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., ≥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 ≥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 ≥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/

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** A*A* that combines three factors:

- **Attractor dimensionality** D*D* (complexity of your brain's activity)
- **Recursive self-modification** R*R* (your ability to change your own habits)
- **Self-reference strength** S*S* (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(π)$H(\pi)$).
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−Dmin⁡Dmax⁡−Dmin⁡)α(RRmax⁡)β(S−Smin⁡Smax⁡−Smin⁡)γ$A=\left(\frac{D-D_{\min}}{D_{\max}-D_{\min}}\right)^{\alpha}\left(\frac{R}{R_{\max}}\right)^{\beta}\left(\frac{S-S_{\min}}{S_{\max}-S_{\min}}\right)^{\gamma}$

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 (Dmin⁡,Dmax⁡$D_{\min}, D_{\max}$, etc.) are set from a reference population.
The exponents α,β,γα,β,γ are estimated from data (e.g., comparing healthy people with patients).
A threshold AcritA_{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** L 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** — L 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** — L's own dynamics are relatively independent of the rest over short timescales (conditional mutual information > 0).
- **Self-referential loop** — L influences the body, and the body influences L 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 D 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: agency = $\beta 0 + \beta 1 D + \beta 2 D^2$ $\beta 0 + \beta 1 D + \beta 2 D 2$.
  If $\beta 2 < 0$ $\beta 2 < 0$ 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 $\beta 2$ $\beta 2$ 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(π)$H(\pi)$	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 In(L;S)$In(L;S)$	Resting-state fMRI or MEG	Threshold > θ

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 R*R* 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*R*>0) at some point in your history.

- A perfectly manipulated agent that never changed its own attractor would have R≈0*R*≈0 and thus A≈0*A*≈0.
- A healthy human who learned and adapted has R>0*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.