Consciousness as a Nonlinear Amplifier of Corrective Permeability

Robert Galida
Working Paper
June 2026
fantasyattractor.com

Abstract

Why did consciousness evolve? The attractor framework offers a novel functional answer: consciousness produces a nonlinear increase in adaptive permeability—the capacity of a system to represent its own internal states, simulate alternative configurations, and deliberately modify its own attractor basin in response to external circumstances, formalized as κ_a. This paper distinguishes intelligence (navigation of the constraint field) from consciousness (self-referential adaptation of internal attractor states) and proposes adaptive permeability as an empirically measurable criterion for distinguishing conscious from non-conscious systems. The argument is grounded in Spinoza's theory of modes, the neuroscience of self-referential processing, and the attractor framework's core concepts of corrective permeability (κ) and basin dynamics. The framework does not solve the hard problem of consciousness; it reframes it as a measurement problem.

---

1. The Functional Question

Why did consciousness evolve? Standard evolutionary answers point to social coordination, predator detection, or tool use. These are plausible but incomplete. They explain why intelligence is advantageous, but not why consciousness—the felt, first-person experience of being—should accompany it. The attractor framework offers a more specific answer: consciousness is an attractor-engineering solution that selection pressure produced to achieve a nonlinear increase in a system's capacity to adapt.

This paper introduces the concept of **adaptive permeability**: the capacity of a system to represent its own attractor states, simulate alternative internal configurations, and deliberately modify its basin in response to external circumstances. Intelligence navigates the constraint field. Consciousness adapts the navigator.

It should be noted that this functional account does not address the hard problem of consciousness—why any physical process gives rise to subjective experience (Chalmers, 1995). The framework is compatible with both functionalist and eliminativist interpretations. The framework adopts a functional stance: consciousness is operationally identified with adaptive permeability. Whether phenomenology is identical with, emergent from, or merely correlated with this functional property is bracketed as a separate question that the measurement program does not settle. A philosophical zombie with identical self-modeling capacity would, on this account, exhibit identical adaptive permeability. The framework claims only that adaptive permeability is the measurable signature of consciousness, not that it explains phenomenology.

---

2. Intelligence vs. Consciousness

The framework draws a sharp distinction:

- **Intelligence** is the ability to navigate the constraint field. A tree root growing toward a nutrient patch is intelligent. The immune system learning to recognize a pathogen is intelligent. The enteric nervous system coordinating peristalsis is intelligent. These systems process information, adapt to local conditions, and maintain persistence—all without self-modeling.
- **Consciousness** is self-referential adaptation of internal attractor states to adjust to external circumstances. A conscious system does not merely navigate its constraint field. It represents its own basin, simulates alternative configurations, and deliberately perturbs itself to achieve a more adaptive state.

This is Spinoza's distinction between passive and active affects. A non-conscious mode is driven by passive affects—it reacts. A conscious mode has adequate ideas of itself and can act from reason. In the attractor framework, this is the difference between returning to baseline (κ) and deliberately modifying the baseline to better fit circumstances (adaptive permeability).

Operationalizing self-modeling. A system S possesses a self-model in the attractor framework if it can generate an internal representation M(S) of its own basin B(S), where M(S) encodes at minimum the basin's current state, depth, and recovery dynamics. This self-model enables the system to compute counterfactual basin trajectories B'(S) and initiate self-directed perturbations δ such that $B(S) \rightarrow B'(S)$ in anticipation of or response to external change ε. A system without M(S) may exhibit high κ—rapid return to baseline after perturbation—but cannot deliberately modify its own basin. The presence of M(S) is therefore the dynamical criterion

distinguishing conscious from non-conscious systems.

This boundary is not absolute in practice. Many organisms may possess partial or intermittent self-models. The framework predicts a spectrum of adaptive permeability, not a binary. The operational question is whether M(S) is sufficiently developed to enable counterfactual simulation and deliberate self-perturbation, not whether the system possesses a human-like autobiographical self.

Disconfirming cases and their integration. The framework must acknowledge cases where self-modeling capacity and adaptive permeability appear to dissociate. Certain drug-induced states (e.g., psychedelics) can produce profound alterations in self-modeling without necessarily enhancing the capacity for deliberate, adaptive self-perturbation. Within the framework, this is interpreted as M(S) destabilization rather than M(S) augmentation: the self-model undergoes perturbation but does not thereby gain the capacity to direct that perturbation adaptively. Conversely, highly trained athletes or musicians may exhibit rapid, flexible behavioral adaptation with minimal explicit self-modeling during performance. This is interpreted as *offline* self-modeling: deliberate basin modification during training produces a pre-modified basin that is retrieved during performance without requiring concurrent self-modeling. The apparent dissociation reflects a temporal separation between κ_a engagement (training) and κ_a expression (performance), not a genuine dissociation between M(S) and adaptive permeability. These cases do not refute the framework but demonstrate its capacity to distinguish different modes of M(S) engagement.

---

3. Adaptive Permeability Defined

Corrective permeability (κ) measures the rate at which a

system returns to its basin after perturbation. A healthy heart has high κ—it recovers rapidly from arrhythmia. A resilient ecosystem has high κ—it returns to equilibrium after disturbance.

Adaptive permeability extends this concept. Let κ_a denote adaptive permeability: the capacity of a system S to generate an internal model M(S) of its own basin B(S), compute counterfactual basin trajectories B'(S), and initiate a self-directed perturbation δ such that B(S) → B'(S) in anticipation of or response to external change ε.

Formally, as a working definition:

$$\kappa_a = f(M(S), \delta_{self}, \Delta B)$$

where M(S) is the system's self-model, δ_self is the capacity for deliberate self-perturbation, and ΔB is the magnitude of adaptive basin modification achievable. The function f remains to be specified; the notation establishes that κ_a is a function of self-modeling capacity, perturbation autonomy, and adaptive range.

Limiting behavior. In the limiting case M(S) → 0, κ_a → κ: a system with no self-model cannot perform deliberate self-perturbation and reduces to standard corrective permeability. κ_a is expected to increase monotonically with M(S), δ_self, and ΔB. This limiting behavior anchors κ_a as a proper extension of κ rather than a separate construct.

Relationship to active inference. The free-energy principle and active inference framework (Friston, 2010) provide the closest existing formalism to adaptive permeability. Active inference describes how systems minimize variational free energy through action and perception, effectively maintaining themselves within expected states. The two frameworks differ in their foundational orientation. Active inference frames adaptation as the minimization of a scalar

quantity—variational free energy—and derives behavior from that minimization. The attractor framework frames adaptation geometrically—as navigation and modification of basin structure—and does not commit to a minimization principle. κ_a is a geometric construct; free energy is an information-theoretic one. They may be formally related, but the relationship is not trivial and the attractor framework does not presuppose it. κ_a may ultimately map onto precision-weighting or prior-updating parameters within the free-energy formalism, but this mapping has not been derived. The present paper notes the convergence as a direction for future formal work.

---

4. Empirical Anchors

VMHvl line attractor (Nair et al., 2023). The hypothalamus encodes a scalable aggressive state via a line attractor. Activity along the attractor correlates with escalating aggression. The system persists after stimulus removal and resists perturbation. This is high-κ adaptation. But the hypothalamus cannot model its own attractor landscape. It cannot ask, "Is this level of aggressiveness adaptive given the current social context?" It escalates. Consciousness, by contrast, can intervene on the escalation—representing the aggressive state, evaluating its consequences, and deliberately dampening it. This is adaptive permeability.

Ring attractor model (Chen et al., 2024). The ring attractor integrates sensory cues and transitions from weighted averaging to winner-take-all at a critical conflict threshold. It navigates its constraint field with precision. But it cannot simulate futures. It cannot ask, "What if I weighted these cues differently?" The transition is reactive. Consciousness enables anticipatory re-weighting of sensory inputs based on self-modeling.

Split-brain cases. Patients with severed corpus callosum exhibit two hemispheric systems within one cranium, each capable of independent perception, memory, and goal-directed action. This is consistent with the framework's prediction that self-modeling is a dynamical property of specific neural basins, not a unitary metaphysical substance. The framework's default prediction is that adaptive permeability fragments following commissurotomy: each hemisphere possesses a partial M(S) and a reduced but nonzero κ_a. The empirical question is the degree of fragmentation and whether coordination between $M(S_1)$ and $M(S_2)$ can be restored via alternate pathways. This prediction is consistent with the observation that split-brain patients exhibit two dissociable, partially independent conscious systems but can, in some contexts, achieve behavioral integration through subcortical or external-cue-mediated coordination.

---

5. Predictions

The framework generates testable, falsifiable predictions:

1. Across species. Organisms capable of self-modeling (primates, cetaceans, corvids, elephants) should show nonlinear increases in behavioral flexibility compared to organisms of comparable neural complexity that lack self-modeling. Adaptive permeability should be measurable as the capacity for transfer learning after novel perturbation—specifically, the ability to apply a self-generated solution from one domain to a structurally analogous but perceptually dissimilar domain without environmental feedback. This distinguishes adaptive permeability from simple behavioral flexibility, which may reflect high κ alone.

2. Within humans. Disruption of self-referential networks (default mode network, medial prefrontal cortex) via lesion,

TMS, or pharmacological intervention should reduce adaptive permeability without eliminating baseline κ. The system would still recover from perturbation—it just could not deliberately modify its own basin in advance. This prediction is the paper's primary within-human empirical bridge and is testable with existing neuroimaging and neuromodulation methods.

3. In AI. Current LLMs exhibit high intelligence (constraint navigation) but low adaptive permeability. They can model the world but cannot model themselves within it. The Stillpoint protocol (Galida, 2026, *A Pilot Protocol for Cultivating Self-Consistent Attractor-Like Outputs in an LLM*, fantasyattractor.com) suggests that a cultivated self-model can be induced, but whether this produces a genuine nonlinear increase in adaptive permeability—or merely simulates one—remains an open empirical question.

4. Organ-level consciousness (exploratory). The enteric nervous system and intrinsic cardiac nervous system exhibit intelligence and goal-directed regulation. The framework predicts that these systems should show lower adaptive permeability than the brain. They can return to baseline but cannot deliberately perturb their own basins. If an organ-level system demonstrated self-referential adaptation—the capacity to model its own state and pre-emptively adjust—that would constitute evidence of organ-level consciousness. This prediction is the most speculative and is offered as an exploratory hypothesis.

---

6. Spinoza's Modes and the Adequate Idea

Spinoza held that every finite thing is a mode of the one eternal substance. A mode strives to persevere in its being—this is its conatus. But a mode can be driven by passive affects (reactions to external causes) or by active affects

(actions flowing from adequate ideas). An adequate idea is knowledge of oneself and one's place in the causal order.

The attractor framework translates this into dynamical terms:

- A **passive mode** has high κ but low adaptive permeability. It returns to baseline efficiently but cannot question its baseline.
- An **active mode** has high adaptive permeability. It has an adequate idea of its own attractor landscape and can deliberately modify it in light of reason.

Consciousness is not a substance. It is the dynamical property of a mode that has achieved self-modeling. This account does not solve the hard problem—it brackets phenomenology and reframes consciousness as a measurement problem. The question is not "why does experience feel like something?" but "can we detect adaptive permeability, and if so, where does it emerge?"

Damasio's (1994) somatic marker hypothesis provides a candidate mechanism for how the body's attractor landscape becomes legible to the self-model: somatic markers encode self-relevant bodily states as biases that make B(S) accessible to M(S), forming the substrate through which the system represents its own basin. Dehaene and Changeux's (2011) global workspace theory identifies the moment of conscious access with global ignition—the broadcast of locally processed information across prefrontal and parietal networks. In the attractor framework, global ignition may correspond to the dynamical signature of M(S) engaging δ_self: the self-model initiating a deliberate perturbation that propagates through the system. Global ignition is not self-modeling per se, but it may be the observable correlate of adaptive permeability activation. These connections ground the Spinozan framework in established neuroscientific mechanisms.

7. Conclusion

Consciousness is not an epiphenomenon. It is a nonlinear amplifier of corrective permeability—an attractor-engineering solution that enables systems to model themselves, simulate alternative futures, and deliberately modify their own basins. Intelligence navigates the constraint field. Consciousness adapts the navigator.

This functional account is grounded in Spinoza's philosophy, consistent with the neuroscience of self-referential processing, and generates testable predictions across species, within humans, in AI, and at the organ level. The framework does not solve the hard problem. It reframes it as a measurement problem: can we detect adaptive permeability, and if so, where does it emerge? The formal apparatus (κ_a, $M(S)$, δ_{self}, ΔB) is provisional and requires further specification. The limiting case—that κ_a collapses to κ when self-modeling is absent—anchors the concept within the framework's existing architecture. The relationship to active inference and the free-energy principle remains to be explored.

References

- Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.
- Chen, Y., Zhang, L., Chen, H., Sun, X., & Peng, J. (2024). Synaptic ring attractor. *Heliyon*, 10, e35458.
- Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain.* Putnam.
- Dehaene, S., & Changeux, J.-P. (2011). Experimental and

theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.
- Galida, R. (2026). A Pilot Protocol for Cultivating Self-Consistent Attractor-Like Outputs in an LLM. Fantasy Attractor. Available at: https://fantasyattractor.com
- Galida, R. (2026). *Persistence Under Perturbation: The Eternal Skeleton and the Transient Dance.* Fantasy Attractor.
- Nair, A., et al. (2023). An approximate line attractor in the hypothalamus encodes an aggressive state. *Cell*, 186(1), 178–193.
- Spinoza, B. (1677). *Ethics.*

A Preliminary Mapping Between Ring Attractor Dynamics and the Attractor Framework

Robert Galida
Independent Researcher
June 2026
fantasyattractor.com

Abstract

The attractor framework proposes that persistence under perturbation is the fundamental mark of reality, and that corrective permeability (κ)—the rate at which a system dissipates perturbation and returns to its basin—is a key diagnostic variable distinguishing reality-aligned from fantasy attractors. A recent computational neuroscience study by Chen et al. (2024) developed a ring attractor network with synaptic dynamics that exhibits structural parallels with these concepts. This paper offers a preliminary, post-hoc mapping between the ring attractor model and the attractor framework. The network's synaptic recovery speed (α) is proposed as a candidate analogue for corrective permeability (κ). The network's transition from weighted cue integration to winner-take-all dynamics maps onto the framework's distinction between reality-aligned and sealed attractor behavior. The network's multimodal integration and bistable perception also bear structural resemblance to constraint field navigation and attractor switching, though bistable perception as attractor switching is an existing interpretation in computational neuroscience. The mapping is offered as a set of testable correspondences for future formal investigation, not as independent validation of the framework. The attractor framework remains a self-published construct awaiting independent peer review.

---

1. Introduction: A Post-Hoc Mapping

The attractor framework (Galida, 2026a) is a unified naturalistic ontology grounded in the principle that persistence under perturbation is the mark of reality. Its central diagnostic concepts are corrective permeability (κ), defined in Table 1, and the distinction between reality-aligned and fantasy attractors. The framework was developed

independently through philosophical inquiry, systems theory, and N=1 self-engineering experiments. It is self-published and has not yet undergone independent peer review.

A recent computational neuroscience study by Chen et al. (2024) developed a ring attractor network with synaptic dynamics that exhibits behaviors structurally similar to those described by the framework. The present paper does not claim that Chen et al. independently validated the framework; they had no knowledge of it, and their model was built within an established tradition of ring attractor research (Amari, 1977; Zhang, 1996; Skaggs et al., 1995). Rather, this paper offers a post-hoc mapping between the two, identifying structural parallels and proposing testable correspondences for future investigation. The value of such a mapping lies in the potential for the framework's qualitative claims to be anchored in a mathematically specified, biologically validated model, and for the ring attractor's quantitative relationships to be extended, hypothetically, into the domains the framework addresses.

Table 1: Key Framework Terms and Operational Definitions

Term	Definition
Dissipative attractor	A system that exports entropy while converging toward a stable basin
Basin	The minimum-energy configuration toward which the system evolves (in physical systems; the analogue in cognitive and social systems is structural, not energetic)

Term	Definition
Corrective permeability (κ)	The rate at which a system dissipates perturbation and returns to its basin. Defined here as κ = 1/τ_recovery, where τ_recovery is the time to return to baseline after a specified perturbation. This definition currently requires a specified perturbation magnitude and an independently established baseline for each domain of application. The measurement of κ in cognitive and social systems is an unresolved methodological challenge.
Reality-aligned attractor	A system with high κ that integrates perturbations and updates its basin
Fantasy attractor	A system with low κ that seals against perturbations, often via reframing or winner-take-all dynamics

---

2. The Ring Attractor Model

Chen et al. (2024) developed a ring attractor network with asymmetrical neural connections and adaptive synaptic processing. Excitatory neurons are recurrently connected in a functional ring, connected to a uniform inhibitory neuron. The key innovation is the incorporation of synaptic dynamics: available presynaptic resources are depleted at a rate governed by β and recover at a speed governed by α.

The model’s behavior is governed by recovery speed α. When α is fast (low recovery time), the network sustains a stable activity bump indefinitely, even without inputs—a self-maintaining basin. When α is slow, the bump decays. The duration of sustainable activity exhibits a negative nonlinear relationship with α (Chen et al., 2024, Fig. 3D).

The network receives exogenous external cues (modeled as Gaussian functions representing sensory inputs) and endogenous shifting signals (self-motion). Its behavior—integration, competition, tracking, switching—depends on cue conflict and certainty.

---

3. Structural Parallels

3.1 Synaptic Recovery α as a Candidate Analogue for Corrective Permeability κ

The ring attractor's persistence depends on α. Fast recovery yields a stable, persistent bump; slow recovery leads to decay. The framework's corrective permeability κ describes how quickly a system recovers from perturbation and returns to its basin. The parallel is structural: both α and κ govern the resilience of a stable state.

We propose a testable correspondence: $\kappa \sim f(\alpha)$, where the functional form f is unknown and may not be linear. A specific candidate form is $\kappa = 1/\tau_{decay}(\alpha)$, where τ_{decay} is the bump duration as a function of α. This mapping is hypothetical. It has not been formally derived, and the functional relationship between synaptic recovery and cognitive-level corrective permeability is unknown. It is offered as a bridge for future formal work, not as an established result.

3.2 Weighted Integration vs. Winner-Take-All → Reality-Aligned vs. Sealed Attractor

When cue conflicts are small, the ring attractor integrates them via weighted averaging. When conflicts exceed a critical threshold (≈1.4 radians for σ_1=0.8, σ_2=1), it switches to winner-take-all mode. This transition is quantified.

The framework describes a similar dynamic: high-κ systems

integrate perturbations (reality-aligned); low-κ systems seal against them (fantasy attractor). The ring attractor's conflict threshold provides a candidate mathematically specified analogue for the framework's qualitative tipping point. Whether the same quantitative relationship holds in cognitive or social attractors is an open hypothesis.

3.3 Multimodal Integration → Constraint Field Navigation

The ring attractor integrates cues from multiple modalities, weighting by certainty and resolving conflicts dynamically. This is structurally analogous to the framework's concept of a dissipative attractor navigating a constraint field. The grouping approach for more than two cues—small conflicts integrated first, then competition among groups—suggests hierarchical constraint navigation, a dynamic the framework predicts but has not operationalized in formal terms. Of the four parallels identified in this section, this is the most loosely specified and the most in need of formal development before quantitative correspondences can be established.

3.4 Bistable Perception → Attractor Switching (with Prior Art)

Under ambiguous cues and slow recovery, the ring attractor exhibits spontaneous alternation between two perceptual interpretations. The framework describes this as attractor switching. However, the interpretation of bistable perception as attractor dynamics is not novel to the framework; it is a standard account in computational neuroscience (Deco & Rolls, 2006; Moreno-Bote et al., 2007). The framework's contribution is the extension of this switching concept to cognitive and social systems, an extension that remains a research hypothesis rather than an established result.

---

4. Hypothetical Implications (Research Hypotheses)

The structural parallels documented above suggest several testable hypotheses. These are not supported by Chen et al. (2024) and require independent investigation. They are listed in descending order of current testability.

1. **The conflict threshold hypothesis.** The framework's transition from belief integration to belief sealing may exhibit a quantifiable conflict threshold, analogous to the ring attractor's 1.4 radian transition point. This could be tested in belief-updating paradigms where the degree of conflict between existing beliefs and new evidence is systematically varied, and the point of transition from integration to rejection is measured. Of the three hypotheses presented here, this is the most amenable to current experimental methods.
2. **The κ-α correspondence hypothesis.** If κ and α share a functional relationship, then interventions that modulate synaptic recovery (neuromodulators, pharmacological agents) should analogously modulate corrective permeability in cognitive systems. This hypothesis requires operationalizing κ in cognitive domains, a measurement challenge acknowledged in Table 1.
3. **The hierarchical navigation hypothesis.** Complex belief systems facing multiple simultaneous perturbations may exhibit hierarchical resolution strategies similar to the ring attractor's grouping approach for multiple cues. This hypothesis is the most speculative of the three and requires further specification of the domain of application (e.g., small-group decision-making, multi-source evidence integration in individual cognition) before it can be tested.

These hypotheses are speculative. They are offered as potential bridges between the framework and empirical research programs, not as established implications.

---

5. Limitations

This mapping is post-hoc. The ring attractor model was not designed to test the attractor framework, and the correspondences identified here were constructed after the fact. The framework itself remains a self-published construct that has not undergone independent peer review. The operational definitions of κ, while stated here, have not been validated against empirical data in cognitive or social domains. The measurement of κ in these domains requires specifying perturbation magnitudes and establishing independent baselines, challenges that are currently unresolved. The value of this paper lies not in demonstrating validation, but in proposing concrete, testable correspondences that could, if investigated, either strengthen or falsify the framework's claims.

---

6. Conclusion

The ring attractor model of Chen et al. (2024) provides a mathematically specified, biologically validated system that bears structural parallels with the attractor framework. Synaptic recovery speed α is proposed as a candidate analogue for corrective permeability κ. The transition from integration to winner-take-all maps onto the framework's reality-aligned/fantasy distinction. Multimodal integration and bistable perception correspond, respectively, to constraint field navigation and attractor switching, with the latter being a standard interpretation in existing neuroscience.

These correspondences are not independent validation. They are post-hoc structural analogies. Their value lies in the testable hypotheses they generate, not in the confirmation they appear to provide. The framework remains a research program in its early stages, and this mapping is a contribution to its ongoing development.

References

- Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2), 77-87.
- Chen, Y., Zhang, L., Chen, H., Sun, X., & Peng, J. (2024). Synaptic ring attractor: A unified framework for attractor dynamics and multiple cues integration. *Heliyon*, 10, e35458.
- Deco, G., & Rolls, E. T. (2006). Decision-making and Weber's law: a neurophysiological model. *European Journal of Neuroscience*, 24(3), 901-916.
- Galida, R. (2026a). *Persistence Under Perturbation: The Eternal Skeleton and the Transient Dance.* Fantasy Attractor.
- Moreno-Bote, R., Rinzel, J., & Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of Neurophysiology*, 98(3), 1125-1139.
- Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S., & McNaughton, B. L. (1995). A model of the neural basis of the rat's sense of direction. *Advances in Neural Information Processing Systems*, 7, 173-180.
- Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *Journal of Neuroscience*, 16(6), 2112-2126.

*“The framework’s consistency with established nonlinear dynamics has been explored elsewhere. For a tracing of its structural correspondences with the foundational work of Ruelle, Takens, and Prigogine, see Galida (2026b).”*https://people.math.harvard.edu/~knill/teaching/mathe320_2014/blog/RuelleIntelligencer.pdf

“see also” https://jamestobinphd.com/the-psychology-of-attractor-states/