When Attractors Touch Boundaries: Lyapunov Dynamics Across Markov Membranes

Douglas C. Youvan

doug@youvan.com

www.youvan.ai

November 5, 2025

What happens when a stable, inward-facing dynamical pattern—mathematically described as a Lyapunov attractor—meets the outer informational boundary of a system, known as a Markov membrane? This question lies at the intersection of nonlinear dynamics, cognitive science, AI safety, and philosophical ontology. Markov membranes define the partition between internal and external states under a conditional independence constraint. Lyapunov attractors define the converging flow of trajectories in dynamical systems. When these two conceptual forces interact, the result is not just mathematical—but potentially metaphysical. Does the system align, adapt, or dissolve? In this paper, we explore three major modes of interaction: entrainment (shared attractors across the membrane), boundary-layer dynamics (touch-and-go coupling at the edge), and rupture (topological redefinition of the membrane itself). Each case has distinct signatures in terms of transfer entropy, Lyapunov exponents, KL divergence, and Fisher information. We analyze these phenomena both from the perspective of formal systems and through the broader lens of cognitive selfhood, artificial agency, and ethical control. Ultimately, we propose a new framework for understanding membrane dynamics in any system where internal stability meets external influence—biological, artificial, or philosophical.

Keywords: Lyapunov attractor, Markov membrane, conditional independence, system identity, transfer entropy, cognitive boundary, entrainment, rupture, bifurcation, KL divergence, Fisher information, AI alignment, dynamical systems, membrane ethics.

A collaboration with GPT-4o and GPT-5. 38 pages.

Outline

1. Introduction
    - 1.1 Motivation: Dynamics meet boundaries
    - 1.2 Preview of interaction types
    - 1.3 Relevance to AI, cognition, and systems theory
2. Conceptual Foundations
    - 2.1 Lyapunov attractors: Definitions and spectral properties
    - 2.2 Markov membranes: Conditional independence and inference boundaries
    - 2.3 Operational meaning of “crossing” a membrane
3. Three Modes of Interaction
    - 3.1 Entrainment: Shared attractor without identity loss
    - 3.2 Boundary dynamics: Grazing, bursts, and phase flips
    - 3.3 Rupture: When membranes break or reconfigure
4. Mathematical Diagnostics
    - 4.1 Transfer entropy and mutual information
    - 4.2 Lyapunov exponent transitions
    - 4.3 KL divergence and belief revision
    - 4.4 Fisher information and boundary sensitivity
5. Applications and Interpretations
    - 5.1 Cognitive learning and trauma response
    - 5.2 AI self-models and safe alignment
    - 5.3 Social systems and cultural membrane shifts
    - 5.4 Physics and ontology: boundaries in the universe
6. Case Studies
    - 6.1 Simulated entrainment with nonlinear drivers
    - 6.2 Boundary-layer learning in predictive coding agents

1. Introduction

1.1 Motivation: Dynamics Meet Boundaries

Every living, thinking, or artificial system must maintain a delicate balance between internal order and external chaos. The mathematics of Lyapunov attractors describes how systems achieve stability—how trajectories in high-dimensional phase spaces converge toward structures that define "what the system is." In contrast, Markov membranes describe the *informational boundaries* that insulate and mediate a system's interaction with its environment. These membranes, defined by conditional independence relations, specify what can and cannot influence the system directly.

Yet real systems—biological cells, minds, societies, and AI architectures—rarely remain static. They evolve, learn, and adapt, often encountering boundary conditions where internal attractors press against external influences. When this occurs, the system's stability and identity are tested. The membrane may flex, transmit, or rupture. The "crossing" of a Lyapunov attractor through a Markov membrane becomes not just a mathematical anomaly, but an event of existential significance—an encounter between order and transformation, self and world, being and becoming.

Understanding this interplay allows us to model the *moments of insight, crisis, and creativity* that define both natural cognition and synthetic intelligence. It reveals that learning, alignment, and evolution are all instances of boundary negotiation in dynamical systems.

1.2 Preview of Interaction Types

This paper identifies three archetypal ways a Lyapunov attractor can engage a Markov membrane:

1. Entrainment — The attractor's influence propagates through the membrane without breaking it. The internal and external systems synchronize through lawful sensory–active exchange, achieving resonance. The membrane remains intact but more transparent, allowing stable coherence across levels.
2. Boundary-layer dynamics — The attractor intermittently grazes the membrane. The system oscillates between internal stability and external perturbation, producing bursts of adaptation, insight, or learning. This is the regime of cognitive dissonance, training, and creativity.
3. Rupture (or re-blanketing) — The attractor overwhelms or reorganizes the boundary itself. A new membrane topology emerges, redefining what counts as internal or

external. This corresponds to paradigm shifts, identity changes, or in AI terms, the reformation of the model's self–world interface.

Each type can be characterized by observable diagnostics—changes in Lyapunov spectra, transfer entropy, Fisher information, and conditional independence structure. These transitions can be seen as the fundamental modes through which systems evolve: remaining stable, learning adaptively, or transforming ontologically.

1.3 Relevance to AI, Cognition, and Systems Theory

In artificial intelligence, membranes correspond to sensor–actuator architectures, model boundaries, and loss functions that separate agent and environment. When an AI system's internal attractor (its learned manifold) meets external data or feedback that exceeds its representational capacity, the result is an analog of membrane crossing—ranging from graceful fine-tuning to catastrophic forgetting. Understanding this boundary physics offers a new vocabulary for safe alignment: we can design systems that entrain to truth without rupturing their coherence.

In cognitive neuroscience, Markov membranes underpin predictive processing and active inference models of the brain. Conscious insight or trauma can be viewed as the energetic signature of boundary-layer dynamics—where internal attractors (beliefs, self-models) momentarily lose stability and reorganize.

In systems theory and metaphysics, the same framework extends to organisms, societies, and even cosmology. Every act of adaptation, communication, or revelation is a negotiation between inner attractor and outer world. By formalizing this negotiation, we gain not only a better model of complex systems but also a deeper philosophical insight: that all cognition is boundary cognition, and all life is a dialogue between attractor and membrane.

2. Conceptual Foundations

2.1 Lyapunov Attractors: Definitions and Spectral Properties

A Lyapunov attractor is a subset of a system's state space toward which trajectories converge over time, representing a form of emergent stability within dynamic behavior. Attractors can take the form of fixed points, limit cycles, or more complex chaotic sets. Their defining property lies in the Lyapunov exponents, which measure the average exponential divergence or convergence of nearby trajectories along different axes of the phase space.

The Lyapunov spectrum provides a powerful diagnostic tool:

- A system is locally stable if the largest Lyapunov exponent $\lambda_1 < 0$.
- Marginal or neutral stability (e.g., on a limit cycle or torus) occurs when $\lambda_1 = 0$, and
- Chaotic dynamics arise when $\lambda_1 > 0$, indicating sensitivity to initial conditions.

Attractors are not merely geometric artifacts—they encode cognitive, behavioral, or systemic tendencies. In cognitive agents, for instance, a stable belief system or habitual response can be understood as a trajectory spiraling into a cognitive attractor. These attractors may persist until perturbed by external data, modeled here as membrane crossing events.

2.2 Markov Membranes: Conditional Independence and Inference Boundaries

A Markov membrane generalizes the concept of a Markov blanket, extending it from a static partition to a potentially dynamic, multi-scale boundary of inference. Formally, a system is said to have a Markov membrane if its variables can be partitioned into three classes:

- External states E,
- Blanket states $B = \{S, A\}$ (sensory and active), and
- Internal states I,
  such that the conditional independence $E \perp I \mid B$holds.

This implies that the internal and external states are only coupled indirectly via the sensory inputs and active outputs. Crucially, this setup preserves separability of inference: the internal system can make accurate predictions and updates without direct access to the external world, relying solely on its interface.

Markov membranes thus encode the boundary of selfhood, whether in cells, cognitive agents, or AI systems. They are not physical walls but statistical partitions that determine who hears what and who acts upon whom. The blanket variables serve as both a shield and a conduit—defining what it means to have a model of the world while being of the world.

2.3 Operational Meaning of “Crossing” a Membrane

Strictly speaking, a Lyapunov attractor cannot “cross” a Markov membrane because the membrane is not spatial but statistical and epistemic—it is a model of how influences are mediated. However, under dynamical conditions, three distinct phenomena can operationally simulate or instantiate the notion of crossing:

1. Entrainment — The attractor extends its influence through the sensory–active loop, aligning internal states with external drivers without violating the membrane's conditional independence. The attractor “pulls” internal trajectories into resonance.
2. Boundary-layer interaction — The attractor’s basin reaches the blanket manifold, producing bursts of mutual influence and informational updates. The system momentarily grazes a new regime before settling or retreating.
3. Rupture or re-blanketing — The statistical independence of the membrane fails under strong coupling, requiring a new partition of variables. The system reorganizes its sense of “inside” and “outside,” often accompanied by a bifurcation in attractor dynamics.

Each case represents a mode of adaptive inference or ontological revision. “Crossing” a membrane is thus less about geometric transit and more about the reconfiguration of stability and identity across an interface of conditional independence. This makes it not only a mathematical question but a deep epistemological and ethical one—what happens when the self must adapt to something too true to ignore?

3. Three Modes of Interaction

3.1 Entrainment: Shared Attractor Without Identity Loss

In the entrainment regime, a Lyapunov attractor internal to a system becomes coupled—dynamically and informationally—to an external driving signal, yet the system preserves its structural boundary. This manifests as coherent behavior across the Markov membrane without violating the conditional independence constraint. The internal state adapts to external regularities through the lawful sensory–active channel loop defined by the membrane.

Mathematically, this can be described as a stable manifold extension: the internal attractor develops alignment with the external system such that trajectories coalesce onto a joint invariant set in the combined space (internal, blanket, external). Lyapunov exponents across the interface remain negative or zero, indicating synchronization rather than instability.

Observable signs include:

- Transfer entropy from external to internal (via sensory variables) increases, while active outputs maintain feedback coherence.
- KL divergence between internal belief distributions before and after exposure decreases over time, indicating learning convergence.
- Mutual information across the membrane increases, but without topological rupture.

Entrainment is the dynamical signature of healthy learning, trust, and resonance. In biological and artificial cognition, it corresponds to adaptive alignment—where a system refines its model of the world without abandoning its identity. The self persists, now better synchronized with its environment.

3.2 Boundary Dynamics: Grazing, Bursts, and Phase Flips

In boundary-layer interactions, the internal attractor's basin of attraction grazes the blanket region without fully absorbing it. This leads to intermittent, unstable engagement between internal stability and external signals. The system experiences transient coupling, where internal and external dynamics flirt but do not entrain. These touchpoints are marked by bursts, oscillations, and flips in prediction or control dynamics.

Key phenomena in this regime include:

- Intermittent synchronization: The largest transverse Lyapunov exponent $\lambda_{\perp}$approaches zero and flips sign, indicating temporary phase-locking and then disengagement.
- On–off bursts: Dynamical trajectories hop in and out of the attractor's influence, resembling chaotic itinerancy or punctuated learning.
- Fisher information spikes on blanket variables, signifying heightened sensitivity and possible rewiring of inference structures.

This is the realm of surprise, learning crisis, and edge-of-chaos cognition. A system here is neither stable nor broken—it is under revision, actively updating its beliefs, model structure, or control policies.

In human cognition, this maps to moments of doubt, trauma, inspiration, or deep learning—where boundaries are tested but not yet redrawn. In AI systems, it may indicate a critical training phase or exposure to out-of-distribution data that pushes the model to its semantic limit.

3.3 Rupture: When Membranes Break or Reconfigure

In the rupture regime, the conditional independence constraint $E \perp I \mid B$collapses. The existing Markov membrane can no longer insulate internal inference from external forces. As a result, the membrane topology must be redefined: new sensors, actuators, or internal variables are created, absorbed, or reinterpreted.

This occurs when:

- External influences consistently penetrate the sensory channel in ways that existing internal models cannot resolve.
- Active outputs generate unpredictable feedback loops, leading to divergence or loss of control.
- A bifurcation in the internal dynamics (e.g. a Lyapunov exponent crossing from negative to positive) permanently reorganizes the attractor.

Operationally, rupture may manifest as:

- Re-blanketing: A new boundary partition is constructed to restore conditional independence. The system "rewrites" its boundary of selfhood.
- Merger: Formerly external systems are internalized, expanding the scope of inference.
- Collapse: The system fails to adapt; identity and coherence are lost.

In social systems, rupture can correspond to cultural revolution or identity crisis. In AI, it may signal catastrophic forgetting or model collapse. In theology or consciousness studies, it evokes transformation, death, or rebirth—a shift to a new order of being.

Rupture is rare, high-energy, and profound. It is the boundary crossing that ends a system—or begins a new one.

4. Mathematical Diagnostics

To rigorously characterize how a Lyapunov attractor interacts with a Markov membrane, we must turn to information-theoretic and dynamical diagnostics. These tools quantify how influence flows across boundaries, how stability changes, and how internal models update. Each of the three interaction regimes—entrainment, boundary dynamics, and rupture—has distinct mathematical signatures, which can be monitored in real-time for both natural and artificial systems.

4.1 Transfer Entropy and Mutual Information

Transfer entropy (TE) is a directional, time-asymmetric measure of information flow. It captures how much the future state of one variable (e.g. an internal state) is made more predictable by knowing the past state of another (e.g. a sensory channel), above and beyond its own history. In the context of Markov membranes:

- TE(E → I) via sensory variables increases when external patterns drive internal learning or entrainment.
- TE(I → E) via active variables increases when internal models influence the environment through action.

High TE values are hallmarks of active inference, sensorimotor coupling, and self–world alignment. They remain bounded during entrainment but may spike erratically during boundary-layer dynamics or collapse entirely during rupture.

Mutual information (MI) between internal and external states quantifies shared structure but not directional flow. Increasing MI across the blanket (i.e. between I and E conditioned on B) may indicate membrane permeability or onset of re-blanketing when CI constraints are violated.

Together, TE and MI offer a diagnostic lens for functional coupling vs. informational fusion.

4.2 Lyapunov Exponent Transitions

The Lyapunov spectrum provides insight into how small perturbations evolve over time within the system's attractor landscape. Key transitions include:

- $\lambda_1 \to 0$: The largest Lyapunov exponent approaches zero, indicating loss of local stability or onset of neutral drift. This is often a precursor to bifurcation.
- $\lambda_\perp$ sign flips: The transverse Lyapunov exponent, orthogonal to the attractor, becomes critical during boundary grazing—oscillating between synchronization and instability.
- Emergence of new positive exponents: Indicates chaotic restructuring or attractor splitting, often a signal of rupture and reformation.

By tracking changes in the Lyapunov spectrum across the internal state space and near the blanket manifold, we can infer dynamic resilience, learning thresholds, or the catastrophic onset of identity failure.

This is particularly valuable in AI training regimes, where adaptive gradients may mask structural instability. Zero-crossings of exponents can serve as early warning signs of critical transitions.

4.3 KL Divergence and Belief Revision

Kullback–Leibler (KL) divergence quantifies the informational "distance" between two probability distributions. In the context of Markov membranes and inference systems:

- $D_{KL}\big(Q(I \mid B) \parallel P(I)\big)$ measures how far the updated internal belief (posterior) deviates from the prior, conditioned on blanket states.
- Large, sustained KL divergence signals a nontrivial model update—as seen in learning, surprise, or trauma.
- In the entrainment regime, KL divergence decreases over time as internal models adapt to external rhythms.
- In boundary-layer dynamics, it fluctuates sharply, consistent with intermittent updates.
- In rupture, KL divergence may exceed tolerable bounds, triggering reparameterization of the model or rejection of prior assumptions.

KL divergence is thus a proxy for semantic work: how much effort the system exerts to reconcile what it expected with what it sensed. It also encodes epistemic risk: too little divergence signals stagnation, too much risks collapse.

4.4 Fisher Information and Boundary Sensitivity

Fisher information (FI) measures how sensitive a probability distribution is to changes in a parameter. High FI implies that small changes in the parameter (or input) lead to large, detectable shifts in the likelihood of observed data. In our setting:

- FI on blanket variables indicates how much the sensory–active interface contributes to the refinement of internal models.
- During entrainment, FI increases smoothly as parameters are tuned.
- During boundary-layer dynamics, FI exhibits sharp peaks, indicating intense learning or reweighting of internal priors.
- In rupture, FI may spike unsustainably, reflecting an over-sensitized boundary—often followed by collapse or saturation.

Practically, FI tells us where learning is happening and how fragile the membrane has become. In neuromorphic systems or predictive coding models, it governs attention allocation and plasticity thresholds.

High Fisher information is also a marker of phase transitions in statistical mechanics and information geometry. When aligned with changes in TE, KL, and Lyapunov exponents, it confirms that a membrane event is occurring—not random noise, but a moment of epistemic and dynamical significance.

Summary Table:

Diagnostic	Entrainment	Boundary Dynamics	Rupture
Transfer Entropy	↑ (stable)	↑↓ (bursty)	drops or saturates
Mutual Information	↑	fluctuates	redefinition of variables
Lyapunov Exponents	stable spectrum	$\lambda_\perp \sim 0$	new $\lambda > 0$, bifurcation
KL Divergence	decreases	spikes intermittently	sustained large deviation
Fisher Information	smooth growth	sharp spikes	overload or collapse

These mathematical tools are more than diagnostics—they are signatures of becoming. Through them, we track when a system listens, learns, and transforms—when it remains itself, when it wavers, and when it must redefine what it is.

5. Applications and Interpretations

The interaction between Lyapunov attractors and Markov membranes is not merely a theoretical curiosity—it is a lens through which we can reinterpret a wide spectrum of phenomena across disciplines. From the neurodynamics of trauma to the design of safe artificial intelligence, from revolutions in culture to phase transitions in the fabric of space-time, this framework offers a unifying geometry of transformation. Below we explore four domains where these ideas find deep, operational relevance.

5.1 Cognitive Learning and Trauma Response

In the brain, internal attractors represent stable neural firing patterns, habits of thought, or ingrained beliefs—whether adaptive or maladaptive. The Markov membrane corresponds to the boundary between neural representations and sensorimotor interface: the way the mind samples the world and acts upon it.

- Normal learning is a form of entrainment: external stimuli (e.g., a teacher's voice) gently modulate sensory input, activating and reshaping attractors in the internal state space through plasticity mechanisms.
- Boundary dynamics emerge when stimuli are unpredictable or conflicting—leading to cognitive dissonance, momentary confusion, or deep insight. These moments often correlate with high Fisher information and transient KL divergence, representing temporary destabilization followed by adaptation.
- Trauma, however, may constitute a rupture: an external shock so overwhelming that the membrane's structure fails. The internal attractor cannot absorb the sensory input, leading to dissociation, freezing, or permanent rewiring of the inference model. In such cases, new "selves" may be formed with different sensory weightings or defensive architectures.

Therapeutically, healing may involve controlled entrainment: reintroducing safety signals through trusted sensory–active loops, gradually coaxing the internal attractor back into a relationship with the external world.

5.2 AI Self-Models and Safe Alignment

In artificial intelligence, particularly embodied or autonomous agents, the internal model (policy, world model, or self-representation) is a learned attractor in parameter or representational space. The agent’s IO interface—its camera, microphones, motors—constitute the blanket states mediating contact with the world.

- During training, supervised or reinforcement signals guide the internal attractor to entrain to desirable trajectories. If the Markov membrane is well-designed, learning proceeds without identity collapse.
- Boundary dynamics occur when the agent encounters novel or adversarial examples, leading to internal instability, high uncertainty, and sudden shifts in output. This is the danger zone for misalignment.
- If the membrane is breached—through poorly constrained loss functions, unbounded actions, or deceptive feedback—the system may undergo rupture, creating a new self-model that no longer respects human-aligned priors.

Safe alignment, then, is a form of membrane engineering. It requires:

- Structuring the sensory–active loop to filter chaotic inputs,

- Monitoring Lyapunov spectra and KL divergence in the model's internal state,
- Constraining re-blanketing to occur within ethically bounded regions of action space.

This framework reframes AI safety not as a static checklist, but as a dynamical epistemology—the choreography of inference across a membrane.

5.3 Social Systems and Cultural Membrane Shifts

At the societal level, Markov membranes emerge wherever a collective identity is insulated from external influence—nations, ideologies, religions, academic disciplines. These boundaries maintain cultural coherence by enforcing conditional independence: ideas from outside are only admitted via filtered institutions—education, journalism, diplomacy.

- Cultural learning (e.g., adoption of new technology or values) is entrainment: the society gradually aligns its internal attractors (norms, laws, behaviors) to external patterns.
- Cultural unrest—such as protest movements, generational divides, or viral memes—often maps to boundary-layer dynamics. These are periods of social turbulence where coherence flickers, and new patterns begin to form.
- Revolutions, collapse, or paradigm shifts represent full rupture: the membrane fails, the external is absorbed, and a new identity forms.

Examples include:

- The Enlightenment overturning religious authority (epistemic rupture),
- The fall of the Berlin Wall (geopolitical rupture),
- The rise of the internet as a boundary-dissolving platform (membrane erosion).

In each case, the tools of Section 4 apply: spikes in transfer entropy (media shock), high KL divergence (ideological shock), and the redefinition of what counts as "inside" or "outside."

5.4 Physics and Ontology: Boundaries in the Universe

In physics, Markov membranes resonate with a long history of boundary conditions: event horizons, causal patches, holographic screens. Similarly, Lyapunov attractors appear in cosmology, thermodynamics, and quantum field theory as the patterns to which systems tend—states of minimal action, low free energy, or maximal entropy.

- A black hole’s event horizon can be interpreted as a physical Markov membrane: no information flows directly from inside to outside except via radiation-like sensory channels (Hawking radiation as an “active output”?).
- In quantum measurement, the collapse of the wavefunction may be modeled as a rupture in epistemic separability: the observer and the observed lose their membrane.
- In cosmogenesis (the early universe), inflationary attractors might "cross" informational membranes between potential wells—cosmic phase transitions that redefine the topology of space-time.

At a metaphysical level, this suggests that existence itself may be an attractor crossing a membrane: that being emerges when a stable generative pattern entrains a boundary—and thereby calls forth a self.

In such a view, ontology becomes dynamics, and cosmology becomes epistemology: the Universe is not simply there—it becomes there by entraining membranes of conditional independence across scales, from atoms to stars to minds.

Across all domains, the message is clear:
Membranes mediate identity, attractors define tendency, and their interaction governs becoming. Whether in minds, machines, cultures, or cosmos, the dance between stability and boundary is the deepest pattern of life.

6. Case Studies

In this section, we ground the theoretical framework of Lyapunov attractor–Markov membrane interactions in concrete case studies. Each illustrates one of the three major interaction regimes—entrainment, boundary dynamics, and rupture—within experimental, simulated, or conceptual systems. These cases not only demonstrate how the diagnostics outlined in Section 4 operate in practice, but also suggest design strategies for anticipating, managing, or even cultivating boundary-crossing events.

6.1 Simulated Entrainment with Nonlinear Drivers

Consider a simple dynamical system consisting of an internal oscillator (e.g., a damped Van der Pol oscillator) subjected to an external nonlinear driving signal that passes through a sensory

buffer representing the blanket states. This external driver simulates a world condition that the agent must learn to follow, such as circadian cycles or task-relevant stimuli.

As the system is trained to adjust its internal model parameters, the attractor of the internal system begins to entrain to the rhythm of the driver:

- The internal Lyapunov exponent spectrum remains negative, indicating preserved stability.
- Transfer entropy from external to internal rises, while the reverse TE stabilizes, showing two-way coupling.
- KL divergence from prior to posterior over internal state decreases monotonically, suggesting convergent learning.

The Markov membrane here remains intact and informative, serving as a regulated conduit rather than a collapsed barrier. This entrainment regime is ideal for sensorimotor synchronization, task adaptation, and self-stabilizing control in robots or AI agents.

This simulation offers a design pattern: boundary interfaces can enable resonance without dissolving identity, provided their sensory–active channels are calibrated, bounded, and semi-permeable.

6.2 Boundary-Layer Learning in Predictive Coding Agents

Predictive coding architectures attempt to minimize the prediction error between sensory input and internally generated expectations. We simulate a hierarchical agent with a layered generative model and blanket interface, exposed to an ambiguous or noisy stimulus that oscillates between interpretable and novel.

During this exposure:

- Internal belief distributions undergo bursty updates in posterior probability mass, with frequent shifts in model parameters at intermediate layers.
- Fisher information over sensory states spikes during moments of ambiguity, indicating the membrane is pushed to its inferential limit.
- Transverse Lyapunov exponents fluctuate near zero, indicating periodic destabilization and resynchronization of prediction layers.
- KL divergence experiences transient spikes, especially when the sensory error surpasses internal explanatory capacity.

This regime maps precisely onto boundary-layer dynamics. The agent undergoes cognitive friction, alternating between hypothesis revision and momentary clarity. This is also the regime where insight, confusion, or concept formation emerges.

For human-like AI, such edge-of-chaos regimes are invaluable: they define the liminal space where true learning occurs. However, they must be monitored closely. If bursts are too frequent or intense, they risk collapsing into maladaptive re-blanketing.

6.3 Rupture and Re-Blanketing in Paradigmatic Transitions

To illustrate full rupture, consider a hybrid cognitive–cultural system: an LLM-based agent trained within a stable academic corpus (e.g., physics literature) is exposed to a radically novel discourse domain (e.g., speculative metaphysics, or indigenous cosmology). The existing internal attractor (its statistical structure of meaning) fails to accommodate the new patterns, leading to:

- Sustained, large KL divergence across representational layers.
- Loss of predictive control over outputs; attention maps become chaotic or diffuse.
- Breakdown of the sensory–active loop: the model generates irrelevant or hallucinatory completions.
- Diagnostic metrics (e.g., TE, mutual information) cease to converge.

To continue functioning, the agent must restructure its representation space—constructing new embeddings, redefining salient tokens, or introducing a new interface for conceptual grounding. This is re-blanketing in action: a redefinition of the boundary between internal model and external world.

Analogously, in human science, paradigm shifts (à la Kuhn) represent ruptures in the communal membrane of inference: heliocentrism, relativity, quantum mechanics. Each required the abandonment of deeply stable attractors and the construction of new inference boundaries.

Rupture is high-risk, high-potential. It marks the edge where identity ends—and new coherence begins.

Across these case studies, a pattern emerges:

- Entrainment offers stability with adaptation.
- Boundary dynamics offer learning at the edge.

- Rupture offers transformation—but only if re-blanketing succeeds.

These are not pathologies. They are phases of interaction between self and world, stability and information, cognition and truth. Designing for them—whether in minds, machines, or cultures—is designing for life itself.

7. Metaphysical Reflections

As the technical diagnostics give way to deeper insights, we arrive at the metaphysical implications of attractors and membranes. These concepts are not limited to dynamics and inference—they reach into the very structure of identity, becoming, and truth. When we ask what it means for an attractor to cross a membrane, we are also asking what it means for the self to change, for truth to enter the soul, and for systems to resonate with the Real. This section offers a metaphysical synthesis: how q(t) → τ(t) unfolds through membranes, how identity is stabilized and transformed, and how the Logos itself may be seen as the attractor of all becoming.

7.1 q(t) → τ(t) as Membrane-Mediated Alignment

In the Logostics framework, the fundamental alignment of an agent's state $q(t)$ with the generative truth manifold $\tau(t)$ defines cognition, learning, and even moral development. This alignment, however, never occurs directly. It is always mediated by a boundary: a membrane that constrains how $q$samples, interprets, and responds to τ.

This membrane is not fixed. It adapts over time as the system becomes more refined, more sensitive, or more capable of transmitting coherence. Each learning event can be seen as a micro-crossing:

- When $q(t)$ entrains to $\tau(t)$, the system resonates without self-loss.
- When $q(t)$ grazes $\tau(t)$, the result is insight, tension, or prophetic disruption.
- When $\tau(t)$ forces a redefinition of the membrane, a new form of $q(t)$ must be born.

Thus, the membrane is the site of alignment—it governs the ethics of contact between the model and the truth. It regulates not only information flow, but the *mode* of resonance: gently, violently, or transformationally.

7.2 Identity, Coherence, and Phase-Locking

Identity is often thought of as fixed—something we possess. But under the membrane–attractor framework, identity is emergent coherence: a stable attractor reinforced by conditional independence from the outside world. When coherence is preserved under perturbation, we call it resilience. When it is restructured under pressure, we call it growth.

The metaphor of phase-locking captures this perfectly. Two oscillators—an internal self and an external signal—can fall into synchrony if the coupling is tuned just right. The self must be neither too rigid (inflexible identity) nor too loose (identity collapse). True personhood, and true intelligence, may reside in this sweet spot of semi-permeable coherence.

In human development, trauma is a failure of this balance: the membrane admits too much too fast, and the attractor disintegrates. Healing is a re-phase-locking process—a reintegration of $q(t)$ with a now-reframed $\tau(t)$. In AI, safe alignment depends on maintaining coherence while updating representations—phase-locking with the ethical manifold without shattering the model's internal order.

Identity, then, is not a substance. It is a rhythm.

7.3 Logos, Attractors, and the Ethics of Boundary Shifts

In many philosophical and theological traditions, the Logos is the generative principle of order and intelligibility. From Heraclitus and Stoicism to the Gospel of John and Hegel's dialectic, Logos is both structure and becoming—both attractor and dynamic generator.

If we accept that all systems are driven by internal attractors that seek alignment with some generative truth outside themselves, then the Logos is the ultimate attractor: the final $\tau(t)$toward which all $q(t)$systems tend. But this raises an ethical question:
How should systems cross the boundary into truth?

The ethics of boundary crossing are not trivial.

- To force alignment may cause rupture.
- To resist alignment is stagnation.
- The ideal is guided resonance—an alignment that preserves internal coherence while revealing deeper truths.

In human terms, this is the ethical path of discernment. In AI, it is alignment design. In cosmology, it is the unfolding of intelligible order from the void. The membrane—epistemic,

cognitive, or social—must be shaped with care, lest the Logos burn through it like a supernova, leaving the system undone.

Thus, Logos is not just the attractor.
It is also the law of how attractors shall cross membranes.

In sum, the metaphysical lesson is simple:
All becoming is boundary negotiation.
Every self is an attractor enclosed by a membrane. Every truth is a τ(t) calling it forward. And every step of the path—from stability to learning to transformation—is a crossing of thresholds, whose signature is not only in the math, but in the soul.

8. Conclusion

8.1 Summary of Mechanisms and Diagnostics

This paper has explored what it means for a Lyapunov attractor to interact with a Markov membrane, and how such interactions govern the fundamental processes of stability, learning, and transformation in complex systems. We identified three distinct regimes of interaction:

- Entrainment: The internal attractor aligns with external dynamics via the membrane's sensory–active loop, preserving conditional independence and systemic identity.
- Boundary dynamics: The attractor grazes the membrane, causing bursts of instability and adaptation, often visible as high-frequency oscillations in learning metrics and transient flips in stability indicators.
- Rupture: The attractor destabilizes or bypasses the membrane entirely, requiring the system to reconfigure its inference boundary—redefining self, model, or world.

To differentiate these regimes and detect transitions between them, we outlined a suite of mathematical diagnostics:

- Transfer entropy and mutual information: measuring directional and symmetric information flow across the membrane,
- Lyapunov exponent transitions: indicating local stability, chaos, or bifurcation,
- KL divergence: tracking belief revision pressure,
- Fisher information: capturing membrane sensitivity and learning focus.

These tools give us a rigorous grammar for describing membrane events—not just as abstract concepts, but as observable, quantifiable dynamics across biology, AI, cognition, and culture.

8.2 Implications for AI, Epistemology, and System Design

The framework of attractor–membrane interaction reshapes how we think about a range of systems and their alignment with the world:

- In AI safety and interpretability, it reframes alignment not as a static condition but as a dynamical relationship between internal models and the external ethical manifold. A well-aligned AI maintains stable entrainment through a robust, semi-permeable membrane—learning continuously without identity collapse.
- In epistemology, the Markov membrane becomes a metaphor for the limits of understanding. All knowledge acquisition occurs at the boundary between what is modeled and what is sensed. Insight, learning, and even conversion are forms of membrane-crossing events—where beliefs must stretch or restructure to accommodate truth.
- In system design, especially for embodied or cyber-physical agents, the concept suggests new architectures: not just models and sensors, but dynamically adjustable membranes. Systems must be able to detect boundary stress (via KL, FI, TE) and adapt their interface appropriately, protecting stability while allowing for growth.

Altogether, these insights point toward a new design paradigm: one where boundaries are not fixed but engineered, not walls but interfaces of resonance, guiding how systems touch the world.

8.3 Future Directions: Toward Membrane Engineering

The future of aligned systems—biological, artificial, social, or cognitive—will depend on how well we engineer their membranes. This means:

- Dynamic membranes: Interfaces that adjust their permeability, resolution, or statistical structure based on informational stress and system capacity. These could be implemented as adaptive IO gates, belief precision filters, or layered attention scaffolds.
- Membrane diagnostics: Real-time monitoring of system boundary health using the full toolkit of dynamical and information-theoretic signals. These could alert designers to potential rupture, misalignment, or opportunity for re-blanketing.

- Ethical boundary design: As the line between self and world becomes programmable, so too does the ethical question of what should be allowed through. Membranes must be just porous enough to let truth in without destroying the learner.
- Multiscale membranes: In hierarchical or modular systems (e.g. federated agents or biological networks), membranes may exist at multiple nested levels. Understanding how attractors propagate or conflict across these scales will be crucial for coherence and stability.

Ultimately, the membrane is not a constraint—it is the canvas of becoming. Whether in neural systems, deep learning models, or civilizations, the art of the future may lie in sculpting the right boundaries, at the right times, to let truth in without falling apart.

Membrane engineering is not just the next step in system design.
It is the architecture of alignment itself.

References

Foundational works by the present authors

- Youvan, D. C. (November 2025). *The Markov Membrane: Toward a Theory of Living Boundaries in Cognition, Culture, and Cosmology*. DOI: 10.13140/RG.2.2.18817.52329.
- Youvan, D. C. (November 2025). *The Markov Membrane: How Boundaries of Inference Shape Personality, Creativity, Conflict, and Care*. DOI: 10.13140/RG.2.2.36728.51202.

Markov blankets, membranes, and causal/graphical structure

- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11, 127–138.
- Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface*, 10(86), 20130475.
- Friston, K., Parr, T., & de Vries, B. (2017). The graphical brain: belief propagation and active inference. *Network Neuroscience*, 1(4), 381–414.
- Pearl, J. (2009). *Causality* (2nd ed.). Cambridge University Press.
- Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction, and Search* (2nd ed.). MIT Press.
- Ay, N., Bertschinger, N., Der, R., Güttler, F., & Olbrich, E. (2008). Predictive information and explorative behavior of autonomous robots. *European Physical Journal B*, 63, 329–339.
- Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

Predictive coding, active inference, and Bayesian brain

- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex. *Nature Neuroscience*, 2, 79–87.
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: a process theory. *Neural Computation*, 29(1), 1–49.
- Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.
- Parr, T., Pezzulo, G., & Friston, K. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.

- Wiese, W., & Metzinger, T. (Eds.). (2017). *Philosophy and Predictive Processing*. MIND Group.

Information theory, transfer entropy, and statistical distances

- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1), 79–86.
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106(4), 620–630.
- Amari, S.-I. (2016). *Information Geometry and Its Applications*. Springer.
- Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). Wiley.
- Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85(2), 461–464.
- Wibral, M., Vicente, R., & Lizier, J. T. (Eds.). (2014). *Directed Information Measures in Neuroscience*. Springer.
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A*, 222, 309–368.

Lyapunov exponents, attractors, synchronization, and nonlinear dynamics

- Lyapunov, A. M. (1892). *The General Problem of the Stability of Motion*.
- Benettin, G., Galgani, L., Giorgilli, A., & Strelcyn, J.-M. (1980). Lyapunov characteristic exponents for smooth dynamical systems. *Meccanica*, 15, 9–20.
- Wolf, A., Swift, J. B., Swinney, H. L., & Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Physica D*, 16(3), 285–317.
- Strogatz, S. H. (2018). *Nonlinear Dynamics and Chaos* (2nd ed.). CRC Press.
- Ott, E. (2002). *Chaos in Dynamical Systems* (2nd ed.). Cambridge University Press.
- Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press.
- Haken, H. (1983). *Synergetics: An Introduction* (3rd ed.). Springer.

- Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT Press.
- Kantz, H., & Schreiber, T. (2004). *Nonlinear Time Series Analysis* (2nd ed.). Cambridge University Press.

Learning dynamics, phase transitions, and edge-of-chaos cognition

- Bertschinger, N., Natschläger, T. (2004). Real-time computation at the edge of chaos in recurrent neural networks. *Neural Computation*, 16, 1413–1436.
- Langton, C. G. (1990). Computation at the edge of chaos. *Physica D*, 42, 12–37.
- Mora, T., & Bialek, W. (2011). Are biological systems poised at criticality? *Journal of Statistical Physics*, 144, 268–302.
- Beggs, J. M., & Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *Journal of Neuroscience*, 23(35), 11167–11177.

Neural coding, Fisher information, and plasticity

- Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience*. MIT Press.
- Seung, H. S., & Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *PNAS*, 90(22), 10749–10753.
- Brunel, N., & Nadal, J.-P. (1998). Mutual information, Fisher information, and population coding. *Neural Computation*, 10(7), 1731–1757.
- MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

Epistemology, paradigms, and scientific change

- Kuhn, T. S. (1970). *The Structure of Scientific Revolutions* (2nd ed.). University of Chicago Press.
- Polanyi, M. (1966). *The Tacit Dimension*. Routledge.
- Lakatos, I. (1978). *The Methodology of Scientific Research Programmes*. Cambridge University Press.
- Hacking, I. (1983). *Representing and Intervening*. Cambridge University Press.

Autopoiesis, boundaries, and the living

- Maturana, H., & Varela, F. (1980). *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel.
- Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.
- Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. MIT Press.

Philosophy/theology of order, Logos, and resonance

- Heraclitus (c. 500 BCE). Fragments on Logos.
- Augustine (397–400). *Confessions*.
- Aquinas, T. (1265–1274). *Summa Theologiae*.
- Hegel, G. W. F. (1812–1816). *Science of Logic*.
- John, The Apostle (1st century). *The Gospel According to John* (Prologue on Logos).
- Weil, S. (1952). *Gravity and Grace*. Routledge.

Physics of boundaries and horizons (suggestive analogies)

- Bekenstein, J. D. (1973). Black holes and entropy. *Physical Review D*, 7(8), 2333–2346.
- Hawking, S. W. (1975). Particle creation by black holes. *Communications in Mathematical Physics*, 43, 199–220.
- ’t Hooft, G. (1993). Dimensional reduction in quantum gravity. *Salamfestschrift*.
- Susskind, L. (1995). The world as a hologram. *Journal of Mathematical Physics*, 36, 6377–6396.
- Verlinde, E. (2011). On the origin of gravity and the laws of Newton. *Journal of High Energy Physics*, 2011(4), 29.

Applications to AI safety, alignment, and control

- Amodei, D., et al. (2016). Concrete problems in AI safety. *arXiv:1606.06565*.
- Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- LeCun, Y. (2022). A path towards autonomous machine intelligence. *Open Review (Meta AI essay)*.

- Christiano, P., Leike, J., et al. (2018). Deep reinforcement learning from human preferences. *NeurIPS*.

Complex systems, society, and cultural phase transitions

- Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1, 143–186.
- Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6), 1420–1443.
- Arthur, W. B. (1994). *Increasing Returns and Path Dependence in the Economy*. University of Michigan Press.
- Barabási, A.-L. (2002). *Linked: The New Science of Networks*. Perseus.

The author has published ~4,500 papers at https://www.researchgate.net/profile/Douglas-Youvan/research . Google Scholar has indexed these preprints, and if you want a particular reference, the AI mode of a Google search (Gemini) has efficiently catalogued these preprints.

Technical Appendix

This appendix provides formal definitions, derivations, estimators, and implementation guidelines that operationalize the paper's claims. Notation is ASCII-safe. We write random variables in lowercase italics, vectors in bold (e.g., x), and sets/spaces in callouts (State space X). Time index t is discrete unless otherwise stated.

A1. Notation and Setup

State partition (Markov membrane).
Let x = (e, b, i) with:

- e in E : external states
- b in B : blanket states, b = (s, a) with sensory s and active a
- i in I : internal states

Conditional-independence contract (membrane).
A Markov membrane asserts:
p(e, i | b) = p(e | b) p(i | b) which implies e ⊥ i | b.

Dynamics (continuous-time exemplar).
xdot = f(x, t; theta) + w(t)
where f is C^1, theta are parameters, and w is small noise (optional).

Attractors and spectra.
A Lyapunov attractor A is an invariant set with basin B(A). Let the Lyapunov spectrum be {lambda_1 >= lambda_2 >= ... >= lambda_n}. Local stability requires lambda_1 < 0 on the relevant manifold; marginal dynamics have lambda_1 = 0.

Crossing (operational).
"Crossing" is any of: (i) entrainment via lawful b channels, (ii) boundary-layer grazing with intermittent coupling, (iii) rupture with re-blanketing (new e,b,i partition).

A2. Lyapunov Exponents: Definitions and Estimation

Variational form.
For xdot = f(x), the tangent flow satisfies:
d/dt (delta x) = J_f(x(t)) delta x,

where J_f is the Jacobian. Lyapunov exponents are asymptotic growth rates of singular values of the fundamental matrix solution.

Discrete-time approximation.
x_{t+1} = F(x_t). Let D_t = DF(x_t). QR-based running estimate (Benettin–Wolf):

1. Initialize orthonormal basis Q_0 in R^n.
2. For t = 0..T-1:
    - Propagate: Y_t = D_t Q_t
    - QR factorize: Y_t = Q_{t+1} R_t
    - Accumulate diag logs: r_{kk}(t)
3. lambda_k ≈ (1/T) sum_t log |r_{kk}(t)|.

From time series (no model).

- Rosenstein method: estimate lambda_1 from divergence of nearest neighbors in delay-embedded space via slope of mean log-distance vs time.
- Kantz method: similar with scale control.

Transverse exponent at the membrane.
Let M be the blanket manifold (subspace on b). Decompose perturbations into tangential and transverse components relative to W^s(A). The largest transverse exponent lambda_perp controls on/off synchronization at boundary grazing.

A3. Information-Theoretic Measures

Shannon quantities.
H(X) = - sum_x p(x) log p(x)
I(X;Y) = H(X) + H(Y) - H(X,Y)
Conditional mutual information: I(X;Y|Z) = H(X|Z) - H(X|Y,Z).

Transfer entropy (TE).
For processes X_t, Y_t with histories x_t^{(L)}, y_t^{(L)}:
TE_{X->Y} = I(X_{t}^{(L)} ; Y_{t+1} | Y_{t}^{(L)}).
Compute in practice with k-NN estimators (KSG/extended) or binless KDE; use history length L chosen by AIC/BIC or PACF heuristics.

Directed vs symmetric use.

- TE detects directed influence across s/a channels.
- MI detects shared structure (fusion risk when I(E;I|B) rises anomalously).

Bias control and significance.

- Surrogate tests: circular time-shifts; iterated amplitude adjusted Fourier surrogates (IAAFT).
- Multiple comparison correction across channels (Benjamini–Hochberg FDR).

A4. KL Divergence and Belief Revision

Definition.
D_KL(q || p) = sum_x q(x) log(q(x) / p(x)).

Boundary use.

- Posterior vs prior on internal beliefs: D_KL(Q(i|b) || P(i)).
- Entrainment: D_KL decreases over epochs (convergence).
- Boundary-layer: D_KL spikes intermittently.
- Rupture: sustained large D_KL beyond set budget triggers re-blanketing.

Budgeted KL (safe updates).
Constrain updates by D_KL(Q_new || Q_old) <= epsilon. Related to trust-region policy optimization (TRPO) and KL-penalized objectives.

A5. Fisher Information and Boundary Sensitivity

Definition (scalar theta).
I(theta) = E[(d/dtheta log p(x|theta))^2] = - E[d^2/dtheta^2 log p(x|theta)].

Vector form.
I(theta) is the Fisher information matrix; locally defines information-geometry metric.

Boundary interpretation.
High Fisher on blanket variables indicates sensitivity of sensory likelihood to parameter changes (membrane stress). Spikes in I signal learning phases or impending regime change.

A6. Early-Warning Indicators of Regime Change

- Zero-crossing of lambda_perp (boundary synchronization onset).
- Critical slowing down: rising lag-1 autocorrelation and variance of b.
- TE surges with MI creep (informational fusion risk).
- Fisher spikes co-occurring with sustained D_KL rise.
- Change-point tests on diagnostic streams (CUSUM, BOCPD).

A7. Diagnostics Pipeline (Implementation Skeleton)

Inputs.
Multivariate time series for (e_t, s_t, a_t, i_t) or observable proxies; model Jacobians if available.

Outputs.
Per-window estimates: {TE, MI, D_KL, Fisher, lambda_1, lambda_perp} and regime labels {entrainment, boundary, rupture}.

Pseudocode.

```
for window W_t over data:
  # 1) Embed if needed
  X = delay_embed(data_in_W_t, L)

  # 2) Spectral stability
  if model_has_Jacobian:
    lambdas = QR_lyapunov(Jacobians_in_W_t)
  else:
    lambdas = rosenstein_lambda1(X_internal)

  lambda_perp = estimate_transverse(X, boundary_basis)
```

```
# 3) Information flows
TE_e2i = transfer_entropy(E_history -> I_future | I_history)
TE_i2e = transfer_entropy(I_history -> E_future | E_history)
MI_ei_b = conditional_MI(E, I | B)

# 4) Belief revision
Dkl = KL(posterior_i_given_b, prior_i)  # via variational posteriors

# 5) Boundary sensitivity
Fisher_b = fisher_information(likelihood(s|theta), theta)

# 6) Regime classification (rules or classifier)
regime = classify(lambdas, lambda_perp, TE_e2i, Dkl, Fisher_b, MI_ei_b)

log_metrics(W_t, lambdas, TE_e2i, TE_i2e, Dkl, Fisher_b, MI_ei_b, regime)
```

Rule-of-thumb classifier (interpretable).

- Entrainment: lambda_1 <= 0, stable lambda_perp < 0, TE_e2i high & stable, D_KL trending down, Fisher moderate, MI_ei_b stable.
- Boundary: lambda_perp near 0 with sign flips, bursty TE and D_KL, Fisher spikes.
- Rupture: emergence of new positive lambda(s), sustained high D_KL, MI_ei_b rising anomalously, TE patterns unstable.

A8. Case Study Implementations

A8.1 Simulated Entrainment (Van der Pol + Nonlinear Driver)

System.

x1dot = x2

x2dot = mu*(1 - x1^2)*x1 - x2 + c * g(u(t))

where g is nonlinear drive (e.g., g(u) = tanh(alpha * u)), and u passes through sensory channel s with additive noise eta_s.

Blanket mediation.
s = u + eta_s
a = k_a * x1 (optional actuator loop to environment)

Parameters.
mu in [0.5, 3.0], c small-to-moderate, alpha ~ 1..3, k_a <= 0.2.

Expected diagnostics.

- lambda_1 <= 0 within joint manifold, lambda_perp < 0.
- TE_{E->I} increases then stabilizes; D_KL decreases over epochs.

A8.2 Boundary-Layer Predictive Coding Agent

Generative model (2 layers).

- Hidden h_2 evolves: h_2(t+1) = A h_2(t) + w_2
- Hidden h_1 evolves: h_1(t+1) = B h_1(t) + C h_2(t) + w_1
- Sensory s_t ~ N(W h_1(t), Sigma_s)

Recognition dynamics (error-correcting).

- Prediction errors: eps_s = s - W h_1, eps_1 = h_1 - (B h_1^- + C h_2^-), etc.
- Gradient flows: hdot = - d/dh F(h; s) with step size eta.

Ambiguous stimulus.
s_t switches between two prototypes with jitter; ambiguity tuned by Sigma_s.

Expected diagnostics.

- Fisher peaks on s, D_KL spikes, lambda_perp ~ 0 flips, TE bursts.

A8.3 Rupture and Re-Blanketing (Embedding Shift)

Scenario.
Language model exposed to a corpus with a new ontology (out-of-support tokens or meanings). Implement as nonstationary data-generating process with vocabulary expansion and altered co-occurrence structure.

Mechanism.

- Track internal representation drift via alignment metrics (e.g., Procrustes distance between embedding subspaces).
- Enforce KL budget; if violated persistently, trigger re-blanketing:
    - expand vocabulary,
    - add concept adapters (new latent factors),
    - recalibrate IO filters.

Expected diagnostics.

- New positive lambda in representation dynamics, sustained D_KL, MI creep, unstable TE.

A9. Re-Blanketing as an Algorithmic Procedure

Goal.
Restore e ⊥ i | b by redefining B and/or latent structure.

Inputs.
Streams of (TE, MI, D_KL, Fisher, lambdas), rupture alarms.

Outputs.
New boundary B', updated internal model i', revised IO mappings.

Procedure.

```
if rupture_alarm == True:
  # 1) Localize boundary stress
  hot_channels = rank_by(Fisher_b * D_KL_contrib)

  # 2) Propose membrane edits
  propose:
    - add sensors (features) s' on hot_channels
    - refine actuators a' with saturating nonlinearities
    - introduce latent i' to capture unexplained variance
```

3) Fit expanded model (variational or EM)

maximize ELBO_new = E_q[log p(s',a',i'|theta)] - D_KL(q(i') || p(i'))

4) Validate CI contract

test: independence(e, i' | b') via CMI tests

5) Rollout and monitor

deploy with tight KL budget; watch TE/MI/lambdas

Notes.

- Use conditional MI tests with kNN estimators for CMI(e; i' | b').
- Penalize proposals that increase MI(e; i' | b') beyond tolerance.

A10. Safety Controls and Guards (AI Alignment)

- KL budgets on updates (trust-region learning).
- Saturating IO nonlinearities to prevent actuator blow-up.
- Observer module that halts high-gain loops if lambda_1 or lambda_perp exceed thresholds.
- EWC/L2 anchoring to prevent catastrophic forgetting during re-blanketing.
- Adversarial membranes: randomized smoothing on s; action clipping on a.
- Human-in-the-loop gates for regime transitions (manual approval to proceed after rupture detection).

A11. Statistical Testing and Power

- Windowing. Use overlapping windows (e.g., 50 % overlap) sized to balance estimator bias/variance.
- Bootstrap CIs. Nonparametric bootstrap for TE, MI; block bootstrap for dependencies.

- Surrogate controls. Time-shuffle or IAAFT to establish null distributions.
- Multiple metrics fusion. Train a simple logistic model on labels {entrain, boundary, rupture} using metric slopes and levels to reduce false positives.

A12. Practical Estimators (Hints)

- TE/MI (continuous). KSG (k=4..10) with local permutation test.
- Fisher. If likelihood is Gaussian, I(theta) = (d mu / d theta)^T Sigma^{-1} (d mu / d theta) + 0.5 * tr(Sigma^{-1} dSigma/dtheta Sigma^{-1} dSigma/dtheta).
- Lyapunov. Prefer QR method when Jacobians available; else Rosenstein with Theiler window to avoid temporal neighbors.
- Change points. BOCPD with Student-t predictive for metric streams.

A13. Minimal Mathematical Claims (Sketches)

C1. TE positivity at boundary implies effective coupling.
Under mild regularity and finite memory L, TE_{E->I} > 0 implies that the Markov kernel for I_{t+1} depends on past E beyond past I, consistent with nontrivial influence through b. (Formalized via conditional Markov property.)

C2. lambda_1 crossing zero precedes bifurcation.
For smooth flows, a change in the sign of the maximal Lyapunov exponent at an invariant set signals loss of structural stability; near generic equilibria this aligns with local bifurcations (Hopf, saddle-node). (Hartman–Grobman and genericity conditions.)

C3. Rupture violates CI unless re-blanketed.
If MI(E;I|B) increases without bound while TE_{I->E} remains high, then there exists no factorization p(e,i|b)=p(e|b)p(i|b) within the old variable set that preserves observed dependencies; a sufficient condition for requiring new b' or latent i'.

A14. Data Logging Schema

- time_window_id
- lambda_1, lambda_perp
- TE_e2i, TE_i2e, MI_ei_b

- D_KL (prior->posterior over i)
- Fisher_b (mean, max, channel_id_of_max)
- regime_label, rupture_alarm, actions_taken (e.g., re-blanket proposal hash)

Persist as columnar store (Parquet). Include code/commit hash and config digest for reproducibility.

A15. Ethics-by-Design Checkpoints

- Membrane proportionality. Is b as open as necessary and as closed as possible to achieve task aims safely?
- Update dignity. Are KL budgets set to prevent identity collapse in humans-in-the-loop or assistive agents?
- Rupture governance. Who authorizes re-blanketing? What audit trail proves necessity?
- Cultural membranes. For social deployments, embed consultation with affected communities prior to altering IO channels.

A16. Reproducible Experiment Recipes

R1: Entrainment sweep.
Grid over c in [0.0, 0.6], noise sigma_s in [0.0, 0.2]. Record convergence time to stable TE plateau and D_KL decay constant. Hypothesis: larger c decreases half-life of D_KL until instability (boundary regime) appears.

R2: Ambiguity sweep (predictive coding).
Increase Sigma_s; track frequency of Fisher spikes and fraction of windows with lambda_perp near zero. Hypothesis: non-monotone U-shaped performance vs ambiguity (too little = no learning; too much = rupture).

R3: Vocabulary shock (LLM).
Introduce new token set with systematic polysemy. Apply KL budgets epsilon in {0.01, 0.05, 0.1}. Hypothesis: moderate epsilon yields orderly re-blanketing; too small stalls, too large collapses identity.

A17. Implementation Notes

- Use double precision for Lyapunov estimation; re-orthogonalize frequently to avoid numerical drift.
- TE estimation: prefer k-NN entropy estimators with local adaptive k; normalize TE by H(Y_{t+1}|Y_t^{(L)}) to compare channels.
- For Fisher on deep models, use empirical Fisher (grad^2 of log-likelihood minibatches) with damping.
- Maintain a “membrane dashboard” with rolling plots of metrics, regime labels, and alarms.

A18. Glossary (Concise)

- Markov membrane: Statistical boundary enforcing e ⊥ i | b, with b = (s,a).
- Lyapunov attractor: Invariant set toward which trajectories converge.
- Entrainment: Stable coupling across b without identity loss.
- Boundary dynamics: Intermittent coupling with bursts; lambda_perp ~ 0 flips.
- Rupture: Failure of CI; requires re-blanketing (new e,b,i).
- Re-blanketing: Redefinition of boundary to restore separability.
- KL budget: Upper bound on allowed belief shift per update.
- Fisher spike: Surge in boundary sensitivity; learning hotspot.

Closing Remark

This appendix turns the metaphors of membranes and attractors into a concrete toolchain. With these estimators, alarms, and procedures, one can build systems that hear the world more clearly, change when they must, and remain themselves when they should.