The Performance Attractor: A Framework for Social Cognition

Robert Galida
July 2026
[A] (Application)

Abstract

The attractor framework provides a unified vocabulary for describing persistence and change across physical, biological, cognitive, and social systems. This paper extends that vocabulary to social cognition. It proposes that social performance — the regulation of behavior in response to an internal model of being evaluated by real or imagined others — can be modeled as an attractor landscape in a high-dimensional social state space. Internal narration does not merely stabilize an attractor—it may actively reshape the attractor landscape over time. Confidence is hypothesized to correspond to a balance of κ, B, and R; insecurity to an imbalance. Happiness is hypothesized to be structurally associated with perceived action capacity and confidence; unhappiness with despondency. The paper formally defines the fantasy attractor of social performance — a self-reinforcing, reality-resistant basin whose update operator exhibits persistent insensitivity to corrective evidence. The Taoist concept of wu wei is interpreted as one computational resolution of the "wu wei paradox." The framework generates testable predictions and is offered as a foundation for empirical investigation.

This paper presents a model hypothesis — that social behavior

can be represented as movement among attractor states — and a philosophical interpretation — that human social existence may be inescapably performative. These are distinct claims. The model hypothesis is the primary contribution; the philosophical interpretation is offered as a generative implication, not a proven conclusion.

---

1. Introduction

Social life involves performance — behavior optimized with respect to an internal model of social evaluation. We adopt roles, manage impressions, curate presentations of self. We monitor ourselves constantly — rehearsing, evaluating, adjusting. And we narrate internally — a running commentary on our own performance.

This is not a bug. It is a feature. Survival depends upon social navigation. Internal narration is practice — rehearsal for future interactions. Without it, there would be far more conflict.

But performance has a cost. Self-awareness becomes acute — and can paralyze. The same mechanism that enables survival can trap the system in a self-reinforcing loop. The performance can become a fantasy attractor — reality-resistant, self-sealing, and ultimately artificial.

A note on the paper's scope: This paper presents a **model hypothesis** — that social behavior can be represented as movement among attractor states in a high-dimensional state space. It also presents a **philosophical interpretation** — that human social existence may be inescapably performative. These are distinct claims. The model hypothesis is the primary contribution; the philosophical interpretation is offered as a generative implication, not a proven conclusion.

A note on the paper's strongest contribution: The central hypothesis is that internal narration does not merely stabilize an attractor — it may actively reshape the attractor landscape over time. This is a novel, testable computational claim.

---

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)	Within this framework, R is operationalized as predictive accuracy — the expected log predictive likelihood	Measures truth-tracking

Note: R is an operational measure of predictive accuracy, not a metaphysical claim about correspondence with reality. It is the expected log predictive likelihood: R=E[log⁡p(y∣X)]$R = E[\log p(y \mid X)]$. When predictions are accurate, R is close to 0 (maximal). When predictions are poor, R is a large negative number (poor alignment).

2.2 Social Performance: A Definition

Social performance is defined as behavior optimized with respect to an internal model of social evaluation.

This definition is:

- **Measurable:** It can be operationalized through self-report, behavioral observation, and physiological measures
- **Distinct:** It distinguishes social performance from other forms of action (e.g., gardening alone, quiet contemplation)
- **Connected to literature:** It aligns with social cognition research on impression management, self-monitoring, and social anxiety

Falsification: If behavior is observed to be independent of internal models of evaluation, the concept is not useful.

2.3 The State Space of Social Performance

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

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

Representation	Form	Domain
Role vector	X=(r1,r2,…,rn)$X = (r_1, r_2, \ldots, r_n)$	Social roles and identities
Self-monitoring vector	X=(a,m,p)$X = (a, m, p)$	Attention to self, monitoring intensity, performance effort
Social feedback vector	X=(f1,f2,…,fn)$X = (f_1, f_2, \ldots, f_n)$	Perceived social feedback

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

2.4 The State Equation (Fixed Landscape)

The dynamics of the social state on a fixed landscape are governed by:$$\dot{X} = -\nabla V(X) + \eta(t) + E(t)$$

where:

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

2.5 The Potential Function

The framework requires a potential function $V(X)$ satisfying:

1. **Differentiability:** V is smooth
2. **Locally stable minima:** Attractors exist
3. **Finite escape barriers:** Basins have finite depth

A convenient illustrative form is:$$V(X) = \frac{1}{2c}\|X - X^*\|^2 + B\left(1 + e^{-\alpha\|X - X^*\|^2}\right)$$

where:

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

Note: This is an illustrative ansatz, not a unique derivation. Other functional forms satisfying the three conditions above

are equally compatible with the framework.

Note on κ/B coupling: Under this specific ansatz, the local curvature at the attractor — and therefore κ — depends on both c and B (and α). Increasing B while holding c fixed also increases κ. This coupling is a property of this particular potential function; other functional forms might decouple them. Whether κ and B can be independently manipulated is an open empirical question.

2.6 Derived Variables

Variable	Derivation	Units
κ	κ=λmin(∇2V(X∗))$\kappa = \lambda_{min}(\nabla^2 V(X^*))$	time−1time−1
B	B=min⁡X∈∂BV(X)−V(X∗)$B = \min_{X \in \partial B} V(X) - V(X^*)$	Energy
R	R=E[log⁡p(y∣X)]$R = E[\log p(y \mid X)]$	Bits (expected log predictive likelihood)

---

3. Adaptive Landscape Dynamics

3.1 From Fixed to Adaptive Landscapes

Sections 2.4–2.6 describe dynamics on a **fixed landscape** — the potential function V(X)$V(X)$ is static. However, Section 3 introduces an **extension** in which the landscape itself evolves through learning, experience, and internal narration.

This is an adaptive landscape:V=V(X,t)$V = V(X, t)$

and the dynamics become:X˙=−∇XV(X,t)+η(t)+E(t)$\dot{X} = -\nabla_X V(X, t) + \eta(t) + E(t)$
V˙=g(narration,learning,experience)$\dot{V} = g(\text{narration}, \text{learning}, \text{experience})$

The landscape evolves over time as a function of internal narration and experience. This distinguishes the framework from fixed-landscape models and makes it genuinely adaptive.

3.2 Internal Narration and Landscape Reshaping

Hypothesis: Internal narration does not merely deepen B — it may reshape the attractor landscape itself.$\dot{V}=g(\text{narration})$

where g captures how narration:

- Deepens existing wells
- Creates new wells
- Splits one basin into multiple identity basins
- Flattens obsolete basins

Empirical anchor: Rumination — a form of repetitive, self-focused narration — is associated with cognitive rigidity, suggesting deeper basins (Nolen-Hoeksema, 1991).

Falsification: If narration frequency does not correlate with B measures or landscape reshaping, the link is unsupported.

3.3 Rehearsal and Performance Improvement

Hypothesis: Internal narration functions as rehearsal — it improves performance under social conditions.

Empirical anchor: Self-talk research shows that strategic internal rehearsal improves public-speaking performance (Hardy, 2006).

Falsification: If narration does not predict performance improvement, the rehearsal hypothesis fails.

3.4 The Bidirectional Loop

The relationship between performance and narration is bidirectional:Performance↔Narration↔V(X,t)Performance↔Narration↔$V(X,t)$

Stage	Description
1. Performance	You adopt a role, manage impressions, curate your presentation
2. Narration	You rehearse, evaluate, adjust, comment on your own performance
3. Reshaping	The landscape evolves — wells deepen, new wells form, obsolete wells flatten
4. Monitoring	You watch yourself constantly
5. Performance improves	The rehearsal makes you a better performer
6. Self-awareness becomes acute	You become hyper-aware of your own performance

The loop is self-reinforcing: performance generates narration, narration reshapes the landscape, and the reshaped landscape generates more performance.

---

4. Confidence vs. Insecurity

4.1 Confidence

Hypothesis: Confidence corresponds to moderate κ + moderate B + moderate R — the system is stable enough to persist, flexible enough to correct, and aligned enough to navigate.

Empirical anchor: Higher self-efficacy correlates with persistence and success in tasks (Bandura, 1997).

Falsification: If confidence does not correlate with the predicted parameter combination, the hypothesis fails.

4.2 Insecurity

Hypothesis: Insecurity corresponds to high error detection (κ_detection) + low behavioral updating (κ_correction) + deep B + low R.

This requires separating two components of corrective permeability:

- **κ_detection:** The rate at which errors are detected
- **κ_correction:** The rate at which behavior is updated in response to errors

Insecurity involves rapid detection but poor updating.

Note: This split into κ_detection and κ_correction is an informal extension to the formal model, introduced to capture the distinction between error detection and behavioral updating. The formal model (see §2.6) defines κ as a single scalar — the slowest-relaxing mode of the Hessian. The two-component decomposition is a heuristic for interpretation, not a derivation from the state equation.

Empirical anchor: Social anxiety involves hyper-vigilance, chronic negative self-monitoring, and low reality-alignment (Clark & Wells, 1995).

Falsification: If insecurity does not correlate with this parameter combination, the hypothesis fails.

4.3 The Difference

State	κ_detection	κ_correction	B	R	Outcome
Confidence	Moderate	Moderate	Moderate	Moderate	Action
Insecurity	High	Low	Deep	Low	Freezing

---

5. Happiness and Unhappiness

5.1 Happiness and Confidence

Hypothesis: Within this framework, happiness is structurally associated with perceived action capacity and confidence. Happiness is hypothesized to correlate with behavioral measures of social engagement, action initiation, and risk-taking.

Empirical anchor: Perceived control correlates negatively with depression (Seligman, 1975). When people feel capable and their actions lead to outcomes, they tend to be happier.

Falsification: If happiness does not correlate with confidence measures, the hypothesis fails.

5.2 Unhappiness and Despondency

Hypothesis: Unhappiness is structurally associated with despondency – the felt sense of being unable to act. Unhappiness is hypothesized to correlate with behavioral measures of withdrawal, inaction, and avoidance.

Empirical anchor: Perceived control correlates negatively with

depression. When people feel powerless, unhappiness rises.

Falsification: If unhappiness does not correlate with despondency measures, the hypothesis fails.

5.3 The Relationships

Relationship	Meaning
Happiness ≈ Confidence	Happiness is structurally associated with the experience of trusting your own basin
Unhappiness ≈ Despondency	Unhappiness is structurally associated with the experience of not trusting your own basin

Note: These are associations, not identities. Happiness includes pleasure, meaning, attachment, physiology, temperament, reward processing, and social connection. Confidence explains part of happiness — not all of it.

---

6. The Fantasy Attractor of Social Performance

6.1 Formal Definition

A **fantasy attractor** is an attractor whose update operator exhibits persistent insensitivity to corrective evidence.

Formally, a fantasy attractor satisfies:

1. **High B:** Deep basin — the system is resistant to leaving
2. **Low effective κ:** Poor correction — the system does not

update in response to evidence
3. **Systematically biased R:** Low reality alignment — the system's models are persistently distorted
4. **Persistent insensitivity to corrective evidence:**

$\partial R \partial E \approx 0 \partial E \partial R \square \approx 0$

despite non-zero prediction error, where E*E* is disconfirming evidence. The system's predictive accuracy does not improve even when errors are present.

6.2 Diagnosis

Hypothesis: The performance-narration system can become a fantasy attractor — a self-reinforcing, reality-resistant basin that persists despite mounting evidence of its artificiality.

Symptom	Description
Low R	The system is aligned with the performance, not with reality
Deep B	The performance is deeply entrenched
Low κ	The system resists correction — any challenge to the performance is a threat
Self-reinforcement	The performance loops back on itself

6.3 Sealing Mechanisms

Mechanism	Description
Confirmation bias	Seeking confirming evidence, ignoring disconfirming cues
Belief perseverance	Beliefs persist after evidence is shown to be false

Mechanism	Description
Counter-evidence discounting	Disconfirming evidence is reframed as an exception
Identity fusion	The performance is tied to self-worth

Falsification: If a person accepts disconfirming evidence readily, the fantasy-attractor model is wrong.

6.4 Attractor Shifts, Not Escape

Hypothesis: The framework predicts that interventions shift individuals between attractor configurations rather than eliminating social regulation entirely.

Empirical anchor: Every intervention tested (mindfulness, therapy, meditation) produces a new cognitive mode, not a blank slate.

Testable prediction: Every intervention preserves some degree of social predictive regulation, even if self-monitoring and explicit narration decrease.

Operationalization: Meditation decreases self-report narration but leaves prediction accuracy above chance. Therapy decreases rumination without eliminating role behaviour. These are measurable quantities.

Falsification: If an intervention produces a state with zero self-monitoring, zero role occupancy, and zero internal narration, the hypothesis fails.

---

7. Testable Predictions

Prediction 1: Narration correlates with B

Frequent internal narration will correlate with measures of role persistence and resistance to social feedback.

Prediction 2: Narration improves performance

Strategic internal narration will predict performance improvement in social tasks.

Prediction 3: Confidence = moderate κ + moderate B + moderate R

High-confidence individuals will show balanced measures of corrigibility, stability, and reality alignment.

Prediction 4: Insecurity = high κ_detection + low κ_correction + deep B + low R

High-insecurity individuals will show rapid error detection, poor behavioral updating, deep role persistence, and poor social prediction accuracy.

Prediction 5: Happiness correlates with confidence

Happiness self-reports will correlate with behavioral measures of social engagement, action initiation, and risk-taking.

Prediction 6: Unhappiness correlates with

despondency

Unhappiness self-reports will correlate with behavioral measures of withdrawal, inaction, and avoidance.

Prediction 7: Taoist practitioners show shallow B + high κ + high R

Taoist practitioners will show shallower role persistence, faster error correction, and higher social prediction accuracy.

Prediction 8: Interventions shift attractors, not eliminate performance

Every intervention preserves some degree of social predictive regulation, even if self-monitoring and explicit narration decrease. Meditation decreases self-report narration but leaves prediction accuracy above chance. Therapy decreases rumination without eliminating role behaviour.

---

8. Philosophical Interpretation: Wu Wei

8.1 Wu Wei as a Distinct Attractor State

Wu wei is a Taoist concept often translated as “non-action” or “effortless action.” Within this framework, we interpret it as a distinct attractor state characterized by shallow B, high κ, and high R — a state of effortless responsiveness, full attunement to reality, and minimal self-monitoring.

The longstanding paradox of deliberate spontaneity (wu wei) has been extensively discussed in the scholarship on early

Chinese thought (Slingerland, 2000). This paper offers one computational resolution of that paradox.

This is one computational interpretation of wu wei, not a definitive reading of the tradition.

Empirical anchor: Taoist practitioners show differences in cognitive flexibility, role persistence, and social prediction accuracy compared to controls.

Falsification: If Taoist practitioners do not show shallower B, higher κ, or higher R, the hypothesis fails.

8.2 The Paradox of Non-Performance

Observation: To claim non-performance is to perform non-performance.

Resolution: The performance of non-performance is not a failure — it is the *only* path. There is no escape from performance; there is only the *choice of which performance to inhabit*.

Performance Type	B	κ	R	Outcome
Social performance (role-playing)	Deep	Low	Low	Trapped in fantasy attractor
Authenticity performance	Moderate	Moderate	Moderate	Closer to reality
Non-performance performance	Shallow	High	High	The closest approximation available

---

8.3 The Taoist's Basin

Claim	Underlying Dynamics
"I am non-performative"	The performance of being non-performative
"I am authentic"	The performance of being authentic
"I have transcended"	The performance of having transcended
"I am at peace"	The performance of being at peace

---

9. What This Paper Does Not Claim

This paper does not claim:

- Performance is inherently pathological
- Escape from performance is possible
- Taoism is a complete solution
- The framework replaces social psychology
- The framework is a theory of everything
- Happiness is *only* confidence
- Wu wei is definitively "performing non-performance"
- The philosophical interpretation is proven

---

10. Limitations

Limitation	Address
κ, B, and R are not yet measured in social contexts	Candidate measures are proposed but not validated
The Taoist mapping is philosophical, not empirical	Empirical testing is required

Limitation	Address
The state space is generic	Specific representations require empirical validation
The potential function is illustrative	Alternative forms are possible

---

11. Conclusion

Social performance can be modeled as an attractor landscape. Internal narration functions as rehearsal, deepening the performance basin or reshaping the landscape. Confidence enables action; insecurity enables freezing. Happiness is structurally associated with confidence; unhappiness with despondency.

The fantasy attractor of social performance is formally defined as an attractor whose update operator exhibits persistent insensitivity to corrective evidence — unifying confirmation bias, belief perseverance, identity-protective cognition, and self-presentation into one dynamical picture.

Wu wei is interpreted as a distinct attractor state characterized by shallow B, high κ, and high R — effortless responsiveness, full attunement to reality.

The framework predicts that adaptive functioning depends less on escaping social performance than on occupying attractor states that remain corrigible, reality-aligned, and resistant to maladaptive self-reinforcement.

References

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.
- Clark, D.M., & Wells, A. (1995). “A cognitive model of social phobia.” In *Social phobia: Diagnosis, assessment, and treatment*.
- Hardy, J. (2006). “Speaking clearly: A critical review of the self-talk literature.” *Psychology of Sport and Exercise*, 7(1), 81–97.
- Nolen-Hoeksema, S. (1991). “Responses to depression and their effects on the duration of depressive episodes.” *Journal of Abnormal Psychology*, 100(4), 569–582.
- Seligman, M.E.P. (1975). *Helplessness: On depression, development, and death*. Freeman.
- Slingerland, E. (2000). “Effortless action: The Chinese spiritual ideal of wu-wei.” *Journal of the American Academy of Religion*, 68(2), 293–328.

Suggested citation: Galida, R. S. (2026). The Performance Attractor: A Framework for Social Cognition. *Fantasy Attractor*.