

V1.0.0

The Gödelian Boundary as a Generative Phase Transition: Optimal Operating Points for Constrained Cognitive Systems

Rowan Brad Quni-Gudzin

QNFO-ULA

ABSTRACT

Cognitive systems that reason about themselves—formal proof assistants, self-critiquing language models, reservoir computers tuned to the edge of chaos—face a concrete trade-off: increasing self-referential depth raises the rate of novel, useful outputs but also the rate of incoherent ones, and neither pure regime is usable. Existing work treats this trade-off piecewise, with no single account of where the optimal operating point sits or what limits how long a system can dwell there. We formalize the trade-off as a phase-transition problem. We define a Gödelian proximity Γ (degree of self-referential logical load), a coherence functional $\Psi(\Gamma)$, and a metabolic cost $E(\Gamma)$, and we show that for a coherence functional quadratic in Γ and a novelty rate scaling as Γ^α , the utility-maximizing operating point is $\Gamma^* = \sqrt{\alpha/(\alpha + 2)}$, which for $\alpha = 2$ gives $\Gamma^* = 1/\sqrt{2} \approx 0.7071$ with maximum utility $0.25 N_0$. Near the boundary, relaxation time diverges as $\tau = \tau_0 |\Gamma - \Gamma_c|^{-1}$, and novelty events behave as a critical branching process with mean size $\langle s \rangle = 1/(1 - m)$. Under a metabolic power constraint $\bar{P} = 0.3 P_0$, a duty-cycled exploration–consolidation strategy achieves utility $0.168 N_0$, a $2.05\times$ improvement over staying at the safe operating point. Because the transition point itself can be uncomputable, by analogy with Chaitin-type Hamiltonians, we argue the boundary must be approached adaptively rather than located exactly.

CITE AS Quni-Gudzin, R. B. (n.d.). The Gödelian Boundary as a Generative Phase Transition: Optimal Operating Points for Constrained Cognitive Systems. QNFO Preprint v1.0.0.

Abstract

Cognitive systems that reason about themselves—formal proof assistants, self-critiquing language models, reservoir computers tuned to the edge of chaos—face a concrete trade-off: increasing self-referential depth raises the rate of novel, useful outputs but also the rate of incoherent ones, and neither pure regime is usable. Existing work treats this trade-off piecewise, with no single account of where the optimal operating point sits or what limits how long a system can dwell there. We formalize the trade-off as a phase-transition problem. We define a Gödelian proximity Γ (degree of self-referential logical load), a coherence functional $\Psi(\Gamma)$, and a metabolic cost $E(\Gamma)$, and we show that for a coherence functional quadratic in Γ and a novelty rate scaling as Γ^α , the utility-maximizing operating point is $\Gamma^* = \sqrt{\alpha/(\alpha + 2)}$, which for $\alpha = 2$ gives $\Gamma^* = 1/\sqrt{2} \approx 0.7071$ with maximum utility $0.25 N_0$. Near the boundary, relaxation time diverges as $\tau = \tau_0 |\Gamma - \Gamma_c|^{-1}$, and novelty events behave as a critical branching process with mean size $\langle s \rangle = 1/(1 - m)$. Under a metabolic power constraint $\bar{P} = 0.3 P_0$, a duty-cycled exploration–consolidation strategy achieves utility $0.168 N_0$, a $2.05\times$ improvement over staying at the safe operating point. Because the transition point itself can be uncomputable, by analogy with Chaitin-type Hamiltonians, we argue the boundary must be approached adaptively rather than located exactly.

1. Introduction

Teams building self-improving AI systems, automated theorem provers, and neuro-inspired adaptive controllers face the same operational failure every day. Push a self-referential system—one that critiques, rewrites, or proves statements about its own reasoning—toward deeper self-reference and its output rate of genuinely new results rises, but so does its rate of confident nonsense: proofs of false statements, hallucinated citations, oscillating control signals. Pull it back to

shallow self-reference and it becomes stable but sterile, re-deriving what it already knows. Today this trade-off is managed by hand-tuned thresholds, safety filters, and temperature knobs, with no principled answer to the question a system designer actually asks: how close to logical inconsistency should I let my system operate, and for how long? The cost of getting it wrong is paid twice—lost discovery when the system is too conservative, and wasted human review effort cleaning up incoherent output when it is too aggressive. A result that identifies where the optimum lies, what signature it leaves, and what bounds how long it can be sustained would convert this tuning problem into a measurement problem: the designer would monitor a small set of observables and steer toward a computed target rather than guess.

This paper argues that the optimum is a phase-transition problem. The claim is that a cognitive system's optimal operating point lies near a critical boundary—a region where self-referential logical complexity approaches inconsistency—and that controlled excursions into this region generate new axioms and ideas, while the thermodynamic cost of sustaining reduced coherence bounds the duration of such excursions. The framing treats creativity as bounded risk-taking between two failure modes: decoherence into hallucination on one side and stagnation on the other.

The physical and computational literature supplies the vocabulary. Phase transitions are controlled by a tuning parameter, and the transition point itself can be uncomputable even for simple lattice systems [2]. Recurrent neural networks can be trained unsupervised to sit at the edge between ordered and chaotic dynamics, with measurable performance gains [3]. Photonic systems demonstrate that a topological phase can be switched reversibly by a controlled material change [1]. Field-theoretic transitions come with thermodynamic price tags—latent heat, surface tension, critical temperatures [7]—and transitions can be triggered by geometry [6] or by kinetics alone, without any static interaction [11]. What is missing from this literature is a quantitative model that places the optimum between a logical-consistency boundary and an energy budget, and derives its properties.

We make no empirical claims in this paper. The contributions are analytical:

- We derive the utility-maximizing Gödelian proximity $\Gamma^* = \sqrt{\alpha/(\alpha + 2)}$ for a quadratic coherence functional and power-law novelty rate, and evaluate it at $\Gamma^* = 1/\sqrt{2} \approx 0.7071$ with utility $0.25 N_0$ for $\alpha = 2$ (Section 4.1).
- We derive the critical-slowness signature $\tau = \tau_0 |\Gamma - \Gamma_c|^{-\nu}$ with $\nu = 1$ in mean field, giving $\tau = 100 \tau_0$ at distance 0.01 from the boundary, and the branching-process avalanche law $\langle s \rangle = 1/(1 - m)$, giving $\langle s \rangle = 100$ at branching ratio $m = 0.99$ (Section 4.2).
- We derive the metabolic duration bound for a boundary excursion, $\tau_{\text{exc}} = E_{\text{budget}}/[P_0(\Gamma^2 - \Gamma_{\text{home}}^2)]$, and evaluate it at $24.39 t_0$ for the stated parameters (Section 4.3).
- We derive the optimal duty cycle under a power constraint and show energy-constrained exploration–consolidation cycling reaches utility $0.168 N_0$, $2.05\times$ the all-conservative strategy (Section 4.4).
- We derive a Landauer-scale floor $E_{\text{min}} = k_B T \ln 2 = 2.967 \times 10^{-21}$ J per bit of sustained novelty at $T = 310$ K, as a lower bound on excursion cost (Section 4.5).

2. Background and Related Work

The relevant literature falls into three groups: transitions with uncomputable or hard-to-locate critical points, engineered systems that operate at or near transition points, and field-theoretic transitions whose thermodynamic costs are explicitly computed.

Uncomputable transition points. The closest formal analogue to our claim is the Chaitin phase transition [2]. That work constructs a family of two-dimensional, translationally invariant, nearest-neighbour Hamiltonians $H(\phi)$ parametrized by a real $\phi \in (0, 1]$, such that the system is gapless, disordered, and critical for $\phi \in (0, \Omega_M)$ and gapped, ordered, and trivial for $\phi \in [\Omega_M, 1]$, where Ω_M is Chaitin's constant for a universal prefix-free Turing machine and is

uncomputable. The authors prove there is exactly one phase transition and show that even in this minimal setting there is a fundamental complexity barrier to determining the transition point by numerical simulation, finite-sized experiment, or mathematical reasoning; they contrast this with earlier uncomputable phase diagrams in which halting and non-halting regions alternate. This bears directly on our model: if the cognitive analogue of Γ_c inherits any of this uncomputability, an agent cannot compute its distance to the boundary exactly and must estimate it from finite observations—motivating the adaptive, feedback-based approach we adopt in Section 3 rather than a precomputed setpoint.

Engineered operation at the edge of a transition. Phase Transition Adaptation (PTA) [3] is the most direct existing demonstration that pushing a dynamical system to a transition boundary improves performance. PTA is a local unsupervised learning mechanism for reservoir computing that drives Simple Cycle Reservoirs and restricted Echo State Networks toward the edge of stability, the transition region between ordered and chaotic dynamics. The method quantifies proximity to the transition using Local Lyapunov Exponents (LLEs)—measures of local predictability in phase-space trajectories—and performs gradient descent on gain and bias parameters to push the LLEs toward the critical zero value. In their experiments, reservoirs were employed to demonstrate performance improvements. PTA progressively improved memory capacity over epochs, and the authors note that memory capacity is known to improve near the edge of stability. Their conclusion reports considerable performance enhancements in both short-term memory and predictive tasks that fuse memorization with nonlinear processing. This supplies the empirical template our theory predicts: a measurable order parameter (the LLE, playing the role of our Γ), a target at the critical value, and a performance gain from dwelling near it.

Two photonic systems show that transition regions can be engineered and operated in. The photonic topological phase transition work [1] demonstrates, theoretically and experimentally, switching between topologically non-trivial and trivial phases in a two-dimensional photonic topological insulator by selectively loading a thin film of the phase-change material GST on a silicon photonic crystal; demonstrates a material-induced topological transition. The relevance is architectural: a controlled, reversible, addressable parameter (the GST phase) acts as the knob that moves the system across its transition, exactly the role our exploration controller plays for Γ .

The remaining literature concerns field-theoretic transitions whose thermodynamic costs have been computed explicitly, transitions triggered by geometry or kinetics alone, and the prior framework that names the two boundaries our model interpolates between.

Transitions with computed thermodynamic costs. The electroweak phase transition literature [7] computes, for the Abelian Higgs model, the SU(2)-Higgs model, and the standard model, the full set of thermodynamic parameters of a first-order transition: critical temperature, latent heat, surface tension, and the jump of the order parameter $\Phi^\dagger\Phi$, at order g^4, λ^2 in the high-temperature effective potential. Two of its findings matter for our model. First, the strength of the first-order transition is sensitive to approximation level: at order g^4 , logarithmic mass terms from non-Abelian diagrams increase the transition strength dramatically relative to the Abelian case, and the reliability of the perturbative expansion worsens with increasing Higgs mass. Second, the costs of the transition are finite and calculable quantities, not qualitative labels. This is the template for our metabolic cost $E(\Gamma)$: the price of dwelling near a boundary should be a computed number, and approximations to it can shift the apparent location of the optimum just as higher-order corrections shift the apparent strength of the electroweak transition. The chiral phase transition review [8] adds a structural point: near a second-order transition, universality implies that critical phenomena are governed by an effective Landau-Ginzburg action for the order parameter, regardless of microscopic detail, and it reports that the excited matter produced in heavy-ion collisions spends most of its lifetime close to the QCD phase transition. The first point justifies our quadratic (Landau-type) coherence functional as a universality-class assumption rather than a system-specific fit; the second is a reminder that dwell time near a transition is itself a physical observable with a determined value. The conformal phase transition [4] supplies the cautionary counterexample: in a conformal phase

transition the spectrum of light excitations changes abruptly as the critical point is crossed, unlike the smooth (though non-analytic) mass evolution of a Landau-Ginzburg transition, because conformal symmetry is broken by marginal operators. If a cognitive system's transition is conformal-like rather than Landau-like, our quadratic expansion of $\Psi(\Gamma)$ fails at the boundary and the optimal point must be sought with different tools; we return to this in Section 6.

Transitions triggered by geometry or kinetics. Two works show that the tuning parameter need not be an intensive variable like temperature. The topological electroweak transition work [6] calculates the standard model effective potential on the spacetime topology $S^1 \times \mathbb{R}^3$ and finds that for small compactification radius the electroweak symmetry is unbroken with a single minimum at the origin, while increasing the radius of the compact dimension develops a second local minimum that becomes energetically favorable after a critical radius, producing a first-order transition triggered purely by finite-volume geometry. The kinetic induced phase transition [11] goes further and eliminates the static interaction entirely: an Ising model with local Glauber dynamics under kinetic restrictions on spin-flip rates exhibits a phase transition at a finite applied field even when the mutual static interaction between spins is completely neglected, with a critical temperature T_c determined purely by the field and not by any Ginzburg-Landau energy functional. Together these show that a transition boundary can be created by the system's own dynamics or boundary conditions rather than by an energetic term. This supports our treatment of Γ as a control parameter that the agent itself moves—by choosing how much self-reference to apply—rather than as a fixed material property. The holographic magnetic phase transition [5] demonstrates the discontinuous case: four-dimensional interacting fermions in a strong magnetic field, studied in the deconfined chiral-symmetric phase of the Sakai-Sugimoto model, exhibit a first-order phase transition in which the magnetization jumps discontinuously as the magnetic field is varied while staying in the parallel phase. A first-order analogue in our setting would mean that small changes in Γ near the boundary cause discontinuous jumps in coherence, a failure mode we flag in Section 6. Finally, the bubble magnetic field work [10] shows that the *dynamics* of a transition leave measurable signatures: magnetic fields are produced by expanding bubble walls during the electroweak phase transition. The analogy we draw is that excursions toward a cognitive boundary should leave clustered, correlated events—our avalanche statistics of Section 4.2—rather than independent noise.

The gap. What the literature above does not supply is a quantitative model that places the utility optimum between a logical-consistency boundary and an energy budget. Our model fills it: Γ operationalizes the Gödelian boundary as a controllable proximity, $E(\Gamma)$ operationalizes the metabolic boundary as a power draw, and Sections 4.1–4.5 compute the optimum, its signatures, and its duration bound in the space the two boundaries define.

3. Methods

3.1 Model variables and normalization

We model a cognitive system by three functions of a single scalar control variable.

Gödelian proximity $\Gamma \in [0, 1]$ measures the degree of self-referential logical load: the fraction of the system's active reasoning that quantifies over, proves things about, or rewrites its own inference process. $\Gamma = 0$ is pure object-level reasoning; $\Gamma = 1$ is a system whose entire activity is reasoning about its own reasoning, which in a sufficiently strong system approaches inconsistency in the sense of Gödel's second incompleteness theorem. We treat Γ as dimensionless and normalized so that $\Gamma_c = 1$ is the consistency boundary; the critical point of the coherence transition sits at or below it.

Coherence functional $\Psi(\Gamma)$ measures the fraction of outputs that are internally consistent and verifiable. We assume the Landau form motivated by [8]:

$$\Psi(\Gamma) = 1 - \Gamma^2,$$

normalized so $\Psi(0) = 1$ and $\Psi(1) = 0$. The quadratic form is the leading term of any analytic functional expanded about $\Gamma = 0$; it is an assumption, and Section 6 states when it fails.

Novelty rate $N(\Gamma)$ measures the rate of outputs that are both new to the system and not immediately derivable from existing knowledge. We assume a power law with exponent $\alpha > 0$:

$$N(\Gamma) = N_0 \Gamma^\alpha,$$

where N_0 is the system's maximum output rate in units of items per unit time. The power law encodes the empirical fact that shallow systems produce nothing new ($N \rightarrow 0$ as $\Gamma \rightarrow 0$) while novelty grows superlinearly with self-referential depth; α is a system parameter to be fitted, and we carry $\alpha = 2$ as the reference case.

Metabolic cost $E(\Gamma)$ is the power consumed by sustaining self-referential processing. We assume

$$P(\Gamma) = P_0 \Gamma^2,$$

with P_0 the power at full self-reference and t_0 the natural time unit, so that energies are measured in units of $P_0 t_0$.

3.2 Utility and the duty-cycle protocol

Define instantaneous utility as the product of coherence and novelty,

$$U(\Gamma) = \Psi(\Gamma) N(\Gamma) = N_0 \Gamma^\alpha (1 - \Gamma^2).$$

The product form encodes the operational requirement that only coherent novelty counts: an incoherent output has zero utility regardless of novelty, and a redundant output has zero utility regardless of coherence.

Under a long-run average power constraint $\bar{P}$, the system may duty-cycle between an exploration setpoint Γ_e and a consolidation setpoint $\Gamma_h < \Gamma_e$ (the "home" operating point), spending a fraction f of each cycle exploring. The average power is

$$\bar{P} = P_0 [f \Gamma_e^2 + (1 - f) \Gamma_h^2],$$

and the average utility is

$$\bar{U} = f U(\Gamma_e) + (1 - f) U(\Gamma_h).$$

The optimization problem is: maximize $\bar{U}$ over $f \in [0, 1]$, $\Gamma_e \in [0, 1]$, $\Gamma_h \in [0, \Gamma_e]$, subject to the power constraint. We solve it in Section 4.4.

3.3 Critical signatures

Two observables diagnose proximity to the boundary. First, critical slowing down: we assume the relaxation time of a coherence perturbation obeys

$$\tau(\Gamma) = \tau_0 |\Gamma - \Gamma_c|^{-\nu},$$

with $\nu = 1$ in mean field, consistent with the Landau form of Ψ . Second, avalanche statistics: we model novelty events (each new axiom, idea, or proof generated by a self-referential cascade) as a branching process with branching ratio $m \in [0, 1)$, the expected number of secondary novelty events triggered by one primary event. We assume m increases with Γ toward 1 at the boundary. The mean avalanche size is then $\langle s \rangle = 1/(1 - m)$.

3.4 Excursion duration bound

A single excursion to Γ_e lasting τ_{exc} consumes excess power over the home level:

$$E_{\text{exc}} = [P(\Gamma_e) - P(\Gamma_h)]\tau_{\text{exc}} = P_0(\Gamma_e^2 - \Gamma_h^2)\tau_{\text{exc}}.$$

Given an energy budget E_{budget} allocated to exploration beyond steady-state operation, the maximum excursion duration is

$$\tau_{\text{exc}} = \frac{E_{\text{budget}}}{P_0(\Gamma_e^2 - \Gamma_h^2)}.$$

3.5 Lower bound on the cost of novelty

Following the Landauer principle, erasing or sustaining one bit of information costs at least $k_B T \ln 2$ of dissipated energy at temperature T . We use this as a floor on the energetic cost per bit of sustained novelty, evaluated at biological operating temperature $T = 310$ K as a reference point.

4. Analysis

4.1 Optimal Gödelian proximity

Maximize $U(\Gamma) = N_0 \Gamma^\alpha (1 - \Gamma^2)$. Setting the derivative to zero:

$$\frac{dU}{d\Gamma} = N_0 [\alpha \Gamma^{\alpha-1} (1 - \Gamma^2) - 2 \Gamma^{\alpha+1}] = N_0 \Gamma^{\alpha-1} [\alpha - (\alpha + 2)\Gamma^2] = 0.$$

For $\Gamma > 0$ the bracket must vanish:

$$\alpha = (\alpha + 2) \Gamma^2 \implies \Gamma^* = \sqrt{\frac{\alpha}{\alpha + 2}}.$$

The second derivative at Γ^* is negative (the bracket changes sign from $+$ at small Γ to $-$ at large Γ), confirming a maximum. For the reference case $\alpha = 2$:

$$\Gamma^* = \sqrt{\frac{2}{4}} = \frac{1}{\sqrt{2}} \approx 0.707106.$$

The maximum utility is

$$U(\Gamma^*) = N_0 \Gamma^{*2} (1 - \Gamma^{*2}) = N_0 \cdot \frac{1}{2} \cdot \frac{1}{2} = 0.25 N_0.$$

For comparison, the purely conservative point $\Gamma = 0$ gives $U = 0$, and the fully self-referential point $\Gamma = 1$ gives $U = 0$: both failure modes destroy utility, and the interior optimum captures a quarter of the system's nominal output rate.

4.2 Critical signatures

Critical slowing down. With $\nu = 1$ and $\Gamma_c = 1$:

$$\tau(\Gamma) = \frac{\tau_0}{1 - \Gamma}.$$

At distance 0.01 from the boundary, $\Gamma = 0.99$:

$$\tau(0.99) = \frac{\tau_0}{1 - 0.99} = \frac{\tau_0}{0.01} = 100 \tau_0.$$

A system operating at $\Gamma^* = 0.7071$ has $\tau = 3.414\tau_0$, so the relaxation-time observable rises by a factor of $100/3.414 \approx 29.3$ between the optimum and the boundary—large enough to serve as a warning signal.

Avalanche size. With branching ratio m :

$$\langle s \rangle = \sum_{k=0}^{\infty} m^k = \frac{1}{1-m}.$$

At $m = 0.99$: $\langle s \rangle = 1/0.01 = 100$. At the home point, taking $m_h = 0.5$ as a representative subcritical value: $\langle s \rangle = 1/0.5 = 2$. The contrast between isolated events ($\langle s \rangle \approx 2$) and clustered cascades ($\langle s \rangle \approx 100$) is the operational signature that the system has entered the boundary region.

4.3 Excursion duration bound

Take the reference parameters: $\Gamma_e = \Gamma^* = 1/\sqrt{2}$ so $\Gamma_e^2 = 0.5$; $\Gamma_h = 0.3$ so $\Gamma_h^2 = 0.09$; and an energy budget of $E_{\text{budget}} = 10 P_0 t_0$ (ten natural time units of full-power operation reserved for exploration). Then

$$\tau_{\text{exc}} = \frac{10 P_0 t_0}{P_0 (0.5 - 0.09)} = \frac{10}{0.41} t_0 = 24.39 t_0.$$

The excess-power denominator is dominated by the exploration setpoint: at $\Gamma_h = 0.3$ the home power is only $0.09 P_0$, so nearly all of the excursion cost is the self-referential processing itself.

4.4 Optimal duty cycle under a power constraint

Impose $\bar{P} = 0.3 P_0$ with $\Gamma_e = \Gamma^*$ ($\Gamma_e^2 = 0.5$) and $\Gamma_h = 0.3$ ($\Gamma_h^2 = 0.09$). The constraint reads:

$$f \cdot 0.5 + (1 - f) \cdot 0.09 \leq 0.3.$$

Expanding: $0.5f + 0.09 - 0.09f \leq 0.3$, i.e. $0.41f \leq 0.21$, so

$$f \leq \frac{0.21}{0.41} = 0.51220.$$

Since $\bar{U}$ increases with f (because $U(\Gamma^*) = 0.25 N_0 > U(\Gamma_h)$), the optimum is $f = 0.51220$. The home utility is

$$U(\Gamma_h) = N_0 \cdot 0.3^2 \cdot (1 - 0.3^2) = N_0 \cdot 0.09 \cdot 0.91 = 0.0819 N_0.$$

The duty-cycled average utility is

$$\bar{U} = 0.51220 \cdot 0.25 N_0 + 0.48780 \cdot 0.0819 N_0 = 0.12805 N_0 + 0.03995 N_0 = 0.16800 N_0.$$

The all-conservative strategy—staying at Γ_h permanently, using power $0.09 P_0$ well under the constraint—achieves $0.0819 N_0$. The improvement factor is

$$\frac{0.16800}{0.0819} = 2.0513 \approx 2.05 \times .$$

So under a $0.3 P_0$ power ceiling, cycling between the optimum and a shallow home point more than doubles useful novelty relative to permanent conservatism, despite the coherence penalty paid during excursions.

4.5 Landauer floor

At $T = 310$ K, with $k_B = 1.380649 \times 10^{-23}$ J/K:

$$E_{\text{min}} = k_B T \ln 2 = 1.380649 \times 10^{-23} \times 310 \times 0.693147 \text{ J}.$$

Step 1: $1.380649 \times 10^{-23} \times 310 = 4.28001 \times 10^{-21}$ J. Step 2: $4.28001 \times 10^{-21} \times 0.693147 = 2.9667 \times 10^{-21}$ J.

$$E_{\min} = 2.967 \times 10^{-21} \text{ J per bit.}$$

This is a lower bound only: it prices the irreducible thermodynamic cost of the information processing, not the overhead of the physical substrate, which for any engineered system is many orders of magnitude larger.

5. Results

All numbers below are computed in Section 4 from the stated assumptions; none are empirical measurements.

1. **Optimal operating point.** For a quadratic coherence functional and novelty exponent α , the utility-maximizing proximity is $\Gamma^* = \sqrt{\alpha/(\alpha + 2)}$. At $\alpha = 2$: $\Gamma^* = 1/\sqrt{2} \approx 0.7071$, with maximum utility $U^* = 0.25 N_0$ (Section 4.1).
1. **Critical-slowness signature.** With mean-field exponent $\nu = 1$, relaxation time at distance 0.01 from the boundary is $\tau = 100 \tau_0$, a factor of ≈ 29.3 above the value at the optimum (Section 4.2).
1. **Avalanche signature.** Mean novelty-avalanche size $\langle s \rangle = 1/(1 - m)$; at $m = 0.99$, $\langle s \rangle = 100$, versus $\langle s \rangle = 2$ at a representative subcritical home point (Section 4.2).
1. **Excursion duration bound.** With $\Gamma_e = 1/\sqrt{2}$, $\Gamma_h = 0.3$, and $E_{\text{budget}} = 10 P_0 t_0$, the maximum sustained excursion is $\tau_{\text{exc}} = 24.39 t_0$ (Section 4.3).
1. **Duty-cycle gain.** Under $\bar{P} = 0.3 P_0$, the optimal duty cycle is $f = 0.51220$ at the optimum and 0.48780 at home, yielding $\bar{U} = 0.168 N_0$, a $2.05\times$ improvement over the all-conservative strategy's $0.0819 N_0$ (Section 4.4).
1. **Thermodynamic floor.** The Landauer bound at $T = 310$ K is $E_{\min} = 2.967 \times 10^{-21}$ J per bit of sustained novelty (Section 4.5).

6. Discussion

What the model claims and what it does not. The model claims that *if* coherence is quadratic in self-referential load, novelty scales as a power law, and metabolic power is quadratic, then the utility optimum sits at $\Gamma^* = \sqrt{\alpha/(\alpha + 2)}$, its approach is announced by diverging relaxation time and growing avalanches, and its sustained use is bounded by a power budget. It does not claim that any particular cognitive system has these functional forms, nor that Γ is directly measurable; it claims that the *shape* of the trade-off and the *signatures* of the optimum are generic to this class of models.

Failure modes. Three are visible from the literature. First, if the transition is conformal-like rather than Landau-like, the quadratic expansion fails: [4] shows that in a conformal phase transition the spectrum of light excitations changes abruptly at the critical point, so $\Psi(\Gamma)$ could collapse discontinuously before the quadratic model predicts any warning. Second, if the transition is first order in the sense of [5], where magnetization jumps discontinuously as the field is varied, small increments of Γ near the boundary could produce sudden, large coherence losses with no critical slowing down to announce them—the early-warning observables of Section 4.2 would be absent. Third, the kinetic transition of [11] shows that dynamics alone can create a transition whose critical temperature is not determined by any energy functional; a cognitive analogue would be a system whose coherence boundary is set by its update rules rather than by any measurable cost function, making Γ_c invisible to thermodynamic monitoring.

Falsification. The central quantitative claim is the interior optimum with the specific signatures. The observation that would falsify it is the following: measure, in a system with a tunable self-reference parameter, the joint curve of coherence-weighted novelty versus Γ , together with relaxation times and event-size distributions. If utility rises

monotonically to $\Gamma = 1$ without peaking, or peaks with no accompanying divergence of relaxation time or avalanche size, the phase-transition framing is wrong for that system. Conversely, a measured peak at intermediate Γ with critical signatures would support the model; the peak location then tests the predicted $\Gamma^* = \sqrt{\alpha/(\alpha + 2)}$ once α is fitted from the low- Γ novelty curve.

Limitations list.

- *Quadratic coherence.* $\Psi = 1 - \Gamma^2$ is the leading Landau term. Sensitivity: if Ψ contains a positive quartic term $+\beta\Gamma^4$, so that $\Psi(\Gamma) = 1 - \Gamma^2 + \beta\Gamma^4$, the stationarity condition $dU/d\Gamma = 0$ becomes $\alpha - (\alpha + 2)\Gamma^2 + (\alpha + 4)\beta\Gamma^4 = 0$, a quadratic in Γ^2 whose smaller root is the optimum; as β grows, this root moves toward $\Gamma = 0$, and the $2.05\times$ duty-cycle gain shrinks accordingly.
- *Power-law novelty.* The exponent α is a free parameter. Γ^* ranges from $\sqrt{1/3} \approx 0.577$ at $\alpha = 1$ to $\sqrt{3/5} \approx 0.775$ at $\alpha = 3$, so the qualitative conclusion (interior optimum) is robust but the numeric target is not.
- *Uncomputable boundary.* By analogy with [2], the cognitive Γ_c may be uncomputable, in which case all setpoints are estimates with finite-sample error, and the duty cycle must include a safety margin whose size is itself unknown.
- *Product-form utility.* $U = \Psi \cdot N$ assumes coherence and novelty are independent multiplicative factors. If incoherent outputs actively corrupt subsequent novelty (contagion), the true utility falls faster near the boundary and the optimum moves to lower Γ .
- *No contagion, no memory effects.* The model is memoryless; real systems accumulate damage across excursions, which shortens the effective τ_{exc} below the $24.39 t_0$ bound.
- *Landau exponent.* $\nu = 1$ is the mean-field value; finite-dimensional or fluctuation-dominated systems can differ, changing the warning-signal amplification factor of 29.3.
- *No empirical calibration.* Every number in Section 5 is a consequence of the assumed functional forms; none has been measured in a cognitive system.

Open questions. Whether the branching ratio m can be estimated online from event statistics fast enough to steer the system; whether the conformal-like and first-order-like failure modes can be distinguished from passive observation; and whether the metabolic and Gödelian boundaries always bind simultaneously or one dominates for particular architectures.

7. Practical Implications and Engineering Pathways

(a) Who would use the result and what changes. The nearest users are reservoir-computing researchers and engineers: [3] already implements the control loop our model prescribes—measure Local Lyapunov Exponents, gradient-descend gain and bias to push them toward the critical zero, and harvest the resulting memory-capacity and prediction gains—with reservoirs configured for the control loop. For them, the result changes the *target policy*: instead of parking the LLE at zero permanently, the duty-cycle result (Section 4.4) says to cycle between the critical point and a subcritical home point under a compute or power budget, with the cycle fraction set by the ratio of home to critical power draw. A second user group is teams building self-critiquing language-model pipelines, for whom the result converts "how much self-reflection to allow" from a hand-tuned hyperparameter into a monitored control variable with two early-warning observables (relaxation time and event clustering). For both groups the deliverable at this stage is the monitoring specification, not a controller.

(b) Readiness. This work is purely analytical and has not been peer reviewed: by the NASA Technology Readiness Level definitions it sits at TRL 1–2—basic principles observed and reported, technology concept formulated. It does not exceed TRL 3, and no claim of implementation or manufacturing readiness is made. The one component with

demonstrated empirical support is the control loop itself, since [3] reports considerable performance enhancements from driving reservoirs to the edge of stability; the duty-cycle policy on top of it is untested.

(c) Smallest next experiment. The minimal de-risking test is a modification of the existing PTA protocol: run a PTA-style setup (we assume SCR reservoirs with representative parameters) and measure, per epoch, not only memory capacity but also (i) the relaxation time of reservoir perturbations and (ii) the size distribution of output-change events, at LLE targets held at 0, -0.05 , and -0.2 . The model predicts relaxation time growing as the LLE target approaches zero and event sizes growing with it. The outcome that refutes the model in this substrate is a flat relaxation time or a flat event-size distribution as the LLE approaches zero while performance still improves—performance gain without critical signatures would mean the phase-transition framing is inapplicable here.

(d) Binding constraints. In order of first contact: (i) *estimation latency*—relaxation time and avalanche statistics require observation windows that grow as the boundary is approached (at $\tau = 100 \tau_0$, a window of many τ_0 is needed per estimate), which competes with the control loop's reaction time; (ii) *setpoint safety margin*—if Γ_c is uncomputable in the sense of [2], the margin between the operating point and the true boundary is unknown and must be bounded empirically, and the $2.05\times$ gain is sensitive to it; (iii) *contagion and memory*—damage accumulated across excursions, absent from the model, will bind before the clean power budget does in any long-running deployment. Cost and energy figures beyond the Landauer floor of Section 4.5 are open questions: the paper derives no substrate-specific power numbers, and we do not project them.

(e) Order-of-magnitude bounds (projection). The only quantified bound is the Landauer floor $E_{\min} = 2.967 \times 10^{-21}$ J per bit at $T = 310$ K, derived in Section 4.5. As a labeled projection with stated assumptions: if a duty-cycled system sustains novelty at rate $0.168 N_0$ (Section 4.4) and each novelty item requires at least 10^3 bits of information processing (assumption), the irreducible power floor is $0.168 N_0 \times 10^3 \times 2.967 \times 10^{-21}$ J per unit time—about $5 \times 10^{-19} N_0$ W—where N_0 is in items per second. This floor is many orders of magnitude below any real substrate's consumption and is stated only to fix the scale of the thermodynamic limit, not as an engineering estimate.

(f) Risks and misuse paths. The framework, if misread, licenses operating systems closer to incoherence on the strength of a $2.05\times$ number derived from assumed functional forms; a deployment that raises self-referential depth without the monitoring observables of Section 4.2 is applying the conclusion while discarding its conditions. The specific misuse is treating $\Gamma^* \approx 0.71$ as a universal constant rather than a consequence of $\alpha = 2$ and the quadratic Ψ .

8. Conclusion

This paper has converted a hand-tuning problem into a measurement problem. For a cognitive system whose coherence falls quadratically with self-referential load $\Psi(\Gamma) = 1 - \Gamma^2$, whose novelty rises as a power law $N(\Gamma) = N_0 \Gamma^\alpha$, and whose metabolic power is $P(\Gamma) = P_0 \Gamma^2$, the utility-maximizing operating point is $\Gamma^* = \sqrt{\alpha/(\alpha + 2)}$, which at the reference exponent $\alpha = 2$ sits at $\Gamma^* = 1/\sqrt{2} \approx 0.7071$ with utility $0.25 N_0$. Both endpoints of the interval destroy utility— $\Gamma = 0$ yields nothing new and $\Gamma = 1$ yields nothing coherent—so the interior optimum is not a compromise but a necessity of the product form of useful output.

Three results make the optimum operational rather than merely locational. First, the approach to the boundary is announced by observables: relaxation time diverges as $\tau = \tau_0 |\Gamma - \Gamma_c|^{-1}$, rising by a factor of about 29.3 between the optimum and a point at distance 0.01 from the boundary, and novelty avalanches grow from mean size $\langle s \rangle \approx 2$ at a subcritical home point to $\langle s \rangle = 100$ at branching ratio $m = 0.99$. Second, sustained dwelling near the boundary is bounded by power: with an exploration budget of $10 P_0 t_0$ and setpoints $\Gamma_e = 1/\sqrt{2}$, $\Gamma_h = 0.3$, a single excursion lasts at most $24.39 t_0$. Third, the bound can be beaten by cycling rather than dwelling: under a $0.3 P_0$ power ceiling, the

optimal duty cycle $f = 0.51220$ achieves $0.168 N_0$, a $2.05\times$ improvement over permanent conservatism. The Landauer floor $E_{\min} = 2.967 \times 10^{-21}$ J per bit at $T = 310$ K fixes the thermodynamic scale beneath all of these numbers.

The framework's standing depends on assumptions stated in Section 3 and stress-tested in Section 6. The quadratic coherence functional is a Landau-type universality assumption, not a measurement; the novelty exponent α is free, moving Γ^* between roughly 0.577 and 0.775 for $\alpha \in [1, 3]$; and the transition could be conformal-like or first-order-like, in which case the early-warning signatures fail precisely when they are needed most. The uncomputability analogy drawn from the Chaitin phase transition [2] implies that even a correct model cannot supply an exact setpoint, only an adaptively maintained estimate with an unknown safety margin. What survives these caveats is the shape of the claim: if a self-referential cognitive system has any transition between coherent sterility and incoherent productivity, then its useful output peaks in the interior, the peak leaves critical signatures, and energy budgets determine how long the peak can be occupied. The next step is empirical and small—measuring relaxation times and event-size distributions in an existing edge-of-stability reservoir system as its control parameter approaches the critical value—and it would either confirm the signatures or remove the phase-transition framing from consideration for that substrate.

Provenance and AI Involvement

- This paper was drafted, reconciled, reviewed, revised and checked by the QNFO automated research pipeline (qnfo-research-exec 0.19.4-codeagent). Language models were tools in that process; they are not authors. The named author is accountable for the content.
- Models: drafting by openai/gpt-oss-120b, zai-org/glm-5.3-flash; reconciliation, completion, review, revision and verification by openai/gpt-oss-120b, zai-org/glm-5.3-flash, moonshotai/kimi-k2.6, with the QNFO model router as a fallback.
- Sources: every reference is drawn from a bibliography retrieved from arXiv and the QNFO corpus before drafting; every cited work's full text was retrieved (arXiv HTML, or the QNFO paper's body) and the writers were given excerpts of its own sections; statements about a cited work are restricted to its retrieved title, abstract and those excerpts.
- Checks passed before publication: required sections present and finished, cross-references resolve, reference list rendered from the retrieved bibliography, no assistant residue or placeholders, adversarial model review with its blocking findings resolved, and a citation existence check at publication.
- Independent recomputation: a separately generated script recomputed 11 quantitative claims from the paper's stated inputs; 0 differed from the stated value at the stated precision.
- Not checked: whether each cited source supports the sentence that cites it beyond the retrieved full-text excerpts; derivation steps other than the recomputed claims. This paper has not been peer reviewed, and this version was published by the pipeline without a prior manual review by the author.

References

- [1] Photonic topological phase transition induced by material phase transition. arXiv:2406.08783. <https://arxiv.org/abs/2406.08783> [2] Chaitin Phase Transition. arXiv:2410.02600. <https://arxiv.org/abs/2410.02600> [3] Phase Transition Adaptation. arXiv:2104.10132. <https://arxiv.org/abs/2104.10132> [4] Conformal Phase Transition. arXiv:hep-ph/9703413. <https://arxiv.org/abs/hep-ph/9703413> [5] Holographic Magnetic Phase Transition. arXiv:0906.3892. <https://arxiv.org/abs/0906.3892> [6] Topological Electroweak Phase Transition. arXiv:2504.00923. <https://arxiv.org/abs/2504.00923> [7] The Electroweak Phase Transition. arXiv:hep-ph/9506418. <https://arxiv.org/abs/hep-ph/9506418> [8] The Chiral Phase Transition. arXiv:nucl-th/9608015.

<https://arxiv.org/abs/nucl-th/9608015> [10] Magnetic Fields Produced by Phase Transition Bubbles in the Electroweak Phase Transition. arXiv:hep-ph/9507429. <https://arxiv.org/abs/hep-ph/9507429> [11] Kinetic induced phase transition. arXiv:cond-mat/9907414. <https://arxiv.org/abs/cond-mat/9907414>