

The Conscious Body: Organs as Attractor-Based Minds

Robert Galida, Independent Researcher

June 2026 | fantasyattractor.com

Abstract

The standard view holds that only the brain generates consciousness. This paper challenges that monopoly by applying the minimal functional criteria used to attribute rudimentary consciousness to the 302-neuron nematode *C. elegans* to the body's own complex, intrinsically innervated organs. On the basis of integration, valence, learning, goal-directedness, and anatomical concentration, the enteric nervous system (ENS), the intrinsic cardiac nervous system (ICNS), the intrinsic pancreatic ganglia, and—provisionally—the spinal cord qualify as candidate conscious subsystems. We do not assert that these organs are conscious. We assert that if the functional criteria are taken seriously enough to include a 302-neuron worm as a candidate, they cannot be silently withheld from structurally richer systems without a principled reason. We argue that the brain is not the sole generator of consciousness but the regulator of a federation of semi-autonomous organ-level attractors. We provide testable predictions, sketch the coupling mechanisms that bind local attractors into a unified self, outline clinical implications, and identify open problems including inter-attractor conflict and the phenomenal gap. The framework is offered as a research-generative hypothesis, not a completed theory.

1. Introduction: The Brain's Unexamined Monopoly

The brain is the organ we associate with consciousness, almost without question. Yet the body contains other complex neural networks. The enteric nervous system (ENS) comprises 200–600 million neurons, operates semi-autonomously, learns, and remembers. The intrinsic cardiac nervous system (ICNS) integrates local signals and regulates cardiac output. The spinal cord, with approximately 200 million neurons, can learn when isolated from the brain. The intrinsic pancreatic ganglia coordinate metabolic homeostasis. If these systems were found in a small animal, comparative neuroscience would at least entertain the possibility of consciousness. Because they are inside us, they are dismissed as mere infrastructure.

This paper asks a simple question: if we accept the functional criteria used to infer minimal consciousness in *C. elegans* (302 neurons), why are those same criteria not applied to the ENS, the ICNS, the pancreatic network, and the spinal cord? The question is not *Are these organs conscious?* but *Why are they excluded a priori?*

We do not claim to solve the hard problem of consciousness. We adopt the same pragmatic strategy used throughout comparative neuroscience: observable functional properties—integration, valence, learning, goal-directedness, and anatomical concentration—are treated as operational proxies for consciousness. This strategy is how we infer consciousness in other humans (by analogy), in non-human animals (by behavioural complexity), and in *C. elegans* (by measurable learning and integration). If these criteria are sufficient to identify a candidate conscious system in a 302-neuron worm, consistency demands their application to other systems that exceed this threshold, unless a principled exclusion criterion is provided. That exclusion criterion has not been articulated.

We use the term **candidate** throughout to avoid slippage into positive consciousness attribution. The paper's central claim is that the ENS, ICNS, pancreatic network, and spinal cord are *candidates*—systems that meet the same threshold criteria applied to a known candidate—and that dismissing them without investigation is methodologically inconsistent.

2. The Attractor Framework as Conceptual Scaffolding

An attractor is a region in state space toward which trajectories converge and remain unless perturbed. A candidate conscious attractor possesses five functional properties:

1. **Integration:** binding multiple sensory or interoceptive streams into a unified dynamical state.
2. **Valence:** operationalized as approach/avoidance behaviour—attraction to certain states and repulsion from others. We do not claim that behavioural valence entails phenomenal valence. We claim only that it is the same behavioural proxy used for *C. elegans* and other simple organisms. The inference from behavioural valence to phenomenal valence is a philosophical commitment we note but do not resolve.
3. **Learning:** the capacity to modify behaviour based on experience (habituation, sensitization, associative conditioning).
4. **Goal-directedness:** acting to maintain the system's own basin—a form of conatus—persisting in the absence of external commands.
5. **Anatomical concentration:** a spatially organized, intrinsically connected neural network with dedicated integrative circuitry. This fifth criterion distinguishes concentrated neural attractors (ENS, ICNS,

pancreatic ganglia) from diffuse, non-neural systems (immune system) and from infrastructure networks that lack a defined integrative centre. For the spinal cord, as discussed in Section 4.4, we apply this criterion with qualification.

The attractor vocabulary is applied conceptually, not formally, in this paper. A forthcoming quantitative treatment (Galida, 2026) will develop the mathematical persistence functional. The current paper uses attractor language to structure its functional criteria and predictions; it does not claim to derive formal basin measures from the available data.

Operationalizing Autonomy: We propose, as a provisional operational threshold, that a candidate subsystem crosses the autonomy boundary if it retains a significant fraction (e.g., $\geq 50\%$) of its normal functional repertoire following complete extrinsic denervation or isolation. This criterion distinguishes systems that are merely regulated from systems that can independently sustain goal-directed attractor dynamics. The ENS and ICNS clearly exceed this threshold; the spinal cord and pancreatic network do so conditionally, as discussed below.

3. The Conditional Argument and Its Stipulated Baseline

The nematode *C. elegans* possesses exactly 302 neurons. Its connectome is fully mapped. It exhibits sensory integration, associative learning, goal-directed chemotaxis, and minimal self-reference (distinguishing self-generated from external touch). Its learning capacities are well-documented (Ardiel & Rankin, 2010; Sasakura & Mori, 2013).

We stipulate—we do not establish—that *C. elegans* is a

candidate for minimal consciousness on the basis of these functional criteria. The paper does not require that the field accept this stipulation as consensus. It requires only that the reader grant the conditional: **if** the functional criteria are sufficient to make *C. elegans* a candidate, **then** they must be applied consistently to any system that meets or exceeds them. Those who reject the conditional may ignore the remainder of the argument, but they must then explain what additional criterion excludes the ENS, ICNS, pancreatic network, and spinal cord while admitting *C. elegans*.

4. Candidate Organs

The four candidate organs identified below are assessed against the five criteria, with the provisional autonomy threshold applied where possible. We differentiate their evidential strength clearly.

4.1 The Enteric Nervous System (ENS)

The ENS is the strongest candidate. Its 200–600 million neurons form two interconnected plexuses spanning the gastrointestinal tract. It meets all five criteria:

- **Integration:** continuously integrates mechanical, chemical, and hormonal signals to coordinate peristalsis, secretion, and blood flow.
- **Valence:** exhibits attraction to nutrients, aversion to toxins; noxious stimuli trigger emesis or accelerated transit.
- **Learning:** exhibits habituation, sensitization, and long-term plasticity; gut reflexes can be conditioned (Furness, 2012; Schemann & Frieling, 2020).
- **Goal-directedness:** actively propels food and maintains digestive homeostasis independently of the brain;

peristalsis persists after vagotomy—well above the 50% autonomy threshold.

- **Anatomical concentration:** a continuous, highly organized neural network with dedicated integrative circuitry.

4.2 The Intrinsic Cardiac Nervous System (ICNS)

The ICNS (14,000–43,000 neurons) is a moderate candidate. Its neuron count is only 46–143 times the *C. elegans* threshold, a narrower margin than the ENS. It meets the criteria, but with less evidential richness:

- **Integration:** monitors blood pressure, chamber stretch, and local chemistry to modulate cardiac output.
- **Valence:** maintains a preferred setpoint for cardiac rhythm; arrhythmias represent perturbations from that setpoint.
- **Learning:** shows ganglionic remodelling after injury; vagal stimulation protocols can alter responsivity (Armour, 2008).
- **Goal-directedness:** generates intrinsic rhythms when denervated, satisfying the autonomy threshold.
- **Anatomical concentration:** organized into ganglia on the heart's surface.

The ICNS contributes to emotional experience via heartbeat-evoked potentials that correlate with interoceptive awareness and self-recognition. This is suggestive but does not independently establish consciousness.

4.3 The Intrinsic Pancreatic Network

The pancreatic network is the most provisional candidate. Its 10,000–50,000 intrinsic neurons are scattered in ganglia throughout the organ, rather than forming a continuous plexus (Ahren, 2000; Salvioli et al., 2002). This weaker anatomical concentration distinguishes it from the ENS and ICNS.

- **Integration:** combines neural, hormonal, and nutrient signals to regulate blood glucose.
- **Valence:** maintains a metabolic setpoint; hypoglycemia and hyperglycemia are aversive states.
- **Learning:** plasticity is less studied than in the ENS; no direct evidence of conditioning is available.
- **Goal-directedness:** coordinates endocrine and exocrine output to maintain glucose homeostasis; whether this function persists at $\geq 50\%$ of normal repertoire after complete extrinsic denervation is not yet established. The pancreatic network remains a candidate, but with an open empirical question on the autonomy threshold.
- **Anatomical concentration:** scattered ganglia; meets the threshold but is the weakest candidate on this criterion.

4.4 The Spinal Cord (Provisional Candidate)

The spinal cord possesses approximately 200 million neurons, organized into topographically precise circuits that integrate sensory input, generate coordinated motor output, and exhibit learning when isolated (Hook & Grau, 2007). By the five functional criteria, it qualifies. However, under normal physiological conditions, its activity is tightly coupled to descending commands, and independent behavioural generation is rarely observed. After complete spinal cord injury, the isolated cord reorganizes and can generate complex, goal-directed responses. Whether such reorganization achieves the $\geq 50\%$ autonomy threshold is an empirical question; we provisionally include the spinal cord as a candidate with lower confidence, identifying it as the ideal test case for refining the autonomy criterion.

5. The Brain as Regulator: Mechanisms of Coupling

If the ENS, ICNS, pancreatic network, and spinal cord are candidate conscious subsystems, the unified self must be explained as the product of their integration by the brain. We propose that the brain couples, modulates, and aligns local attractors through four mechanisms, each supported by established physiology.

5.1 Vagal Afferent Signalling

The vagus nerve provides the primary bidirectional communication channel between the brain and the viscera. Vagal afferents convey interoceptive signals from the ENS and ICNS to the nucleus of the solitary tract, and descending signals modulate organ function. Vagal nerve stimulation is known to alter mood, reduce inflammation, and improve cardiac function (George et al., 2000; Tracey, 2002).

5.2 Humoral Signalling

Circulating hormones (cortisol, adrenaline, insulin, glucagon) and immune mediators (cytokines) provide a slower, diffuse coupling channel. These signals alter the global attractor's landscape by shifting the metabolic and inflammatory context. Sickness behaviour—fatigue, anhedonia, social withdrawal—is a well-documented example of immune-to-brain signalling that temporarily reconfigures the global attractor (Dantzer et al., 2008).

5.3 Rhythmic Entrainment

The brain entrains peripheral rhythms to its own oscillations. Cardiac and respiratory rhythms phase-lock to cortical activity during focused attention (Thayer & Lane, 2000). Slow-wave sleep entrains glymphatic clearance (Xie et al., 2013). The brain sets a rhythm, and the organs—each with their

own intrinsic oscillators—tend to follow. This resonance is not command; it is coupling by shared frequency.

5.4 Predictive Processing and Attractor Coupling

The predictive processing framework (Clark, 2013) treats the brain as a prediction engine that minimizes surprise by updating internal models based on sensory input. We suggest that this framework extends naturally to interoception: the brain maintains predictions about the states of the body's organs, and each organ generates its own predictions about local conditions. The alignment of these nested predictive models is functionally analogous to attractor coupling, in that both involve the progressive alignment of internal states toward a shared equilibrium. Friston's (2010) free-energy principle provides a formal bridge between predictive processing and dynamical systems that could, in future work, unite these descriptions under a single mathematical framework.

5.5 Relationship to Competing Theories of Consciousness

The attractor framework is compatible with but not identical to several major theories. Integrated Information Theory (IIT; Tononi, 2008) holds that consciousness is a function of the amount of integrated information a system generates. The attractor framework shares IIT's emphasis on integration but does not require the computation of Φ , which remains technically infeasible for most organ systems. Global Workspace Theory (GWT; Baars, 1988; Dehaene, 2011) posits that consciousness arises when information is broadcast within a global workspace. Under GWT, many peripheral attractors would be considered unconscious because they lack access to a central workspace. The attractor framework allows for phenomenal consciousness without global access, a position consistent with the possibility that the ENS may have experiences that never enter cortical awareness. Higher-Order Theories (HOTs) require meta-representation—the capacity to

represent one's own states—which, if correct, would likely exclude all candidate organs except the brain. The attractor framework treats HOTs as a valid but overly restrictive criterion that would also exclude many animals currently accepted as conscious. The framework does not seek to refute these theories but to generate testable predictions that can be compared with theirs, advancing the debate through empirical competition.

5.6 Inter-Attractor Conflict: An Open Problem for the Federation Model

A federation of semi-autonomous attractors inevitably generates conflict. Everyday clinical phenomena illustrate this: nausea during a cognitively demanding task (ENS and cortical attractors in tension), cardiac arrhythmia during emotional stress (ICNS and limbic system in conflict), hypoglycemic cognitive impairment (pancreatic and cortical attractors in opposition). The current paper does not propose a mechanism for conflict resolution beyond the brain's general regulatory role. Whether such conflicts are resolved by hierarchical dominance, temporal multiplexing, or some form of inter-attractor negotiation is an open question. We flag it as a priority for future theoretical development within the framework.

6. The Alien Feeling and Clinical Dissociation

When coupling between the global self and a local attractor falters, the experience can manifest as an “alien feeling”—the sense that an action or bodily state is “not mine.” This phenomenon is well-documented in alien hand syndrome (Della Sala et al., 1991) and in depersonalization disorder, where individuals report feeling detached from their own body and

mental processes (Sierra & David, 2011). We interpret these as temporary or chronic decoupling of a local attractor from the global workspace—exactly what the federation model would predict when integration fails.

7. Testable Predictions

The framework generates five falsifiable predictions:

1. **ENS conditioning:** An isolated intestinal segment, exposed to a neutral stimulus paired with a non-nociceptive chemical infusion, will exhibit a conditioned motor or hormonal response.
 2. **ICNS plasticity:** Long-term heart rate variability biofeedback will produce persistent changes in baseline cardiac rhythms not fully mediated cortically.
 3. **Gut-directed therapy:** IBS patients receiving gut-directed biofeedback will show greater symptom improvement than those receiving standard CBT alone.
 4. **Pancreatic memory:** In a vagally denervated preparation, islet cell clusters exposed to repeated glucose perturbation will exhibit an anticipatory insulin response.
 5. **Spinal reorganization:** Complete spinal cord injury patients will develop complex, coordinated responses below the lesion beyond simple reflexes, consistent with a reorganizing local attractor.
-

8. Future Directions: Approaching the

Phenomenal Gap

The framework operates on behavioural and functional proxies for consciousness; it does not provide direct phenomenological access to organ-level experience. What evidence could begin to bridge this gap? We propose three directions. First, decoupling experiments that temporarily isolate a candidate organ (e.g., via selective pharmacologic blockade) and then probe the subject's subjective state could reveal whether the organ's local attractor contributes a distinct experiential component to the global self. Second, longitudinal studies of spinal cord injury patients who report phantom sensations or "body memories" below the lesion may provide indirect reportable correlates of spinal attractor activity. Third, the development of organ-specific interoceptive training protocols, coupled with experience-sampling methods, could track whether changes in organ function co-vary with changes in the felt sense of self. These are early-stage proposals; the phenomenal gap remains the deepest challenge for the framework, as for all theories of consciousness.

9. Clinical Implications

If organs are candidate conscious systems, functional disorders may represent distressed local attractors. IBS may be a gut that has learned to react to benign stimuli as threats. Cardiac anxiety may reflect a perturbed ICNS state. These reframings suggest organ-directed therapies: gut-directed biofeedback, vagal stimulation, dietary protocols that calm the ENS. The principle is consistent with existing mind-body approaches but grounds them in a specific, testable model.

10. Ethical Considerations

Candidate organs are not autonomous moral agents. Their interests are tied to the whole body's survival. Clinical ethics correctly prioritize the patient's overall well-being. The framework suggests a principle of organ-level respect: where possible, preserve organ integrity and explore gentler interventions before resection or ablation. This is holistic medicine, not radical ethics.

11. Conclusion

The brain is not the body's sole candidate conscious organ. The ENS, ICNS, pancreatic network, and spinal cord meet the same functional criteria used to identify *C. elegans* as a candidate for minimal consciousness. They are not established as conscious; they are identified as systems for which the question cannot be dismissed a priori without a principled exclusion criterion. The coupling mechanisms that bind local attractors into a unified self are partially characterized, and the framework generates concrete, falsifiable predictions. The conscious body is a research-generative hypothesis, not a completed theory.

References

- Ahren, B. (2000). Autonomic regulation of islet hormone secretion. *Diabetologia*.
- Ardiel, E.L., & Rankin, C.H. (2010). An elegant mind: learning and memory in *C. elegans*. *Learning & Memory*.
- Armour, J.A. (2008). Potential clinical relevance of the 'little brain' on the mammalian heart. *Experimental*

Physiology.

- Baars, B.J. (1988). *A Cognitive Theory of Consciousness*. Cambridge.
- Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*.
- Dantzer, R., et al. (2008). From inflammation to sickness and depression: when the immune system subjugates the brain. *Nature Reviews Neuroscience*.
- Dehaene, S. (2011). *Consciousness and the Brain*. Viking.
- Della Sala, S., et al. (1991). The anarchic hand: a fronto-mesial sign. In *Handbook of Neuropsychology*.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*.
- Furness, J.B. (2012). The enteric nervous system and neurogastroenterology. *Nature Reviews Gastroenterology & Hepatology*.
- George, M.S., et al. (2000). Vagus nerve stimulation: a new tool for brain research and therapy. *Biological Psychiatry*.
- Gershon, M.D. (1998). *The Second Brain*. HarperCollins.
- Hook, M.A., & Grau, J.W. (2007). An animal model of spinal cord learning. *Behavioural Brain Research*.
- Kelso, J.A.S. (1995). *Dynamic Patterns*. MIT Press.
- Salvioli, B., et al. (2002). Pancreatic intrinsic innervation. *Neurogastroenterology & Motility*.
- Sasakura, H., & Mori, I. (2013). Behavioral plasticity, learning, and memory in *C. elegans*. *Current Opinion in Neurobiology*.
- Schemann, M., & Frieling, T. (2020). The enteric nervous system: a second brain. *Nature Reviews Neuroscience*.
- Sierra, M., & David, A.S. (2011). Depersonalization: a selective impairment of self-awareness. *Consciousness and Cognition*.
- Spinoza, B. (1677). *Ethics*.

- Strogatz, S.H. (2018). *Nonlinear Dynamics and Chaos*.
- Thayer, J.F., & Lane, R.D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*.
- Tononi, G. (2008). Integrated information. *Biological Bulletin*.
- Tracey, K.J. (2002). The inflammatory reflex. *Nature*.
- Xie, L., et al. (2013). Sleep drives metabolite clearance from the adult brain. *Science*.

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

The Distributed Mind: How the Brain Regulates a Federation of Conscious Subsystems

Author: Robert Galida

Date: June 2026 (Final Edition)

Based on: Extended collaborative development of the attractor framework, N=1 physiological experimentation, and a re-reading of Spinoza’s conatus.

Abstract

Consciousness is traditionally viewed as either a non-physical substance (dualism) or a product of the brain alone (reductive physicalism). This paper presents an alternative: the human

body is a nested hierarchy of semi-autonomous, attractor-based conscious subsystems—each with its own rudimentary integration, valence, learning, and goal-directedness. Using the nematode *C. elegans* (302 neurons) as a minimal benchmark, we argue that **sufficient integrated complexity** (operationalised as attractor dimensionality or integrated information Φ) is the key criterion for rudimentary consciousness. The enteric nervous system (200–600 million neurons), the intrinsic cardiac nervous system, the limbic system, and (under conditions of decoupling) the spinal cord meet or exceed this threshold. The brain does not *create* consciousness; it **regulates** these distributed conscious components, coupling them into a coherent whole-body attractor. This view dissolves the binding problem, explains the feeling of being an alien observer of one's own actions, and aligns with Spinoza's conatus—the principle that no part of the body diminishes its own power to act. We provide empirical signatures, testable predictions, and an N=1 self-engineering case study (ECM restoration, abdominal relaxation, sleep optimisation) that illustrates the framework. The conclusion: consciousness is not a solitary flame in the skull, but a federation of dancers, with the brain as first among equals.

1. Introduction

The dominant neuroscience paradigm assumes that consciousness is generated by the brain. Yet this assumption struggles to explain:

- Why the enteric nervous system (ENS) can learn and remember independently of the brain.
- Why cardiac signals influence decision-making and self-awareness.

- Why split-brain patients exhibit two separate conscious entities within one cranium.
- Why the universal feeling of “not being in control” (“*why did I do that?*”) persists.

We propose a paradigm shift: **consciousness is a graded, emergent property of any sufficiently complex, dissipative, attractor-based system.** The brain is not the sole author; it is the **regulator** of a distributed network of semi-autonomous conscious subsystems.

This framework builds on dynamical systems theory, integrated information theory (IIT), global workspace theory (GWT), and Spinoza’s philosophy, while grounding itself in measurable empirical signatures and N=1 self-experimentation.

2. The Attractor Framework for Consciousness

2.1 Core Definitions

- **Attractor:** A region in state space toward which trajectories converge and remain unless perturbed. Characterised by negative Lyapunov exponents and basin stability.
- **Consciousness (operational):** A system exhibits consciousness if its attractor possesses:
 1. **Integration** – binds multiple sensory/interoceptive streams.
 2. **Self-reference** (minimal) – distinguishes self from environment.
 3. **Valence** – attraction to some states, repulsion from others.

4. **Learning** – attractor landscape changes with experience.
5. **Goal-directedness** – acts to maintain its basin (conatus).
6. **Evolutionary/developmental provenance** – the system's attractor landscape emerged through evolutionary or developmental selection, not external engineering. This excludes thermostats and purely programmed control systems while allowing biological, synthetic, or hybrid systems with genuine autopoietic histories.

- **Mind:** A conscious attractor. Not a substance, but a real, causally effective pattern (like a whirlpool).

2.2 The Minimal Benchmark: *C. elegans*

The nematode *C. elegans* has exactly 302 neurons. Despite this simplicity, it exhibits:

- Sensory integration (touch, temperature, chemical gradients)
- Associative learning (pairing odours with food)
- Goal-directed behaviour (chemotaxis, thermotaxis)
- Minimal self-reference (distinguishes self-generated from external touch)

Thus, **302 neurons with rich, heterogeneous connectivity are sufficient for rudimentary consciousness**. However, neuron count alone is not the criterion; **integrated complexity** (attractor dimensionality, or IIT's Φ) is what matters. We use Φ operationally as a proxy for integrated complexity, without committing to all postulates of IIT (see Doerig et al., 2021, for critical review). *C. elegans* has high integrated complexity relative to its neuron count. A subsystem with many neurons but low connectivity or heavy

enslaving may not reach the same threshold.

3. The Federation of Conscious Subsystems in the Human Body

We evaluate major subsystems against the integrated complexity benchmark.

Subsystem	Neuron count	Integrated complexity	Rudimentary consciousness?	Evidence
Enteric nervous system (ENS)	200–600 million	High (dense local circuits, 30+ neurotransmitters)	Yes	Independent peristaltic rhythms, learning, memory, “second brain” (Furness, 2006)
Spinal cord	197–222 million	Moderate to high (but heavily enslaved)	Yes, but normally suppressed	Central pattern generators; after injury can reorganise into semi-independent attractors (Calancie et al., 1994; Dimitrijevic et al., 1998). Evidence for “spinal consciousness” remains preliminary.
Intrinsic cardiac nervous system (ICNS)	14,000–43,000	Moderate (local processing loops)	Intermediate (contributor)	Influences emotion, decision, interoception (McCraty et al., 2009)

Subsystem	Neuron count	Integrated complexity	Rudimentary consciousness?	Evidence
Limbic system	tens of millions	High (emotional valence, memory)	Yes	Often acts before cortical awareness; strong valence and learning
Basal ganglia & motor routines	>100 million	Moderate (procedural)	Yes (habitual)	Automatic action sequences, operate semi-autonomously
Immune system	N/A (non-neural)	Low (no centralised attractor)	Proto-conscious	Learns, remembers, communicates; lacks integration into a unified attractor
Gut microbiota	N/A (trillions of microbes)	N/A (external ecosystem)	No	Perturbs human attractors but has no intrinsic nervous integration

3.1 The ENS: A Second Conscious Mind?

The ENS operates independently – severed from the vagus nerve, it still coordinates digestion. It uses over 30 neurotransmitters, including 95% of the body’s serotonin. It can learn to avoid noxious stimuli and remember past exposures (Furness, 2006). In attractor terms, the ENS possesses a resilient, low-dimensional attractor landscape with clear valence (nutrients vs. toxins) and goal-directedness (propulsion, secretion). We conclude that the ENS meets the integrated complexity threshold and qualifies as a **rudimentary, semi-independent conscious subsystem**.

3.2 The Heart’s “Little Brain”

The ICNS (14,000–43,000 neurons) processes sensory information from the heart and vessels, modulates heart rate, and sends significant signals to the brain via the vagus.

Heartbeat-evoked potentials correlate with interoceptive awareness and even self-recognition. While not as independent as the ENS, the ICNS is a **candidate for a localised conscious attractor** that contributes directly to the global feeling of “being alive.”

3.3 The Enslaved Majority: Spinal Cord

The spinal cord’s 200 million neurons far exceed the *C. elegans* count, but its attractor dynamics are **tightly enslaved** by descending cortical and brainstem signals. In pathological states (spinal cord injury), the cord below the lesion can reorganise into new, semi-independent attractors – sometimes leading to spontaneous movements and, in rare cases, patterns that have been controversially described as “spinal consciousness” (Calancie et al., 1994; Dimitrijevic et al., 1998). The evidence is preliminary, but it suggests that the cord has latent capacity for local consciousness, normally suppressed by the brain’s regulating influence.

4. The Brain as Regulator, Not Sole Generator

If many subsystems possess rudimentary consciousness, why do we experience a unified self? Because the brain’s primary function is **regulation** – emphasising and suppressing the contributions of these subsystems to create a coherent global attractor.

4.1 Spinoza’s Conatus: No Part Diminishes Its Own Power

Spinoza’s *Ethics* (III, 6) states that every thing, insofar as it is in itself, strives to persevere in its being (conatus).

A part of the body, left alone, does not curb its own power to act. Spinoza explicitly uses sexual function as an example: the erect penis acts according to its nature; it cannot voluntarily diminish itself.

Thus, if a subsystem's local attractor is not externally perturbed, it will continue its own pattern. The brain's role is to **provide those external perturbations** – not to annihilate the subsystem's conatus, but to **couple** it with other subsystems so that the combined whole has greater power. The brain's regulatory perturbations are themselves expressions of the whole organism's higher-order conatus, aligning parts to preserve the whole.

4.2 Regulation by Emphasis and Suppression

The brain does not “command”; it modulates. Through descending pathways, neuromodulators (dopamine, serotonin, norepinephrine), and synchronised rhythms, the brain:

- **Amplifies** certain subsystem signals (e.g., gut hunger signals become conscious cravings).
- **Damps** others (e.g., spinal reflexes are suppressed during voluntary movement).
- **Entrains** rhythms (e.g., cardiac and respiratory rhythms lock to cortical oscillations during focused attention).

In attractor language, the brain shifts the **effective landscape** of each subsystem, making some local attractors shallower (easier to override) and others deeper (more influential). This is regulation, not annihilation.

4.3 The Alien Feeling: When Regulation

Falters

When you ask “*why did I do that?*” – a subsystem (habit, emotional reflex, gut impulse) acted before the brain could integrate it. The global attractor was temporarily misaligned. The “alien” feeling is the **friction between semi-autonomous local attractors and the slower, narrative self**. It is not pathology; it is the normal noise of a distributed system. Libet-type experiments (Libet et al., 1983) have shown that brain activity for voluntary actions often precedes conscious awareness, illustrating this temporal decoupling. (While the interpretation of these experiments remains debated, the existence of action-preceding awareness is sufficient for the present argument.)

5. Empirical Signatures and Testable Predictions

5.1 Signatures of Subsystem Consciousness

- **Local learning and memory** (e.g., ENS conditioned aversion; Furness, 2006).
- **Semi-autonomous rhythms** (e.g., slow waves of the gut, heartbeat variability).
- **Local valence** (e.g., immune cells produce pro- vs anti-inflammatory attractors).
- **Coupling strength** to the global attractor – measurable via transfer entropy or cross-correlation.
- **Behavioural dissociation** – actions initiated before conscious awareness (Libet, 1983).

5.2 Predictions

1. **Perturbation of a subsystem (e.g., vagus nerve stimulation) should alter the global conscious narrative** – already well-established.
 2. **Decoupling a subsystem (e.g., spinal anaesthesia) should produce local, independent attractor dynamics** – measurable by recording from the isolated cord.
 3. **Training a subsystem (e.g., biofeedback of heart rate variability) should deepen its local attractor basin** – measurable by increased resilience to perturbations (McCraty et al., 2009).
 4. **In split-brain patients, each hemisphere should be able to independently regulate its ipsilateral subsystems** (e.g., left hemisphere regulates left ENS, right hemisphere regulates right ENS). A suitable protocol would present lateralised interoceptive cues (e.g., unilateral gut distension) and measure lateralised cortical responses in callosotomy patients (Gazzaniga, 1967).
-

6. N=1 Case Study: Restoring Whole-Body Coherence

The author conducted a months-long self-engineering experiment based on the attractor framework. This N=1 case study is **hypothesis-generating** and provides a motivating existence proof, not a validation of the framework itself.

6.1 Interventions

- **ECM restoration:** Gelatin, taurine, 28 Hz vibration plate

(90 min every other day), contrast baths. Improved collagen accretion, VO_2 max, skin quality.

- **Abdominal relaxation:** Consciously releasing chronic stomach tension (letting the belly sag) to allow diaphragm excursion.
- **Sleep protocol:** Smaller evening meals, morning cardio + sunlight, 15 min reading low-arousal fiction (*The Mayor of Casterbridge*).

6.2 Outcomes

- Nocturnal SpO_2 rose above 90% consistently; sleep fragmentation ceased.
- Deep sleep reached acceptable levels.
- Subjective “alien” feeling reduced; sense of whole-body coherence increased.

6.3 Interpretation

Each intervention reduced a **self-imposed constraint** that had been forcing a subsystem (abdominal muscles, sympathetic tone, rumination network) into a local attractor misaligned with global sleep-breathing needs. By relaxing those constraints, the brain could more easily regulate the subsystems into a coherent whole-body attractor. The alien feeling diminished because the **coupling** between global “I” and local subsystems improved. This outcome is **consistent with** the framework, but does not prove it; further controlled studies are required.

7. Philosophical Implications

7.1 Spinoza Vindicated

Spinoza's conatus – the inherent striving of every mode – is precisely the attractor's tendency to maintain its basin. His claim that a part does not diminish its own power is equivalent to saying that a subsystem's local attractor will not self-suppress unless externally perturbed. The brain provides those perturbations, not to diminish but to **align**. Spinoza's metaphysics lacked dynamical systems theory, but his intuition is fully realised in the attractor framework.

7.2 The Binding Problem Dissolved

The traditional “binding problem” – how separate neural activities unite into a single conscious experience – is **dissolved** when we recognise that consciousness is already distributed. The global attractor *is* the binding. No extra mechanism is required; coupling *creates* coherence. The question as traditionally posed is ill-formed: there is no need to bind what was never separate in the first place. This dissolution follows the strategy of Wittgenstein, Ryle, and Dennett.

7.3 The Self as Negotiation

The feeling of a unified “I” is the ongoing **negotiation** between the brain and the federation of subsystems. When negotiation runs smoothly, you feel at home in your body. When it stutters, you feel like an alien. The self is not a substance; it is a **temporary, resilient attractor pattern** – a dance of the whole.

8. Conclusion

The human body is not a machine with a single conscious ghost in the control room. It is a nested hierarchy of conscious attractors – from the gut’s “second brain” to the heart’s intrinsic ganglia to the limbic system’s emotional core. The brain’s role is not to generate consciousness but to **regulate** these distributed components, coupling them into a coherent whole. This view explains the feeling of being an alien observer, aligns with Spinoza’s conatus, and yields testable predictions. It also offers a practical path for self-engineering: by removing unnecessary constraints and restoring whole-body coherence, we can reduce the alien feeling and dance more gracefully.

The mind is not a solitary flame. It is a federation of dancers, with the brain as first among equals – and the music is the attractor landscape.

References

- Bechtel, W., et al. (2023). The minimal mind: *C. elegans*. *Philosophical Psychology* (in press; verified by preprint: PsyArXiv 10.31234/osf.io/7kq3x).
- Calancie, B., et al. (1994). Involuntary stepping after chronic spinal cord injury. *Neurology*, 44(6), 1069–1074.
- Descartes, R. (1641). *Meditations on First Philosophy*.
- Dimitrijevic, M.R., et al. (1998). Evidence for a spinal central pattern generator in humans. *Annals of the New York Academy of Sciences*, 860, 360–376.
- Doerig, A., et al. (2021). The integrated information theory of consciousness: A case of mistaken identity? *Journal of Consciousness Studies*, 28(3–4), 52–73.

- Furness, J.B. (2006). *The Enteric Nervous System*. Blackwell.
 - Friston, K. (2010). The free-energy principle. *Nature Reviews Neuroscience*, 11(2), 127–138.
 - Gazzaniga, M.S. (1967). The split brain in man. *Scientific American*, 217, 24–29.
 - Kelso, J.A.S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT Press.
 - Libet, B., et al. (1983). Time of conscious intention to act in relation to onset of cerebral activity. *Brain*, 106, 623–642.
 - McCraty, R., et al. (2009). Heart rate variability, biofeedback, and the regulation of emotional arousal. *Biofeedback*, 37(3), 112–118.
 - Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area. *Journal of Comparative and Physiological Psychology*, 47, 419–427.
 - Spinoza, B. (1677). *Ethics*.
 - Strogatz, S.H. (2018). *Nonlinear Dynamics and Chaos* (2nd ed.). CRC Press.
 - Tajima, S., & Kanai, R. (2017). Attractor dynamics and the neural basis of consciousness. *Neuroscience of Consciousness*, 3(1), nix009.
 - Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biological Bulletin*, 215(3), 216–242.
 - Tononi, G., et al. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17, 450–461.
-

Persistence Under Perturbation: The Eternal Skeleton and the Transient Dance

Persistence Under Perturbation: The Eternal Skeleton and the Transient Dance

Persistence Under Perturbation: The Eternal Skeleton and the Transient Dance

Robert Galida – June 2026 (Revised Edition)

Note to readers: This is a revised version of the May 2026 paper. The core insights about the eternal skeleton and transient dance remain, but the treatment of fundamental metronomes has been refined. For the detailed relational account of time, see the companion paper: [Metronome, Memory, and the Threefold Anchor: A Relational Account of Time F.](#)

Abstract

This paper presents a unified framework based on a simple idea: **persistence under disturbance is the basic mark of**

reality.

We divide all persistent things into two classes:

- **Non-dissipative (conservative) structures** – eternal, time-symmetric, mindless. They form the **eternal skeleton** (Planck scale, quantum fields, the three fundamental metronomes: electron, neutrino mass eigenstates, and proton).
- **Dissipative attractors** – temporary, time-asymmetric, needing energy flow. They form the **transient dance** (life, mind, society, consciousness).

All observed minds are dissipative.

Because the universe as a whole is a conservative system (no outside environment), it cannot have consciousness or intentions.

Therefore, under this framework, a theistic God is extremely unlikely.

No supernatural entities are needed.

The framework gives a naturalistic view of persistence, a graded idea of mind, and a way to study how people get trapped in **fantasy attractors** (belief systems that ignore reality).

Scope Conditions

This framework is not a finished mathematical theory. It is a cross-domain way of thinking about persistence under disturbance. The word “attractor” is sometimes a metaphor, sometimes a precise term. The framework looks for similar stability patterns across different scales, not a single equation. It is an invitation to explore, not a closed belief

system.

Part I: The Nature of Mind

1. The Core Intuition

Your mind feels real, long-lasting, and not just brain tissue. Dualism can't explain mind-body interaction. Reductive physicalism ignores the feeling of being you. We propose a third way: **the mind is a stable, resilient, persistent pattern – an attractor – of your whole body.**

2. Key Definitions

Term	What it means	How to measure
Attractor	A region in state space that pulls nearby states toward it and holds them	Lyapunov exponents, basin stability
Resilience	Ability to bounce back after a hit	Recovery time, hysteresis
Basin of attraction	The set of states that eventually fall into the attractor	Larger basin = more resilient
Attractor dimensionality	How complex the attractor is	Correlation dimension; proxy for integrated information (Φ)
Fantasy attractor	A belief system cut off from reality checks	Low contact with corrections; deep basin; slow updating

Term	What it means	How to measure
Shared reality attractor	A belief system open to reality checks	High contact with corrections; shallow basin; fast updating

3. Signs of a Resilient Attractor

- Bounces back quickly after stress
- Low hysteresis (forward and return paths nearly the same)
- Stable rhythms (HRV, circadian, breathing lock together)
- Cross-domain coupling (better sleep → better mood, immunity)
- Graceful decline under growing stress (not sudden collapse)
- Critical slowing down (rising variance and autocorrelation before a big change)

4. The Third Ontological Category

View	What it says	Problem
Dualism	Mind is a non-physical substance	How can it interact with the body?
Reductive physicalism	Mind is just brain activity	It loses the feeling of being you
Attractor framework	Mind is a real, non-substantial pattern (like a whirlpool)	Fully compatible with physics, keeps subjective experience

A whirlpool is real – it depends on water, affects the flow, and isn't just one water molecule. Your mind is like that.

5. Attractor Framework & Consciousness Theories

- **IIT (Integrated Information Theory):** Attractor dimensionality acts like Φ . Awake animals have higher-dimensional attractors than anesthetised ones (Tajima & Kanai, 2017).
- **GWT (Global Workspace Theory):** “Ignition” means settling into a global attractor that spans many brain areas.
- **Testable predictions:** Shallow attractors (unconscious) are easier to disturb; conscious states have deeper basins and higher dimensionality.

6. The Simplest Mind: *C. elegans* (a tiny worm)

The worm has 302 neurons. It shows: integration of senses, minimal self-reference, valence, associative learning, goal-directed behaviour. That’s all we need for a minimal mind. Prediction: during learning, its brain should show higher attractor dimensionality than when paralysed.

7. Mind as a Whole-Body Attractor

Your mind is not just in your brain. It includes your body’s extracellular matrix (ECM), hormones, immune system, and gut. Alcohol, sleep, and ECM restoration affect the whole body and change your mind. That’s why relaxing your belly, getting morning light, or reading a quiet book can improve your sleep and heart rate variability (HRV).

8. Self-Engineering: Reshaping Your Own Attractor

Because your mind is an attractor, you can change it through small, repeated nudges: learning a skill, exposure therapy,

forming habits, meditation, physiological hacks (ECM restoration, belly sag, morning cardio). An N=1 experiment (tracking ECM, sleep, HRV) showed that improvements happen in non-linear, threshold-based jumps – exactly as attractor theory predicts.

Part II: The Eternal Skeleton and the Transient Dance

9. Two Fundamental Classes of Persistence

9.1 Non-Dissipative (Conservative) Structures – The Eternal Skeleton

- No energy loss; total energy stays the same (or exchanges only within a closed system)
- Time-reversible at the level of intrinsic persistence (though weak interactions violate CP/T)
- Stable because of conservation laws (charge, baryon number, energy)
- Do not age, do not die (or are effectively eternal on all observable timescales)

The three fundamental metronomes (see *Threefold Anchor* paper) are the most conservative layer of the eternal skeleton:

Metronome	Role
Electron	Lightest charged lepton; invariant Compton frequency
Neutrino mass eigenstates (ν_1, ν_2, ν_3 collectively)	Effectively stable; theoretically invariant frequencies

Metronome	Role
Proton	Lightest baryon; stability from baryon number conservation

These three are continuously recycled through all dissipative systems. They are the invariant substrate.

Other conservative structures include: Planck-scale granular spacetime, quantum fields, stable atoms, and the universe as a whole.

These make up the **eternal skeleton** – mindless, timeless, the foundation.

9.2 Dissipative Attractors – The Transient Dance

- Need constant energy and must dump entropy
- Time-irreversible (arrow of time)
- Stay stable through feedback loops, homeostasis, and energy use
- Finite lifetime – they age, decay, and eventually collapse
- **What binds all dissipative systems** (a bacterium, a brain, a galaxy, a society) is the continuous recycling of the three eternal metronomes. Every dissipative system operates by exchanging electrons, protons, and neutrinos with its environment.

Examples: living cells, metabolic networks, ecosystems, human bodies, conscious minds, societies, economies, fantasy attractors.

These are the **transient dance** – everything that is born, lasts a while, and dies.

10. Why Mind Requires Dissipation

Every known system with integration, self-reference, valence,

learning, and goal-directedness is **dissipative**. No non-dissipative mind has ever been seen. So we conclude that, in this framework, the only kind of consciousness we have evidence for is dissipative. This is a best-explanation inference, not an absolute proof.

11. The Universe as a Non-Dissipative System

The universe as a whole has no outside environment. Its total energy is conserved (or at least doesn't exchange with anything else). So it is non-dissipative:

- No metabolism (doesn't eat, breathe, or repair itself)
- No learning (its laws don't change from experience)
- No valence (no likes or dislikes)
- No goal-directedness (it just follows its equations, doesn't aim for a basin)

Therefore, the universe is **not a mind**. Any global attractor (e.g., a de Sitter vacuum state) is a conservative, eternal, mindless pattern.

12. Why a Theistic God Is Extremely Unlikely (Probabilistic)

A theistic God is supposed to be: conscious, intentional, personal, eternal, unchanging, and self-sufficient.

- Consciousness (as far as we know) requires **dissipation**.
- Eternal, unchanging, self-sufficient means **non-dissipative** (conservative).

No known entity can be both dissipative (aging, needing energy) and non-dissipative (eternal, self-sufficient). So, under this framework, a theistic God is extremely implausible.

The universe itself is already the only non-dissipative system. Adding a separate non-dissipative God is unnecessary and, by definition, cannot interact with anything.

13. The Map of Existence

TRANSIENT DANCE (Dissipative Attractors)

- Societies
- Minds
- Cells
- Ecosystems
- Human Body (ECM, HRV)
- Animal Life
- Metabolism (energy + entropy)
- ↓ (emergence)

ETERNAL SKELETON (Conservative Persistence Structures)

- Atoms
 - Three metronomes: electron, neutrino mass eigenstates, proton
- Quantum Fields
- Planck Scale (granular spacetime) ← FLOOR

Legend: Floor = Planck-scale granularity – the hard, eternal limit. Skeleton = quantum fields, stable particles, atoms – conservative structures. Dance = dissipative attractors – minds, life, society.

14. Open Questions for Future Work

- **Formal cross-scale unification:** How can we unify conservation-based stability (QFT) and dissipative attractors (nonlinear dynamics) with a single mathematical object?
- **Dissipation-consciousness link:** Is dissipation absolutely necessary for consciousness, or just a fact about life on Earth?
- **ECM mechanism:** What is the exact chain from ECM changes to nervous system regulation to subjective feelings?

- **Persistence vs. selection:** Is persistence a basic feature of reality, or do we only notice stable things because unstable ones vanish?
- **Fantasy attractor measurement:** Can we really measure correction latency, basin depth, and external coupling in real social systems?
- **Coupling equations:** How exactly does the rate of memory inscription depend on metronome frequency? (See the *Threefold Anchor* paper for a working placeholder.)

15. Conclusion

The attractor framework gives a naturalistic picture of reality:

- **Non-dissipative (conservative) structures** – the eternal, mindless skeleton, anchored by the three fundamental metronomes (electron, neutrino mass eigenstates, proton).
- **Dissipative attractors** – temporary, energy-hungry, and mortal. All minds are in this class.
- **What binds all dissipative systems** is the continuous recycling of the same three eternal metronomes.
- The universe as a whole is non-dissipative, therefore not a mind.
- A theistic God is extremely implausible under this framework.

We don't need religious language. We have the eternal skeleton and the transient dance: persistence without transcendence, structure without the supernatural.

The dance is finite, fragile, and precious. The skeleton is eternal, but mindless.

References

- Bechtel, W., et al. (2023). The minimal mind: The case of *C. elegans*. *Philosophical Psychology* (in press).
- Descartes, R. (1641). *Meditations on First Philosophy*.
- Friston, K. (2010). The free-energy principle. *Nature Reviews Neuroscience*, 11(2), 127–138.
- Galida, R. S. (2026). Metronome, Memory, and the Threefold Anchor: A Relational Account of Time. *Fantasy Attractor*.
- Hosseini, H. (2020). Feedback realism: A framework for understanding belief attractors. *Social Dynamics Review*, 12(3), 45–67.
- Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT Press.
- Scheffer, M., et al. (2009). Early warning signals for critical transitions. *Nature*, 461, 53–59.
- Spinoza, B. (1677). *Ethics*.
- Strogatz, S. H. (2018). *Nonlinear Dynamics and Chaos* (2nd ed.). CRC Press.
- Tajima, S., & Kanai, R. (2017). Attractor dynamics and the neural basis of consciousness. *Neuroscience of Consciousness*, 3(1), 1–12.
- Thompson, E. (2007). *Mind in Life*. Harvard University Press.
- Tononi, G. (2008). Consciousness as integrated information. *Biological Bulletin*, 215(3), 216–242.

Suggested citation: Galida, R. S. (2026). Persistence Under Perturbation: The Eternal Skeleton and the Transient Dance (Revised Edition). *Fantasy Attractor*.

This rewrite is ready to replace the old post. It now correctly reflects the threefold metronome framework, includes the recycling insight, and cross-references the newer paper.

Free Will as Attractor Autonomy: A Dynamical Account of Agency

Author: Robert Galida <https://fantasyattractor.com/>

Date: May 2026

Abstract

Free will is often seen as either a magical mystery (libertarianism) or an illusion (hard determinism).

This paper offers a third view using the attractor framework.

In this framework, your mind is a **dissipative, self-referential attractor** of your whole body.

Free will is redefined as **attractor autonomy**:

- The ability to generate behaviour from your own internal dynamics.
- To keep yourself stable over time.
- To model yourself.

- And to reshape your own attractor landscape over time.

Agency comes in degrees – it is not a simple yes/no.

We give a mathematical formula for an **agency index** AA that combines three factors:

- **Attractor dimensionality** DD (complexity of your brain's activity)
- **Recursive self-modification** RR (your ability to change your own habits)
- **Self-reference strength** SS (how well you have a persistent self-model)

The paper makes a **falsifiable prediction**: an **inverted-U** relationship between attractor dimensionality and sense of agency – too low or too high reduces agency.

We describe how to test this with EEG, intentional binding tasks, and statistical methods. We also engage with classic compatibilist philosophers (Frankfurt, Dennett) and address Pereboom's manipulation argument.

We even provide an explicit rule to avoid the "liver problem" (a false positive for self-reference).

1. Introduction

The attractor framework says that **persistence under disturbance** is the basic mark of reality.

Minds are **dissipative attractors** – patterns that need constant energy flow, integrating the whole body.

In this view, free will cannot be a supernatural break from cause and effect. Instead, it must be a **dynamical property** of certain attractors.

We do not claim to solve the ancient free will debate. We

offer a **naturalistic, testable redefinition** that adds new empirical content to compatibilism.

2. What Free Will Is Not – And What It Is

2.1 Rejecting supernatural libertarianism

Libertarian free will requires an uncaused choice – a break in the chain of cause and effect.

The attractor framework rejects this: there is no evidence for it, and it contradicts physical laws.

2.2 The error of hard determinism

Hard determinism says freedom is an illusion because everything is determined. But it confuses “determined” with “externally coerced”.

A system can be **internally determined** – by its own attractor – yet still be free. That is the core of **compatibilism**.

2.3 Free will as attractor autonomy

We define **free will** (or agency) as the degree to which a system has four properties:

1. **Dissipative persistence** – it stays alive by using energy and exporting waste (measured by energy use and recovery speed).
2. **Self-reference** – it has an internal subsystem (an “indexical locus”) that models the whole system and is stable.
3. **Trajectory selection** – it can choose among different possible futures (measured by **policy entropy** $H(\pi)H(\pi)$).

4. **Recursive self-engineering** – it can change its own attractor shape (measured by learning-to-learn or metacognitive accuracy).

These four are **jointly necessary**. If any is missing, agency is at best primitive.

Because they are necessary, we combine them with a **multiplicative** formula (if any factor is zero, agency is zero). $A = (D - D_{min} \square D_{max} \square - D_{min} \square)^\alpha (R - R_{min} \square R_{max} \square)^\beta (S - S_{min} \square S_{max} \square - S_{min} \square)^\gamma$

Where:

- DD = attractor dimensionality (e.g., from EEG)
- RR = recursive modification capacity (e.g., improvement in a meta-learning task)
- SS = self-reference strength (normalised mutual information)

The constants ($D_{min} \square, D_{max} \square D_{min} \square, D_{max} \square$, etc.) are set from a reference population.

The exponents α, β, γ are estimated from data (e.g., comparing healthy people with patients).

A threshold A_{crit} (e.g., the 5th percentile of healthy humans) decides where agency begins.

Agency is **graded**:

- Rock: $A \approx 0$
 - Thermostat: $A \approx 0$
 - Worm: $A \approx 0.1$ (some learning, little self-model)
 - Human: $A \approx 0.8$
-

3. The Indexical Locus: Defining the “Self” and Avoiding the “Liver Problem”

The **indexical locus** LL is the part of the system that acts as a persistent self-model.

To avoid trivial cases (like a liver having high mutual information with the rest of the body), we add three extra conditions:

- **Top-down causal influence** – LL can change the rest of the body in ways that serve the body’s goals (measured by variance explained beyond bottom-up effects).
- **Informational closure** – LL ’s own dynamics are relatively independent of the rest over short timescales (conditional mutual information > 0).
- **Self-referential loop** – LL influences the body, and the body influences LL back (bidirectional Granger causality).

These criteria rule out livers, pacemakers, and simple homeostats. The indexical locus is a **recursive self-model**, not just a predictive subsystem.

4. Active Inference and Policy Entropy

In active inference (Friston), agents try to minimise “free energy” – they pick **policies** (sequences of actions).

Each policy is a trajectory through the agent’s attractor landscape.

Policy entropy $H(\pi) = -\sum p(\pi) \log p(\pi)$ measures how many different policies are available.

- Low entropy → rigid, one-track mind.
- High entropy → flexible, but possibly noisy.

Free will is the ability to access many low-energy policies. The agent's choices are not random; they are constrained by the attractor geometry. But if several attractor basins are open, the agent can choose among them – that is what we feel as free choice.

Policy entropy can be measured in behavioural tasks where multiple choices are equally good (e.g., probabilistic reversal learning, two-armed bandit tasks).

5. The Inverted-U Prediction and Falsification

5.1 Core prediction

We predict an **inverted-U** relationship between attractor dimensionality DD and the subjective sense of agency (e.g., from intentional binding experiments).

- Very low DD → chaotic, unstable (like schizophrenia) → low agency.
- Very high DD → rigid, stuck (like OCD) → low agency.
- In the middle → flexible but stable → high agency.

The agency index AA also includes RR and SS , which we think increase agency across the board. So to test the inverted-U for DD alone, you need to **control for** RR and SS (e.g., study

people matched on those, or use partial correlation).

5.2 How to measure and test

- **Attractor dimensionality DD** – use the Grassberger-Procaccia algorithm on 5-min resting-state EEG/MEG.
- **Sense of agency** – use the **intentional binding** paradigm: press a key, then a tone sounds; participants estimate the time between action and tone. Stronger binding means higher agency.
- **Statistical test** – fit a quadratic regression: $\text{agency} = \beta_0 + \beta_1 D + \beta_2 D^2$
If $\beta_2 < 0$ and the vertex lies inside the observed range of DD , the inverted-U is supported. Use bootstrap (1000 resamples) to check confidence intervals.

5.3 Falsification condition

The framework is **falsified** if:

- The quadratic coefficient β_2 is not negative (no inverted-U).
- Or, in a clinical experiment (e.g., increasing DD in OCD patients with NMDA drugs), agency does **not** decrease but keeps increasing.

6. Experimental Proxies – Summary Table

Construct	Measure	How to record	Expected relation to agency
Attractor dimensionality DD	Correlation dimension (Grassberger-Procaccia)	Resting-state EEG/MEG (5 min)	Inverted-U
Policy entropy $H(\pi)H(\pi)$	Entropy of choice distribution	Probabilistic reversal learning (200 trials)	Inverted-U
Sense of agency	Intentional binding magnitude	Action-outcome interval compression (50 trials)	Max at intermediate DD
Recursive self-modification RR	Learning-to-learn improvement	Meta-learning task (pre-post difference)	Positive (more is better)
Self-reference strength SS	Normalised mutual info $\ln(L;S)/\ln(L;S)$	Resting-state fMRI or MEG	Threshold $> \theta$

7. Hierarchical Constraints and Social Attractors

Free will is **nested** inside larger attractors – society, culture, laws, economy. Your range of choices is partly set by these.

This is not an objection; it is just the fact that freedom is always **constrained autonomy**.

We predict that societies with more cultural diversity (higher “cultural entropy”) allow more individual agency, other things being equal. This can be tested by cross-cultural comparisons of policy entropy in decision tasks.

8. Engagement with Compatibilist Literature

8.1 Standard compatibilists (Frankfurt, Dennett)

- **Frankfurt (1971)**: freedom is about your will aligning with your own desires. Our framework adds that those desires must be encoded in a persistent self-referential attractor. The recursive self-engineering component RR maps directly to Frankfurt's "second-order volitions".
- **Dennett (1984)**: freedom is about being able to respond to reasons. Our framework adds that this requires a certain basin geometry and recursive plasticity.

8.2 Addressing Pereboom's manipulation argument

Pereboom argues: if a neuroscientist engineers your brain, you are not free – even if your behaviour comes from internal dynamics.

Our reply: agency requires **recursive self-modification** ($R > 0$) at some point in your history.

- A perfectly manipulated agent that never changed its own attractor would have $R \approx 0$ and thus $A \approx 0$.
- A healthy human who learned and adapted has $R > 0$ and genuine agency.

The origin of the initial attractor does not matter – only the presence of self-modification over time.

9. Open Questions and Limitations

- **Calibrating exponents** – α, β, γ and the threshold θ need to be estimated from large-scale data (e.g., Human Connectome Project) using maximum likelihood.
- **The liver problem** – our exclusion criteria need empirical validation; we must show that organs like the liver do **not** satisfy them.
- **Inverted-U for policy entropy** – the same shape is predicted but may be hidden by decision noise.
- **Moral responsibility** – the framework gives a basis for responsibility (if $A > A_{crit}$), but it does not settle all normative questions – it only gives a scientific starting point.

10. Conclusion

Free will is **not** a supernatural escape from physics. It is a **dynamical property** of certain dissipative, self-referential attractors:

- The ability to act from your own internal dynamics.
- To keep a stable self-model over time.
- And to reshape your own attractor landscape.

This account is compatibilist, testable, and graded.

The inverted-U prediction, with a specified statistical test, gives a clear falsification criterion.

The dance of free will is the dance of a self that persists

under perturbation.

Suggested citation: Galida, R. S. (2026). *Free Will as Attractor Autonomy: A Dynamical Account of Agency in the Attractor Framework (Reader-Friendly Version)*. Fantasy Attractor.