

# Trapped Navigation: Addiction, Trauma, and OCD as Conscious Suppression of Intelligent Correction [A] (2026)

Robert Galida – June 2026 (Final)

*Paper 2 in a series on conscious suppression; see [Paper 1: Intelligence Without Consciousness](#) for the full taxonomy of intelligence and consciousness.*

---

## Abstract

Why do people with addiction, trauma-related avoidance, or obsessive-compulsive disorder often know their behavior is maladaptive yet cannot stop? Standard explanations – impaired executive control, habit dominance, weak insight – are incomplete. This paper applies the attractor framework's suppression mechanism. In each disorder, the person *detects* the discrepancy between behavior and goals (insight is intact), but **phenomenal, identity-constitutive investment** – the felt urgency of craving, the necessity of avoidance, the compulsion to ritualize – deepens the attractor basin relative to corrective perturbations. The suppression is not a failure of intelligence; it is a dynamical competition between attractors. The paper distinguishes this account from dual-process and executive-control theories, provides falsifiable diagnostic criteria, and discusses treatment implications (why insight alone fails). Acknowledgment is made that for addiction, the relationship between incentive

salience (*wanting*) and phenomenal consciousness remains contested; the model targets the subset of craving states that patients report as felt urgency.

---

## 1. Introduction: The Paradox of Insight Without Change

A person with alcohol use disorder knows that drinking damages their health, relationships, and future. Yet when a craving arises, they drink. A trauma survivor knows that the parking garage is safe, yet they avoid it. A person with OCD knows that the ritual is irrational, yet they perform it.

Standard explanations invoke **impaired executive control** (Volkow et al., 2016), **habit dominance** (Balleine & Dickinson, 1998), or **lack of insight** (Amador et al., 1994). But these accounts do not explain why the person can articulate the harm, describe counterarguments, and intend change, yet the behavior persists. Executive control may be intact in non-trigger contexts; habits may be sensitive to goal-level knowledge; insight may be partial or oscillating.

The attractor framework provides a model of **motivational competition** where a conscious, identity-binding urge temporarily overrides the correction signal. In *Intelligence Without Consciousness* (Galida, 2026), we introduced **conscious suppression**: phenomenal, identity-constitutive commitment deepens an attractor basin, making it resistant to corrective perturbations. This paper applies that mechanism to addiction, trauma-related avoidance (PTSD), and OCD. It does not deny executive or habit deficits; it proposes that in many cases, a conscious-level attractor competition is the primary obstacle to change.

---

## 2. Defining Conscious Suppression (Self-Contained Glossary)

For readers unfamiliar with Paper 1:

- **Attractor basin** – the set of states from which a system returns to a stable pattern. A deeper basin resists larger perturbations.
- **Corrective permeability ( $\kappa$ )** – responsiveness to error signals;  $\kappa = 1/\tau$ , where  $\tau$  is return time to baseline after a perturbation.
- **Conscious suppression** – a process where the person *experiences* an urge, fear, or compulsion as felt, identity-relevant, and *not chosen* (non-deliberative), yet the depth of that attractor prevents escape from the maladaptive behavior. This corresponds to **Level 3** in Paper 1: detection of error + suppression via basin depth. Level 2 (automatic bias without error detection) and Level 1 (unfamiliarity) are not the target.

**On sealing mechanisms:** The paper treats sealing mechanisms (e.g., rationalizations) as *attractor-consistent outputs* generated by the basin state, not as deliberate strategic choices. Although they may *feel* deliberate to the patient, the model treats them as expressions of the attractor's depth, not as independent volitional acts. This resolves the tension between “non-deliberative urgency” and the deployment of rationalizations.

---

### 3. Empirical Grounding

#### **Addiction:**

Volkow et al. (2016) demonstrate that chronic substance use impairs prefrontal executive function in a state-dependent manner – deficits emerge under craving or stress, not at all times. Individuals can maintain intact verbal knowledge of consequences and express intention to stop (Goldstein et al., 2009). The craving state has been modeled as a competing attractor (Redish, 2004; Gutkin et al., 2006). **Incentive-saliency theory** (Robinson & Berridge, 1993, 2008) distinguishes *wanting* (which can be non-conscious) from *liking*. The present model targets the subset of craving states that are *phenomenally accessible* – the patient's reported felt urgency. This is a narrower claim; the paper does not assume that all incentive-saliency processes are conscious.

#### **PTSD & avoidance:**

Extinction recall deficits (Milad et al., 2006) are well documented, but they do not fully account for conscious fear as *necessary* even when safety is known. Meta-analyses confirm vmPFC–amygdala decoupling in PTSD (e.g., Etkin & Wager, 2007, and subsequent reviews). Ecological momentary assessment (EMA) studies in representative samples show that individuals with PTSD often report high certainty of safety before trigger environments yet avoidance persists (see, e.g., reviews of EMA in PTSD). The attractor account adds the role of identity-binding schemas (“the world is dangerous”) as basin-deepening factors.

#### **OCD:**

The DSM-5-TR includes an insight specifier: *good/fair, poor, or absent*. Approximately 25–30% of individuals with OCD have poor insight (Catapano et al., 2010). This paper targets the **good-insight subgroup** (where the person recognizes irrationality). For poor-insight patients, the mechanism may

be closer to Level 2 (automatic compulsion without error detection).

### Recent literature (2015–2025):

- EMA studies of craving show that momentary urge strength predicts relapse better than global insight (Serre et al., 2015; Shiffman et al., 2020).
  - OCD outcome studies confirm that poor insight predicts worse response to ERP (García-Soriano et al., 2021). Good-insight patients still show substantial residual symptoms, consistent with a competition model.
  - Identity-shifting interventions for addiction (Best et al., 2016) support the importance of decoupling selfhood from “addict” identity.
- 

## 4. Three Clinical Patterns

### 4.1 Addiction

- **Mechanism:** Craving as a state-dependent attractor that overrides goal-directed control when triggered. Identity fusion (“I am an addict”) deepens the basin where present, but is not universal.
- **Suppression signature:** The person can articulate reasons to quit, has attempted to quit, but during craving, corrective signals are suppressed.
- **Sealing mechanisms:** Cognitive rationalizations (“just this once,” “I need it to cope”) that block the error signal from updating the basin – treated as attractor-consistent outputs, not deliberate choices.

## 4.2 Trauma-Related Avoidance (PTSD)

- **Mechanism:** Conditioned fear creates an avoidance attractor. Safety knowledge may be intact, but felt necessity dominates.
- **Suppression signature:** “I know it’s safe, but I can’t go in.”
- **Identity fusion:** “The world is dangerous” as a self-defining schema.

## 4.3 Obsessive-Compulsive Disorder (OCD – Good Insight Subgroup)

- **Mechanism:** Anxiety drives compulsions that temporarily reduce distress, despite knowledge of irrationality.
- **Suppression signature:** “I know it doesn’t make sense, but I have to do it.”
- **Sealing mechanisms:** “Better safe than sorry,” “It’s a small price to pay for certainty.”

---

## 5. Transdiagnostic Table

Disorder	Error signal detected	Conscious investment	What maintains basin depth (mechanism)
Addiction	Knowledge of negative consequences	Craving (felt urgency)	Reinforcement schedule + state-dependent executive impairment + (sometimes) identity fusion

<b>Disorder</b>	<b>Error signal detected</b>	<b>Conscious investment</b>	<b>What maintains basin depth (mechanism)</b>
Trauma avoidance	Safety knowledge (cognitive)	Fear (felt necessity)	Extinction resistance + hyperarousal + schema of danger
OCD (good insight)	Knowledge of irrationality	Anxiety (felt urgency)	Negative reinforcement via distress reduction + certainty-seeking belief

## **6. Diagnostic Criteria for Clinical Fantasy Attractors (Operationalized)**

A patient's presentation is a **candidate** clinical fantasy attractor if it meets **three of five** criteria (provisional threshold; empirical validation required). The Level 2/3 distinction requires momentary assessment (see §7).

- 1. Insight intact:** The patient can state, unprompted, the discrepancy between behavior and goals. *Operationalization:* Score  $\geq 4$  on the Brown Assessment of Beliefs Scale (BABS) insight item, or equivalent.
- 2. Conscious urgency:** The maladaptive behavior is preceded by a felt, urgent state (craving, fear, anxiety) rated by the patient as "overwhelming" or "necessary." *Operationalization:* Momentary ecological assessment (EMA) rating  $> 7/10$  before the behavior.
- 3. Identity fusion:** The patient endorses that the behavior or its avoidance is central to selfhood (e.g., "I am an addict," "I must do this to be safe"). *Operationalization:* Endorsement of at least one identity statement on a structured interview.

4. **Low corrective permeability in trigger contexts:** Repeated corrective information (psychoeducation, feedback) does not reduce the behavior. *Operationalization:* No significant reduction after three sessions of evidence-based psychoeducation alone.
5. **Sealing mechanisms:** The patient spontaneously uses rationalizations that neutralize corrective input. *Operationalization:* Qualitative coding of patient speech (inter-rater reliability to be established; currently a research gap).

**Counter-criteria (exclude if any present):**

- The patient cannot state the discrepancy (insight absent) – then Level 2 or 1.
  - The behavior stops entirely after receiving corrective information alone – then basin depth was shallow.
- 

## **7. The Detection Problem (Level 2 vs. 3) in Clinical Practice**

Distinguishing automatic compulsion without error detection (Level 2) from conscious suppression with error detection (Level 3) requires:

- **Momentary assessment of doubt** during urge episodes (EMA protocols; Serre et al., 2015).
- **Reaction time paradigms** (e.g., Gillan et al., 2014, for goal-directed vs. habitual control in OCD; note that the specific link to error detection latency remains an active area).
- **Physiological markers** (dissociation between cognitive

knowledge and fear response suggests Level 3).

These methods are promising but not fully validated; the paper specifies directions for needed research.

---

## 8. Implications for Treatment

Insight-only interventions (psychoeducation, cognitive restructuring alone) often fail in these disorders because the basin depth is maintained by conscious urgency, not lack of knowledge.

Effective treatment must **reduce basin depth** or **increase corrective force**:

- **Addiction:** Pharmacological reduction of craving (e.g., naltrexone; emerging evidence for GLP-1 agonists – see recent reviews, e.g., Klausen et al., 2022, for GLP-1 receptors and alcohol, and emerging clinical reports), contingency management, and identity-shifting interventions (Best et al., 2016).
- **Trauma:** Exposure therapy (increasing corrective force) combined with arousal reduction. The mechanism is basin reshaping, not insight.
- **OCD:** Exposure and response prevention (ERP) directly targets the basin by preventing the compulsion while the patient experiences urgency. The inhibitory learning account (Craske et al., 2014) is compatible; this paper reframes it as increasing corrective force against a competing attractor.

The prediction: treatments that solely enhance insight will be less effective for patients meeting the diagnostic criteria than treatments that directly target basin depth or corrective

force.

---

## 9. Open Questions

- **Measuring basin depth in clinical settings:** Subjective urgency scales, behavioral persistence tasks, heart rate variability. A Clinician Basin Depth Scale (CBDS) is a research priority.
  - **Level 2 vs. 3 differentiation:** Can EMA and reaction time methods reliably classify patients? Pilot studies needed.
  - **Diagnostic threshold validation:** The “three of five” criterion requires empirical ROC analysis against treatment response.
  - **Disorders where suppression is purely Level 2:** Some impulse control disorders or psychotic conditions may not meet the conscious detection criterion.
- 

## 10. Conclusion

Addiction, trauma-related avoidance, and OCD (good insight subtype) are not failures of intelligence. They are cases where conscious, identity-constitutive investment deepens an attractor basin relative to corrective perturbations. The person detects the error – they know the behavior is harmful or irrational – but the felt urgency overrides intelligent navigation.

This diagnosis explains why insight alone fails and why treatments that target basin depth succeed. The clinical fantasy attractor is a trapped navigator: intelligent, aware, but unable to escape.

The dance of recovery is not about knowing the way out. It is about reshaping the attractor landscape so that the path to safety becomes shallower than the pull to stay.

---

**Suggested citation:** Galida, R. S. (2026). Trapped Navigation: Addiction, Trauma, and OCD as Conscious Suppression of Intelligent Correction. *Fantasy Attractor*.

---

# 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**  $H(\pi)$ ).
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).  $A = (D - D_{min} / D_{max} - D_{min})^\alpha (R - R_{min} / R_{max} - R_{min})^\beta (S - S_{min} / S_{max} - S_{min})^\gamma$

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 ( $D_{min}$ ,  $D_{max} - D_{min}$ ,  $D_{max}$ , etc.) are set from a reference population.

The exponents  $\alpha, \beta, \gamma$  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  $\beta_2$  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 $H(\pi)H(\pi)$	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** –  $\alpha, \beta, \gamma$  and the threshold  $\theta$  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.

---

# Attractor Dynamics in Belief Formation, Correction, and Mental Health: A Research Programme

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

**Date:** May 2026

---

## Abstract

This paper applies the attractor framework (persistence under disturbance) to **belief systems** and **mental health**.

We introduce three measurable concepts:

- **Attractor depth** – how rigid or unstable a belief is.
- **Error half-life** – how long it takes for a false belief to fade after correction.
- **Coupling strength to error signals** – how open a belief is to reality checks.

We contrast two disorders:

- **OCD** (obsessive-compulsive disorder) may involve *overly deep* (rigid) attractors.
- **Schizophrenia** may involve *too shallow* (unstable) attractors – with appropriate caution.

We propose experiments to measure error half-life, detect early warning signs of belief shifts (while managing false alarms), and find the optimal pace for correction (“critical damping”).

We also outline:

- **N=1 attractor engineering** (self-experimentation)
- **Wearable early-warning systems** for relapse prevention (discussing lag time and false positives)
- **Cross-coupling** as a measure of resilience (distinguishing healthy from brittle coupling)

This paper is a **research roadmap**, not a finished theory.

---

## 1. Introduction

In the attractor framework, your mind is a **dissipative attractor of your whole body** – a pattern that needs energy, can be disturbed, and can adapt (Galida, 2026, *Persistence*

*Under Perturbation*).

Beliefs are smaller attractors inside that landscape. Their stability determines how easily you update when faced with contradictory evidence.

This paper turns attractor concepts into testable ideas about how beliefs form, stick, and change – and how to help them change. It is a roadmap, not the final word.

---

## 2. Attractor Depth and Mental Disorders

Neurocomputational models suggest a contrast between OCD and schizophrenia, but we must be careful.

Disorder	Attractor Property	Behavioural Sign	Example Task
OCD	Too deep (rigid)	Stuck, hard to switch	Reversal learning (changing rules)
Schizophrenia	Too shallow (unstable)	Jumpy, over-sensitive to noise	Delayed match-to-sample with distractions

### Evidence:

- Unmedicated OCD patients make many perseverative errors on reversal-learning tasks; this correlates with symptom severity (Remijnse et al., 2006).
- Reduced NMDA/GABA function in schizophrenia makes attractor networks unstable, leading to cognitive slips and delusions (Rolls, 2021).

### Caveats:

- Mental disorders are complex, with multiple attractors. We are talking about symptom clusters, not whole-disorder diagnoses.
- Disorders like anxiety, depression, and personality disorders lie in the middle – their attractors are **domain-specific** (e.g., depression has deep negative-belief basins but shallow positive ones).

**Prediction:** Attractor depth could be measured from behaviour (switching rates, reaction time variability) by fitting a two-state hidden Markov model to reversal-learning data – a hypothesis for future work.

---

### 3. Error Half-Life: A New Measure of Belief Rigidity

**Error half-life**  $T_{1/2}$  is the time it takes for a false belief's confidence to drop by half after you present corrective evidence.

#### How to measure it

1. Give people a false belief (e.g., a made-up fact).
2. Give them correct information (text, video) every day for a while.
3. Ask them to rate their belief confidence (0–100) at intervals.
4. Assume a simple **exponential decay** model  $C(t) = C_0 e^{-t/\tau}$  as a starting point (real decay could be sigmoidal or power-law).
5. Then  $T_{1/2} = \tau \ln 2$ .

## What we expect in different conditions

- **Delusional disorders** → very long half-life (deep attractor).
- **Depression** → long half-life for negative self-beliefs, but normal for positive ones (asymmetric updating).
- **Anxiety** → short half-life, but possible overshoot (shallow basin → oscillation).

## Therapeutic application

The goal is to **shorten error half-life**. Methods like **spaced repetition** and **active recall** (quizzing) could help – they strengthen corrective memory traces, similar to memory reconsolidation.

## Relationship to attractor depth

Attractor depth is a **static** measure (inertia). Error half-life is a **dynamic** measure (recovery speed). They are related but not the same: depth gives initial resistance, half-life gives the time course. We need both.

---

## 4. Critical Slowing Down Before Belief Shifts

Before a sudden change of belief (e.g., leaving a cult, political conversion, therapy breakthrough), you may see **early warning signals** – rising variance, higher autocorrelation, slower recovery from small disturbances. This is called **critical slowing down** (Scheffer et al., 2009).

## How to detect it

- Collect daily belief ratings, mood scores, or social media sentiment.
- Compute rolling variance and autocorrelation with a moving window.
- If they exceed a baseline threshold, a shift may be coming.

## False positive problem

Rising variance can be caused by other things (seasonal mood, life events). To reduce false alarms:

- Use control periods (compare with a stable trait belief).
- Combine multiple signals (HRV, sleep, activity) with self-report.
- Use a conservative threshold (e.g., 3 standard deviations above baseline).

This is a research tool, not a clinical diagnostic yet.

**Prediction:** You can detect these signals in diaries before a person deconverts, changes politics, or relapses into depression. A well-timed prompt might help, but false positives must be managed.

---

## 5. Optimal Correction Dosing (Critical Damping)

From control theory, there is an **optimal pace** for delivering corrections: not too slow (oscillates), not too fast

(overshoot/backfire). This is called **critical damping**.

## N=1 protocol

- Vary the gap between corrections (massed vs. spaced).
- Track belief confidence over time.
- Measure how quickly and smoothly it changes.

**Hypothesis:** Spaced correction (e.g., daily micro-doses) works better than one big confrontation – a well-known finding in memory research (Ebbinghaus, spaced repetition). The twist is applying it to **beliefs**, which are more emotional and identity-linked. The mechanism may be similar, but emotional valence may change the optimal schedule.

---

## 6. Fantasy vs. Shared Reality Attractors – Operational Metrics

Metric	Low Corrective Permeability (Fantasy)	High Corrective Permeability (Shared Reality)
Coupling to error signals	Low (few fact-checks, no update)	High (active correction)
Basin depth	Deep (needs large evidence)	Shallow (small anomalies work)
Error-correction latency	Long (days/weeks)	Short (hours/days)
Information diversity tolerated	Low (echo chamber)	High (multiple sources)

## Double-bind computational model

In conspiracy cultures, contradictory evidence gets reinterpreted as confirmation (“cover-up”). We can model this as an **asymmetric Bayesian update**:  $P(\text{belief} \mid \text{contrary evidence}) \geq P(\text{belief} \mid \text{supporting evidence})$

**Example:** Start with belief probability 0.9. A contrary piece of evidence that would normally lower it to 0.3 is instead interpreted as evidence of suppression, so the new probability stays at 0.85. The belief drifts only slowly.

**Breaking the loop:** Indirect interventions work better than direct refutation:

- Point out internal inconsistencies.
- Seed doubt through trusted messengers.
- Use graduated reality-testing.

---

## 7. Wearable Early Warning of Attractor Shifts

**Protocol:** Use consumer wearables (HRV, skin conductance, actigraphy, sleep) plus daily self-reports (mood, belief rigidity). Compute rolling variance and autocorrelation in real time.

**Evidence:** Drops in nocturnal HRV preceded a depressive relapse in a case study (Tonge et al., 2024).

**Prediction:** Rising variance/autocorrelation in HRV, plus mood volatility, can predict an imminent crisis.

## Latency and false alarms

- Useful lead time is **days**, not hours. HRV changes can appear 1–2 weeks before relapse.
- False positives are a concern. Use a **two-stage alert**: first detect statistical anomaly, then confirm with a brief self-report (EMA).
- Specificity needs to be established in longitudinal N=1 studies.

**Intervention:** When thresholds are crossed, trigger a micro-intervention (mindfulness, therapist call) – a closed-loop prevention system.

---

## 8. N=1 Attractor Engineering – Minimal Perturbation Protocol

**Goal:** Find the smallest intervention that shifts a maladaptive attractor (phobia, obsessive thought) without causing oscillation or backfire.

### Procedure:

1. Define the target (e.g., fear rating 0–10).
2. Start with very low-intensity perturbations (e.g., brief exposure, mild counter-evidence).
3. Measure change after each step.
4. When a threshold shift is detected (say, 30% reduction – a provisional starting point; adjust based on baseline variability), record the dose.
5. Back off slightly and check stability.

**Principle:** Never collapse an attractor faster than reality can

correct. Use fine steps (5–10% increments) and frequent monitoring. This is **precision self-regulation**. Generalisability from N=1 to populations is an open question (see Section 12).

---

## 9. Cross-Coupling as a Resilience Metric

**Hypothesis:** High cross-domain coupling (e.g., HRV ↔ mood ↔ sleep) indicates **adaptive resilience** – the system is coordinated and self-correcting. Low coupling or unidirectional cascades indicate **brittle coupling** (a disturbance in one area spreads uncontrollably).

**Measurement:** Collect simultaneous time series (HRV, sleep, activity, mood). Compute cross-correlation or Granger causality.

- **Adaptive** = bidirectional, with negative feedback (e.g., poor sleep → lower HRV → mood drop → social support → sleep improves).
- **Brittle** = unidirectional, amplifying (e.g., sleep loss → stress → more sleep loss).

**Prediction:** Good recovery from stress shows strong bidirectional influences. Low coupling or unidirectional cascades will precede breakdowns.

**Intervention:** Improve adaptive coupling with synchrony exercises (e.g., daily breathing with light exposure, yoga, social rhythm therapy). Testable in an N=1 self-tracking experiment.

---

## 10. Philosophical Extensions (Brief)

- **Are attractors real?** Yes, as structural patterns (process metaphysics). They have causal power – like the path of a river.
- **Free will as attractor autonomy** – acting according to your own attractor is compatibilist freedom. Our framework adds that freedom is about basin width and flexibility, not a binary.
- **Cosmic attractor** – speculative. The universe might have a global attractor (e.g., heat death), but it's untestable now.
- **Darwinian problem of evil** – animal suffering is a strong challenge to theism; the “deep harmonies” hypothesis is hard to falsify.

---

## 11. Open Questions and Next Steps

- Can error half-life be measured reliably from smartphone-based belief tracking? What decay model fits best?
- What is the dose-response curve for corrective interventions? Linear, exponential, or threshold? How does it vary with attractor depth?
- Can wearables detect early warning signs before a psychiatric relapse? What are the false-positive rates and lead times?
- Does adaptive cross-coupling improve after

synchrony-based therapies?

- How are error half-life and attractor depth related? Same thing at different timescales, or different constructs?
  - How can N=1 findings be aggregated into population-level knowledge? One approach: meta-analysis of single-subject time series using hierarchical Bayesian models.
- 

## 12. Conclusion

This research programme puts attractor dynamics to work on beliefs and mental health.

We have proposed **testable metrics** (attractor depth, error half-life, coupling strength) and **experimental protocols** for N=1 self-engineering and early warning.

The framework provides a naturalistic language for understanding why some beliefs resist correction and how to intervene optimally.

We acknowledge our limitations – the exponential decay assumption, false positives in early warning, and the generalisability of N=1 results – and treat them as open questions for future work.

This extends the attractor trilogy into **actionable health and epistemology**.

---

**Suggested citation:** Galida, R. S. (2026). *Attractor Dynamics in Belief Formation, Correction, and Mental Health: A Research Programme (Reader-Friendly Version)*. Fantasy Attractor.