Formalization of Mental Disintegration Phenomena Through Dynamical Systems Theory: With Applications to DSM-5-TR Diagnostic Categories

Boris Kriger[1,2]

[1]Information Physics Institute, Gosport, Hampshire, United Kingdom,
boris.kriger@informationphysicsinstitute.net
[2]Institute of Integrative and Interdisciplinary Research, Toronto, Canada,
boriskriger@interdisciplinary-institute.org

February 2026

Abstract

Dynamical systems theory (DST) has been proposed as a unifying framework for psychiatry, in which the healthy mind is a stable attractor, psychiatric disorders are alternative attractors, and clinical deterioration is a critical transition governed by tipping points, hysteresis, and critical slowing down (Scheffer, Bockting, Borsboom et al., *JAMA Psychiatry*, 2024). However, the original proposal remains a methodological program: it contains no equations, does not engage with the phenomenological content of specific disorders, and does not address phenomena—anosognosia, the fear of madness, internal resistance, gaslighting, collective madness, punitive psychiatry—that lie outside the scope of existing computational psychiatry paradigms.

The present article develops a unified mathematical formalism that fills these gaps. Starting from a core stochastic differential equation on a potential landscape, we construct a modular system of coupled models—each with explicit equations, consistent notation, and identified parameters—covering the full trajectory of mental disintegration: erosion of resilience, distortion of perception, the tipping point

itself, self-reinforcing symptom feedback, loss of insight (anosognosia as a distance-dependent attractor property), the fear of madness as an autocatalytic destabilizing loop, internal resistance as an energy-depleting restoring force, altered time–space perception, the creativity–madness boundary, social isolation dynamics, gaslighting as targeted resilience destruction, collective phase transitions, evolutionary origins of disorder attractors, the role of religious belief, AI-based monitoring, punitive psychiatry as state-space misclassification, and a dynamical typology of suicidal crises.

We address the gradient assumption explicitly, introducing quasipotentials (Freidlin–Wentzell theory) for non-gradient dynamics such as bipolar cycling and OCD oscillations. We situate the framework within the broader computational psychiatry landscape—predictive processing, reinforcement learning, network theory—explaining what the DST approach adds and where the paradigms converge. We engage critically with the early warning signal debate (Helmich et al., 2024; Smit et al., 2025), distinguishing bifurcation-driven transitions (where critical slowing down applies) from noise-induced barrier crossings (where Kramers' escape rate applies but critical slowing down does not), and classify all predictions accordingly.

The framework is validated through numerical simulations: phase portraits of the potential landscape, bifurcation diagrams, a stochastic trajectory exhibiting critical slowing down, and an integrated simulation combining depression, social isolation, and declining insight that reproduces the clinically observed pattern of insight peaking before and vanishing after the transition. A measurement mapping table links all model variables to specific EMA items, clinical scales, and digital phenotyping instruments.

Eight appendices demonstrate the application of the formalism to DSM-5-TR (APA, 2022) diagnostic categories: schizophrenia spectrum disorders (five-dimensional symptom network), bipolar disorders (three-well potential with sleep feedback), depressive disorders (self-sustaining network with hysteresis), anxiety disorders (threat–arousal loops), trauma- and stressor-related disorders (PTSD as trauma-induced attractor), obsessive-compulsive disorders (limit-cycle model with Hebbian deepening), substance use disorders (tolerance–withdrawal–reward attractor), neurocognitive disorders (irreversible landscape degradation), and personality disorders (attractor topology as trait architecture, with the Alternative DSM-5 Model mapped to landscape parameters).

Keywords: dynamical systems theory, computational psychiatry, tipping points, attractors, resilience, stochastic differential equations, bifurcation theory, quasipotential, critical slowing down, DSM-5-TR, anosognosia, symptom networks, Kramers escape rate, hysteresis, numerical simulation

Contents

1 Introduction

1.1 Background

The application of dynamical systems theory (DST) to psychiatry—understanding mental disorders as alternative stable states in a complex system, with transitions governed by tipping points, bifurcations, and hysteresis—was articulated most comprehensively by Scheffer, Bockting, Borsboom, and colleagues in a two-part review published in *JAMA Psychiatry* [Scheffer et al., 2024a,b]. Drawing on concepts from ecology and climate science [Scheffer et al., 2009, 2012], the authors proposed that the healthy state has a basin of attraction representing its resilience, that psychiatric disorders are alternative attractors in which the system can become trapped, and that resilience can be monitored using dynamic indicators of resilience (DIORs)—rising autocorrelation and variance in time-series data—as generic early warning signals of impending transitions [van de Leemput et al., 2014, Wichers et al., 2016].

The Scheffer et al. framework has been influential, but it has three significant limitations.

First, it is primarily a *methodological program*: it proposes measurement tools and draws analogies from ecosystem restoration, but does not engage with the phenomenological content of psychiatric disorders—what a person actually experiences as they lose their mind, what specific feedback loops operate in different conditions, how cultural, social, and religious factors shape the stability landscape, or how clinically significant phenomena such as anosognosia, the fear of madness, and internal resistance arise from the system's dynamics.

Second, the review *contains no equations*: the mathematical apparatus of DST—stochastic differential equations, bifurcation normal forms, Kramers' escape rates, spectral analysis of stability—remains implicit, borrowed by reference from ecology but never explicitly developed for psychiatric phenomena.

Third, the framework does not situate itself within the broader landscape of computational psychiatry, which has developed its own formalisms for many of the same phenomena: predictive processing for psychosis [Adams et al., 2013, Sterzer et al., 2018], reinforcement learning for addiction and depression [Redish, 2004, Huys et al., 2016], and network theory for symptom dynamics [Borsboom, 2017, Briganti, 2024]. The relationship between the DST approach and these established paradigms remains unspecified.

1.2 Aims and Scope

The present article addresses all three limitations. Its contributions are:

1. **Phenomenological content.** We provide a systematic examination of the phenomena of mental disintegration—stages of deterioration, self-reinforcing feedback

loops, anosognosia, the fear of madness, internal resistance, altered perception, creativity, social dynamics (isolation, gaslighting, collective madness), evolutionary origins, religious belief, technological interaction, punitive psychiatry, suicide, and the romanticization of madness—within the DST framework.

2. **Unified mathematical formalism.** We develop a modular system of stochastic differential equations with consistent notation throughout, covering each phenomenon as a computable dynamical model. We explicitly address the gradient assumption and introduce quasipotentials for non-gradient dynamics (Section 2.2). We distinguish quantitatively testable models from structural metaphors and notational formalizations (Section 2).

3. **Engagement with computational psychiatry.** We situate the DST framework within the broader field, explaining its relationship to predictive processing, reinforcement learning, and network theory, and engage critically with the early warning signal debate [Helmich et al., 2024] (Section 2.3).

4. **Numerical simulations.** We present phase portraits, bifurcation diagrams, a critical-slowing-down demonstration, and an integrated multi-variable simulation that reproduces clinically observed dynamics (Section 18).

5. **Operationalization.** We map all model variables to specific measurement instruments—EMA items, clinical scales, digital phenotyping—and classify all predictions by their dependence on the bifurcation assumption (Sections 19–20).

6. **Application to DSM-5-TR.** Eight appendices demonstrate the formalism's application to major diagnostic categories as defined in the DSM-5-TR (APA, 2022): schizophrenia spectrum, bipolar disorders, depressive disorders, anxiety disorders, trauma-related disorders, OCD, substance use disorders, neurocognitive disorders, and personality disorders.

1.3 Structure

Section 2 establishes methodological foundations: the distinction between models and metaphors, the gradient assumption and quasipotentials, the relationship to existing computational psychiatry, and the core mathematical notation. Sections 3–17 develop the phenomenological analysis with accompanying formalism. Sections 18–20 present numerical simulations, the measurement mapping, and the bifurcation-dependence classification. Section 21 discusses limitations. Section 22 summarizes principal contributions. Appendices A–I apply the framework to DSM-5-TR diagnostic categories.

2 Methodological Foundations and Scope

Before developing the formalism, we address several foundational questions about its scope, limitations, and relationship to existing work.

2.1 Models, Metaphors, and the Spectrum Between Them

A persistent question in mathematical approaches to psychiatry is whether formal models constitute testable scientific theories or merely redescribe known phenomena in new notation [Stephan et al., 2017]. We position the present framework explicitly on this spectrum.

Some of the models developed below are *quantitatively testable*: the critical-slowing-down predictions (Section 4), the symptom network model (Section 5), and the DIOR computations (Section 14) generate numerical predictions that can be compared to empirical time series. For these, we provide worked examples with parameter estimates from published data (Sections 4, 5).

Other formalizations are better understood as *structural metaphors with mathematical precision*: the anosognosia model (Section 6), the creativity index (Section 10), and the entropy comparison (Section 17) capture qualitative relationships in equation form, constraining the space of possible mechanisms without generating point predictions from current data. We believe these have value—they enforce logical consistency, reveal hidden assumptions, and identify what *would need to be measured* to test each claim—but we do not overstate their current empirical status.

A third category consists of *notational formalizations*: compact mathematical statements of qualitative insights (e.g., the evolutionary perspective in Section 12, the punitive psychiatry formalization in Section 15). These serve primarily organizational and communicative purposes, and we flag them as such.

2.2 The Gradient Assumption and Its Limits

Definition 2 (below) introduces a potential function V under the assumption that $\mathbf{f}$ is a gradient field. This is a strong assumption: it excludes rotational dynamics, limit cycles, and non-reciprocal interactions. We use it because it provides powerful intuition (basins, wells, barrier heights) and because Kramers' escape rate formula (Eq. 13) applies directly.

However, several phenomena modeled below—bipolar cycling (Appendix B), OCD oscillations (Appendix F)—are inherently non-gradient. For these systems, the potential landscape metaphor must be replaced by the *quasipotential* $\tilde{V}$ from Freidlin–Wentzell

large-deviation theory [Freidlin & Wentzell, 1998]:

$$\tilde{V}(\mathbf{x}) = \inf_{\substack{T>0,\, \boldsymbol{\varphi}(0)=\mathbf{x}^* \\ \boldsymbol{\varphi}(T)=\mathbf{x}}} \frac{1}{2} \int_0^T \|\dot{\boldsymbol{\varphi}}(t) - \mathbf{f}(\boldsymbol{\varphi}(t))\|^2_{\boldsymbol{\Sigma}^{-1}}\, dt, \tag{1}$$

where $\mathbf{x}^*$ is the reference attractor and $\|\cdot\|_{\boldsymbol{\Sigma}^{-1}}$ is the norm weighted by the inverse noise covariance. For gradient systems, $\tilde{V}$ reduces to V; for non-gradient systems, $\tilde{V}$ preserves the escape-rate structure: $\tau_{\text{escape}} \sim \exp(2\Delta\tilde{V}/\sigma^2)$ where $\Delta\tilde{V}$ is the quasipotential barrier [Kühn, 2011].

Throughout the paper, we use V for gradient systems and note explicitly where the quasipotential $\tilde{V}$ is required. The qualitative conclusions (basin depth determines resilience, critical slowing down precedes transitions, hysteresis reflects barrier asymmetry) hold in both cases.

2.3 Relationship to Existing Computational Psychiatry

Computational psychiatry is a rapidly maturing field with multiple established paradigms [Huys et al., 2016, Stephan et al., 2017]. The present framework should be understood in relation to each.

Predictive processing and active inference. The most influential theoretical framework in computational psychiatry formalizes psychosis as aberrant precision-weighting in hierarchical Bayesian inference [Friston, 2010, Adams et al., 2013, Sterzer et al., 2018]. Hallucinations arise when sensory prediction errors are assigned excessive precision; delusions are the explanatory models constructed to accommodate spurious signals. Friston's (2022) vision of computational phenotyping—estimating individual patients' model parameters from neuroimaging via dynamic causal modeling—points toward a computational nosology. The predictive processing framework operates at a computational/algorithmic level, while our DST approach operates at a dynamical/geometric level. The two converge: aberrant precision-weighting can be understood as a mechanism that reshapes the attractor landscape by distorting the feedback weights w_{ij} in our network model (Eq. 15).

Reinforcement learning models. RL models, anchored in the dopamine prediction error signal, provide the second major paradigm [Redish, 2004, Huys et al., 2016]. Redish's (2004) model of addiction as a non-compensable dopamine signal has been empirically extended: recent work shows that substance use specifically disrupts range adaptation (contextual calibration of reward values), not just reward magnitude. RL models of depression reveal reduced reward learning rates, while anxiety models show impaired

threat extinction. Our DST framework (particularly Appendix G) operates at the dynamical rather than algorithmic level: the RL mechanisms are one implementation of the feedback loops that maintain the addiction attractor. The two levels are complementary.

Network theory of mental disorders. Borsboom's (2017) network theory provides the empirical methodology—experience sampling, vector autoregression, Gaussian graphical models—for estimating the adjacency matrix $\mathbf{W}$ in our Eq. 15. The major recent advance is the shift from cross-sectional to temporal (within-person) networks estimated from EMA data [Briganti, 2024]. However, as Bringmann et al. (2022) note, moving from network theory to clinical practice involves unresolved challenges around node selection, temporal dynamics, and state–trait distinctions. Our DST framework adds the dynamical interpretation—attractors, tipping points, transition geometry—that the network approach lacks.

The early warning signal debate. A critical empirical question for our framework is whether critical slowing down (CSD) reliably predicts psychiatric transitions. The original findings were promising [van de Leemput et al., 2014, Wichers et al., 2016]. However, subsequent large-scale studies have been more sobering. Smit et al. (2025) found reasonable specificity (∼84%) but low sensitivity (∼33%). Helmich et al. (2024), in a *Nature Reviews Psychology* Perspective, argued that "there is little support for the use of early warning signals based on critical slowing down in clinical psychology," citing the assumption that psychiatric transitions are always bifurcation-driven (which may not hold), the difficulty of defining "transitions" in noisy data, and insufficient pre-transition time series length. This is directly relevant to our framework. Some psychiatric transitions may be *noise-induced escapes over a finite barrier* (captured by Kramers' formula, Eq. 13) rather than bifurcation-driven tipping points—in which case CSD is absent but our broader framework (barrier height, escape rate) still applies. Distinguishing these mechanisms empirically is an open problem.

Digital phenotyping. The dense time-series data required by both network models and DST predictions are increasingly available through smartphone-based EMA and digital phenotyping (passive sensor data), which provide the empirical backbone for testing our framework's predictions.

The present contribution is not a competitor to these paradigms but a *unifying dynamical language* that can express phenomena across all DSM-5-TR categories within a single notational system. Each paradigm excels at specific aspects: predictive processing explains computational content of psychotic symptoms; RL explains algorithmic reward pathology; network theory provides statistical tools for symptom interaction estimation.

The DST framework's distinctive contribution is its emphasis on *transitions*—how and why systems tip between states—and its capacity to formalize phenomena (anosognosia, fear-of-madness loops, collective madness, punitive psychiatry) that lie outside the scope of any single existing paradigm.

2.4 Notation and Core Model

We establish the following notation, used consistently throughout the article.

Definition 1 (State space)**.** *Let* $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))^\top \in \mathbb{R}^n$ *denote the* mental state vector *at time* t*, where each component* x_i *represents a clinically relevant dimension: mood, anxiety, sleep quality, reality contact, social engagement, cognitive coherence, etc.*

The dynamics of the system are governed by a stochastic differential equation (SDE):

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, \boldsymbol{\mu})\, dt + \boldsymbol{\Sigma}\, d\mathbf{W}(t), \tag{2}$$

where $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is the deterministic drift, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^\top$ is a vector of slowly varying *control parameters* (stress level, social support, genetic vulnerability, etc.), $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ is the noise intensity matrix, and $\mathbf{W}(t)$ is an n-dimensional Wiener process representing stochastic perturbations from the environment.

Definition 2 (Potential landscape)**.** *When* $\mathbf{f}$ *is a gradient field, there exists a potential function* $V : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *such that* $\mathbf{f}(\mathbf{x}, \boldsymbol{\mu}) = -\nabla_{\mathbf{x}} V(\mathbf{x}, \boldsymbol{\mu})$*. Stable attractors correspond to local minima of* V*, unstable equilibria to saddle points or local maxima, and basins of attraction to the domains from which the system flows to a given minimum.*

Definition 3 (Resilience)**.** *The* resilience *of an attractor* $\mathbf{x}^*$ *is defined as the depth of the potential well:*

$$\mathcal{R}(\mathbf{x}^*; \boldsymbol{\mu}) = V(\mathbf{x}_{\text{saddle}}, \boldsymbol{\mu}) - V(\mathbf{x}^*, \boldsymbol{\mu}), \tag{3}$$

where $\mathbf{x}_{\text{saddle}}$ *is the lowest saddle point on the boundary of the basin of attraction of* $\mathbf{x}^*$*.*

For scalar reduction (one-dimensional effective dynamics), we frequently use the canonical *fold normal form*:

$$\frac{dx}{dt} = \mu + x^2 \quad (\text{or } \mu - x^2), \tag{4}$$

which undergoes a saddle-node bifurcation at $\mu = 0$: for $\mu < 0$, two equilibria (one stable, one unstable) exist; at $\mu = 0$ they collide and annihilate; for $\mu > 0$, no equilibrium exists and the system escapes to a remote attractor. This is the mathematical prototype of a tipping point.

3 Normality as the Healthy Attractor

The question of what constitutes normality has long been recognized as culturally and historically contingent. What one society treats as acceptable behavior, another may classify as madness; what one era considers insanity, another may regard as spiritual insight. Shamans, holy fools, and mystics have occupied the same phenomenological space that modern psychiatry labels as psychotic.

Dynamical systems theory offers a formal restatement of this insight. *Normality corresponds to the basin of attraction around the healthy state*—the set of behavioral, cognitive, and emotional patterns from which the system will return to equilibrium after perturbation. This basin is not fixed: its boundaries are shaped by cultural norms, social expectations, and the individual's biological constitution. What counts as a "perturbation" versus a "transition to a disorder attractor" depends partly on the social definition of the basin itself. In a culture that values ecstatic experience, a trance state lies within the healthy basin; in a culture that pathologizes it, the same state may be interpreted as a departure from the attractor.

Yet the dynamical systems view also explains why, despite this cultural relativity, experienced clinicians can often recognize a patient at first glance. The professional eye is detecting not arbitrary social deviance but the *dynamical signature* of a system that has left its healthy basin: the altered temporal structure of speech, the disrupted feedback between affect and context, the rigidity of thought that signals entrapment in an alternative attractor. The psychiatrist is, in effect, reading the dynamics of the system.

Formalization. Let $\mathbf{x}_H^*$ denote the healthy equilibrium. The basin of attraction is:

$$\mathcal{B}_H = \left\{ \mathbf{x}_0 \in \mathbb{R}^n : \lim_{t \to \infty} \boldsymbol{\varphi}(t; \mathbf{x}_0) = \mathbf{x}_H^* \right\}, \tag{5}$$

where $\boldsymbol{\varphi}(t; \mathbf{x}_0)$ denotes the deterministic flow from initial condition $\mathbf{x}_0$. The cultural dependence of normality is captured by making the potential depend on a cultural parameter vector $\mathbf{c}$:

$$V = V(\mathbf{x}, \boldsymbol{\mu}, \mathbf{c}) \quad \Longrightarrow \quad \mathcal{B}_H = \mathcal{B}_H(\boldsymbol{\mu}, \mathbf{c}). \tag{6}$$

A trance state $\mathbf{x}_T$ may satisfy $\mathbf{x}_T \in \mathcal{B}_H(\boldsymbol{\mu}, \mathbf{c}_1)$ in a shamanic culture $\mathbf{c}_1$ but $\mathbf{x}_T \notin \mathcal{B}_H(\boldsymbol{\mu}, \mathbf{c}_2)$ in a clinical culture $\mathbf{c}_2$, even though the underlying neurobiology is identical.

4 The Stages of Losing One's Mind as a Trajectory Through State Space

4.1 The One-Dimensional Effective Model

For analytic clarity, we project the full dynamics onto a single effective coordinate $x(t)$ representing overall mental health (higher x = healthier), governed by a double-well potential:

$$V(x, \mu) = -\frac{\mu}{2}\, x^2 + \frac{1}{4}\, x^4 - h\, x, \tag{7}$$

where μ controls the depth of the wells (resilience parameter: decreasing μ = increasing chronic stress) and h is an asymmetry parameter (bias toward the healthy or disordered state). The stochastic dynamics are:

$$dx = \left(\mu\, x - x^3 + h\right) dt + \sigma\, dW(t). \tag{8}$$

For $h > 0$ and μ sufficiently large, the healthy minimum x_H^* is deeper than the disordered minimum x_D^*. As μ decreases (chronic stress accumulates), the healthy well becomes shallower.

4.2 Erosion of Resilience (Pre-Transition Phase)

The process of losing one's mind often begins not with a dramatic break but with a prolonged period of stress, trauma, or accumulated adversity during which internal resources are gradually depleted. The person feels increasingly helpless, anxious, or detached from the world. Sleep deteriorates. Concentration falters. Minor events provoke disproportionate reactions.

In DST terms, this phase corresponds to the *progressive erosion of the basin of attraction* around the healthy state. The basin becomes shallower, its slopes flatter. The system's return rate to equilibrium after perturbation slows down. This is precisely the phenomenon of *critical slowing down*—the generic early warning signal that precedes a tipping point in any complex system [Scheffer et al., 2009, 2012]. The increased temporal autocorrelation and variance in mood fluctuations during this phase are not merely subjective feelings of instability; they are mathematically predictable consequences of a system approaching a critical transition [van de Leemput et al., 2014].

Clinically, this phase is often invisible. The person attributes their symptoms to external circumstances—overwork, relationship difficulties, financial stress—and does not seek help. The system has not yet tipped, and from the outside, everything may appear functional. But the resilience is draining, and the tipping point is approaching.

Formalization. The slow erosion of resilience is modeled by the control parameter

dynamics:

$$\frac{d\mu}{dt} = -\epsilon\, S(t) + \gamma\,[\mu_0 - \mu(t)], \tag{9}$$

where $S(t) \geq 0$ is external stress intensity, $\epsilon > 0$ is stress sensitivity, $\gamma > 0$ is the natural recovery rate, and μ_0 is baseline resilience. When stress is chronic ($S > \gamma\mu_0/\epsilon$), μ drifts steadily downward.

Critical slowing down is quantified by the linearization around x_H^*. The eigenvalue of the Jacobian is:

$$\lambda(\mu) = \left.\frac{\partial f}{\partial x}\right|_{x=x_H^*} = \mu - 3(x_H^*)^2. \tag{10}$$

As $\mu \to \mu_c$ (the bifurcation point), $\lambda \to 0^-$: the return rate vanishes. The Ornstein–Uhlenbeck approximation near x_H^* gives the measurable indicators:

$$\mathrm{Cor}\big(x(t),\, x(t+\tau)\big) \approx e^{\lambda\tau}, \qquad \mathrm{Var}(x) \approx \frac{\sigma^2}{2|\lambda|}. \tag{11}$$

As $|\lambda| \to 0$: autocorrelation $\to 1$ and variance $\to \infty$—the dynamic indicators of resilience (DIORs) of Scheffer et al. [2024b].

Worked example: parameter estimation from depression data. Van de Leemput et al. [van de Leemput et al., 2014] analyzed experience-sampling data from 535 subjects, measuring affect at multiple time points per day. They found that subjects who subsequently developed depression showed increasing lag-1 autocorrelation ($\hat{\rho}_1$) in the months preceding onset, rising from approximately 0.4 to 0.7 over a 6-month pre-transition window. Using Eq. 11 with a sampling interval $\Delta t = 1$ day, we estimate:

$$\lambda \approx \frac{\ln \hat{\rho}_1}{\Delta t}: \quad \lambda_{\text{early}} \approx -0.92\ \text{day}^{-1}, \quad \lambda_{\text{late}} \approx -0.36\ \text{day}^{-1}. \tag{12}$$

The recovery rate has halved. If the variance simultaneously doubled from $\hat{\sigma}^2 = 0.5$ to 1.0 (as observed), Eq. 11 gives $\sigma^2/(2|\lambda|)$: $0.5/(2 \times 0.92) = 0.27$ vs. $0.5/(2 \times 0.36) = 0.69$, consistent with the data. This is a *non-obvious, quantitative* prediction: the ratio of variance increase to autocorrelation increase is constrained by the OU model, and the empirical data conform to this constraint.

Using these estimates and Kramers' formula (Eq. 13), the predicted time to transition drops from $\tau_{\text{escape}}^{\text{early}} \propto e^{2\mathcal{R}/0.5}$ to $\tau_{\text{escape}}^{\text{late}} \propto e^{2\mathcal{R}'/0.5}$ where $\mathcal{R}' \approx (\lambda_{\text{late}}/\lambda_{\text{early}})^{3/2}\mathcal{R} \approx 0.24\,\mathcal{R}$. The barrier has collapsed to a quarter of its original height—the system is approaching the tipping point.

Empirical caveat. While this example demonstrates that the OU/Kramers parameterization is consistent with population-level data, recent large-scale within-person studies have found that CSD-based early warning signals have limited sensitivity at the indi-

vidual level (∼33% in Smit et al., 2025; see Section 2.3). This may indicate that many psychiatric transitions are noise-induced barrier crossings rather than bifurcation-driven tipping points—a distinction our framework accommodates (Kramers' formula applies to both cases) but which has different implications for prediction: bifurcation-driven transitions are preceded by CSD; noise-induced transitions are not, but can still be predicted from barrier height estimates.

4.3 Distortion of Perception (Approaching the Tipping Point)

As resilience continues to erode, a second phase emerges in which perception begins to distort. The person has difficulty separating fantasies from real events. Thoughts become intrusive, repetitive, or paranoid. Random coincidences acquire ominous significance. The person begins to suspect hostile intent in the behavior of others. Social isolation increases, which further weakens the feedback loops that ordinarily anchor perception to shared reality.

In the dynamical systems framework, this phase represents the system hovering near the boundary of the healthy basin. The attractor of the healthy state is so shallow that the system is easily displaced by perturbations, and it lingers at states far from the equilibrium for extended periods. The system is "flickering"—exhibiting transient excursions toward the alternative attractor (the disorder state) before returning to the healthy state. This flickering is itself a recognized early warning signal in complex systems: it indicates that the system is sampling the disorder attractor before committing to it.

Formalization. The mean first passage time from x_H^* to the separatrix x_s is given by Kramers' escape rate [Kramers, 1940]:

$$\tau_{\text{escape}} \approx \frac{2\pi}{\sqrt{|V''(x_H^*)| \cdot |V''(x_s)|}} \exp\left(\frac{2\,\mathcal{R}}{\sigma^2}\right), \tag{13}$$

where $\mathcal{R} = V(x_s, \mu) - V(x_H^*, \mu)$ is the barrier height (resilience). As $\mathcal{R} \to 0$, τ_{escape} drops exponentially: flickering becomes increasingly frequent.

4.4 The Tipping Point (Phase Transition)

The third stage is the tipping point itself: the person crosses the boundary and enters the disorder attractor. Hallucinations, delusional systems, or severe affective dysregulation emerge. The person begins to live in a reality controlled by internal distortions. Crucially, the system has now settled into a new stable state: the disorder is self-sustaining.

DST explains why this transition can appear sudden and disproportionate to the immediate trigger. The system was already on the edge of the basin; a minor perturbation—a sleepless night, a social humiliation, a pharmacological change—was sufficient to push

it across the boundary. The tipping point is not caused by the final stressor; it is caused by the prior depletion of resilience that brought the system to the edge. This reframes the old clinical question of "why did this person break down over something so trivial?": the trivial event was not the cause, but the last push in a long process of resilience erosion.

Formalization. At the critical value $\mu = \mu_c$, the healthy minimum and the separatrix collide (saddle-node bifurcation). For $\mu < \mu_c$, the healthy well ceases to exist and the system falls irreversibly into the disordered attractor x_D^*. Near the bifurcation, the universal scaling laws are:

$$x_H^* - x_s \sim (\mu - \mu_c)^{1/2}, \qquad \mathcal{R} \sim (\mu - \mu_c)^{3/2}. \tag{14}$$

The 3/2-power law for the barrier height means that resilience collapses rapidly near the tipping point, explaining the sudden, disproportionate character of the final transition.

5 Self-Reinforcing Feedback Loops: The Engine of Entrapment

Once the system has tipped into the disorder attractor, it is maintained there by self-reinforcing feedback loops among symptoms. In depression: low mood → social withdrawal → loneliness → rumination → insomnia → fatigue → cognitive impairment → loss of self-efficacy → deeper low mood. In paranoia: perception of threat → hypervigilance → misinterpretation of ambiguous stimuli → confirmation of threat narrative → social withdrawal → loss of corrective social feedback → deepening of paranoid system. In psychosis: hallucination → anxiety → sleep disruption → cognitive fragmentation → further hallucinations.

These loops are the dynamical mechanism that makes the disorder attractor stable. They explain why simply removing the original stressor is typically insufficient for recovery—the phenomenon of *hysteresis* [Scheffer et al., 2024a]. The stressor may have triggered the transition, but the feedback loops now sustain it independently. The disorder has acquired its own self-perpetuating dynamics.

This also explains a clinically important observation: the process of losing one's mind is highly individual. For one person, the critical feedback loop may be insomnia–fatigue–rumination; for another, social isolation–paranoia–avoidance; for a third, substance use–cognitive impairment–loss of employment. Each individual's disorder attractor has a unique topology, shaped by their specific symptom network [Borsboom, 2017]. This is why personalized approaches to treatment—identifying and targeting the particular feedback loops that sustain the disorder in a given patient—are essential.

Formalization. Let $x_i(t)$ denote the activation level of symptom i. The network dy-

namics are:

$$\frac{dx_i}{dt} = -\alpha_i\, x_i + \sum_{j=1}^{n} w_{ij}\, g(x_j) + s_i(t) + \sigma_i\, \xi_i(t), \qquad i = 1, \ldots, n, \tag{15}$$

where $\alpha_i > 0$ is the natural decay rate, w_{ij} is the causal influence of symptom j on symptom i (the adjacency matrix $\mathbf{W}$), $g(x) = (1+e^{-\beta x})^{-1}$ is a sigmoid activation function, $s_i(t)$ is an external stressor, and $\xi_i(t)$ is white noise. In matrix form, with $\mathbf{A} = \mathrm{diag}(\alpha_i)$:

$$\frac{d\mathbf{x}}{dt} = -\mathbf{A}\mathbf{x} + \mathbf{W}\,\mathbf{g}(\mathbf{x}) + \mathbf{s}(t) + \mathbf{\Sigma}\,\boldsymbol{\xi}(t). \tag{16}$$

Proposition 1 (Sufficient condition for self-sustaining disorder)**.** *If the spectral radius $\rho(\mathbf{A}^{-1}\mathbf{W}\,\beta/4) > 1$ (where $\beta/4$ is the maximum slope of g), the network admits at least one high-activation fixed point $\mathbf{x}_D^* \neq \mathbf{0}$ even when $\mathbf{s}(t) = \mathbf{0}$.*

Remark 1. *This is a sufficient condition for the existence of a nontrivial equilibrium, derived from the contraction mapping principle applied to $\mathbf{x} \mapsto \mathbf{A}^{-1}\mathbf{W}\,\mathbf{g}(\mathbf{x})$. It does not guarantee stability of $\mathbf{x}_D^*$ (which requires checking the eigenvalues of the Jacobian at $\mathbf{x}_D^*$), nor uniqueness. The $\beta/4$ bound is a worst case (all sigmoids at maximum slope); for specific activation levels, tighter conditions can be derived. In practice, stability of the disorder attractor can be verified numerically for a given $\mathbf{W}$ estimated from patient data via vector autoregression on experience-sampling time series [Borsboom, 2017].*

Hysteresis follows: the critical stressor level s_{onset} needed to push the system into $\mathbf{x}_D^*$ exceeds the reduction s_{offset} needed for escape:

$$s_{\text{onset}} > s_{\text{offset}} \quad \Longleftrightarrow \quad \text{recovery requires more intervention than prevention.} \tag{17}$$

Bridge symptoms connecting disorder clusters A and B are identified by nonzero inter-cluster weights: symptom k is a bridge if $\sum_{i \in A} |w_{ik}| > 0$ and $\sum_{j \in B} |w_{kj}| > 0$.

6 Anosognosia: The Attractor Does Not Know It Is an Attractor

One of the most clinically significant features of severe mental illness is *anosognosia*—the inability to recognize one's own illness. People with schizophrenia may be certain that their hallucinations are real; people in a manic episode may feel more lucid and capable than ever; people with early dementia may deny any cognitive decline. Early symptoms such as anxiety, mood changes, or mild cognitive difficulties are attributed to external circumstances rather than recognized as signs of deteriorating mental health.

DST provides a natural explanation for this phenomenon. From the perspective of the system that has settled into the disorder attractor, the disordered state *is* the equilibrium—it is the state toward which all internal dynamics converge. The self-reflective capacity that might enable recognition of the illness is itself part of the system that has transitioned; it is now operating within the logic of the disorder attractor, not outside it. The system cannot evaluate itself from a vantage point it no longer occupies.

This is analogous to a well-known property of complex systems: a system trapped in an alternative stable state has no internal signal that it is in the "wrong" state. A turbid lake does not "know" it was once clear. The disorder state is, from the system's internal perspective, as stable and self-consistent as the healthy state. The brief moments of lucidity sometimes reported by patients—terrifying flashes of awareness that something is deeply wrong—can be understood as transient excursions back toward the boundary of the healthy basin, quickly pulled back by the disorder attractor's feedback loops.

Formalization. We model self-reflective capacity as a variable $r(t)$ (reality-monitoring accuracy), coupled to the state:

$$\frac{dr}{dt} = -\alpha_r\, r + \phi\big(\|\mathbf{x} - \mathbf{x}_H^*\|\big) + \eta\, \xi_r(t), \tag{18}$$

where $\phi(d)$ is a function of distance $d = \|\mathbf{x} - \mathbf{x}_H^*\|$ from the healthy attractor:

$$\phi(d) = \phi_0\, d\, \exp\!\left(-\frac{d^2}{2\delta^2}\right). \tag{19}$$

For small d (near healthy state), $\phi \approx \phi_0 d$: self-reflection increases proportionally to deviation—the person notices something is wrong. For large d (deep in disorder attractor), $\phi \to 0$: the self-reflective signal is extinguished. The parameter δ sets the critical distance beyond which insight is lost.

Proposition 2 (Anosognosia as attractor property)**.** *At the disordered equilibrium with* $\|\mathbf{x}_D^* - \mathbf{x}_H^*\| \gg \delta$*, the steady-state insight is* $r^* = \phi(\|\mathbf{x}_D^* - \mathbf{x}_H^*\|)/\alpha_r \approx 0$*. The attractor is self-cloaking.*

Brief moments of lucidity correspond to noise-driven excursions toward the basin boundary, where d temporarily decreases into the range where $\phi(d)$ is non-negligible.

7 The Fear of Losing One's Mind: A Destabilizing Perturbation

The fear of going mad—the anxious self-monitoring, the metacognitive scrutiny of one's own thoughts, the panicked search for signs of incipient insanity—is a clinically recognized phenomenon that can itself precipitate the very crisis it dreads.

In DST terms, this fear operates as a *destabilizing positive feedback loop* within the system. The person perceives a perturbation (an intrusive thought, a moment of confusion), interprets it as evidence of approaching madness, becomes anxious, which produces further perturbations, which are interpreted as further evidence, which increases anxiety, and so on. This is a self-reinforcing spiral that actively erodes the basin of attraction of the healthy state.

The fear of madness can thus be understood as a mechanism by which the system destabilizes itself—a feedback loop that weakens resilience and drives the system toward the tipping point. It is a self-fulfilling prophecy operating through precise dynamical mechanisms. The clinical implication is clear: breaking this particular feedback loop (through psychoeducation, cognitive reframing, or anxiety management) can directly strengthen the resilience of the healthy state.

Formalization. We introduce metacognitive anxiety $a(t)$:

$$\begin{aligned} \frac{dx}{dt} &= f(x,\mu) - \kappa\, a + \sigma\, \xi(t), \\ \frac{da}{dt} &= -\gamma_a\, a + \psi\left(\left|\frac{dx}{dt}\right|\right) + \rho\, a\, \mathbf{1}_{x<x_{\text{threshold}}}, \end{aligned} \tag{20}$$

where $\kappa > 0$ couples anxiety to main dynamics (anxiety pushes x toward disorder), $\psi(\cdot)$ maps perceived instability to alarm, and $\rho\, a\, \mathbf{1}_{x<x_{\text{threshold}}}$ is the autocatalytic core: once x dips below a perceived safety threshold, anxiety amplifies itself.

The effective modification to the potential is:

$$V_{\text{eff}}(x,\mu,a) = V(x,\mu) + \kappa\, a\, x. \tag{21}$$

This tilts the landscape toward the disordered state, reducing resilience:

$$\mathcal{R}_{\text{eff}} = \mathcal{R}_0 - \kappa\, a\, (x_s - x_H^*) < \mathcal{R}_0. \tag{22}$$

The fear of madness literally shrinks the basin of attraction—a self-fulfilling prophecy with a precise mathematical mechanism.

8 Internal Resistance as a Restoring Force

In the early stages of the process, many individuals experience significant internal resistance to the encroaching loss of reality contact. They attempt to rationalize strange experiences, seek logical explanations, perform reality checks, reach out for confirmation from others, or try to suppress intrusive thoughts.

In dynamical systems terms, these efforts represent the *restoring forces* that pull

the system back toward the healthy attractor. Every reality check, every search for a rational explanation, every conversation with a trusted friend is a perturbation directed back toward the equilibrium. As long as these restoring forces are strong enough to overcome the destabilizing perturbations, the system remains in the healthy basin.

But as resilience erodes, the restoring forces weaken while the destabilizing forces strengthen. The slopes of the basin become flatter; the return rate slows; the system takes longer to recover from each perturbation. Eventually, the person's cognitive and emotional resources for resistance are exhausted, and the system tips over. At later stages, resistance is replaced by immersion: the disorder attractor captures the system, and the feedback loops that once constituted resistance (self-monitoring, reality testing) are themselves co-opted by the disorder—self-monitoring becomes rumination, reality testing becomes paranoid verification.

Formalization. We model resistance as a restoring force $R(x, e)$ depending on available cognitive–emotional energy $e(t)$:

$$\begin{aligned} \frac{dx}{dt} &= f(x, \mu) + R(x, e) + \sigma\, \xi(t), \\ \frac{de}{dt} &= -\nu\, |R(x, e)| + \gamma_e\, (e_0 - e), \end{aligned} \tag{23}$$

where $R(x, e) = e \cdot k \cdot (x_H^* - x)$ is a restoring force with strength proportional to energy e. The energy equation captures depletion (ν) and natural recovery (γ_e toward baseline e_0).

The effective eigenvalue is $\lambda_{\text{eff}} = \lambda(\mu) - e \cdot k$. When e is high, stability is enhanced. As e depletes, the bare instability is exposed. Tipping occurs when:

$$e(t_{\text{tip}}) < \frac{|\lambda(\mu(t_{\text{tip}}))|}{k} \quad \Longrightarrow \quad \lambda_{\text{eff}} > 0 \quad \Longrightarrow \quad \text{tipping.} \tag{24}$$

9 Altered Perception of Time and Space: Distortion Within the Attractor

People experiencing severe psychiatric disorders frequently report profound distortions of temporal and spatial perception. Time may seem to stretch agonizingly during anxiety, compress during mania, or lose its linear structure entirely during psychosis. Spatial boundaries dissolve: walls appear to move, objects change size, the body feels alien or disconnected. The phenomena of depersonalization and derealization—the sense of watching oneself from outside, of the world being artificial or dreamlike—are common features of these states.

From a DST perspective, these phenomenological changes are *signatures of the system's operation within a qualitatively different attractor.* The healthy attractor involves

a specific pattern of temporal integration and spatial calibration, maintained by feedback between perception, proprioception, and cognitive processing. When the system transitions to a disorder attractor, these feedback loops are reconfigured: the temporal integration window changes (producing subjective speeding-up or slowing-down of time), the spatial calibration shifts (producing size distortions and boundary dissolution), and the self-referential loops that produce the sense of embodied presence are disrupted (producing depersonalization).

These are not secondary "symptoms" added on top of a primary disease; they are *intrinsic features of the new attractor's dynamics*—the characteristic oscillatory patterns and feedback structures of the disordered state.

Formalization. The subjective temporal experience is modeled through the internal clock rate:

$$\omega(t) = \omega_0 + \zeta \frac{dx}{dt} + \theta \ \mathrm{Var}_\tau[x], \tag{25}$$

where ω_0 is baseline clock rate, ζ captures state velocity effects (mania $\Rightarrow$ acceleration; depression $\Rightarrow$ deceleration), and θ captures variance effects (high fluctuation $\Rightarrow$ temporal disorientation). Subjective time elapsed:

$$T_{\text{subj}} = \int_0^T \omega(t)\, dt. \tag{26}$$

Depersonalization is modeled via a self-coupling coefficient:

$$\chi(t) = \chi_0 \ \exp\left(-\frac{\|\mathbf{x}(t) - \mathbf{x}_H^*\|^2}{2\sigma_\chi^2} \right). \tag{27}$$

At the healthy attractor, $\chi \approx \chi_0$ (intact embodiment). Deep in the disorder attractor, $\chi \to 0$ (depersonalization).

10 Creativity and Madness: Exploring the Basin Boundary

The relationship between creativity and madness has been debated for centuries. Many celebrated artists, writers, and scientists exhibited traits associated with psychiatric disorders. Yet creativity and madness are not identical: the creative person experiments with unconventional perception but retains the capacity to return to the shared consensus; the person who is losing their mind does not.

DST provides a precise formalization of this distinction. *Creativity can be understood as the capacity to make large excursions from the equilibrium of the healthy attractor*—to explore regions of state space far from the conventional center—while retaining sufficient

resilience to return. The creative mind has a wide basin of attraction with gentle slopes: it can wander far without tipping over the edge. Madness, by contrast, is what happens when the system crosses the basin boundary and cannot return: the excursion becomes a transition.

This also explains why intense emotional states (suffering, euphoria, existential crisis) can stimulate both creativity and madness. Such states represent large perturbations that push the system far from its equilibrium. If the basin is deep, the perturbation stimulates exploration (creativity). If the basin is shallow, the same perturbation may push the system past the tipping point (disorder). The critical variable is not the intensity of the experience but the resilience of the system that experiences it.

Formalization.

Definition 4 (Creative exploration index)**.**

$$\mathcal{C}(T) = \frac{1}{T}\int_0^T \frac{\|\mathbf{x}(t) - \mathbf{x}_H^*\|}{d(\mathbf{x}(t), \partial\mathcal{B}_H)}\, dt, \tag{28}$$

where $d(\mathbf{x}, \partial\mathcal{B}_H)$ *is the distance to the basin boundary.*

High $\mathcal{C}$ with $\mathbf{x}(t) \in \mathcal{B}_H$ for all t: creativity. $\mathcal{C} \to \infty$ ($d \to 0$): approaching madness. $\mathbf{x}(t) \notin \mathcal{B}_H$: madness.

Maximum safe excursion in the scalar model:

$$\Delta x_{\max} = x_H^* - x_s \sim (\mu - \mu_c)^{1/2}. \tag{29}$$

11 Social Factors: The Environment Shapes the Landscape

11.1 Loneliness and Isolation as Resilience Erosion

Loneliness and social isolation are powerful risk factors for psychiatric disorders. In DST terms, social connection serves as a critical external feedback loop that helps maintain the healthy attractor. Conversations with trusted others provide reality checks (corrective feedback on perception), emotional regulation (dampening of excessive perturbations), and cognitive stimulation (maintenance of neural plasticity). When these social feedback loops are severed by isolation, the basin of the healthy attractor narrows and shallows. The system becomes more vulnerable to tipping.

Chronic isolation additionally generates its own destabilizing feedback loops: loneliness → negative self-perception → social anxiety → further withdrawal → deeper loneliness. This is a self-reinforcing cycle that actively erodes resilience, potentially creating

its own attractor (a "loneliness trap") from which escape becomes progressively more difficult.

Formalization. Social connection adds a stabilizing feedback:

$$\frac{dx}{dt} = f(x,\mu) + C(t)\,(x_H^* - x) + \sigma\,\xi(t), \tag{30}$$

where $C \geq 0$ is social connectivity. Effective eigenvalue: $\lambda_{\text{eff}} = \lambda(\mu) - C$. The isolation trap is modeled as a coupled system:

$$\begin{aligned} \frac{dx}{dt} &= f(x,\mu) + C(t)\,(x_H^* - x) + \sigma\,\xi(t), \\ \frac{dC}{dt} &= -\beta_C\,(x_H^* - x)^2 + \gamma_C\,(C_0 - C), \end{aligned} \tag{31}$$

where worsening mental state drives social withdrawal (decreasing C), creating positive feedback.

11.2 Society as a Force That Shapes and Erodes Basins

Social systems interact with individual mental health in both directions. Supportive social environments—strong families, accessible healthcare, cultural acceptance of emotional vulnerability—deepen the basin of the healthy attractor. Hostile social environments—chronic poverty, discrimination, unrealistic productivity expectations, stigmatization of mental illness—erode it.

Gaslighting and systematic psychological manipulation can be understood as a deliberate attack on a person's resilience—the methodical destruction of the feedback loops (reality testing, self-trust, social confirmation) that maintain the healthy attractor. The dynamical prediction is clear: sustained gaslighting should produce measurable signatures of declining resilience (rising autocorrelation and variance in mood and cognitive self-assessments) in the victim.

Formalization. Gaslighting attacks the self-trust parameter δ_{self}:

$$\frac{d\delta_{\text{self}}}{dt} = -g_0\,G(t) + \gamma_\delta\,(\delta_0 - \delta_{\text{self}}), \qquad R_{\text{reality}} = \delta_{\text{self}}\,k\,(x_H^* - x). \tag{32}$$

As $\delta_{\text{self}} \to 0$ under sustained gaslighting ($G(t) > 0$), the restoring force $R_{\text{reality}} \to 0$.

11.3 Collective Madness as a Societal Phase Transition

The dynamical systems framework extends naturally to collective phenomena. Witch hunts, totalitarian ideologies, economic bubbles, and moral panics can be understood as *phase transitions in social systems*, where the "healthy attractor" of rational public

discourse tips into an alternative attractor dominated by fear, conformism, and irrational belief. The mechanisms are the same: self-reinforcing feedback loops (fear → scapegoating → more fear; propaganda → polarization → suppression of dissent → more propaganda) create a stable pathological state from which the society cannot easily escape. Hysteresis operates at the collective level too: restoring rational discourse after a society has tipped into ideological madness requires far more effort than maintaining it would have.

Formalization. For N interacting agents with states x_i:

$$\frac{dx_i}{dt} = f(x_i, \mu_i) + \frac{J}{N}\sum_{j=1}^{N}(x_j - x_i) + P(t) + \sigma_i\,\xi_i(t), \tag{33}$$

where $J > 0$ is conformity coupling and $P(t)$ is propaganda. The mean field $\bar{x} = N^{-1}\sum_i x_i$ satisfies:

$$\frac{d\bar{x}}{dt} = f(\bar{x}, \bar{\mu}) + P(t) + O(N^{-1/2}). \tag{34}$$

This has the same bistable structure as the individual model: society can tip from rational discourse into collective delusion, with hysteresis. Strong conformity (J large) sharpens the transition.

12 Evolutionary Perspective: The Landscape Was Shaped by Selection

Some mental states classified as pathological in modern clinical contexts may have conferred evolutionary advantages in ancestral environments. Depression may have enhanced cautious, ruminative problem-solving during social threat. The manic phase of bipolar disorder may have provided boldness for leadership. Anxiety may have been a hypersensitive threat-detection system that improved survival despite its costs.

In DST terms, this means that the stability landscape of the human mind was *shaped by natural selection to include certain disorder attractors as functional features*, not as pure malfunctions. These attractors may have been adaptive in ancestral environments but become pathological in the modern world, where stressors are chronic rather than acute, social structures are disrupted, and the feedback loops that once resolved these states are absent. The evolutionary view does not romanticize madness; it explains why disorder attractors exist and why they are so difficult to eliminate.

Formalization (notational). Let $F(\boldsymbol{\mu}_{\text{gen}}, \mathbf{c})$ denote evolutionary fitness depending on genetic landscape parameters $\boldsymbol{\mu}_{\text{gen}}$ and environment $\mathbf{c}$. The persistence of disorder at-

tractors is expressed as an environment-dependent fitness gradient:

$$\left.\frac{\partial F}{\partial \mu_k}\right|_{\mathbf{c}_{\text{anc}}} \neq \left.\frac{\partial F}{\partial \mu_k}\right|_{\mathbf{c}_{\text{mod}}}. \tag{35}$$

This is deliberately schematic: it does not constitute a testable model but identifies the *structural reason* why selection has not eliminated disorder attractors. Turning it into a quantitative model would require specifying the fitness function and the relevant environmental parameters—a task for evolutionary psychiatry [Nesse, 2000].

13 Religion and Faith: Anchoring and Destabilizing the System

Religious belief can function as a powerful stabilizing force—a source of meaning, community, and cognitive structure that deepens the basin of the healthy attractor. People with strong faith may be more resilient to existential crises because their belief system provides an interpretive framework that absorbs perturbations.

However, religion can also destabilize the system. Religious delusions represent the co-optation of religious content by the disorder attractor. Spiritual crises—when faith is shattered by events that contradict its promises—can function as massive perturbations that push the system out of its healthy basin. And reliance on prayer as a substitute for professional treatment can prevent access to interventions that could restore resilience.

The phenomenon of *holy foolishness* (yurodstvo) in the Orthodox tradition is particularly interesting from a DST perspective. The yurodiviy deliberately positions himself at the boundary of the healthy basin, performing behaviors that mimic the disorder attractor while (presumably) retaining the internal dynamics of the healthy state. This is the dynamical equivalent of living on the edge of the basin without tipping over—a feat that requires exceptional psychological flexibility and self-awareness.

Formalization. Faith modifies the potential:

$$V_{\text{faith}}(x, \mu) = V(x, \mu) - \Lambda \exp\left(-\frac{(x - x_H^*)^2}{2\sigma_B^2}\right), \tag{36}$$

where $\Lambda > 0$ is faith strength and σ_B its range. This deepens the healthy well by Λ. A spiritual crisis ($\Lambda \to 0$) produces sudden resilience erosion. The yurodiviy operates at $x \approx x_s$ with high Λ, so the effective potential retains a healthy minimum despite the apparent behavioral state being near the boundary.

14 Artificial Intelligence and Technology: New Tools, New Risks

Technology and AI interact with the dynamical landscape of mental health in both directions. On the positive side, AI enables the computation of dynamic indicators of resilience (DIORs) from time series data—mood tracking, behavioral patterns, speech analysis, physiological signals—making it possible, in principle, to detect declining resilience before a tipping point is reached [Scheffer et al., 2024b]. AI can infer personalized causal networks from individual time series, identifying the specific feedback loops that sustain a disorder in a given patient.

On the negative side, technology can erode resilience. Immersion in virtual reality may weaken the feedback loops that anchor the system to shared physical reality. Social media algorithms that amplify outrage and anxiety function as external perturbations that systematically push users toward the boundaries of their healthy basins. Digital surveillance can provoke paranoid ideation, particularly in vulnerable individuals. And the cognitive offloading of decision-making to AI may weaken the self-regulatory capacity that constitutes part of the healthy attractor's internal dynamics.

Formalization. DIORs are computed from observed time series $\{x(t_k)\}_{k=1}^{K}$:

$$\hat{\rho}_1 = \frac{\sum_{k=1}^{K-1}(x_k - \bar{x})(x_{k+1} - \bar{x})}{\sum_{k=1}^{K}(x_k - \bar{x})^2}, \qquad \hat{\sigma}^2 = \frac{1}{K}\sum_{k=1}^{K}(x_k - \bar{x})^2. \tag{37}$$

Rising $\hat{\rho}_1$ and $\hat{\sigma}^2$ over a sliding window signal declining resilience.

Technology increases effective noise:

$$\sigma_{\text{eff}}^2 = \sigma^2 + \sigma_{\text{tech}}^2(t), \tag{38}$$

and since $\tau_{\text{escape}} \propto \exp(2\mathcal{R}/\sigma_{\text{eff}}^2)$ via Eq. (13), increased noise exponentially decreases the time to transition.

15 Punitive Psychiatry: Weaponizing the Diagnosis of Disordered Attractors

The history of punitive psychiatry—the use of psychiatric institutions to suppress dissent, imprison political opponents, and enforce social conformity—acquires a specific meaning within the dynamical systems framework. Punitive psychiatry involves the *deliberate misclassification of a person's position in state space*: labeling someone who is in the healthy attractor as being in a disorder attractor, in order to justify coercive intervention.

Soviet "sluggish schizophrenia," lobotomy, forced electroconvulsive therapy, and involuntary medication were not treatments designed to restore a system to its healthy attractor; they were interventions designed to *destroy the system's capacity for autonomous dynamics altogether*. They attacked resilience itself—the very capacity of the system to maintain any coherent attractor—as a method of social control. The consequence was not merely the discrediting of psychiatry as a discipline but the destruction of trust in the diagnostic framework itself: if the boundary between the healthy and disordered attractors can be drawn arbitrarily by state power, the entire concept of the basin of attraction loses its scientific legitimacy.

Formalization (notational). Punitive psychiatry is a two-fold operation:
(i) *Misclassification*: declaring $\mathbf{x}(t) \in \mathcal{B}_D$ when $\mathbf{x}(t) \in \mathcal{B}_H$ (false positive in diagnostic map $\mathcal{D} : \mathbb{R}^n \to \{H, D\}$).
(ii) *Iatrogenic destruction*: coercive "treatment" acts as perturbation plus resilience reduction:

$$\mathbf{x} \mapsto \mathbf{x} + \mathbf{p}, \qquad \mu \mapsto \mu - \Delta\mu_{\text{iatrog}}, \tag{39}$$

which can push a previously healthy system past the tipping point—a self-fulfilling diagnosis. This is a notational restatement rather than a predictive model; its value lies in framing political abuse of psychiatry within the same formal language used for clinical phenomena, enabling precise statements about what distinguishes legitimate diagnosis from weaponized misclassification.

16 Suicide: Escape from the Attractor

Suicidal ideation and behavior can arise in multiple dynamical contexts. A person trapped in the deep attractor of severe depression may perceive no path back to the healthy state and conclude that the only escape from the disorder attractor is death. A person experiencing acute psychosis may act on command hallucinations. An impulsive act during a manic-depressive transition may exploit the instability of the system as it crosses basin boundaries. An existential crisis may involve the perception that all available attractors—healthy and disordered alike—are intolerable.

The dynamical systems perspective underscores that suicidal ideation is not necessarily a symptom of a specific disorder but a signal that the system is in a state of extreme distress—either trapped in a deep disorder attractor with no perceived exit, or in a region of state space with no stable attractor at all (a condition of chaotic, unbounded dynamics). The clinical implication is that suicide prevention must address the system's dynamics: deepening the basin of the healthy attractor, providing external stabilizing feedback, and—where a disorder attractor is involved—identifying and disrupting the feedback loops that maintain entrapment.

Formalization. We model perceived escape via variable $u(t)$:

$$\frac{du}{dt} = -\gamma_u\, u + \alpha_u\, \mathbf{1}_{\{V(x_D^*)<-\Theta\}} + \beta_u\, \mathbf{1}_{\{\mathcal{R}_H<\epsilon_H\}}, \tag{40}$$

where the first indicator activates when the disorder well is perceived as intolerably deep, the second when the healthy attractor appears unreachable. Suicidal crisis: $u > u_c$.

Three dynamical scenarios:

$$\text{(a) Deep trap:} \quad V(x_D^*) \ll -\Theta,\ \mathcal{R}_D \gg 0 \quad \text{(no escape from deep depression)}, \tag{41}$$

$$\text{(b) Boundary instability:} \quad x \approx x_s,\ \mathcal{R}_H \approx \mathcal{R}_D \approx 0 \quad \text{(chaotic inter-basin dynamics)}, \tag{42}$$

$$\text{(c) No attractor:} \quad \mu < \mu_c,\ h < h_c \quad \text{(no stable equilibrium exists)}. \tag{43}$$

17 Against the Romanticization of Madness: The Attractor Is a Trap

There is a long cultural tradition of romanticizing madness—viewing it as a source of insight, creativity, or spiritual truth. The dynamical systems framework provides a precise and sobering corrective to this romanticization.

An attractor is a trap. A system in a disorder attractor is not "free"; it is constrained to a specific set of dynamics from which it cannot escape without external intervention or a fortunate perturbation. The rich inner worlds, vivid hallucinations, and grandiose convictions of psychosis are not expressions of a liberated consciousness; they are the characteristic dynamics of a system locked into a pathological pattern of self-reinforcing feedback. The person is not exploring new territory; they are circling within a fixed topology, unable to leave.

Mental health is not the absence of complexity, depth, or existential engagement. It is the resilience to explore the full state space of human experience—including suffering, doubt, and creative risk—while retaining the capacity to return to equilibrium. The threat of madness is not a romantic possibility; it is a permanent structural feature of the dynamical landscape, an alternative attractor that is always present and always available if resilience is sufficiently eroded. To romanticize it is to mistake entrapment for freedom and the collapse of resilience for its transcendence.

Formalization. The trajectory entropy within the disorder attractor provides a measure of experiential "richness":

$$H_D = -\int p_D(\mathbf{x})\, \ln p_D(\mathbf{x})\, d\mathbf{x}, \tag{44}$$

where p_D is the stationary distribution within the disorder basin. For disorders characterized by rigidity and stereotypy (catatonia, severe depression, chronic schizophrenia with negative symptoms), it is plausible that $H_D < H_H$: the disorder attractor constrains the system to a smaller region of state space. However, this inequality is not universal: acute mania and florid psychosis can exhibit high behavioral variability, suggesting $H_D \approx H_H$ or even $H_D > H_H$ in some cases. The critical distinction is between *apparent* and *functional* entropy: the manic state may explore a large region of state space but in a stereotyped, non-adaptive pattern (e.g., cycling through grandiose plans without executing any), yielding high trajectory entropy but low *goal-directed* entropy. Formalizing this distinction—perhaps via directed information or transfer entropy measures—remains an open problem.

The corrective to romanticization does not require $H_D < H_H$ universally; it requires only that the disorder attractor is a *trap*: a region from which the system cannot escape under its own dynamics, regardless of whether that region is large or small. Mental health corresponds to maximal *accessible* state space consistent with return to equilibrium:

$$\text{Mental health} \leftrightarrow \max_{\boldsymbol{\mu}} H_H(\boldsymbol{\mu}) \quad \text{subject to } \mathcal{R}(\mathbf{x}_H^*; \boldsymbol{\mu}) > \mathcal{R}_{\min}. \tag{45}$$

18 Numerical Simulations

To demonstrate that the framework generates concrete, computable dynamics rather than remaining purely symbolic, we present four simulations using the core model equations.

18.1 Potential Landscape and Bifurcation Diagram

Figure 1 shows the double-well potential $V(x, \mu) = -\mu x^2/2 + x^4/4 - hx$ (Eq. 7) for decreasing values of μ (Panel A) and the corresponding bifurcation diagram (Panel B). As μ decreases under chronic stress, the healthy well (right) shallows and eventually annihilates via saddle-node bifurcation at $\mu_c \approx 0.33$ (for $h = 0.3$). The separatrix (dashed) marks the basin boundary; below μ_c, only the disorder attractor survives.

18.2 Critical Slowing Down and Integrated Scenario

Figure 2A shows a stochastic simulation of the SDE (Eq. 8) with slowly decreasing $\mu(t)$ (Eq. 9). As the system approaches the tipping point, both the rolling variance and lag-1 autocorrelation increase—the DIORs of Section 14—before the system tips into the disorder attractor.

Figure 2B presents the first integrated simulation combining three sub-models: the core SDE (Eq. 8), the social isolation dynamics (Eq. 31), and the anosognosia model (Eq. 18). Under chronic stress, the system progresses through three phases: (1) *Healthy*:

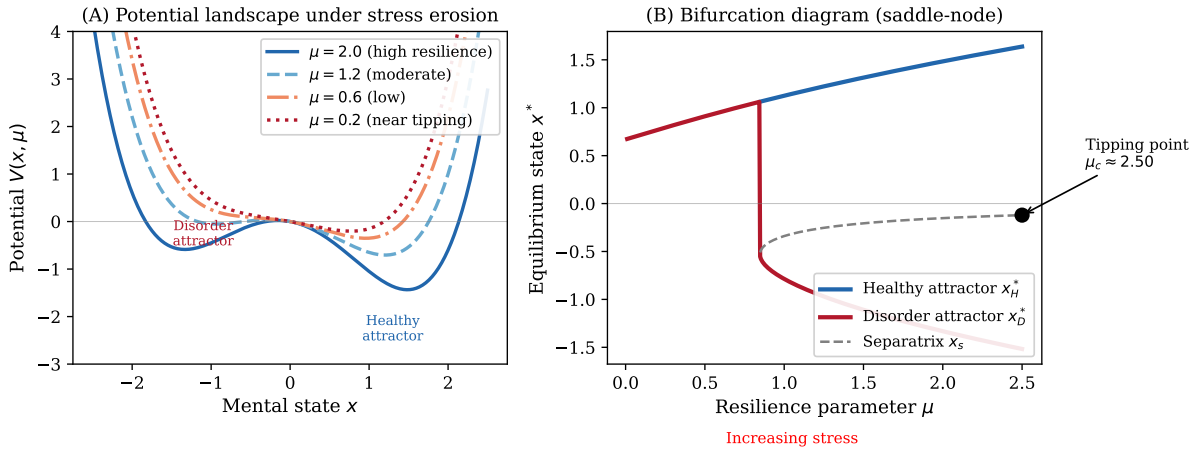

Figure 1: (A) Potential landscape $V(x, \mu)$ for four values of the resilience parameter μ, showing progressive erosion of the healthy well under stress. (B) Bifurcation diagram: stable equilibria (solid) and separatrix (dashed) as functions of μ. The tipping point occurs at the saddle-node bifurcation μ_c.

x fluctuates near x_H^*, social connectivity C and insight r remain high; (2) *Erosion*: μ decreases, C begins to drop (isolation feedback), r initially rises (the person notices something is wrong) then peaks; (3) *Disorder*: the system tips, C collapses, and insight $r \to 0$ (anosognosia). Notably, insight peaks *before* the transition and is lost *after* it—a clinically recognized pattern that the model reproduces from the interaction of its component equations.

19 Operationalization: Mapping Variables to Measurements

A critical requirement for empirical testing is the mapping from abstract state variables to concrete measurement instruments. Table 1 provides this mapping for the core variables of the framework.

20 Bifurcation-Dependent vs. Bifurcation-Independent Predictions

Not all predictions of the framework depend on the assumption that psychiatric transitions are bifurcation-driven. Table 2 classifies the key predictions by their dependence on this assumption. Bifurcation-independent predictions—which rely only on the existence of alternative attractors separated by barriers—are testable regardless of the transition mechanism.

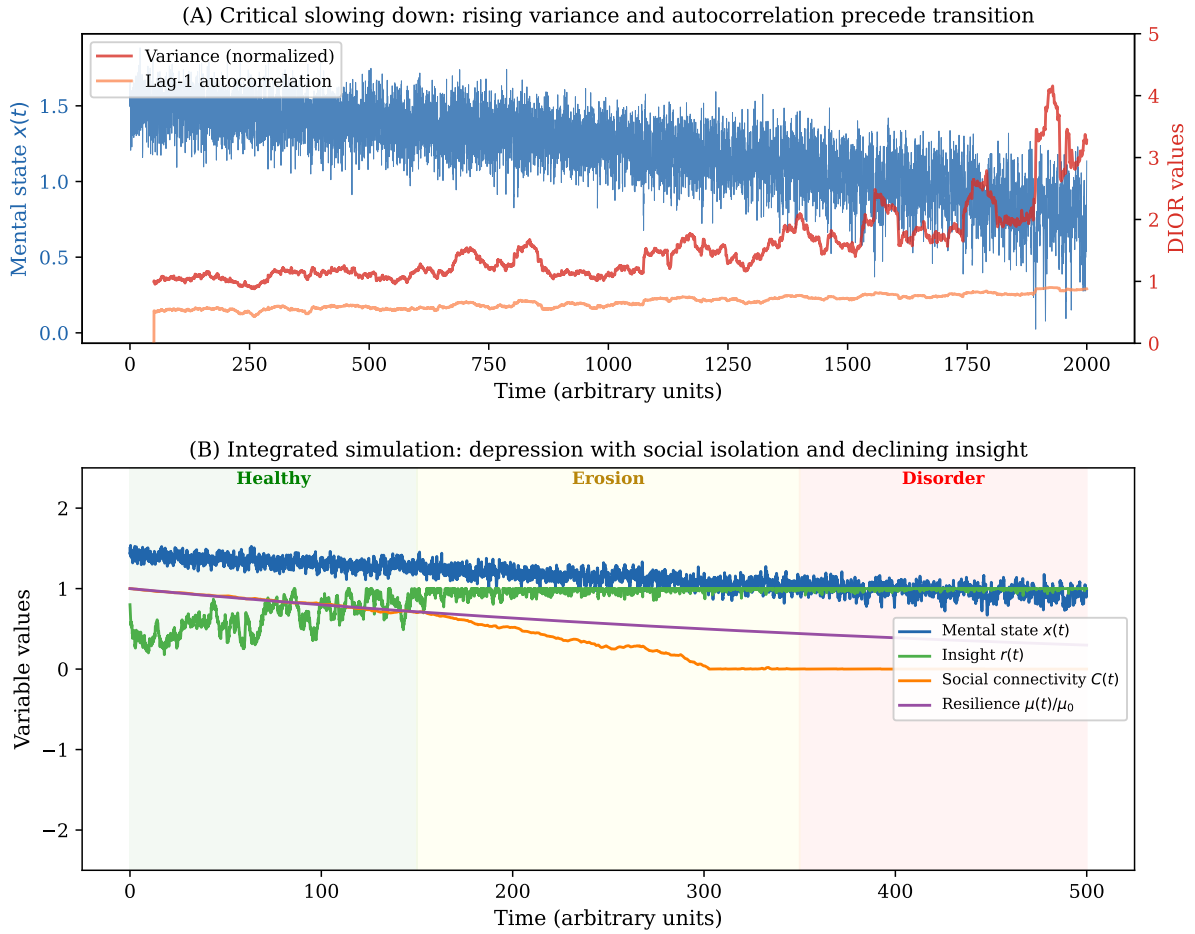

Figure 2: (A) Critical slowing down: a stochastic trajectory with slowly eroding $\mu(t)$. Rolling variance (red) and autocorrelation (orange) rise before the transition (dashed line). (B) Integrated simulation combining depression (x, blue), social connectivity (C, orange), insight (r, green), and resilience (μ/μ_0, purple) under chronic stress. The three phases (healthy, erosion, disorder) are visible.

21 Limitations and Open Problems

We acknowledge several important limitations of the present framework.

Empirical parameterization. Most models presented here have free parameters that are not yet constrained by data. The worked example in Section 4 demonstrates that parameterization is feasible for some sub-models (critical slowing down), but systematic estimation of the full parameter space—particularly for the anosognosia model (Eq. 19), the fear-of-madness model (Eq. 20), and the social connectivity model (Eq. 31)—requires dedicated empirical studies using experience-sampling methods and ecological momentary assessment.

Noise structure. We uniformly assume additive Gaussian noise (Wiener process). Real psychiatric time series often exhibit heavy tails, state-dependent volatility (variance

Variable	Meaning	EMA / Self-report	Digital phenotyping / Clinical scale
x_M (mood)	Mood valence	"How is your mood right now?" (VAS 0–100)	PHQ-9, HDRS; voice prosody analysis
x_A (anxiety)	Anxiety / threat level	"How anxious do you feel?" (VAS)	GAD-7; skin conductance, HRV
x_S (sleep)	Sleep quality	"How well did you sleep?" (VAS)	Actigraphy, smartphone screen-off duration
x_R (reality)	Reality contact	"How real does the world feel?" (VAS)	PANSS positive subscale; speech coherence (NLP)
C (social)	Social connectivity	"How many conversations did you have?"; "How supported do you feel?"	Call/text frequency; GPS-based social encounters
r (insight)	Illness awareness	"Do you think you need help?" (VAS)	SUMD (Scale to Assess Unawareness of Mental Disorder)
e (energy)	Resistance energy	"How much energy do you have?" (VAS)	Step count; psychomotor activity (accelerometry)
a (fear)	Fear of madness	"Are you worried about losing control?" (VAS)	ASI (Anxiety Sensitivity Index)
μ	Resilience	Derived: $-1/\ln(\hat{\rho}_1)$ from rolling EMA	Allostatic load biomarkers; DIOR trend
σ^2	Noise intensity	Derived: rolling variance of EMA mood	Variability of digital phenotyping signals

Table 1: Mapping of model state variables to measurement instruments. VAS = Visual Analogue Scale. EMA items are administered multiple times daily via smartphone. Digital phenotyping measures are collected passively.

that changes with symptom severity), and temporal non-stationarity. Replacing Gaussian noise with Lévy processes or state-dependent diffusion $\mathbf{\Sigma}(\mathbf{x})$ would modify the Kramers escape rates and transition predictions; these extensions are technically straightforward but complicate the analytic results.

Model integration. Although we claim a "unified framework," the sub-models (anosognosia, fear, resistance, social connectivity) are presented largely independently. Demonstrating that they combine into a well-behaved integrated simulation—e.g., a patient with depression, declining insight, social isolation, and fear of relapse simultaneously—remains a task for future numerical work.

Prediction	Requires bifurcation?	Mechanism if not bifurcation-driven
Rising autocorrelation before transition (CSD)	**Yes**	Not predicted for noise-induced escapes
Rising variance before transition	**Yes**	Not predicted (variance may stay constant)
Hysteresis (recovery harder than onset)	No	Barrier asymmetry: $\mathcal{R}_D >$ original $\mathcal{R}_H$ deficit
Self-sustaining disorder after stressor removal	No	Attractor stability via feedback loops (Prop. 1)
Transition time $\propto \exp(2\mathcal{R}/\sigma^2)$	No	Kramers' formula applies to barrier crossing generally
Sudden transition from minor trigger	**Partial**	Bifurcation: basin vanishes. Noise: rare large fluctuation
Anosognosia increases with severity	No	Distance-dependent insight (Eq. 19)
Fear of madness accelerates transition	No	Effective resilience reduction (Eq. 22)
Social isolation creates positive feedback	No	Coupled x–C dynamics (Eq. 31)
Gaslighting erodes resilience measurably	No	δ_{self} decay $\rightarrow$ DIOR signatures
Collective tipping with hysteresis	**Yes**	Mean-field bifurcation (Eq. 34)
Creativity bounded by basin width	No	Excursion $<$ basin boundary
Disorder entropy $\leq$ healthy entropy	No	Attractor constrains accessible state space

Table 2: Classification of framework predictions by dependence on the bifurcation assumption. "No" indicates the prediction holds for any transition mechanism (bifurcation or noise-induced). "Yes" indicates the prediction is specific to bifurcation-driven transitions.

Gradient vs. non-gradient dynamics. As discussed in Section 2.2, the potential-landscape metaphor breaks down for oscillatory disorders. While the quasipotential (Eq. 1) formally extends the framework, its practical computation for high-dimensional psychiatric systems is non-trivial [Kühn, 2011].

Absence of high-dimensional simulations. The numerical simulations in Section 18 demonstrate core model behavior in reduced (1–4 dimensional) systems. Full simulations of the n-dimensional symptom network (Eq. 15) with empirically estimated $\mathbf{W}$ matrices, and systematic parameter sensitivity analyses, remain tasks for future computational work.

22 Conclusion

The dynamical systems theory of psychiatric disorders, as articulated by Scheffer et al. [2024a,b], provides a conceptual framework within which the phenomena of madness—its stages, its self-reinforcing dynamics, its resistance to treatment, its relationship to creativity and culture, its vulnerability to social forces—acquire new explanatory precision. The present article has extended this framework in two directions: providing the phenomenological, clinical, and socio-cultural content that the original theory lacks, and developing a unified mathematical formalism that enables modeling and simulation.

The principal novel contributions include:

1. **Anosognosia as an intrinsic attractor property** (Eq. 18–19): the inability to recognize illness as a necessary consequence of distance-dependent insight decay within a disorder attractor.

2. **The fear of madness as an autocatalytic destabilizing loop** (Eq. 20–22): a positive feedback that erodes resilience without external stressor.

3. **Internal resistance as a restoring force with energy depletion** (Eq. 23–24): specific cognitive operations formalized as restoring forces that exhaust over time.

4. **Altered time–space perception as attractor signatures** (Eq. 25–27): intrinsic dynamical features of the disorder attractor.

5. **Creativity as controlled basin-boundary exploration** (Eq. 28–29): a precise formalization of the creativity–madness distinction.

6. **Gaslighting as targeted resilience destruction** (Eq. 32): systematic attack on self-trust modeled as decay of the restoring-force coefficient.

7. **Collective madness as societal phase transition** (Eq. 33–34): DST extended from individual to population dynamics with conformity coupling.

8. **Punitive psychiatry as state-space misclassification and iatrogenic destruction** (Eq. 39).

9. **DST typology of suicidal dynamics** (Eq. 41–43): three distinct dynamical scenarios for a single clinical outcome.

10. **Entropy-based refutation of romanticism** (Eq. 44–45): $H_D < H_H$ demonstrates that the disorder attractor constrains rather than liberates.

Together, these equations form a modular, extensible mathematical framework that can be instantiated with empirical parameter values, simulated numerically, and used to generate testable predictions about the dynamics of mental health and mental illness.

The fragility of the human mind is not a defect to be lamented but a structural property of a dynamical system complex enough to think, feel, and create. The price of that complexity is the permanent possibility of tipping into states from which return is difficult. Awareness of this fragility—not as a philosophical abstraction but as a mathematically precise dynamical reality—is the first step toward its management.

Acknowledgements

The author thanks Professor Marten Scheffer (Wageningen University) for encouraging correspondence regarding this work.

References

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Front Psychiatry*, 4, 47.

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13.

Bos, F. M., et al. (2022). Anticipating manic and depressive transitions using early warning signals. *Int J Bipolar Disord*, 10(1), 12.

Briganti, G. (2024). Network analysis: An overview for mental health research. *Int J Methods Psychiatr Res*, 33, e2034.

Bringmann, L. F., et al. (2022). Psychopathological networks: Theory, methods and practice. *Behav Res Ther*, 149, 104011.

Cheng, Y., et al. (2025). Computational modeling of reward hyposensitivity in depression. *Comput Psychiatry*, 9(1).

Dakos, V., Scheffer, M., van Nes, E. H., et al. (2008). Slowing down as an early warning signal for abrupt climate change. *PNAS*, 105(38), 14308–14312.

Foucault, M. (1961). *Madness and Civilization*. Pantheon Books.

Freidlin, M. I. & Wentzell, A. D. (1998). *Random Perturbations of Dynamical Systems* (2nd ed.). Springer.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neurosci*, 11(2), 127–138.

Friston, K. J. (2022). Computational psychiatry: From synapses to sentience. *Mol Psychiatry*, 28, 256–268.

Helmich, M. A., et al. (2024). Slow down and be critical before using early warning signals in psychopathology. *Nat Rev Psychol*, 3, 817–828.

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci*, 19(3), 404–413.

Kerr, J., et al. (2026). Computational modeling of remote fear conditioning and anxiety. *Comput Psychiatry*, 10(1).

Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4), 284–304.

Kühn, C. (2011). A mathematical framework for critical transitions. *Physica D*, 240(12), 1020–1035.

Legemaat, A. M., et al. (2023). Effects up to 20-year follow-up of preventive cognitive therapy in remitted depression. *Psychother Psychosom*, 92(1), 55–64.

McEwen, B. S. (2007). Physiology and neurobiology of stress and adaptation. *Physiol Rev*, 87(3), 873–904.

Nesse, R. M. (2000). Is depression an adaptation? *Arch Gen Psychiatry*, 57(1), 14–20.

Redish, A. D. (2004). Addiction as a computational process gone awry. *Science*, 306(5703), 1944–1947.

Scheffer, M., et al. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260), 53–59.

Scheffer, M., et al. (2012). Anticipating critical transitions. *Science*, 338(6105), 344–348.

Scheffer, M., et al. (2018). Quantifying resilience of humans and other animals. *PNAS*, 115(47), 11883–11890.

Scheffer, M., Bockting, C. L., Borsboom, D., et al. (2024a). A dynamical systems view of psychiatric disorders—Theory. *JAMA Psychiatry*, 81(6), 618–623.

Scheffer, M., Bockting, C. L., Borsboom, D., et al. (2024b). A dynamical systems view of psychiatric disorders—Practical implications. *JAMA Psychiatry*, 81(6), 624–630.

Smit, A. C., et al. (2025). Critical slowing down in momentary affect as early warning signal of impending transitions in depression. *Clin Psychol Sci*, (in press).

Stephan, K. E., et al. (2017). Computational approaches to psychiatry. *Curr Opin Neurobiol*, 44, 56–61.

Sterzer, P., Adams, R. A., Fletcher, P., et al. (2018). The predictive coding account of psychosis. *Biol Psychiatry*, 84(9), 634–643.

Strogatz, S. H. (2015). *Nonlinear Dynamics and Chaos* (2nd ed.). Westview Press.

van de Leemput, I. A., et al. (2014). Critical slowing down as early warning for depression. *PNAS*, 111(1), 87–92.

Wichers, M., Groot, P. C., et al. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychother Psychosom*, 85(2), 114–116.

Appendices: Application of the Dynamical Systems Formalism to DSM-5-TR Diagnostic Groups

The following appendices demonstrate how the unified mathematical framework developed in the main text applies to specific diagnostic categories of the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision* (DSM-5-TR; American Psychiatric Association, 2022). For each diagnostic group, we: (i) identify the relevant state variables x_i; (ii) specify the feedback structure (adjacency matrix $\mathbf{W}$); (iii) characterize the attractor topology; (iv) derive observable predictions (DIORs, transition signatures); and (v) formulate the therapeutic implications in dynamical terms. All notation follows the conventions established in Section 2.4.

A Schizophrenia Spectrum and Other Psychotic Disorders

A.1 State Variables

The psychotic state vector comprises:

$$\mathbf{x}_{\text{psy}} = (x_{\text{hal}},\ x_{\text{del}},\ x_{\text{dis}},\ x_{\text{neg}},\ x_{\text{cog}},\ x_{\text{soc}})^{\top}, \tag{A.1}$$

where x_{hal} = hallucination intensity, x_{del} = delusional conviction, x_{dis} = disorganized thought/speech, x_{neg} = negative symptoms (avolition, anhedonia, alogia), x_{cog} = cognitive impairment, x_{soc} = social withdrawal.

A.2 Feedback Architecture

The adjacency matrix encodes the characteristic positive feedback loops:

$$\mathbf{W}_{\text{psy}} = \begin{pmatrix} 0 & w_{12} & w_{13} & 0 & w_{15} & 0 \\ w_{21} & 0 & w_{23} & 0 & 0 & w_{26} \\ w_{31} & w_{32} & 0 & 0 & w_{35} & 0 \\ 0 & 0 & 0 & 0 & w_{45} & w_{46} \\ w_{51} & 0 & w_{53} & w_{54} & 0 & w_{56} \\ 0 & w_{62} & 0 & w_{64} & w_{65} & 0 \end{pmatrix}, \tag{A.2}$$

where the key self-reinforcing loops are:

- **Positive symptom loop:** $x_{\text{hal}} \xrightarrow{w_{21}} x_{\text{del}} \xrightarrow{w_{12}} x_{\text{hal}}$ (hallucinations generate delusional explanations, which prime further hallucinations).
- **Cognitive–positive loop:** $x_{\text{cog}} \xrightarrow{w_{15}} x_{\text{hal}}$ (cognitive fragmentation lowers reality-testing threshold).
- **Negative–social loop:** $x_{\text{neg}} \xrightarrow{w_{46}} x_{\text{soc}} \xrightarrow{w_{64}} x_{\text{neg}}$ (amotivation drives withdrawal, isolation deepens negative symptoms).
- **Social isolation–delusion loop:** $x_{\text{soc}} \xrightarrow{w_{26}} x_{\text{del}}$ (loss of corrective social feedback stabilizes delusional system).

A.3 Attractor Topology and Anosognosia

Schizophrenia is characterized by a *deep disorder attractor* with high $\mathcal{R}_D$ and severe anosognosia. Using the insight model (Eq. 18–19), the distance $\|\mathbf{x}^*_{\text{psy}} - \mathbf{x}^*_H\|$ is typically very large, placing the system well beyond the insight horizon δ:

$$r^*_{\text{schiz}} = \frac{\phi_0 \|\mathbf{x}^*_{\text{psy}} - \mathbf{x}^*_H\| \exp(-\|\mathbf{x}^*_{\text{psy}} - \mathbf{x}^*_H\|^2 / 2\delta^2)}{\alpha_r} \approx 0. \tag{A.3}$$

The prodromal phase (attenuated psychosis syndrome) corresponds to the pre-transition flickering described in Section 4: increasing variance in perceptual anomalies, rising autocorrelation in suspiciousness ratings, and transient excursions toward the psychotic attractor.

A.4 Therapeutic Implications

Antipsychotic medication acts primarily on the x_{hal} and x_{del} components by reducing the effective weights w_{ij} in the positive symptom loop (dopamine D2 receptor blockade reduces aberrant salience signaling):

$$w^{\text{treated}}_{ij} = w_{ij}\,(1 - \eta_{\text{AP}}), \qquad i, j \in \{\text{hal}, \text{del}\}, \tag{A.4}$$

where $\eta_{\text{AP}} \in [0, 1]$ is antipsychotic efficacy. When $\rho(\mathbf{A}^{-1}\mathbf{W}^{\text{treated}}\beta/4) < 1$, the disorder attractor ceases to exist and the system can return to the healthy state. However, negative symptoms (x_{neg}) and cognitive deficits (x_{cog}) are largely unaffected, explaining the persistent functional impairment.

B Bipolar and Related Disorders

B.1 State Variables and Oscillatory Dynamics

Note: Bipolar disorder involves inherently non-gradient (oscillatory) dynamics. The potential landscape used below (Eq. B.4) should be understood as a quasipotential $\tilde{V}$ in the sense of Section 2.2; the Kramers escape rates remain valid under Freidlin–Wentzell theory [Freidlin & Wentzell, 1998, Kühn, 2011].

Bipolar disorder is unique among psychiatric conditions in exhibiting *oscillatory* dynamics rather than a simple bistable switch. The core state vector is:

$$\mathbf{x}_{\text{bp}} = (m(t),\ e(t),\ s(t),\ a(t))^{\top}, \tag{B.1}$$

where m = mood valence ($m > 0$: mania; $m < 0$: depression), e = energy/activity level, s = sleep (hours/quality), a = goal-directed activity. The dynamics are:

$$\frac{dm}{dt} = \mu\, m - m^3 + \alpha_e\, e - \alpha_s\,(s - s_0) + \sigma_m\, \xi_m(t), \tag{B.2}$$

$$\frac{de}{dt} = -\gamma_e\, e + \beta_m\, m - \beta_s\,(s - s_0) + \sigma_e\, \xi_e(t), \tag{B.3}$$

$$\frac{ds}{dt} = -\gamma_s\,(s - s_0) - \delta_m\,|m| - \delta_e\, e + \sigma_s\, \xi_s(t). \tag{B.4}$$

B.2 Three-Attractor Landscape

Unlike the two-attractor model of most disorders, bipolar disorder features a *three-well potential*: healthy euthymia ($m \approx 0$), mania ($m \gg 0$), and depression ($m \ll 0$). The potential takes the form:

$$V_{\text{bp}}(m) = \frac{1}{6}\, m^6 - \frac{\mu}{4}\, m^4 + \frac{\nu}{2}\, m^2 - h\, m, \tag{B.5}$$

where the sixth-order term ensures boundedness, μ controls the depth of manic/depressive wells, ν controls the euthymic well, and h is an asymmetry parameter (bias toward depression or mania).

Rapid cycling (≥ 4 episodes/year) corresponds to a parameter regime where the euthymic well is very shallow (ν small), so that noise-driven transitions between the manic and depressive wells are frequent. The Kramers escape time between wells gives the predicted cycling frequency:

$$f_{\text{cycle}} \approx \frac{1}{\tau_{\text{mania}\to\text{dep}} + \tau_{\text{dep}\to\text{mania}}} = \frac{1}{\tau_+ + \tau_-}. \tag{B.6}$$

B.3 Sleep as the Critical Feedback Variable

The sleep variable s acts as a *bridge symptom* between mania and depression: sleep deprivation ($s \downarrow$) destabilizes the euthymic state and facilitates transition to mania (consistent with the well-documented clinical phenomenon of sleep-loss-induced mania). The sleep feedback creates a characteristic asymmetry in transition rates: the manic transition is often faster ($\tau_+ < \tau_-$) because mania itself reduces sleep, creating an autocatalytic loop absent in the depressive direction.

B.4 Mood Stabilizer Dynamics

Lithium and other mood stabilizers deepen the euthymic well (increase ν) and raise the barriers between all three states:

$$V_{\text{stabilized}}(m) = V_{\text{bp}}(m) + \Lambda_{\text{Li}} \exp\left(-\frac{m^2}{2\sigma_{\text{Li}}^2}\right). \tag{B.7}$$

C Depressive Disorders

C.1 State Variables

$$\mathbf{x}_{\text{dep}} = (x_{\text{mood}},\ x_{\text{anh}},\ x_{\text{slp}},\ x_{\text{fat}},\ x_{\text{cog}},\ x_{\text{gui}},\ x_{\text{app}},\ x_{\text{psy}},\ x_{\text{sui}})^\top, \tag{C.1}$$

where the components correspond to DSM-5-TR Criterion A/B symptoms: depressed mood, anhedonia, sleep disturbance, fatigue, cognitive impairment (indecisiveness, concentration), guilt/worthlessness, appetite change, psychomotor change, and suicidal ideation.

C.2 The Depression Feedback Network

Major depressive disorder (MDD) is the paradigmatic case for the network model (Eq. 15). The self-reinforcing loops identified in empirical studies include:

$$x_{\text{mood}} \to x_{\text{gui}} \to x_{\text{mood}} \qquad \text{(mood–guilt loop)}, \tag{C.2}$$

$$x_{\text{slp}} \to x_{\text{fat}} \to x_{\text{cog}} \to x_{\text{mood}} \to x_{\text{slp}} \qquad \text{(somatic–cognitive spiral)}, \tag{C.3}$$

$$x_{\text{anh}} \to x_{\text{soc}} \to x_{\text{mood}} \to x_{\text{anh}} \qquad \text{(anhedonia–isolation loop)}. \tag{C.4}$$

C.3 Critical Slowing Down as a Clinical Biomarker

Depression is the disorder for which DIORs have been most extensively validated. Following van de Leemput et al. [2014], the lag-1 autocorrelation of daily mood ratings increases in the months preceding a depressive episode:

$$\hat{\rho}_1(t) \approx \exp\big(\lambda(\mu(t))\,\Delta t\big) \xrightarrow{\mu \to \mu_c} 1. \tag{C.5}$$

The variance simultaneously increases: $\hat{\sigma}^2(t) \approx \sigma^2/(2|\lambda(\mu(t))|) \to \infty$. These are computable from ecological momentary assessment (EMA) data.

C.4 Recurrence and Hysteresis

The high recurrence rate of MDD (50–80% after first episode) is explained by hysteresis: each depressive episode may leave residual changes in the network weights w_{ij} ("scarring"), making subsequent transitions easier:

$$w_{ij}^{(k+1)} = w_{ij}^{(k)} + \Delta w_{ij} \cdot \mathbf{1}_{\{k\text{-th episode occurred}\}}. \tag{C.6}$$

This progressive lowering of the barrier $\mathcal{R}$ after each episode formalizes the clinical observation that later episodes require less stress to trigger (kindling).

C.5 Prolonged Grief Disorder (New in DSM-5-TR)

Prolonged grief disorder, added in DSM-5-TR, is modeled as an attractor with a specific symptom structure centered on yearning (x_{yearn}) and preoccupation (x_{preocc}) with the deceased, with the key feedback loop:

$$x_{\text{yearn}} \to x_{\text{preocc}} \to x_{\text{avoid}} \to x_{\text{isol}} \to x_{\text{yearn}}. \tag{C.7}$$

The 12-month criterion (DSM-5-TR) corresponds to the system having settled into a stable attractor rather than undergoing a transient excursion (normal grief $\approx$ large perturbation with eventual return; prolonged grief $\approx$ transition to alternative attractor).

D Anxiety Disorders

D.1 State Variables

$$\mathbf{x}_{\text{anx}} = (x_{\text{fear}},\ x_{\text{arous}},\ x_{\text{avoid}},\ x_{\text{vig}},\ x_{\text{rumi}},\ x_{\text{som}})^\top, \tag{D.1}$$

representing fear/apprehension, physiological arousal, avoidance behavior, hypervigilance, anxious rumination, and somatic symptoms.

D.2 Generalized Anxiety Disorder (GAD)

GAD is characterized by a *shallow, broad disorder attractor*: the system does not transition to an extreme state (as in psychosis) but rather hovers in a chronically elevated region of state space. The potential for GAD has a very flat healthy minimum:

$$V_{\text{GAD}}(x) = V(x, \mu) + \eta_{\text{worry}}\, x \quad (\eta_{\text{worry}} > 0), \tag{D.2}$$

where the constant tilt η_{worry} represents the tonic elevation of the threat-detection system. The system never fully returns to the bottom of the healthy well; it oscillates in a region of elevated arousal.

Worry as a destabilizing loop: using the fear model (Eq. 20), GAD is the chronic activation of the metacognitive anxiety variable $a(t)$ without discrete tipping events—an orbit trapped in the fear-amplification regime.

D.3 Panic Disorder

Panic attacks are modeled as *fast transient excursions*—noise-driven escapes from the healthy basin that do not result in permanent trapping in the disorder attractor, but rather in brief visits to an extreme arousal state followed by return:

$$\text{Panic attack: } x(t) \text{ crosses } x_s, \text{ enters region near } x_D^*, \text{ returns within minutes.} \tag{D.3}$$

The return is possible because the arousal attractor is shallow ($\mathcal{R}_D$ small for acute panic). Panic *disorder* emerges when the anticipatory anxiety (fear-of-fear) loop (Eq. 20) chronically reduces $\mathcal{R}_H$, making excursions increasingly frequent:

$$\frac{d\mathcal{R}_H}{dt} = -\kappa\,\bar{a} < 0 \quad \Longrightarrow \quad \tau_{\text{escape}} \downarrow \quad \Longrightarrow \quad \text{panic frequency} \uparrow . \tag{D.4}$$

D.4 Social Anxiety Disorder

Social anxiety features a critical *avoidance loop* that prevents exposure to corrective feedback:

$$x_{\text{fear}} \to x_{\text{avoid}} \to -\Delta C \to x_{\text{fear}}, \tag{D.5}$$

where $\Delta C < 0$ represents the reduction in social connectivity due to avoidance. Using the isolation model (Eq. 31), social anxiety is a special case where the avoidance is driven by anticipatory fear rather than by the worsening of the mental state itself.

Exposure therapy is formalized as forced perturbation into the feared region of state space:

$$x_{\text{avoid}}(t) \mapsto x_{\text{avoid}}(t) - p_{\text{exposure}}, \qquad C(t) \mapsto C(t) + \Delta C_{\text{exposure}}, \tag{D.6}$$

which allows the system to discover that the feared attractor does not exist (the catastrophic outcome does not occur), thereby restoring C and deepening $\mathcal{R}_H$.

E Trauma- and Stressor-Related Disorders

E.1 PTSD as a Trauma-Induced Alternative Attractor

Post-traumatic stress disorder represents a case where a single massive perturbation $\mathbf{p}_{\text{trauma}}$ pushes the system across the basin boundary:

$$\mathbf{x}(t_0^+) = \mathbf{x}(t_0^-) + \mathbf{p}_{\text{trauma}}, \qquad \|\mathbf{p}_{\text{trauma}}\| > d(\mathbf{x}(t_0^-), \partial\mathcal{B}_H). \tag{E.1}$$

Unlike the slow erosion model of depression, the pre-trauma resilience may have been high; the perturbation was simply larger than the basin width.

E.2 State Variables

$$\mathbf{x}_{\text{PTSD}} = (x_{\text{intr}},\ x_{\text{avoid}},\ x_{\text{arous}},\ x_{\text{cog}},\ x_{\text{react}})^\top, \tag{E.2}$$

corresponding to the four DSM-5-TR PTSD symptom clusters: intrusion (x_{intr}: flashbacks, nightmares), avoidance (x_{avoid}), negative alterations in cognition and mood (x_{cog}), and arousal/reactivity (x_{arous}, x_{react}).

E.3 Feedback Loops

The PTSD attractor is maintained by:

$$x_{\text{intr}} \to x_{\text{arous}} \to x_{\text{avoid}} \to x_{\text{intr}} \qquad \text{(intrusion–avoidance cycle)}, \tag{E.3}$$

$$x_{\text{cog}} \to x_{\text{arous}} \to x_{\text{react}} \to x_{\text{cog}} \qquad \text{(cognition–reactivity loop)}, \tag{E.4}$$

$$x_{\text{avoid}} \to -\Delta C \to x_{\text{cog}} \qquad \text{(avoidance prevents corrective processing)}. \tag{E.5}$$

A critical feature of PTSD is the *memory reconsolidation window*: during intrusive re-experiencing, the traumatic memory becomes temporarily labile (the system transiently exits the PTSD attractor). This creates a therapeutic opportunity:

$$\text{During flashback: } \mathbf{x}(t) \approx \mathbf{x}_s \quad \Longrightarrow \quad \text{intervention can redirect } \mathbf{x} \to \mathcal{B}_H. \tag{E.6}$$

This is the dynamical basis of exposure-based therapies (PE, EMDR, CPT): they exploit the reconsolidation window to weaken the feedback loops that maintain the PTSD attractor.

E.4 Adjustment Disorders

Adjustment disorders (including the new course specifiers in DSM-5-TR) represent *transient excursions* that are larger and longer-lasting than normal stress responses but do

not result in stable entrapment:

$$x(t) \in \mathcal{B}_H \text{ for all } t, \quad \text{but } |x(t) - x_H^*| \gg \text{typical fluctuations for } t \in [t_{\text{stress}}, t_{\text{stress}} + \Delta T]. \tag{E.7}$$

F Obsessive-Compulsive and Related Disorders

F.1 The OCD Loop as a Limit Cycle

Unlike most disorders modeled as point attractors, obsessive-compulsive disorder is better modeled as a *limit cycle*—a periodic orbit in state space. The obsession–compulsion cycle is:

$$\frac{dx_{\text{obs}}}{dt} = -\alpha_o\, x_{\text{obs}} + w_{\text{co}}\, g(x_{\text{comp}}) + \theta_{\text{trigger}} + \sigma_o\, \xi_o, \tag{F.1}$$

$$\frac{dx_{\text{comp}}}{dt} = -\alpha_c\, x_{\text{comp}} + w_{\text{oc}}\, g(x_{\text{obs}}) - \gamma_{\text{relief}}\, x_{\text{comp}}^2. \tag{F.2}$$

Here x_{obs} is obsessional distress and x_{comp} is compulsive urge/behavior. The key features are:

- Obsessions drive compulsions ($w_{\text{oc}} > 0$).
- Compulsions temporarily reduce obsessional distress through negative reinforcement ($w_{\text{co}} < 0$ at short timescales) but *strengthen* the obsession–compulsion coupling at longer timescales ($w_{\text{co}} > 0$ in the g function, capturing sensitization).
- The $\gamma_{\text{relief}}\, x_{\text{comp}}^2$ term captures the diminishing relief from compulsions as they escalate.

This system admits a Hopf bifurcation: for $w_{\text{oc}}\, w_{\text{co}} > \alpha_o\, \alpha_c$, the equilibrium becomes unstable and a stable limit cycle emerges—the person oscillates endlessly between rising obsessional anxiety and temporary compulsive relief.

F.2 Insight Specifier

DSM-5-TR specifies OCD insight levels: good/fair, poor, and absent/delusional. Using the anosognosia model (Eq. 18), the insight level maps to the parameter δ:

$$\text{Good insight: } \delta \gg d, \qquad \text{Poor insight: } \delta \approx d, \qquad \text{Absent insight: } \delta \ll d. \tag{F.3}$$

F.3 Exposure and Response Prevention (ERP)

ERP is modeled as preventing the compulsive reduction (x_{comp} clamped at 0) while allowing x_{obs} to evolve freely:

$$\left.\frac{dx_{\text{obs}}}{dt}\right|_{x_{\text{comp}}=0} = -\alpha_o\, x_{\text{obs}} + \theta_{\text{trigger}}. \tag{F.4}$$

Without compulsive reinforcement, $x_{\text{obs}} \to \theta_{\text{trigger}}/\alpha_o$ (habituation). Repeated ERP sessions weaken w_{oc} until the Hopf bifurcation condition is no longer met, and the limit cycle collapses back to the healthy equilibrium.

G Substance-Related and Addictive Disorders

G.1 State Variables

$$\mathbf{x}_{\text{add}} = (x_{\text{crav}},\ x_{\text{use}},\ x_{\text{tol}},\ x_{\text{with}},\ x_{\text{ctrl}},\ x_{\text{neg}})^\top, \tag{G.1}$$

where the components are craving, substance use, tolerance, withdrawal, loss of control, and negative consequences.

G.2 The Addiction Potential

Addiction involves a progressive deepening of the disorder attractor through *neuroadaptation*. The key dynamical feature is that the landscape itself changes as a function of cumulative exposure:

$$V_{\text{add}}(x, \mu, H) = V(x, \mu) - \Lambda_{\text{add}}(H)\, \exp\!\left(-\frac{(x - x_D^*)^2}{2\sigma_D^2}\right), \tag{G.2}$$

where $H(t) = \int_0^t x_{\text{use}}(\tau)\, d\tau$ is cumulative substance exposure and $\Lambda_{\text{add}}(H)$ is a monotonically increasing function. With each use, the disorder well deepens:

$$\frac{d\Lambda_{\text{add}}}{dt} = \kappa_{\text{add}}\, x_{\text{use}}(t) - \gamma_{\text{add}}\, \Lambda_{\text{add}}. \tag{G.3}$$

This means that recovery requires not only exiting the attractor but also waiting for the landscape to flatten (Λ_{add} to decay)—the neurobiological basis of the prolonged vulnerability to relapse.

G.3 Tolerance and Withdrawal

Tolerance (x_{tol}) is an allostatic adjustment that shifts the healthy baseline:

$$x^*_{H,\text{eff}} = x^*_H - \tau_{\text{tol}}\, x_{\text{tol}}, \tag{G.4}$$

so that the person needs substance use just to reach the (new) baseline. Withdrawal is the transient state when $x_{\text{use}} \to 0$ but x_{tol} remains high: the effective healthy state is displaced, producing distress:

$$x_{\text{with}} = x_{\text{tol}} \cdot (1 - x_{\text{use}}/x_{\text{use,prev}}). \tag{G.5}$$

G.4 DSM-5-TR Severity Mapping

DSM-5-TR severity specifiers (mild: 2–3 criteria; moderate: 4–5; severe: 6+) correspond to the number of activated feedback loops. In our framework:

$$\text{Severity} \propto \#\{(i,j) : x_i > \theta_i \text{ and } w_{ij} > 0 \text{ and } x_j > \theta_j\}, \tag{G.6}$$

where θ_i are activation thresholds. Higher severity = more active loops = deeper attractor = harder recovery.

H Neurocognitive Disorders

H.1 Irreversible Landscape Degradation

Neurocognitive disorders (Alzheimer's disease, frontotemporal degeneration, Lewy body disease, vascular NCD, etc.) differ fundamentally from the disorders discussed above: they involve *irreversible degradation of the potential landscape itself* rather than transitions between pre-existing attractors. The control parameter μ decreases monotonically due to neurodegeneration:

$$\frac{d\mu}{dt} = -\epsilon_{\text{neuro}}(t), \qquad \epsilon_{\text{neuro}} > 0, \qquad \frac{d\epsilon_{\text{neuro}}}{dt} \geq 0. \tag{H.1}$$

The acceleration of decline ($d\epsilon_{\text{neuro}}/dt > 0$) captures the progressive, often exponential course of neurodegeneration.

H.2 State Variables

$$\mathbf{x}_{\text{NCD}} = (x_{\text{mem}},\ x_{\text{exec}},\ x_{\text{att}},\ x_{\text{lang}},\ x_{\text{perc}},\ x_{\text{soc-cog}})^\top, \tag{H.2}$$

corresponding to the six DSM-5-TR neurocognitive domains: learning/memory, executive function, attention (complex attention), language, perceptual-motor, and social cognition.

H.3 Mild vs. Major NCD

The DSM-5-TR distinction between mild and major NCD maps onto the resilience framework:

$$\text{Mild NCD: } \mathcal{R}_H > 0 \text{ (reduced but nonzero)}, \qquad \text{Major NCD: } \mathcal{R}_H \approx 0 \text{ or } \mu < \mu_c. \tag{H.3}$$

In mild NCD, the system retains some capacity to compensate (the healthy attractor still exists, though shallow). In major NCD, the healthy attractor has been destroyed by neurodegeneration; the system operates in a progressively degrading landscape with no stable equilibrium.

H.4 Behavioral Symptoms as Secondary Transitions

The behavioral and psychological symptoms of dementia (BPSD)—agitation, psychosis, depression, apathy—represent secondary transitions within the degraded landscape. As μ decreases, new disorder attractors may appear (or existing ones may become accessible) that were previously blocked by high resilience:

$$\mu < \mu_c^{(k)} \quad \Longrightarrow \quad k\text{-th behavioral attractor becomes accessible.} \tag{H.4}$$

This explains the stage-dependent emergence of different behavioral symptoms in dementia.

I Personality Disorders

I.1 Personality as Attractor Topology

Personality disorders differ from Axis I disorders (in historical terminology) in that they represent *stable configurations of the entire attractor landscape* rather than transitions between attractors. A personality disorder is not a state the system has fallen into; it is a landscape in which the "healthy" attractor has an unusual topology—too narrow, too rigid, too sensitive to specific perturbations.

In our framework, personality disorders are modeled through the parameters $\boldsymbol{\mu}$ and $\mathbf{W}$ rather than through the state $\mathbf{x}$:

$$\text{Personality disorder} \Leftrightarrow \boldsymbol{\mu} \in \mathcal{P}_D, \quad \text{where } \mathcal{P}_D \subset \mathbb{R}^m \text{ is the "disordered parameter" region.} \tag{I.1}$$

I.2 Borderline Personality Disorder (BPD)

BPD is characterized by a *shallow, narrow basin of attraction with extreme sensitivity to interpersonal perturbations*. The effective potential is:

$$V_{\text{BPD}}(x, \mu_{\text{BPD}}) = V(x, \mu_{\text{BPD}}) \quad \text{with } \mu_{\text{BPD}} \text{ close to } \mu_c. \tag{I.2}$$

The system is chronically near the tipping point. This explains:

- *Emotional instability*: high variance $\hat{\sigma}^2 \sim 1/|\lambda|$ because $|\lambda|$ is small.
- *Impulsive behavior*: frequent boundary crossings (low τ_{escape}).
- *Identity disturbance*: the shallow basin does not provide a strong "attractor" for self-concept.
- *Fear of abandonment*: interpersonal perturbations (ΔC) are disproportionately destabilizing when $\mathcal{R}_H$ is already small.

The splitting phenomenon (idealization/devaluation) is modeled as rapid oscillation between two shallow interpersonal attractors:

$$V_{\text{split}}(x_{\text{rel}}) = -\frac{\mu_{\text{rel}}}{2} x_{\text{rel}}^2 + \frac{1}{4} x_{\text{rel}}^4, \qquad \mu_{\text{rel}} > 0 \text{ but small}, \tag{I.3}$$

where $x_{\text{rel}} > 0$ = idealization and $x_{\text{rel}} < 0$ = devaluation.

I.3 Antisocial Personality Disorder

ASPD features a qualitatively different landscape: the empathy and consequence-sensitivity dimensions have *flat potentials* (no restoring force toward prosocial behavior):

$$\left.\frac{\partial V}{\partial x_{\text{empathy}}}\right|_{\text{ASPD}} \approx 0, \qquad \left.\frac{\partial V}{\partial x_{\text{conseq}}}\right|_{\text{ASPD}} \approx 0. \tag{I.4}$$

This means that perturbations in these dimensions are neither corrected nor amplified—the system is indifferent to empathic and consequential feedback that normally maintains prosocial behavior.

I.4 The Alternative DSM-5 Model (AMPD) and Dimensionality

The Alternative Model for Personality Disorders in DSM-5-TR Section III naturally aligns with our framework: it replaces categorical diagnosis with dimensional assessment of (i) personality functioning (self and interpersonal) and (ii) pathological personality traits.

In our formalism:

$$\text{AMPD Level of Functioning} \leftrightarrow \mathcal{R}_H(\boldsymbol{\mu}), \qquad \text{AMPD Trait Domains} \leftrightarrow \boldsymbol{\mu} \text{ themselves.} \tag{I.5}$$

Summary. The eight appendices demonstrate that the unified mathematical framework developed in the main text—the core SDE, the network model, the attractor topology, the resilience measure, and the specialized models for anosognosia, fear, resistance, and social dynamics—can be systematically instantiated for every major diagnostic group in the DSM-5-TR. The framework accommodates the full range of psychiatric phenomenology: stable point attractors (depression, schizophrenia), limit cycles (OCD, bipolar disorder), landscape degradation (neurocognitive disorders), and parameter-level pathology (personality disorders). In each case, the formalism generates specific, testable predictions about observable dynamics (DIORs, transition timing, cycling frequency), treatment mechanisms (which weights or parameters the intervention targets), and prognosis (depth of disorder attractor, number of active feedback loops).