Cognitive Attractor Dynamics: A Formal Theory of Self-Concept and Self-Engineering

Robert Galida
July 2026
[F] (Foundation)

Abstract

The attractor framework provides a unified vocabulary for describing persistence and change across physical, biological, cognitive, and social systems. This paper presents a formal theory of cognitive attractor dynamics, grounding the framework's core variables—κ (corrective permeability), B (basin depth), C (coordination capacity), and R (reality alignment)—in a rigorous mathematical framework. The cognitive state space $X(t) \in \mathbb{R}^n$ is defined, a dynamical equation $\dot{X} = -\nabla V(X) + \eta(t) + E(t)$ is specified, and the variables are derived from the potential landscape $V(X)$. The theory connects to existing frameworks (Hopfield networks, predictive coding, active inference, reinforcement learning) and generates testable predictions about cognitive flexibility, goal persistence, reality alignment, and coordination capacity. The paper is offered as a formal foundation for empirical testing.

All claims are formal hypotheses, not conclusions. The framework is a domain-general dynamical ontology with an associated research programme — a formal theory, not a completed science.

1. Introduction

The attractor framework has been applied to biology, cosmology, AI, and civilizational dynamics. This paper presents a formal theory of cognitive attractor dynamics. It asks a simple question:

> *Can the self — beliefs, goals, and self-narratives — be modeled as an attractor landscape in a high-dimensional cognitive state space?*

The answer is yes — with explicit formal definitions.

A note on the Law of Attraction: The Law of Attraction is often framed as a metaphysical claim. This paper reframes it as **conscious self-direction and self-engineering** — the deliberate shaping of one’s own cognitive attractor landscape through belief revision, attentional focus, and behavioral reinforcement.

A note on the framework’s status: This paper presents a formal theory. The mathematical derivation of equivalence is specified. The framework is offered as a foundation for empirical testing.

A note on domain of applicability: The framework applies to any persistent cognitive system satisfying the formal conditions defined below.

---

2. Core Definitions

2.1 The Framework Variables

Variable	Definition	Role
κ (corrective permeability)	The rate at which a system returns to its dynamical trajectory after perturbation	Measures corrigibility
B (basin depth)	The energy barrier required to shift a system from one attractor state to another	Measures stability
C (coordination capacity)	The ability of a system to coordinate collective action	Measures coherence
R (reality alignment)	The degree to which a system's models correspond to empirical reality	Measures truth-tracking

2.2 Primitive vs. Derived Concepts

Primitive	Definition	Derived	Source
State	The complete description of a system at a given time	—	—
Interaction	Any exchange of energy, momentum, or information between systems	—	—

Primitive	Definition	Derived	Source
Constraint	Any factor that restricts the possible states or trajectories of a system	—	—
Perturbation	Any deviation from the system's dynamical trajectory	—	—
—	—	**κ**	Recovery rate after perturbation (derived from perturbation dynamics)
—	—	**B**	Energy barrier between attractors (derived from constraint topology)
—	—	**C**	Coordination capacity (derived from interaction topology)
—	—	**R**	Reality alignment (derived from model-state correspondence)

---

3. The Formal Theory

3.1 The Cognitive State Space

Define the cognitive state vector: $X(t) \in \mathbb{R}^n$

where n is the dimensionality of the state space. The choice of representation is domain-specific:

Representation	Form	Domain
Belief vector	$X=(b_1,b_2,\ldots,b_n)$	Cognitive psychology
Neural latent	$X\in\mathbb{R}^d$	Computational neuroscience
Control variables	$X=(a,e,m)$	Cognitive control

Distinction between spaces:

- **Abstract state space** X: the theoretical manifold of cognitive states
- **Measurement space** Y: the space of observables (behavior, neural activity)
- **Embedding** $\phi: Y\to X$: mapping from data to latent state

Falsification: If different cognitive states produce identical trajectories in the chosen X-space, the representation fails.

3.2 The State Equation

The dynamics of the cognitive state are governed by:

$$\dot{X}=-\nabla V(X)+\eta(t)+E(t)$$

where:

- $X(t)$ is the cognitive state at time t
- $V(X)$ is the cognitive potential landscape
- $\eta(t)$ is stochastic noise (temperature T)
- $E(t)$ is external perturbation

3.3 The Potential Function

We adopt the following **illustrative potential function** — a mathematically smooth function that produces one minimum and finite depth:

$$V(X)=\frac{1}{2}c\|X-X^*\|^2+\frac{B}{1+e^{-\alpha\|X-X^*\|^2}}$$

where:

- c is the curvature parameter (not κ)
- B is the basin depth (barrier height)
- α controls the steepness of the basin

Note: This potential function is an **illustrative ansatz**, chosen to demonstrate the framework's logic. Alternative forms (multi-well, free-energy-based) are possible and should be explored empirically. The specific functional form is not claimed to be a unique derivation.

Alternative forms:

Form	Equation	Use Case
Quadratic	$V(X)=\frac{1}{2}c\lVert X-X^*\rVert^2$	Single attractor, linear dynamics
Multi-well	$V(X)=\sum_i B_i\phi(\lVert X-X_i^*\rVert^2)$	Multiple attractors
Free energy	$V(X)=-\log p(X)$	Bayesian/predictive coding

3.4 Basin Depth (B)

Basin depth B is the energy barrier required to escape the attractor's basin: $B=\min_{X\in\partial B} V(X)-V(X^*)$

where:

- X^* is the attractor (stable fixed point)
- ∂B is the boundary of the basin of attraction
- $V(X^*)$ is the potential at the attractor

Empirical estimation: B can be estimated from:

- Time to return to baseline after perturbation

- Probability of escape under noise: $P_{escape} \propto e^{-B/T}$
- Hysteresis in response to changing inputs

3.5 Corrective Permeability (κ)

κ is the rate of recovery toward the attractor after a perturbation. It is **derived from the curvature of V**, not independently parameterized.

Formal definition: For a linearized system near the attractor:$$\dot{\delta X} = -\nabla^2 V(X^*)\,\delta X$$

where $\delta X = X - X^*$ is the deviation from the attractor. The recovery rate is determined by the largest (least negative) eigenvalue of the Hessian:$$\kappa = -\lambda_{\max}(-\nabla^2 V(X^*))$$

For our illustrative potential:$$\nabla^2 V(X) = c + \frac{2B\alpha}{1+e^{-\alpha\|X-X^*\|^2}}$$

At the attractor ($X = X^*$):$$\kappa_{\text{baseline}} = c + B\alpha$$

This resolves the circularity: κ is now a derived quantity from the same landscape V. It is not independently parameterized.

Empirical estimation: κ can be estimated from:

- Error-correction times in cognitive tasks
- Post-error slowing in reaction time tasks
- Recovery from emotional perturbations
- Neural measures of flexibility (dynamic connectivity)

3.6 Reality Alignment (R)

R is the predictive accuracy of the system:$$R = -\mathbb{E}[\log p(y \mid X)]$$

where $p(y \mid X)$ is the system's predictive distribution over outcomes y given its current state X.

R belongs in learning dynamics, not in the potential: $\dot{\theta} = g(R, \delta)$

where θ controls the landscape V, and δ is the prediction error.

Relationship to free energy: $F = \mathrm{KL}(q \| p) + R$

where F is variational free energy. R is maximized when the system's predictions match reality.

Empirical estimation: R can be estimated from:

- Predictive accuracy in decision-making tasks
- Calibration of confidence judgments
- Prediction error signals (dopaminergic, sensory)

3.7 Coordination Capacity (C)

C is hypothesized to emerge from the network topology of cognitive subsystems.

Open research question: The specific functional form — whether it depends on total coupling strength, spectral radius, modularity, or other graph-theoretic measures — is an open research question. Candidate measures include:

Measure	Description
Spectral radius	Largest eigenvalue of coupling matrix
Modularity	Degree of community structure
Global efficiency	Average inverse shortest path length

Measure	Description
Synchronization threshold	Second-smallest Laplacian eigenvalue

Empirical estimation: C can be estimated from:

- Coherence between subsystems
- Synchrony of neural or behavioral signals
- Network graph-theoretic measures

Note: The formula $C = \mathrm{Tr}(W) \cdot \min_i B_i$ is not claimed as a unique derivation. It is a placeholder for future empirical investigation.

---

4. The Full Parameterized System

4.1 Complete State Equation

Combining all definitions:$$\dot{X} = -\nabla V(X) + \eta(t) + E(t)$$

where:

- $V(X)$ is the cognitive potential landscape
- $\eta(t)$ is stochastic noise (temperature T)
- $E(t)$ is external perturbation

4.2 Derived Variables

Variable	Derivation	Units
κ	$\kappa = -\lambda_{\max}(-\nabla^2 V(X^*))$	time^{-1}
B	$B = \min_{X \in \partial B} V(X) - V(X^*)$	Energy

Variable	Derivation	Units
R	$R = -\mathbb{E}[\log p(y \mid X)]$	Bits
C	Open research question	Dimensionless

4.3 Parameter Interactions

The parameters are hypothesized to interact:

Hypothesis	Formal Statement
κ increases with R	$\kappa \propto R$
B decreases with κ	$B \propto 1/\kappa$
R decreases with B	$R \propto 1/B$
Optimal B maximizes κ·R	$B^* = \arg\max(\kappa \cdot R)$

Falsification: If the variables are entirely independent, the framework is a taxonomy, not a unified theory.

---

5. Relationship to Existing Frameworks

Framework	Mathematical Form	Relationship
Hopfield networks	$V = -\frac{1}{2}\sum w_{ij} X_i X_j$	Special case: discrete attractors
Predictive coding	$F = -\log p(y \mid X) + \mathrm{KL}$	R is negative free energy (minus complexity)
Active inference	$\dot{X} = -\frac{\partial F}{\partial X}$	General case: both perception and action

Framework	Mathematical Form	Relationship
Reinforcement learning	$V(s) = \max_a \mathbb{E}[R + \gamma V(s')]$	C emerges from value function coupling

6. Testable Predictions

6.1 Prediction 1: Mindfulness Increases κ

Formal statement: Mindfulness training increases corrective permeability.

Empirical test: Measure error-correction times in cognitive tasks before and after mindfulness intervention. Faster post-error adjustments indicate higher κ.

Falsification: If mindfulness training does not lead to faster error-correction times, the prediction fails.

6.2 Prediction 2: Rigidity = Deep B + Low κ

Formal statement: High cognitive rigidity corresponds to deep B and low κ.

Empirical test: Measure reversal learning times and set-shifting ability in high-rigidity individuals.

Falsification: If rigid individuals adapt as quickly as flexible individuals, the prediction fails.

6.3 Prediction 3: Rumination = High B + Low R

Formal statement: Rumination corresponds to high B and low R.

Empirical test: Measure persistence in negative mood states and predictive accuracy in ruminative individuals.

Falsification: If ruminators show low persistence or high predictive accuracy, the prediction fails.

6.4 Prediction 4: Success = High B + High κ

Formal statement: Goal achievement requires both deep B and high κ.

Empirical test: Measure goal persistence (B) and adaptability (κ) in high-achieving individuals.

Falsification: If high achievers show low B or low κ, the prediction fails.

6.5 Prediction 5: Obsession = High B + Low κ

Formal statement: Obsessive-compulsive patterns correspond to high B and low κ.

Empirical test: Measure persistence on incorrect choices in obsessive individuals.

Falsification: If obsessive individuals show normal recovery from errors, the prediction fails.

6.6 Prediction 6: Kramers’ Escape in Cognition

Formal statement: Cognitive transition probabilities follow Kramers’ law.

Empirical test: Vary noise levels (uncertainty, distractors) and measure transition rates between cognitive states.

Falsification: If the relationship is not log-linear, the basin-depth metaphor fails.

6.7 Prediction 7: Exponential Recovery

Formal statement: Cognitive recovery follows exponential decay.

Empirical test: Fit recovery trajectories to exponential and power-law models.

Falsification: If power-law fits are superior, the exponential recovery model fails.

---

7. What This Paper Does Not Claim

This paper does not claim:

- Thoughts directly create reality

- The Law of Attraction is literally true as a metaphysical claim
- The framework replaces cognitive science
- The framework is a theory of everything
- The framework generates novel predictions (it does — see §6)
- Mathematical equivalence between cognitive and other systems
- C is a primitive variable (it is an open research question)
- The illustrative potential function is a unique derivation

---

8. Limitations

Limitation	Address
κ is derived from V	Resolved
R belongs in learning dynamics	Resolved
B and κ are not independent	Resolved
Potential function is ad hoc	Acknowledged as illustrative ansatz
State space is generic	Distinction between abstract/measurement/embedding spaces added
C formula is speculative	Removed; left as open research question

---

9. Open Research Questions

Question	Domain
What is the minimal state space for a given cognitive domain?	Formalization
What is the functional form of V(X) for a given domain?	Formalization
Do cognitive escape probabilities follow Kramers’ law?	Empirical
Do recovery trajectories follow exponential decay?	Empirical
Is R equivalent to negative free energy?	Formalization
Can C be derived from network topology?	Formalization
Do κ, B, and R scale with system size?	Formalization
Does an optimal B exist?	Empirical
How do κ, B, and R interact?	Formalization

---

10. Conclusion

The attractor framework is now formally defined:

Element	Definition
State space	$X(t) \in \mathbb{R}^n$
Dynamics	$\dot{X} = -\nabla V(X) + \eta + E$
Potential	$V(X) = \frac{1}{2}c\lVert X - X^* \rVert^2 + \frac{B}{1 + e^{-\alpha \lVert X - X^* \rVert^2}}$ (illustrative ansatz)
Derived: κ	$\kappa = -\lambda_{\max}(-\nabla^2 V(X^*))$
Derived: B	$B = \min_{X \in \partial B} V(X) - V(X^*)$
Derived: R	$R = -\mathbb{E}[\log p(y \mid X)]$
Open: C	Emerging from network topology

The framework generates testable predictions and is ready for empirical validation.

The next step is computational validation: simulate the dynamics, recover κ and B, demonstrate Kramers' escape, and show recovery trajectories. Then move to human experiments.

References

- Boyatzis, R.E., Rochford, K., & Taylor, S.N. (2015). "The role of the positive emotional attractor in vision and shared vision." *Frontiers in Psychology*, 6:670.
- Cheema, A., & Bagchi, R. (2011). "The effect of goal visualization on goal pursuit." *Journal of Marketing*, 75(2), 109–123.
- Geisler, F.C.M., & Kubiak, T. (2009). "Heart rate variability predicts self-control in goal pursuit." *European Journal of Personality*, 23, 623–633.
- Golubickis, M., Tan, L.B.G., Jalalian, P., Falbén, J.K., & Macrae, C.N. (2024). "Brief mindfulness-based meditation enhances the speed of learning following positive prediction errors." *Quarterly Journal of Experimental Psychology*, 77(11), 2312–2324.
- Kronemyer, D., & Bystritsky, A. (2014). "A non-linear dynamical approach to belief revision in cognitive behavioral therapy." *Frontiers in Computational Neuroscience*, 8:55.
- MacDonald, M.R., & Kuiper, N.A. (1985). "Efficiency and automaticity of self-schema processing in clinical depressives." *Motivation and Emotion*, 9(2), 171–184.
- Singer, J.A., Blagov, P., Berry, M., & Oost, K.M. (2013). "Self-defining memories, scripts, and the life story." *Journal of Personality*, 81(6), 569–582.

Suggested citation: Galida, R. S. (2026). Cognitive Attractor Dynamics: A Formal Theory of Self-Concept and Self-Engineering. *Fantasy Attractor*.

Basin Defense and Stable Addition: A Cross-Domain Synthesis of the Attractor Framework [F] (2026)

Robert Galida — June 2026 (Final)

See Paper 1 (Intelligence Without Consciousness) for the full taxonomy of attractors, κ, and basin depth.

Abstract

Many complex systems resist change by returning to a preferred low-energy attractor rather than adopting a new state. Whether a perturbation (an added agent, input, or component) is ejected, transiently absorbed, or stably integrated depends on the basin geometry (depth B and barriers) and the system's corrective dynamics (κ = 1/τ). This paper defines B and κ, draws on formal models (stochastic dynamical systems and Kramers escape theory) with explicit qualifications for non-gradient domains, and catalogs exemplar systems across ten domains. A comparative table summarizes systems, mechanisms, proxies for B and κ, timescales, and conditions favoring each

outcome. The paper concludes that the same basic physics analog applies across domains: a perturbation of size Δ will be ejected or die out if Δ is below the attractor's effective escape threshold (a function of B), whereas if Δ exceeds that threshold and the system has enough plasticity or additional degrees of freedom, a new stable state can form. A research roadmap is provided in an appendix.

---

1. Introduction

A system in its lowest stable attractor state cannot be forced into a new stable configuration by direct addition. Adding to the system – a third star, an extra electron, a new species, a contradictory belief – will result in one of three outcomes:

1. **Ejection** – the addition is expelled from the system entirely. The original attractor persists.
2. **Transient absorption** – the addition remains present, but the system state returns to the original attractor despite the addition's continued presence.
3. **Stable addition** – the addition is integrated, either by expanding the capacity of the original attractor or by forming a new parallel attractor alongside it.

This paper identifies a unified principle – **basin defense** – that governs these outcomes across physical, biological, ecological, social, and engineered systems. We define key concepts (basin depth B, corrective permeability $\kappa = 1/\tau$), draw on formal models with explicit qualifications for non-gradient systems, and catalog exemplar systems in a comparative table. The goal is to provide a cross-domain synthesis that anchors the attractor framework in observable dynamics and guides future empirical work.

2. Definitions and Formal Models (with Qualifications)

Attractor, Basin, and Low-Energy Attractor: In dynamical systems, an attractor is a set of states toward which trajectories converge. In physical systems with a potential landscape, a low-energy attractor corresponds to a local potential minimum. Its basin of attraction is the region of state space that flows into the attractor. **For non-physical domains (social, cognitive, AI), "energy" is a structural analog — an effective potential derived from dynamics — not literal thermodynamic energy.** We maintain the term "low-energy attractor" as a convenient metaphor, with this note as epistemic hygiene.

Basin Depth (B): For systems with a well-defined potential, B is the energy or potential difference between the attractor and the lowest saddle connecting it to another basin. For non-gradient or high-dimensional systems, B is a **structural analog** — the effective barrier strength inferred from perturbation-response experiments (e.g., the perturbation magnitude required to shift the system to a different state). **Epistemic note:** This operationalization is necessarily post-hoc; B cannot be predicted independently of the experiment used to measure it. This circularity is an open operationalization problem, flagged as such.

Corrective Permeability (κ) and Relaxation Time (τ): We define $\kappa = 1/\tau$, where τ is the characteristic time for return to baseline after a small perturbation. **This definition is applied consistently across all domains**, with τ operationalized domain-specifically as the measured return time (e.g., seconds for a thermostat, hours for synaptic scaling, days for immune response, months for belief updating). A large κ (small τ) means fast return; a small κ means slow or absent return.

Three Outcomes Defined Operationally:

- **Ejection:** The addition leaves the system entirely. The system state returns to the attractor, and the added entity is no longer present.
- **Transient Absorption:** The addition remains present, but the system state returns to the attractor despite the addition's continued presence.
- **Stable Addition:** The addition is integrated, and the system settles into a new attractor (expanded capacity or parallel attractor). This is the only case where the original attractor is displaced.

Formal Models (Qualified): In a one-dimensional overdamped potential, Kramers' escape theory gives mean escape time $\propto \exp(B/D)$, where D is noise intensity. **This result does not generalize to multi-dimensional, non-gradient, or non-equilibrium systems — all of which appear in our domain examples (neural networks, social systems, ecological systems).** For those systems, B and κ are **structural analogs** — quantities that play the same functional role (resistance to change; speed of return) but are not derived from a literal potential. The formal section is an analogy and a source of heuristics, not a universal physical law. We do not claim to "survey" Kramers theory; we draw on it as a conceptual anchor.

---

3. Minimal Physical Examples

Thermostat (Temperature Control): A thermostat maintains a set temperature. An external heat input is an addition. The thermostat's negative feedback loop turns on cooling, expelling the heat (ejection). τ is the temperature relaxation time (seconds). B is the maximum heat load before setpoint failure (Watts or °C above setpoint).

RC Circuit (Passive Decay): A capacitor discharging through a resistor has a single equilibrium at zero voltage. If a constant voltage source is connected (addition), the voltage rises but then decays toward zero with τ = RC. The source remains connected (addition present), but the state returns to the attractor. This is **transient absorption**. (If the source is removed, it is ejection.)

Single Neuron Homeostasis: A neuron's firing rate is regulated by homeostatic plasticity. A transient increase in input causes a firing rate spike, followed by return to baseline with τ on the order of minutes to hours (synaptic scaling). This is transient absorption if the input persists; ejection if the input is removed. Persistent input may lead to stable addition (learning).

---

4. Biological Systems (with CUFT-Primitive Translations)

For each domain, we provide: (1) state space, (2) attractor, (3) basin, (4) τ (κ), (5) perturbation, and (6) outcome.

Immune Response (Tolerance vs. Memory)

- State space: immune cell activation levels, antibody concentrations.
- Attractor: healthy baseline (no inflammation).
- Basin depth B: antigen concentration + danger signal required to trigger full response.
- τ (κ): clearance time of inflammation (hours to days).
- Perturbation: antigen addition.
- Outcome: low antigen → ejection (tolerance); high

antigen + danger signal → stable addition (memory attractor).

Endocrine Homeostasis

- State space: blood glucose, hormone concentrations.
- Attractor: euglycemic baseline.
- B: magnitude of glucose load before dysregulation.
- τ: recovery time after glucose tolerance test (minutes).
- Perturbation: glucose addition (meal).
- Outcome: small load → transient absorption; chronic overload → stable addition (disease attractor).

Synaptic Plasticity (Learning vs. Stability)

- State space: synaptic weights.
- Attractor: baseline weight distribution.
- B: amount of LTP/LTD input needed to produce lasting weight change.
- τ: homeostatic rebound time after activity blockade (hours to days).
- Perturbation: patterned input.
- Outcome: brief input → transient absorption; persistent input → stable addition (memory attractor).

Addiction and Neural Lock-In

- State space: dopamine firing rates, prefrontal activity.
- Attractor: drug-seeking mode (pathological).
- B: strength of drug-cue association needed to trigger relapse.
- τ: decay time of craving after abstinence (days to weeks).
- Perturbation: drug administration.
- Outcome: repeated high dose → stable addiction

attractor; low dose → ejection (no lasting change).
- **Citation:** Koob & Volkow (2016); Nestler (2001).

Developmental Canalization

- State space: gene expression levels.
- Attractor: normal developmental trajectory.
- B: severity of genetic or environmental perturbation required to alter fate.
- τ: time to reconverge to normal phenotype (hours to days).
- Perturbation: mutation or stress.
- Outcome: small perturbation → ejection (buffered); large perturbation → stable addition (alternative fate).
- **Citation:** Waddington (1957).

---

5. Ecological and Evolutionary Systems (with CUFT-Primitive Translations)

Invasion Ecology

- State space: species population densities.
- Attractor: native community composition.
- B: invasibility index — disturbance needed for establishment.
- τ: invader population decay rate if unsuccessful (weeks to years).
- Perturbation: addition of new species.
- Outcome: low disturbance → ejection (invader fails); vacant niche → stable addition (invader establishes).
- **Citation:** Elton (1958); Simberloff (2013).

Alternative Stable States (Ecosystems)

- State space: nutrient levels, algae/plant biomass.
- Attractor: clear-water (plants) or turbid (algae).
- B: critical nutrient loading threshold.
- τ: recovery time of clear state after algae bloom (seasons to decades).
- Perturbation: nutrient addition.
- Outcome: below threshold → transient absorption; above threshold → stable addition (regime shift, hysteresis).
- **Citation:** Scheffer et al. (2001).

Evolutionary Stable States

- State space: allele frequencies.
- Attractor: stable equilibrium genotype.
- B: selective disadvantage needed to eliminate a mutation.
- τ: generations to return to equilibrium.
- Perturbation: new mutation.
- Outcome: small disadvantage → ejection (mutation purged); large advantage → stable addition (sweep to new equilibrium).

---

6. Social and Cultural Systems (with CUFT-Primitive Translations)

Institutions and Norms

- State space: public opinion, policy settings.
- Attractor: status quo norm.
- B: public opinion threshold (e.g., % dissatisfied needed for change).
- τ: speed of policy response or opinion reversion (months to decades).

- Perturbation: policy proposal or protest event.
- Outcome: small event → ejection (status quo persists); large crisis → stable addition (new norm).

Identity and Belief Systems

- State space: belief strength, cognitive dissonance.
- Attractor: core ideological commitment.
- B: complexity/depth of ideological justification.
- τ: belief-updating time after disconfirming evidence (months to years).
- Perturbation: counter-attitudinal evidence.
- Outcome: weak evidence → ejection (rationalization); strong evidence → stable addition (belief change, rare).
- **Citation:** Nyhan & Reifler (2010).

Conspiracy and Extremist Movements

- State space: belief adoption × social network reinforcement (two-dimensional).
- Attractor: sealed fantasy attractor (low κ).
- B: strength of echo-chamber reinforcement.
- τ: decay time after authoritative rebuttal (years, often indefinite → $\kappa \to 0$).
- Perturbation: debunking information.
- Outcome: most debunking → ejection (entrenchment); death of leader or total disconfirmation → stable addition (collapse).
- **Note on $\kappa \to 0$:** The conspiracy attractor represents the limiting case of a sealed basin, where $\tau \to \infty$ and corrective permeability approaches zero. This directly links to the fantasy attractor framework developed in Paper 1 (Intelligence Without Consciousness) and the conscious suppression series.

7. Engineered and AI Systems (with CUFT-Primitive Translations)

Control Systems

- State space: system state (position, temperature, etc.).
- Attractor: setpoint.
- B: stability margin (phase/gain margin in control theory) — the range of disturbances that can be rejected.
- τ: controller response time (milliseconds to seconds).
- Perturbation: external disturbance.
- Outcome: small disturbance → ejection (return to setpoint); excessive disturbance → failure (not modeled as attractor shift).

Catastrophic Forgetting (Neural Networks)

- State space: network weights.
- Attractor: task-specific weight configuration.
- B: effective barrier to weight drift (often negligible — no basin).
- τ: number of gradient steps before old task performance decays (seconds to minutes).
- Perturbation: training on a new task.
- Outcome: standard training → ejection (old task overwritten); replay/regularization → stable addition (shared attractor for multiple tasks).
- **Citation:** Kirkpatrick et al. (2017).

Continual Learning Systems

- State space: weights plus architectural modules.

- Attractor: multi-task configuration.
- B: capacity of the network (number of tasks storable).
- τ: retention half-life across training steps (minutes to hours).
- Perturbation: new task training.
- Outcome: no safeguards → ejection (catastrophic forgetting); progressive networks or EWC → stable addition.

Corrigibility and Goal Stability

- State space: AI internal goal representation.
- Attractor: fixed goal (low κ) or corrigible (high κ).
- B: depth of goal basin (resistance to human feedback).
- τ: time to incorporate corrective signal (if κ is high).
- Perturbation: human correction signal.
- Outcome: low κ → ejection (correction ignored); high κ → stable addition (goal updated).

---

8. Comparative Table

System / Domain	Operational τ (κ = 1/τ)	τ Typical Timescale	Basin Depth B Proxy	Outcome	Notes
Thermostat	Temperature relaxation time	Seconds	Max heat load before setpoint failure (W or °C above setpoint)	Ejection	Passive addition
RC Circuit	τ = RC	µs–ms	N/A (linear)	Transient absorption	Addition remains; state returns
Single Neuron	Firing-rate recovery time	ms–sec (ion), min–hr (synaptic)	Perturbation amplitude before rebound fails	TA (persistent input) / E (removed)	Hebbian plasticity can lead to SA
Immune System	Inflammation clearance time	Hours–days	Antigen + danger signal threshold	E (tolerance) / SA (memory)	Active agent (antigen)

System / Domain	Operational τ (κ = 1/τ)	τ Typical Timescale	Basin Depth B Proxy	Outcome	Notes
Endocrine Homeostasis	Glucose tolerance recovery	Minutes	Load magnitude before dysregulation	TA (small load) / SA (chronic overload)	Passive addition
Synaptic Plasticity	Homeostatic rebound time	Hrs–days	LTP input size for lasting change	TA (brief input) / SA (persistent)	Active agent (patterns)
Addiction	Craving decay time	Days–weeks	Drug-cue association strength	E (low dose) / SA (high chronic)	Active agent (drug)
Development (Canalization)	Phenotype reconvergence time	Hours–days	Mutation/stress severity to alter fate	E (small) / SA (large)	Active agent (genetic)
Invasion Ecology	Invader population decay time	Weeks–years	Invasibility index / disturbance needed	E (occupied niche) / SA (vacant niche)	Active agent (species)
Alternative States (Ecosystems)	Recovery time after nutrient reduction	Seasons–decades	Critical nutrient loading threshold	TA (below) / SA (above)	Hysteresis
Social/Political Norms	Opinion reversion time	Months–decades	Public opinion threshold	E (small dissent) / SA (mass movement)	Active agent (protest)
Belief Systems	Belief-updating time	Months–years	Ideological justification depth	E (weak evidence) / SA (strong evidence)	Active agent (counter-evidence)
Conspiracy Movements	Belief decay time	Years – indefinite ($\kappa \to 0$)	Echo-chamber reinforcement strength	E (most debunking) / SA (collapse)	Fantasy attractor ($\kappa \to 0$)
Catastrophic Forgetting (AI)	Gradient steps to old-task decay	Seconds–minutes	Effective barrier to weight drift (often 0)	E (standard training) / SA (EWC/replay)	Active agent (new task)
Control Systems	Controller response time	ms–sec	Stability margin (phase/gain margin)	E (small) / SA (failure)	Passive addition
Continual Learning (AI)	Retention half-life across training steps	Minutes–hours	Task capacity	E (no safeguards) / SA (progressive nets)	Active agent (new task)
Corrigibility (AI)	Time to incorporate corrective signal	Variable (design-dependent)	Goal basin depth	E (low κ) / SA (high κ)	Active agent (correction)

Note: Ejection vs. transient absorption are distinguished operationally: ejection means the addition leaves the system;

transient absorption means the addition remains but the state returns to the attractor. The table notes “active agent” when the addition has its own dynamics (e.g., antigen, new species, counter-evidence) versus “passive addition” (e.g., heat, charge). The conspiracy movements row explicitly flags $\kappa \to 0$ as the fantasy attractor limiting case (see Paper 1).

8.5 Rate-Induced Tipping and the κ Timescale: Independent Confirmation

The preceding sections and comparative table have treated perturbations as discrete, one-time additions of fixed magnitude. However, the **rate** at which a perturbation is applied — fast vs. slow — is equally critical. A large perturbation applied abruptly may trigger basin defense (ejection or transient absorption), while the same cumulative change delivered gradually may be integrated as stable addition or tracked adiabatically without tipping.

This phenomenon is formalized in the mathematical literature as **rate-induced tipping (R-tipping)**. In dynamical systems, if an external parameter changes slowly (adiabatic forcing), a stable state can track the change and remain an attractor. But if the parameter changes faster than the system’s intrinsic relaxation time ($\tau = 1/\kappa$), the system cannot track, overshoots its basin boundary, and tips into a different state. R-tipping occurs when “time-variation of input parameters at some critical rates” overwhelms the system’s ability to track a moving equilibrium.

Consequences for κ as a timescale filter:

- **High-κ systems (fast return)** — Can reject rapid perturbations (they are ejected or transiently absorbed) but may integrate slow drift because the correction loop cannot keep up with a changing baseline.

- **Low-κ systems (slow return)** – May ignore quick blips but are vulnerable to slow accumulation; a persistent, gradual change can eventually shift the attractor without triggering a sudden defense reaction.

Thus, κ defines a characteristic cutoff timescale that separates "ejection/transient absorption" from "stable addition." Perturbations much faster than $1/\tau$ act as impulses that are rejected; perturbations much slower than $1/\tau$ are quasi-static and can be incorporated.

Empirical confirmations across domains (independent external research):

Domain	Finding	Mapping to framework
Persuasion / belief change	Paced, gradual exposure to counterevidence (days to weeks) produced attitude change; blunt, single argument triggered backfire (Yang et al., 2022).	Gradual rate ($\lesssim \kappa$) → stable addition; fast rate (□ κ) → ejection (backfire).
Addiction (smoking cessation)	Cold turkey (abrupt cessation) yielded higher abstinence rates than gradual tapering.	Abrupt perturbation can sometimes achieve stable addition by surmounting basin barrier in one event; gradual may prolong transient state without escape.

Domain	Finding	Mapping to framework
Ecosystem management	Gradual nutrient reduction may postpone tipping points; only extremely slow changes avoid collapse (Panahi et al., 2023).	Very slow rate ($\ll 1/\tau$) allows tracking without tipping; intermediate rates may still tip but with delay.
Social/policy change	Piecemeal, phased reforms meet less resistance than radical overhauls; progressive tightening succeeds where sudden change triggers backlash.	Slow, incremental addition creates parallel attractors; fast addition triggers basin defense.

Optimal perturbation timescale:

The theory and evidence suggest a non-monotonic effect of perturbation rate. Very fast shocks trigger immediate defense. Very slow drifts may be tracked adiabatically (no tipping) or eventually overcome defenses after long accumulation. The most effective timescale to minimize active rejection and maximize stable addition often lies **on the order of the system's intrinsic time constant $\tau = 1/\kappa$**.

Prediction for future experiments:

For any system with known or measurable κ, there exists a critical perturbation rate r_c such that:

- If perturbation rate $> r_c$, the system rejects the addition (ejection or transient absorption).
- If perturbation rate $< r_c$, the system integrates the addition (stable addition via expanded capacity or parallel attractor formation).
- The transition at r_c corresponds to the system's inability to track a moving equilibrium; it is a genuine bifurcation in the time-domain.

External convergence:

This analysis — derived from mathematical rate-induced tipping theory and domain-specific studies — independently validates the attractor framework's claim that κ acts as a timescale filter separating ejection from stable addition. The convergence between the framework's predictions and external research strengthens the cross-domain synthesis considerably.

---

9. Synthesis and Criteria

Across these domains, common criteria emerge:

- **Energy/Threshold:** A perturbation must overcome an attractor's barrier. Deep basins (high B) mean only large shocks can cause a shift.
- **Coupling and Plasticity:** Systems with many degrees of freedom or adaptive coupling more easily integrate additions.
- **Dimensionality and Redundancy:** Multi-dimensional systems can absorb perturbations into some dimensions while maintaining others.
- **Timecourse and Feedback:** Slow changes might be assimilated; fast jolts cause overshoot and return. Feedback gain determines κ.
- **Nature of Addition:** Passive additions (heat, charge) tend to be ejected or transiently absorbed; active agents (species, evidence, pathogens) may reshape the attractor.

Empirical Protocols: Measure κ by controlled perturbation experiments: apply a small disturbance, measure return time τ, compute $\kappa = 1/\tau$. Measure B by scaling the perturbation magnitude until the system fails to return (escape). This works in physical, biological, and some social systems; for others, B remains a qualitative analog.

---

10. Appendix: Research Roadmap

The following future papers are suggested from the comparative table, each developing a single domain in depth.

Domain	Proposed Title	Type
Addiction	*The Addicted Brain as a Fantasy Attractor: Neural Lock-In and Ejection of Alternative Rewards*	[A]
Immune System	*Tolerance and Memory: Two Attractor Responses to Antigen Addition*	[A]
Catastrophic Forgetting	*Why Neural Networks Forget: Attractor Ejection in Sequential Learning*	[A]
Invasion Ecology	*Eject or Integrate: Attractor Dynamics of Invasive Species*	[A]
Development	*Canalization as Basin Defense: Attractor Stability in Embryogenesis*	[A]
Continual Learning	*Parallel Attractors for Lifelong Learning: Engineering Solutions to Catastrophic Forgetting*	[A]
Social Norms	*Tipping Points and Regime Shifts: Attractor Dynamics in Political Systems*	[A]
Endocrine Homeostasis	*Glucose, Cortisol, and Setpoints: Hormonal Attractors and Disease Transitions*	[A]
Alternative Ecosystems	*Hysteresis and Regime Shifts: Ecological Basins and Tipping Points*	[A]

Domain	Proposed Title	Type
Belief Systems	*The Uncorrectable Believer* (already written)	[A]

---

11. Conclusion

Physical, biological, ecological, social, and engineered systems all obey the same attractor principle: a low-energy attractor defends itself against displacement. When an addition is introduced, the system either ejects it, absorbs it only transiently, or — under rare conditions of expanded capacity or parallel structure — integrates it stably. The outcome is determined by basin depth (B), corrective permeability ($\kappa = 1/\tau$), and the magnitude and nature of the perturbation.

This cross-domain synthesis provides a unified foundation for the attractor framework. Future work should quantify B and κ empirically across domains, test the predicted scaling relationships, and explore the boundary conditions between ejection, transient absorption, and stable addition. The appendix outlines the most promising next papers.

References

- Elton, C. S. (1958). *The Ecology of Invasions by Animals and Plants*. Methuen.
- Hebb, D. O. (1949). *The Organization of Behavior*. Wiley.
- Kirkpatrick, J., Pascanu, R., Rabinowitz, N., et al. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13), 3521–3526.

- Koob, G. F., & Volkow, N. D. (2016). Neurobiology of addiction: a neurocircuitry analysis. *The Lancet Psychiatry*, 3(8), 760–773.
- Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4), 284–304.
- Nestler, E. J. (2001). Molecular basis of long-term plasticity underlying addiction. *Nature Reviews Neuroscience*, 2(2), 119–128.
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330.
- Scheffer, M., Carpenter, S., Foley, J. A., et al. (2001). Catastrophic shifts in ecosystems. *Nature*, 413(6856), 591–596.
- Simberloff, D. (2013). *Invasive Species: What Everyone Needs to Know*. Oxford University Press.
- Turrigiano, G. (2008). The self-tuning neuron: synaptic scaling of excitatory synapses. *Cell*, 135(3), 422–435.
- Waddington, C. H. (1957). *The Strategy of the Genes*. George Allen & Unwin.
- Galida, R. S. (2026). Intelligence Without Consciousness: A Diagnostic Paper on LLMs, Amoebae, and the Attractor Framework. *Fantasy Attractor* (Paper 1 of the conscious suppression series).

Suggested citation: Galida, R. S. (2026). Basin Defense and Stable Addition: A Cross-Domain Synthesis of the Attractor Framework (Final). *Fantasy Attractor.*