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	$X=(b_1,b_2,\dots,b_n)$	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:

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

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: $\delta \dot{X} = -\nabla^2 V(X^*)\,\delta X$

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

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	$V(s)=\max_a E[R+\gamma V(s')]$	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*.