse (e.g. by pulling top-down signals into the global state) or selectively modulates the content within the collapsed nucleus. But attention is not the same as consciousness – we would say that attention facilitates global collapse, but does not guarantee it (e.g. you can also suddenly consciously perceive something inattentively, or process something unconsciously attentively). GWT takes a similar view, but is often equated in common parlance. Here, our model maintains a clear formal line: **one is conscious when collapse; whether with or without focused attention is secondary.**
- **Empirical tests:** GWT has strong predictions such as the existence of the "P3 wave" in the EEG only with conscious processing, which has been experimentally supported. The collapse model can easily absorb these findings, since such late waves are indicators of global feedback – i.e. indicators of collapse. An interesting potential distinction: There could be states of high Φ (i.e. integration) without typical GWT "ignition" signatures. For example, meditative state: Highly integrated, but no specific P3 because no specific stimulus. GWT wouldn't say much here (no concrete prediction in the absence of stimulus), but our model would (system collapsed anyway, ergo conscious). Such cases could separate theories – but previous tests (mostly stimulus-related) show no direct conflicts.

In summary, it can be said that the **causal collapse model and GWT/GNW are largely consistent**. GWT provides a **functional story** of how collapse occurs

mechanistically (through global recurrent processing and working memory formation), and our model gives GWT an **ontological sharpness** (why this global availability constitutes subjective experience – namely, because the brain then forms a single causal unit). Thus, both perspectives benefit from each other. An empirical difference could be that our model strictly demands that a unity really emerges. GWT is satisfied when information is disseminated – but theoretically you could have a globally distributed system that still consists of two independent parts that are merely synchronized. GWT would treat it as a consciousness, our model would check whether indivisibility actually exists. However, such details are rather philosophical; in practice, both address the same phenomena (integration, recurrence, DMN/executive involvement).

4.3 Comparison with plasticity-based approaches:

As a third category, we consider theories that focus less on momentary network dynamics and more on the *long-term structural adaptation* of the brain – i.e. **learning and plasticity** – as key factors for consciousness. These approaches are heterogeneous, but examples would be: Gerald Edelman's theory of *neural Darwinism* and the *Theory of Neuronal Group Selection*, which sees consciousness as an emergent product of selected, reentrant networks; or hypotheses that consciousness is required for certain forms of synaptic consolidation (Crick & Koch's idea that attention/consciousness helps to detect relevant synaptic changes). consolidate); or that conscious processing is closely linked to **Hebbian learning**, because repeated firing in the global workspace causes synaptic reinforcement. Newer ideas such as *Entropic Brain* (Carhart-Harris) also refer to network "flexibility" and thus indirectly plasticity (where acute entropy is meant).

Our model, as presented so far, focuses on the **current state** of the brain. Nevertheless, it considers plasticity as a *condition* to achieve and maintain this state. In Section 2.1, for example, it has already been suggested that **Hebb's adaptations** in the network shape the attractor landscape in which a collapse of consciousness can take place. Model 3 (BBCG) outlines, for example, that learning by modifying the attractor landscape (Hebb) is part of the concept. So we agree with plasticity-oriented theories to the extent that **consciousness is not statically given, but is shaped by experience**. The development of consciousness (ontogeny) according to our model (Section 2.2, Ontogeny) shows how higher forms of consciousness emerge with growing *plasticity and structural complexity*. For example, only through the **acquisition of language and narrative coherence** (a learning process) does a stable reflected self-awareness become possible. Such statements are in complete agreement with plasticity-oriented views, which emphasize that **experience differentiates consciousness**.

What's the difference? Plasticity models would sometimes claim that consciousness *itself* is a product or by-product of learning mechanisms. The collapse model, on the other hand, sees consciousness as a *current dynamic property* that is prepared by learning mechanisms, but is not reduced to it. One could say: plasticity builds the *stage* on which the causal collapse can take place, but the performance itself (the current conscious experience) is a separate phenomenon of real-time dynamics. However, there are interesting overlaps: Edelman's concept of a **"dynamic nucleus"** in the thalamo-cortical system, corresponding to consciousness, arose within the framework of his theory of neural groups and selection by reentry. He emphasized both **reentrant dynamics** (comparable to our collapse) and **selection** (comparable to our reference to resistance and stabilization through adaptation). The collapse model can virtually unite Edelman's theory: The dynamic core = causal core; the plasticity processes = mechanisms that allow the brain to build and maintain this nucleus.

In terms of empirical evidence: Plasticity-oriented models point to phenomena such as **critical periods** (that a brain may only be able to develop consciousness if it learns correct stimuli in early childhood), or **neurochemical modulation** (which influence wakefulness and plasticity together – for example, acetylcholine high in wakefulness promotes learning and awareness). The collapse model can classify all these factors as *modulators of the readiness for collapse* . For example, a freshly formed neuronal ensemble (after learning) could be more easily integrated into future collapse nuclei because synaptic weights have been adjusted. Conversely, prolonged absence of consciousness (such as in coma) can lead to the breakdown of synaptic connections, which limits the ability to collapse – hence the often difficult return from long comas. Our model would predict that a **certain minimum connectivity** and **plasticity** is required for a system to be able to generate a global collapse at all (see CC criterion: minimum complexity and self-generation are necessary).

One point where we differ is possibly the **immediacy of plasticity for the act of consciousness**. Some researchers (e.g., Lamme in Recurrent *Processing Theory*) argue that the difference between conscious and unconscious perception is that only conscious ones get into late recurrent phases, where they also become relevant for *synaptic changes* (i.e., consciousness marks information for storage). Our model would not contradict this, but see it as a consequence of the collapse: If the global network fires, then Hebb rules will also have a far-reaching effect – for example, the consciously experienced stimulus will be increasingly anchored in episodic memory (known as the consciousness advantage for memory). To interpret consciousness here as a *state of maximum willingness to learn* is interesting: in collapse, all relevant synapses

are in play and can be modulated. Unconscious processing, on the other hand, remains local, so changes remain limited. Thus, our model also integrates with learning functions: causal collapse allows the brain to react coherently to new situations and at the same time reconsolidate the networks.

In summary, plasticity-based models provide important **developmental and learning-related perspectives**, and the collapse model is thus **complementary**. It focuses on the *active state*, while the latter focuses on the *emerging structure*. In the big picture, both unite: consciousness arises in a brain that has been shaped by plastic processes, and this conscious experience in turn leaves traces (plasticity). Our model is based on systems theory perspectives that consider the brain and consciousness as a **circular causal system** (enactivism, autopoietic systems). Causal collapse also has an autopoietic note – it is **self-generating** and **self-sustaining**. Plasticity helps to stabilize this autopoiesis over time (e.g., through narrative integration of the self through life events). In the competition of theories, our model thus takes an **integrating position** : it tries to combine the strengths of IIT (formality of integration), GWT (functional global availability) and plasticist perspectives (development and learning) in a consistent framework, without limiting itself to a single explanatory level.

5. Conclusion

In this paper, we have substantially expanded the model of causal collapse and put it in a new light. Based on the idea that consciousness corresponds to a state of maximum causal inseparability, the following core findings were gained:

- **Neurophysiological anchoring:** Causal collapse can be linked to empirical indicators. High **PCI values** and global, complex TMS-EEG responses signal a collapsed causal nucleus (consciousness), while low values indicate separate, locally remaining dynamics (unconsciousness). Characteristic connectivity patterns underpin this: In waking consciousness, there is an integrative brain network (especially DMN as a hub) that falls apart during deep sleep or under anesthesia. Extraordinary states such as **dreams**, **flow** or **psychedelic experiences** could be interpreted as special modulations of the collapse dimensions (temporal, functional, intensity-related), in line with known neurophysiological changes (e.g. DMN decoupling under psychedelics or DMN attention coupling in meditation). These findings demonstrate the **model's compatibility with empirical research**: Future experiments could specifically search for the "degree of collapse" in different states – for example, by means of combined **perturbation** and **connectivity**

measurements to map the phase space between full consciousness and complete unconsciousness.

- **Differentiated forms of consciousness:** Through the structural-theoretical lens of causal collapse, **primary consciousness**, **self-consciousness**, **intentionality** and **affective consciousness could be** traced back to a common mechanism, but differentiated on the basis of different subnetworks involved. For example, primary consciousness was characterized as *a sensory-limbic collapse without self-reflection* (e.g., an animal's feeling here and now), while self-consciousness represents a *recursive collapse with the inclusion of one's own self-model* (e.g., stable first-person perspective in humans). Intentionality arises when the collapse includes *representational and prospective content* (directedness towards the absent, goals), and affective consciousness when *emotional circuits* are an integral part of the core state (emotional experience). This analysis shows that the model is flexible enough to map qualitative differences in consciousness to quantitative/different **collapse profiles**. In doing so, it provides a framework for understanding pathological changes, for example: A patient with asomatognosia (loss of body awareness) could be described as someone whose collapse nucleus no longer includes the interoceptive areas; someone with depersonalization may have an unstable self-collapse; Autism could be considered as a preponderance of a specific mode of collapse (little involvement of social intentionality, strong focus on sensory patterns). Through this point of view, one can conceptually classify **clinical phenomena and variants of consciousness** and get hypotheses for neuronal correlates (e.g. "DMN integration is disturbed in depersonalization" etc., which would be empirically testable).
- **Theoretical integration and demarcation:** Compared to IIT, GWT and plasticity-oriented approaches, the extended model proves to be **integrative and synergistic**. It shares with IIT the emphasis on irreducible integration ($\Phi \sim$ collapse measure Σ), but extends this concept to include dynamic aspects (critical attractor at the edge of chaos) and more concrete neural location (incl. brainstem, frontal lobe). It agrees with GWT in terms of global availability and feedback, but provides an ontological basis for why global broadcast is synonymous with subjective experience – namely, because it forms a causally indivisible whole. Plasticity models are not negated, but complemented: learning and structural adaptation form the conditions for the collapse core, which in turn provides the content that is learned (circular process). The collapse model thus positions itself not as an antithesis, but rather as a **meta-framework** in which the evidence-based parts of the other theories can

be embedded. A unique selling point is **interdisciplinarity**: findings from theoretical computer science (network complexity, critical dynamics), neuroscience (connectivity, PCI, DMN), psychology (attention, working memory), philosophy of mind (intentionality, phenomenal qualia) and evolutionary theory (phylogenetic development of consciousness) flow together here. This broad foundation makes the model connectable and at the same time consistent.

Finally, it should be emphasized that the extended model of causal collapse not only describes retrospectively, but also provides **predictions and practical implications**. For example, there is a criterion for *artificial consciousness*: An artifact would be conscious if it has a self-generating, integrative causal core that exceeds a certain complexity threshold and is behaviorally relevant. This could be operationalized by measuring in AI systems whether perturbation triggers global recurrent activity and whether a measure is above threshold, analogous to PCI. First architectural principles were outlined in earlier work (e.g. recursive interconnection, critical connectivity, meta-modeling) and could now be pursued in a more targeted manner. The collapse concept could also be applied clinically: For example, one could better assess the **level of consciousness of comatose patients** by performing targeted stimulation and network analyses to check whether a global collapse can be used (which would be an indicator of possible recovery).

All in all, this thesis is an independent, structurally complete paper that **theoretically consolidates and expands** the model of causal collapse without losing sight of the empirical connection. The model combines disparate pieces of the puzzle into a holistic view of consciousness: a *causal core* as a dynamic center, distributed in the brain, which, once formed, makes the world appear to the system. This brings us a further step closer to the long-standing goal of consciousness research, which is to close the gap between subjective experience and objective brain activity. Future work will have to test the hypotheses and structures presented here in concrete experiments and simulations. But it is already becoming apparent that the path via **integration, feedback and emergent causal structures** is a promising one to further explore the mystery of consciousness.

Open tasks include in particular the experimental validation of the Σ(t) measure, its differentiation from established complexity measures such as PCI and Φ, and the further specification of dynamic thresholds (Σ_c). In addition, the formalized distinction from the IIT approach opens up new avenues for empirical theory competition.

Literature:

Baars, Bernard J. *A Cognitive Theory of Consciousness*. Cambridge University Press, 1988.

Brentano, Franz. *Psychology from an Empirical Standpoint*. Routledge, 1874.

Carhart-Harris, Robin L., et al. The Entropic Brain: A Theory of Conscious States Informed by Neuroimaging Research with Psychedelic Drugs." *Frontiers in Human Neuroscience* 8 (2014): 20.

Casali, Adenauer G., et al. A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior.*Science Translational Medicine* 5, no. 198 (2013): 198ra105.

Damasio, Antonio. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt, 1999.

Damasio, Antonio. *Self Comes to Mind: Constructing the Conscious Brain*. Pantheon, 2010.

Dehaene, Stanislas, Hakwan Lau, and Sid Kouider. What Is Consciousness, and Could Machines Have It?*Science* 358, no. 6362 (2017): 486–492.

Dehaene, Stanislas, and Jean-Pierre Changeux. Experimental and Theoretical Approaches to Conscious Processing.*Neuron* 70, no. 2 (2011): 200–227.

Edelman, Gerald M., and Giulio Tononi. *A Universe of Consciousness: How Matter Becomes Imagination*. Basic Books, 2000.

Feinberg, Todd E., and Jon M. Mallatt. *The Ancient Origins of Consciousness: How the Brain Created Experience*. MIT Press, 2016.

Ginsburg, Simona, and Eva Jablonka. *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. MIT Press, 2019.

Laureys, Steven, et al. Cortical Connectivity Changes in Coma and Related States.*Brain* 129, no. 10 (2006): 2636–2649.

Lutz, Antoine, et al. Long-Term Meditators Self-Induce High-Amplitude Gamma Synchrony during Mental Practice.*PNAS* 101, no. 46 (2004): 16369–16373.

Mashour, George A., et al. Conscious Processing and the Global Neuronal Workspace Hypothesis.*Neuron* 105, no. 5 (2020): 776–798.

Massimini, Marcello, et al. Breakdown of Cortical Effective Connectivity during Sleep." Science 309, no. 5744 (2005): 2228–2232.

Nagel, Thomas. What Is It Like to Be a Bat?" *The Philosophical Review* 83, no. 4 (1974): 435–450.

Oizumi, Masafumi, Larissa Albantakis, and Giulio Tononi. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0." PLoS Computational Biology 10, no. 5 (2014).

Panksepp, Jaak. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press, 1998.

Raichle, Marcus E. The Brain's Default Mode Network." *Annual Review of Neuroscience* 38 (2015): 433–447.

Schröder, Tobias, and Christian Schilling. Recursive Self-Modeling Architectures for Synthetic Consciousness." *Frontiers in Artificial Intelligence* 6 (2023).

Searle, John R. *The Rediscovery of the Mind*. MIT Press, 1992.

Stegemann, Wolfgang. Integrated model of causal collapse. https://doi.org/10.5281/zenodo.15350050 (2025).