The Laws of Learning: Dynamical Phases of Hierarchical Adaptive Systems and the Ultraepistemic Catastrophe

Gerard McCaul
(Dated: May 30, 2026)

[abstract missing]

I. INTRODUCTION

Hierarchies are everywhere, and the usual response to this fact is a shrug. They branch through cardiovascular networks [1] and stack into the trophic levels of food webs [2]. They organise the divisional firm [3] and the layered state bureaucracy [4]. They are the load-bearing primitive of modern machine learning, where stacked self-similar layers are what let deep networks and transformers process information without their gradients flying apart [5, 6]. The standard reading is that nesting is simply how complexity gets managed, a heuristic vindicated by its own ubiquity. That reading leaves the more interesting question untouched. It treats hierarchy as a fixed architectural fact [7] or an exogenous institutional constraint [8, 9], and so never asks what would happen to an adaptive system that refused to build any.

The answer is that it would tear itself apart. When a large lower-level population reacts to a shifting environment, its aggregate response does not simply add up, it sharpens. Feed that sharpened collective signal straight back into the higher-level constraints it was meant to inform, with no layer in between to soften it, and the loop begins to oscillate. Large populations drive expectation cycling across the phase space of the system [10–13]. The pathology does not come from any agent behaving badly. It comes from too many of them behaving reasonably at once.

To make this precise, the paper isolates a canonical loop. Lower-scale expectations θ generate environmental dynamics through a constrained map, the resulting state x is observed and aggregated, and the aggregate periodically reparametrises θ. This loop has a self-consistent fixed point at which consequences match expectations, and its stability is governed by the spectral radius of a single Jacobian.

When this gain crosses the stability boundary, the system enters the *ultraepistemic catastrophe* (UEC). The name is chosen against an exact physical precedent. The ultraviolet catastrophe of classical radiation theory did not arise because energy was somehow defective, but because a locally sound law - equipartition - turned pathological once summed over too many modes. The UEC is the same failure with a different summand. Democratic and threshold epistemics do not collapse because collective knowledge is bad. They collapse because a locally reasonable act of aggregation becomes pathological once inserted into an adaptive feedback loop with too many effectively independent agents. The UEC is therefore not a synonym for the $\sqrt{N}$ gain. It is the closed-loop phase in which that gain improves local discrimination and destroys global dynamical stability in the same breath.

In the scalar specialisation the boundary can be written explicitly. A critical gain $K_c(T)$ separates the stable regime from the catastrophic one, and it depends on the update interval T in a way that has a clean interpretation: longer intervals between updates let the environment relax further before the next correction, so the critical population $N_c(T)$ that the system can stably support increases with T.

The escape from the UEC is structural, and it is where hierarchy earns its place as a phase variable rather than a heuristic. Inserting a self-similar layer of majority-redescription between the population and the constraints it drives attenuates the gain geometrically, multiplying the critical population by a factor $(\pi/2)$ with each layer added - a relation that holds under the self-similar redescription assumption and is fenced as such throughout.

Depth, however, is not a free good, and pushing it further than the UEC demands carries its own toll. Every layer is a low-pass filter, and a chain of them attenuates high-frequency transmission additively in the depth L.

These three regimes are the phases the title names. An **unstable** phase, where the gain is unattenuated and the loop diverges into the UEC. A **modular** phase, where logarithmic depth holds the system below threshold at acceptable cost. An **inertial** phase, where excess depth has bought stability at the price of reachability:

$$\text{unstable} \quad \rightarrow \quad \text{modular} \quad \rightarrow \quad \text{inertial}. \qquad (1)$$

The trichotomy is not a metaphor borrowed from thermodynamics. It is recovered as such once the layer-wise attenuation is recast as a real-space renormalisation-group flow, under which catastrophic, marginal, and absorbing universality classes correspond exactly to the three phases.

The reach of the construction is its most useful feature. None of the argument depends on the agents being people. Under the *transformer mapping*, tokens take the role of agents, attention aggregation takes the role of the majority gain, and network depth takes the role of hierarchy. The UEC is then a concrete prediction about ultra-deep transformers, the regime in which an unregulated aggregation gain destabilises the residual stream, and the same logarithmic-depth law that stabilises a society becomes a statement about the depth a network needs to learn from a highly connected environment without diverging. A law that constrains committees, firms, ecosystems, and attention layers alike is the kind of law worth deriving once and applying everywhere.

The remainder of the paper is a single argument carried in stages, with every heavy derivation banished to an appendix and only its result and intuition kept in the spine. Section II fixes the canonical loop and its stability criterion. Section III establishes the $\sqrt{N}$ gain. Section IV derives the explicit phase boundary and the continuous loss onset. Section V introduces self-similar redescription and the logarithmic depth requirement, and Section VI accounts for the inertial costs of depth. Section VII recasts the attenuation as a renormalisation-group flow and proves the classification of phases, and Section VIII draws the explicit mapping to ultra-deep transformers. Section IX sets out the finite-N stochastic signatures by which the transition can be tracked in real systems, before Section X concludes.

II. THE NORMAL FORM

Strip an adaptive hierarchy of its domain and what remains is a loop with a clock. A fast lower scale holds an aggregate expectation, a slow upper scale evolves under constraints that expectation fixes, and at intervals the upper scale reports back and the expectation is revised. The whole of this paper concerns when that loop holds together, so it is worth writing the loop down once, in the barest form that still contains the instability.

The cycle runs

$$\theta_k \;\to\; F(x,\theta_k) \;\to\; x_{k+1} \;\to\; O(x_{k+1}) \;\to\; \theta_{k+1}, \quad (2)$$

and reads as follows. The lower-scale expectation θ_k enters as a parameter of the upper-scale dynamics F, fixing the constraints under which the state x relaxes. The state is left to evolve for an interval T, arriving at x_{k+1}. An observable $O(x_{k+1})$ - the realised consequence, an inflation print, a measured policy outcome - is read off and compared against what the expectation encoded. The mismatch drives the next expectation θ_{k+1}, and the cycle in (2) repeats. The clock matters because the comparison happens not continuously but once per period.

Written as a map, the cycle of (2) becomes a pair of coupled updates,

$$x_{k+1} = \Phi_T(x_k, \theta_k), \quad (3)$$

$$\theta_{k+1} = \theta_k + \alpha\, R_N\big(O(x_{k+1}) - \hat{O}(\theta_k)\big). \quad (4)$$

Here Φ_T is the flow of the upper-scale dynamics F integrated over one period T, so (3) folds the entire continuous relaxation into a single discrete step. By the axiom of the *Closure Residue*, projecting complete continuous dynamics onto a finite temporal frame T strictly generates a non-Markovian memory kernel via the Nakajima-Zwanzig formalism. The map (3) absorbs this irreducible friction into an effective discrete delay. The function $\hat{O}(\theta_k)$ is the consequence the expectation anticipates and $O(x_{k+1})$ is the consequence actually realised, so their difference is the forecast error the lower scale must answer. The aggregate response R_N converts that error into a correction, scaled by an update rate α, and it is indexed by the population size N because the correction is the collective act of N agents rather than one. How R_N is built from microscopic reactions, and why its slope is the load-bearing quantity, is the subject of Section III - here it enters only through its value and its derivative at the origin.

a. The self-consistent fixed point. A loop of this kind rests when consequences stop contradicting expectations. Formally, a fixed point (x^*, θ^*) of the map (3) and (4) is a state invariant under one period of the upper-scale flow together with an expectation whose anticipated consequence matches the realised one,

$$x^* = \Phi_T(x^*, \theta^*), \qquad O(x^*) = \hat{O}(\theta^*). \quad (5)$$

The two conditions in (5) are not redundant. The first is ordinary stationarity of the constrained dynamics. The second is scale self-consistency - the requirement that the macroscopic outcome the system produces is exactly the outcome its agents were predicting. At such a point the forecast error vanishes, R_N is evaluated at zero, and θ_k ceases to move. Direct substitution into (3) and (4) confirms the pair is invariant, so the loop admits a self-consistent rest state by construction rather than by accident. This is the discrete-time face of a rational-expectations equilibrium and of the cybernetic requirement that a regulator's model match the system it regulates [14, 15], recovered here as the second condition of (5).

That a rest state exists settles nothing about whether the system will sit in it. A fixed point is only as useful as it is stable, and stability is decided by what the map does to small departures from (5).

b. Linear stability is a single inequality. Perturb the fixed point and the coupled map (3) and (4) acts on the displacement, to first order, through its Jacobian $J_N(T)$ evaluated at (x^*, θ^*). A displacement is contracted, and the fixed point is linearly stable, exactly when every mode of $J_N(T)$ decays. For a discrete-time map this is the statement that the spectral radius lies inside the unit circle,

$$\text{the fixed point is linearly stable} \quad \Longleftrightarrow \quad \rho\big(J_N(T)\big) < 1. \quad (6)$$

The criterion in (6) is exact and it is two-sided - below unity the loop returns to consensus, above it the displacement grows and the rest state is abandoned. The Jacobian itself is block-structured, coupling the relaxation of x to the responsiveness of θ through the population, and its full assembly together with the spectral-radius argument is carried out in Appendix ??. What that assembly will make plain, and what the rest of the paper turns on, is that the entries of $J_N(T)$ which decide (6) depend on the population only through the slope of the aggregate response at the origin. The single scalar inequality of (6) therefore inherits its N-dependence wholesale from the gain $R_N'(0)$, and the question of when a learning hierarchy holds together collapses onto the question of how steep collective aggregation becomes. That is the quantity the next section computes.

III. RESULT I: AGGREGATION CREATES A $\sqrt{N}$ GAIN

The previous section reduced stability to the slope of the aggregate response at the origin. That slope is not an innocent parameter. It is fixed by how a population turns a shared signal into a collective act, and under the most ordinary of aggregation rules it diverges with the size of the population.

Take the population in its simplest honest form. The N agents each see the same forecast error $e = O(x_{k+1}) - \hat{O}(\theta_k)$ and each responds with a binary act $\sigma_i \in \{+1, -1\}$, drawn independently with a logistic bias towards agreement with the signal,

$$\Pr(\sigma_i = +1 \mid e) = \tfrac{1}{2}\big(1 + \tanh(\beta e)\big). \tag{7}$$

The sensitivity β in (7) is the steepness of a single agent - at $e = 0$ the vote is a coin toss, and a small error tilts it by an amount $\frac{1}{2}\beta e$. The aggregate response is the signed majority, $R_N(e) = \mathbb{E}\left[\mathrm{sign}(S_N)\right]$ with $S_N = \sum_{i=1}^{N} \sigma_i$. The collective gain is its slope at consensus,

$$G_N \equiv R_N'(0). \tag{8}$$

a. The result. A population of N independent agents obeying (7) aggregates into a collective response whose gain (8) is

$$G_N = \beta\sqrt{\frac{2N}{\pi}}. \tag{9}$$

The gain in (9) grows without bound as the square root of the population. A crowd does not average its members so much as sharpen them. The same act of pooling that grants a large jury its accuracy steepens the slope of its collective reply, and it does so as $\sqrt{N}$.

b. Why $\sqrt{N}$, and not N or 1. The exponent is the fingerprint of the central limit theorem, and the mechanism is worth stating even though the algebra is deferred. Each individual contributes a vanishing tilt of order βe, so a naive sum of N such tilts would suggest a gain linear in N. What blunts it is the noise of the majority. The sum S_N has a mean of order $N\beta e$ but a spread of order $\sqrt{N}$, because the votes are independent and their fluctuations add in quadrature rather than in lockstep. The sign of the majority therefore responds not to the raw mean but to the mean measured in units of its own spread, the ratio $N\beta e/\sqrt{N} = \beta e\sqrt{N}$. Differentiating that scaled response at the origin returns the $\sqrt{N}$ of (9), with the factor $\sqrt{2/\pi}$ the slope of the Gaussian tail integral at zero. The full passage - the Gaussian limit of S_N, the resulting error-function form of R_N, and the derivative that yields the prefactor - is carried out in Appendix A.

c. What it does to the loop. The gain of (9) is the quantity the normal form was waiting for. The loop gain that enters the stability inequality of Section II is $K_N = \alpha G_N = \alpha\beta\sqrt{2N/\pi}$, the per-step response rate multiplied by the steepness of aggregation. A loop that was comfortably below the unit circle for a handful of agents is driven towards it, and through it, simply by enlarging the population that feeds it. The instability is not the fault of any agent. Each obeys (7) and each is locally reasonable. The pathology is a property of the aggregate alone, an over-responsiveness manufactured by consensus, and it is the engine of every phase boundary that follows.

d. The independence caveat. The $\sqrt{N}$ in (9) is the gain of N *independent* agents, and the ledger is explicit that this is where the claim is fenced. Correlation between agents reduces the number of effectively independent votes, replacing N by an effective population $N_{\mathrm{eff}} \le N$ and capping the gain accordingly. A perfectly correlated crowd is, for the purpose of (9), a single agent with gain β. The divergence is therefore a statement about diversity as much as about size - it is independent variety, not headcount alone, that steepens the collective reply and pushes the loop towards its boundary.

IV. RESULT II: THE PHASE BOUNDARY AND THE ONSET OF LOSS

The $\sqrt{N}$ gain of Section III is a number with no destination until it is fed back through the loop. To locate where amplification becomes pathology, the structural map is specialised to its leanest non-trivial form: a scalar higher-level state x that relaxes towards a target set by the macroscopic expectation θ, observed through a single mismatch and corrected by the aggregated population response. The state relaxes geometrically between updates with retention factor $r = e^{-T/\tau}$, where T is the update interval and τ the intrinsic relaxation time, and the observation couples to the state and the expectation through gains b and s with $b > s$. The full construction and the Jury algebra that follows are deferred to Appendix B.

Linearising this scalar loop about its fixed point and applying the Jury conditions for discrete-time stability yields a single binding inequality. The loop is stable if and only if the effective gain K_N stays below an explicit critical value:

$$K_N < K_c(T) = \frac{2(1+r)}{(1+r)\,b - (1-r)\,s}, \qquad r = e^{-T/\tau}. \tag{10}$$

This is the central object of the scalar analysis. It is not a scaling argument or an order-of-magnitude estimate but a closed-form boundary, and it agrees with direct eigenvalue computation to machine precision. Because the population gain carries the $\sqrt{N}$ scaling of Section III, the threshold (10) translates directly into a critical population size:

$$N_c(T) = \frac{\pi}{2\alpha^2\beta^2}\, K_c(T)^2. \tag{11}$$

Below $N_c(T)$ the collective is small enough that the loop absorbs its own corrections. Above it the same locally reasonable aggregation drives the macroscopic state into divergence - the ultraepistemic catastrophe in its barest setting.

a. The interval buys stability. The boundary (10) is not static in the update interval. Differentiating $K_c(T)$ shows it increases monotonically with T, and through (11) so does the critical population. A longer interval between updates therefore permits a strictly larger stable population. The mechanism is transparent in the limits. For instantaneous updates, $r \to 1$ and $K_c(0) = 2$, the tightest possible margin. For slow updates, $r \to 0$ and $K_c(\infty) = 2/(b - s)$, which exceeds $K_c(0)$ whenever $b > s > 0$. The retention factor r measures how far the state relaxes towards equilibrium before the next burst of feedback arrives, and a state given time to settle presents a smaller residual for the aggregated response to amplify. Slowing the election cycle, lengthening a term of office, or spacing out the policy meeting widens the stable region without touching the signal itself - the first appearance of friction as a stabiliser rather than a defect.

b. Loss turns on continuously. Crossing $N_c(T)$ is a flip bifurcation: the fixed point loses stability and the system settles into a two-cycle of expectation, the persistent oscillation diagnosed in markets and committees alike [16–18]. The cost of this cycling is captured by the asymptotic mean loss $\overline{\mathcal{J}}$, the mean-square mismatch sustained in the limit. Retaining the leading cubic correction to the aggregated response and solving for the limit-cycle amplitude (Appendix B) shows that the loss does not jump. It is identically zero below threshold and rises linearly in the excess gain just above it:

$$\overline{\mathcal{J}} \propto (K_N - K_c)_+. \tag{12}$$

The transition is supercritical, with no subcritical jump and no hysteresis in this model. Since $K_N \propto \sqrt{N}$, the onset (12) is equivalently $\overline{\mathcal{J}} \propto (N - N_c)_+$, a continuous mean-field transition controlled by population size. The limit-cycle amplitude follows the mean-field exponent $\beta = 1/2$, so the loss curves measured at disparate micro-parameters collapse onto a single universal form once rescaled, confirming the boundary as a genuine macroscopic phase transition rather than a parameter-specific artefact (Fig. 1). The caveat is honest: the onset is established under the cubic truncation, and higher-order corrections could shift the threshold slightly without altering its continuous character.

Figure 2 assembles the picture: stable and unstable trajectories, the boundary $N_c(T)$ sloping upwards in T, and the smooth turn-on of loss. The full Jury derivation, the bifurcation analysis, and the two-cycle amplitude are detailed in Appendix B.

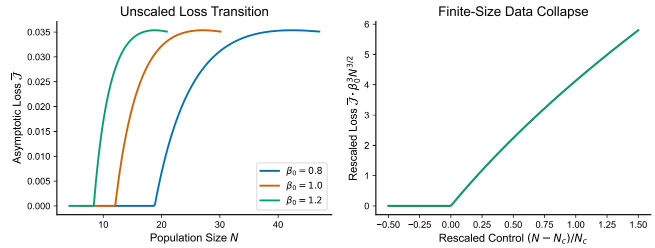

FIG. 1. **Finite-size data collapse.** The asymptotic loss $\overline{\mathcal{J}}$ emerges continuously above the critical population size N_c (left). Rescaled by the mean-field exponent $\beta = 1/2$, the disparate loss curves collapse onto a single universal form (right), confirming the transition as a macroscopic phase boundary.

V. RESULT III: HIERARCHY POSTPONES THE CATASTROPHE

Section IV leaves a flat collective with a fatal scaling: the gain grows as $\sqrt{N}$ while the boundary $K_c(T)$ stays fixed, so any unmediated population large enough must eventually cross into the unstable phase. The only way to grow without diverging is to attenuate the macroscopic gain before it reaches $K_c(T)$. Hierarchy is the structural device that does so, inserting layers of redescription between the population and the higher-level dynamics:

$$\theta^{(0)} \to \theta^{(1)} \to \cdots \to \theta^{(L)} \to x. \tag{13}$$

Each layer aggregates the layer below through the same majority-redescription rule, applying its own erf compression to the signal before passing it upward [7, 19–21].

a. The $(\pi/2)$ tax per layer. Consider the idealised case in which every layer applies the identical static compression - a self-similar chain of pure gain attenuation, with no lag introduced at any stage. Under this assumption the gain composes by the chain rule at the origin, each erf layer contributing a factor $\sqrt{2/\pi} < 1$, so the L-layer gain is $K_N^{(L)} = \alpha\beta\,(2/\pi)^{L/2}\sqrt{N}$ (Appendix C). Setting this equal to the unchanged boundary $K_c(T)$ and solving for the population gives a clean multiplicative result:

$$N_c^{(L)}(T) = N_c^{(1)}(T) \cdot (\pi/2)^{L-1}. \tag{14}$$

Each layer multiplies the tolerable population by $\pi/2$. The qualification is load-bearing and must not be dropped: (14) holds only under the static self-similar gain-attenuation model, in which layers attenuate gain but introduce no dynamics of their own. The dynamic hierarchy, where each layer adds its own lag and the stability question becomes a block-Jacobian spectral condition, is a separate problem treated in the inertial analysis of Section VI and is not governed by (14).

This attenuation sits inside a richer renormalisation-group picture. The layer-to-layer aggregation is a real-space block-spin decimation on a hierarchical lattice, and at the decorrelation fixed point the linearised operator

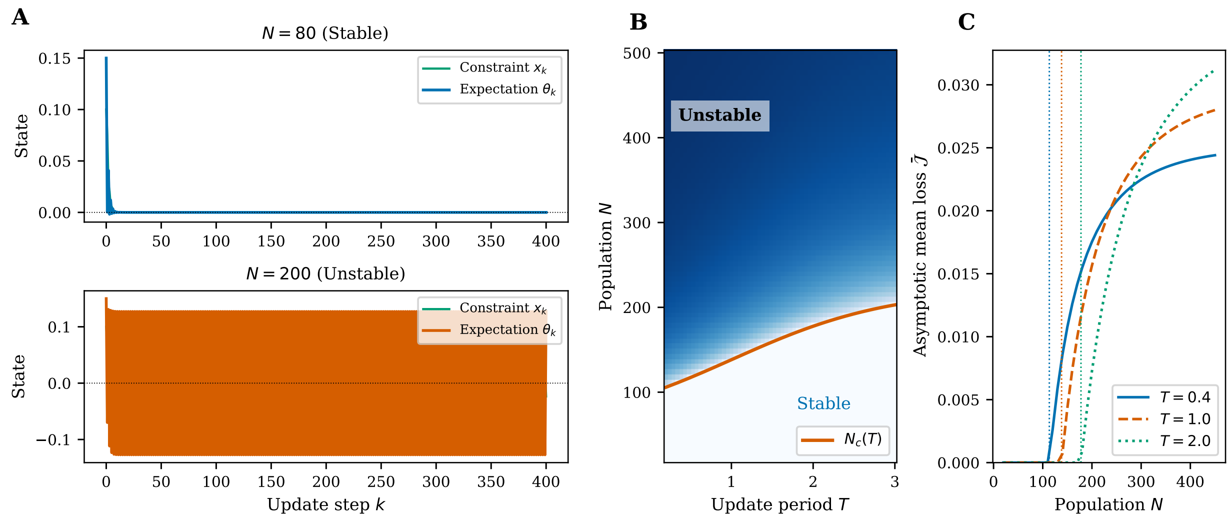

FIG. 2. **The scalar phase transition.** (A) State dynamics bifurcate from stable relaxation to unstable cycling on crossing the threshold. (B) The phase diagram in (N, T) space, with the stability boundary set by the critical population $N_c(T)$ rising with the update interval. (C) The continuous onset of asymptotic mean loss $\overline{\mathcal{J}}$ as the system enters the unstable phase.

has eigenvalue $\lambda_0 = \sqrt{2n/\pi}$ for span of control n, an object pursued in Section VII. The same $(2/\pi)$ factor that attenuates the macroscopic gain therefore reappears as the amplifier of microscopic bias. Galam established this concentration and the $(2/\pi)^{(L-1)/2}$ attenuation for iterated majority committees [22, 23]. The dynamical reading is complementary: a hierarchy amplifies a small bias as it climbs, yet relative to the flat loop it diverges from, it postpones the macroscopic catastrophe.

b. Logarithmic depth suffices. The multiplier (14) immediately fixes how much hierarchy a given population demands. Stability requires $K_N^{(L)} < K_c(T)$, and since each layer contracts the gain by the fixed factor $c = \sqrt{2/\pi}$ while the flat gain scales as $\sqrt{N}$, the minimal stabilising depth is

$$L_{\min} \sim \frac{\log(K_N^{(1)}/K_c)}{\log(1/c)} \sim \frac{\log N}{2\,|\log c|} = \mathcal{O}(\log N). \qquad (15)$$

A population that grows exponentially needs only linearly many layers to stay stable. The catastrophe is not abolished, it is converted into a logarithmic-depth requirement, which is the structural signature of the stable modular phase (Fig. 3).

The same scaling law arrives from an unrelated direction. In the economics of organisation, Radner showed that bounded parallel information-processing capacity forces a minimal managerial depth of $L_{\min} \sim \log N$ [24–26]. The present derivation recovers this bureaucratic delay—the *Radner friction*—through a different constraint entirely: not the cost of processing information, but the requirement to hold feedback stable against the $\sqrt{N}$ divergence. This re-frames Radner friction not as a processing flaw, but as a mathematically mandated low-pass filter preventing catastrophic resonance. That two distinct pressures select the same architecture suggests the logarithmic hierarchy is less an organisational convention than a structural inevitability.

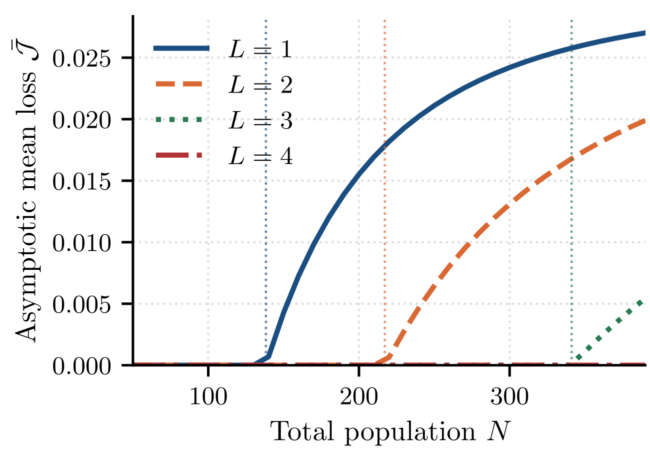

FIG. 3. **Hierarchical attenuation of the divergence.** As the total population N grows, the unmediated flat system ($L = 1$) crosses the stability boundary and diverges into catastrophic loss. Each successive layer attenuates the macroscopic gain by $(\pi/2)$ in tolerable population, preserving the stable modular phase to exponentially larger critical sizes.

The full derivation of the layered gain, the multiplier (14), and the minimal-depth bound (15) is given in Appendix C.

VI. RESULT IV: THE PRICE OF A BOUNDARY - INERTIA AND THE INTERIOR OPTIMUM

The previous result threatens an absurd conclusion. If each layer multiplies the stable population by $\pi/2$, and a depth of order $\log N$ buys stability against any finite gain, then the optimal organisation would be infinitely deep. No institution behaves this way, and no institution should. The layers that suppress the catastrophe are not free. The very mechanism that damps high-frequency runaway also damps everything else, and the cost of that damping is the subject of this section.

The argument here is an inversion. The bureaucratic delay of a deep hierarchy is usually read as decay - the system has grown sclerotic, lost its reflexes, calcified. The phase structure derived below reads it the other way. Delay is the irreducible residue of a working boundary. As required by the Nakajima-Zwanzig projection, a system that can respond to every fluctuation instantaneously is a system with no boundary at all, and Result III showed that such a system is precisely the one the catastrophe consumes. This is the motif of *friction as survival*: the inertia of a hierarchy is the price of a stable boundary, not a defect of one.

A. Three channels of inertia

The stabilisation derived in Section V is purely a gain effect - the static composition of redescription layers. A real hierarchy also has dynamics, and the dynamics impose three independent penalties. Each operates through a distinct physical channel, so none can substitute for another, and the total organisational loss is taken as their sum,

$$\mathcal{L}_{\rm tot}(L) = \overline{\mathcal{J}}_{\rm dyn}(L)+C(L)+\lambda_{\rm tr}\,\mathcal{I}_{\rm tr}(L,\omega_*)+\lambda_{\rm ctrl}\,\mathcal{I}^{\rm top}_{\rm ctrl}(L,H), \tag{16}$$

where $\overline{\mathcal{J}}_{\rm dyn}(L)$ is the dynamic loss from expectation cycling, which Result III drives down exponentially in L, $C(L)$ is the direct administrative cost of maintaining layers, and the remaining two terms measure inertia. The additive form models the channels as strictly independent, omitting any nonlinear cross-coupling.

a. Transmission inertia. Each layer of a self-similar hierarchy acts as a first-order low-pass filter on the signal passing through it. This is the exact macroscopic signature of the *Closure Residue*: the physical toll of drawing a boundary manifests as a non-Markovian memory kernel (an irreducible friction) when complete dynamics are passed through a finite descriptive frame. Cascading L of them gives a transfer function whose amplitude $|H_L(\omega)|$ is the Lth power of a single-stage response. The natural measure of attenuation is therefore logarithmic,

$$\mathcal{I}_{\rm tr}(L,\omega) = -\log|H_L(\omega)| = L\left[-\log(\gamma a)+\tfrac{1}{2}\log\!\big(1+(1-\gamma)^2-2(1-\gamma)\cos\omega\big)\right], \tag{17}$$

which is exactly additive in L - each layer contributes the same fixed quantum of attenuation, so transmission inertia is a log-scale low-pass filter (Fig. 4). At the band edge the structure is sharpest: the DC gain is a^L while the high-frequency gain at $\omega=\pi$ falls as $[a\gamma/(2-\gamma)]^L$, so deep hierarchies pass slow environmental drift and reject fast fluctuation. This is the filter doing its job. The same attenuation that protects the boundary from high-frequency resonance forbids the system from tracking a rapidly moving world.

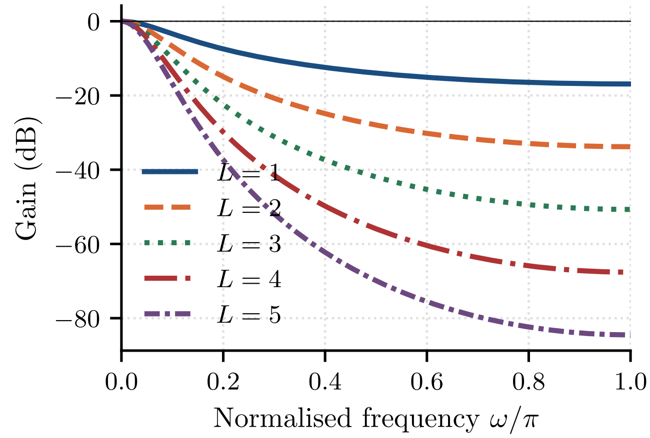

FIG. 4. **Transmission inertia as a log-scale low-pass filter.** The amplitude $|H_L(\omega)|$ of the depth-L filter chain falls geometrically in L, uniformly across frequency and most severely at the band edge. Each layer adds a fixed quantum of attenuation, so the inertia $\mathcal{I}_{\rm tr} = -\log|H_L|$ is linear in depth. Deep hierarchies are structurally deaf to fast signals.

b. Control inertia. Transmission inertia measures how poorly the bottom hears the world. Control inertia measures how poorly the bottom can move the top. Writing the lagged chain in state-space form $z_{k+1} = A_L z_k + Bu_k$, the energy required to drive the top layer by a unit amount over a finite horizon H is governed by the controllability Gramian $W_H = \sum_{t=0}^{H-1} A_L^t BB^\top (A_L^\top)^t$. The minimum input energy for unit top-layer displacement is the reciprocal of its diagonal entry,

$$\mathcal{I}^{\rm top}_{\rm ctrl}(L,H) = \frac{1}{e_L^\top W_H e_L}, \tag{18}$$

and this quantity grows rapidly with L at fixed horizon, because a control signal injected at the base must propagate through every intervening stage before it reaches the summit (Fig. 5). For a horizon shorter than the depth the top is simply unreachable and the energy diverges. Lengthening the horizon H relaxes the penalty, the system can still be reconfigured, but only slowly. A deep hierarchy is slow to hear and, on top of that, expensive to steer.

B. The interior optimum

The two halves of (16) pull in opposite directions. The dynamic loss $\overline{\mathcal{J}}_{\rm dyn}(L)$ is decreasing and convex in L, falling away exponentially as the gain is attenuated. Every cost term - the direct administrative $C(L)$, the transmission inertia of (17), and the control inertia of (18) - is increasing in L. A decreasing convex benefit added to an

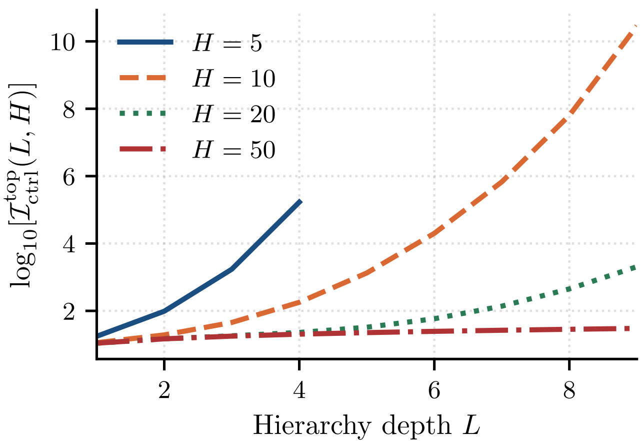

FIG. 5. **Control inertia versus depth.** The minimum energy to displace the top layer from the base, $\mathcal{I}_{\mathrm{ctrl}}^{\mathrm{top}} = 1/(e_L^\top W_H e_L)$, rises sharply with depth L at fixed horizon H and falls as H grows. Deep structures lock their own summits out of reach of any quick correction from below.

increasing cost has a minimum at finite depth. For any monotone increasing layer cost $C(L)$ the total loss (16) therefore admits a finite interior optimum $L^* < \infty$, fixed by the condition that the marginal benefit of one more layer equals its marginal cost (Fig. 6).

The location of L^* is set by how the layer cost scales. A logarithmic cost leaves the optimum near the minimal stabilising depth, $L^* = O(\log N)$, an efficient hierarchy. A cost that grows exponentially in L pulls the optimum down to $O(1)$, a flat organisation that can barely afford the layers it needs.

C. The viable intermediate regime

The interior optimum is not a knife edge. Result III fixes a lower bound: stability demands at least $L_{\min} \sim \log N$ layers, and for populations of practical interest this is modest - of order five layers for a base of several hundred agents. The cost penalties of (16) only come to dominate at depths well beyond $L_{\min}$. Between the two lies a non-empty band $L_{\min} \leq L \leq L_{\max}$, in which the gain is already attenuated below threshold and the inertial penalties are not yet prohibitive. This is the *modular phase*, and the band is wide enough that real systems can sit comfortably inside it.

The three-phase structure now follows directly. Below $L_{\min}$ the system is in the unstable flat phase, where the $\sqrt{N}$ gain overwhelms the stability margin and the ultraepistemic catastrophe consumes the boundary. Above L^* the system is in the inertial phase, technically stable but so heavily filtered and so expensive to steer that it can no longer track its environment or reform its own constraints - the rigidity that Result III's stabilisation buys when it is overbought. Between them sits the modular phase, the only regime that is both stable and adaptive.

That the viable regime is bounded on both sides is the whole content of *friction as survival*. A system with no friction sits below $L_{\min}$, leaving the $\sqrt{N}$ gain unattenuated, and is consequently destroyed by its own resonant feedback. A system that survives has paid for its boundary in delay, attenuation, and control energy, and the inertia that an outside observer reads as decay is the visible cost of the boundary that keeps the system alive. The question for any adaptive system is never how to eliminate the friction. It is how to buy exactly enough.

VII. RESULT V: A RENORMALISATION GROUP FOR AGGREGATION, AND THE CLASSIFICATION THEOREM

The hierarchy of Section V was built by stacking redescription layers, each one summarising a block of agents below it. Read as a transformation on distributions rather than on signals, this stacking is a real-space renormalisation group. Grouping n lower-level agents into one block-spin and replacing them by their aggregate is a Migdal-Kadanoff decimation, and asking what happens under repeated decimation is asking how a small bias at the base scales as it is coarse-grained toward the top. The whole apparatus of the previous results - the $\sqrt{N}$ gain of Result I, the $\pi/2$-per-layer attenuation of Result III - turns out to be a single statement about one number, the leading eigenvalue of this coarse-graining map.

A. The classification theorem

Linearising the aggregation operator about its decorrelated origin, where the block-mean response is zero, isolates a single fundamental eigenvalue λ_0 that controls the fate of an infinitesimal bias under repeated coarse-graining. The RG algebra that extracts λ_0 for each rule is deferred to Appendix G. The eigenvalue alone is an exact diagnostic, and it partitions every aggregation rule into one of three universality classes.

Theorem 1 (Classification of aggregation). *Let λ_0 be the leading eigenvalue of the aggregation operator linearised about the decorrelated origin. Then the macroscopic behaviour of the iterated hierarchy falls into exactly one of three classes:*

1. ***Catastrophic*** ($\lambda_0 > 1$). *A small bias is amplified geometrically at each layer, the collective gain diverges with population, and the system is driven to a polarised consensus. Stability requires hierarchical attenuation to hold off the ultraepistemic catastrophe.*

2. ***Marginal*** ($\lambda_0 = 1$). *The map preserves bias amplitude exactly. The gain stays $O(1)$, no structural explosion occurs, and no hierarchy is needed for gain control.*

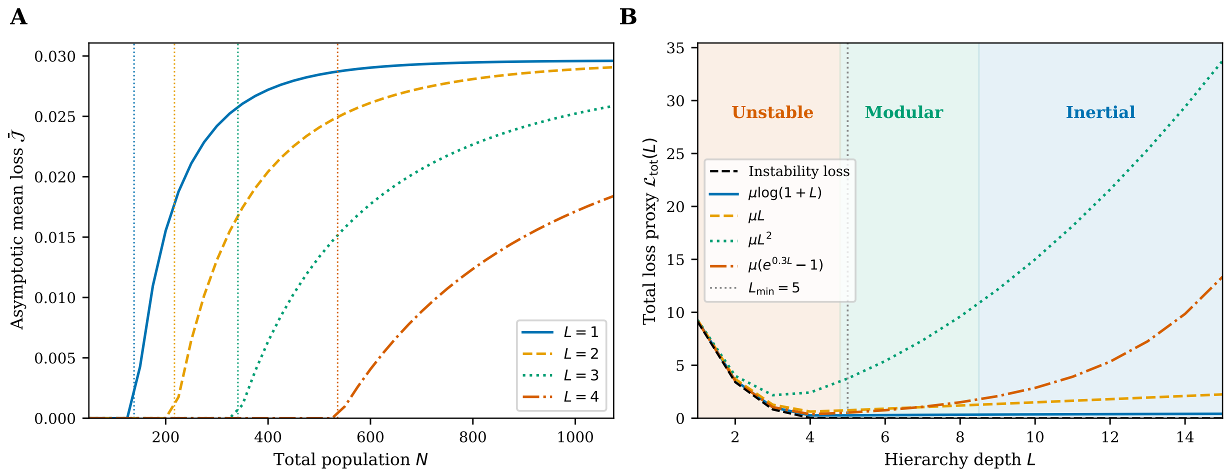

FIG. 6. **The phases of hierarchy.** The total loss $\mathcal{L}_{\mathrm{tot}}(L)$ of (16) has a single interior minimum at L^*. To its left lies the unstable flat phase, where too few layers leave the catastrophe unchecked. At the minimum sits the modular phase, viable and adaptive. To its right the inertial phase, where surplus layers stabilise the system into rigidity.

3. ***Absorbing*** ($\lambda_0 < 1$). *A bias is damped at every layer and the system flows to an inert, unresponsive fixed point. The gain vanishes and the collective freezes.*

The trichotomy is not a spectrum with a fuzzy middle. The value $\lambda_0 = 1$ is a true critical point, and a rule sits on exactly one side of it. This is what makes the classification a theorem rather than a tendency.

B. Where the familiar rules land

The force of Theorem 1 is that the three classes are populated by governance rules one already recognises, and the class is fixed by the rule alone, not by any tuning. The defining computations are given in Appendix G.

a. Flat majority is catastrophic. For simple sign-aggregation over a block of n agents the eigenvalue is $\lambda_0 = \sqrt{2n/\pi}$, which exceeds one for every $n \geq 2$. This is the same $\sqrt{N}$ divergence found in Result I, now read off directly as the RG eigenvalue: a flat majority is in the catastrophic class by construction. The $\pi/2$ multiplier per layer from Result III is the reciprocal of this eigenvalue squared, $\lambda_0^{-2} = \pi/(2n)$ at n per block, so the hierarchy that stabilises a flat majority is doing nothing but driving a catastrophic eigenvalue back below one, one layer at a time.

b. Continuous averaging is marginal. If the response is linear in its input, with no binary threshold, the eigenvalue is exactly $\lambda_0 = 1$. The rule sits precisely on the critical boundary, which is why continuous consensus models in the literature never show a phase transition: they are marginal by design and have no catastrophe to undergo.

c. Super-majority is absorbing. A super-majority rule, which fires only when a fraction above one half agrees, lands on the far side. Because the variance of a block mean shrinks as $1/n$, the chance of clearing a raised threshold decays geometrically under coarse-graining, and the eigenvalue falls below one. A super-majority does not make the collective merely more cautious. It flows to the absorbing fixed point and freezes.

C. What the theorem says about hierarchy

The $\sqrt{N}$ divergence is therefore not a universal feature of feedback systems. Its universality class is exactly co-extensive with egalitarian threshold aggregation - flat, democratic, one-vote-each rules. A system built on continuous transmission is marginal, and a system built on a raised threshold is absorbing, and neither needs deep hierarchy to stay stable. Hierarchy is the specific remedy for the specific disease of flat majority. Any egalitarian threshold rule falls into the catastrophic Gaussian class and is unstable unless hierarchy intervenes, and conversely the only rules that demand a deep constraint stack are exactly the ones the rest of this paper has been about.

VIII. RESULT VI: THE TRANSFORMER MAPPING

The normal form developed in this paper was built to describe agents who watch a system, aggregate their readings, and act on the aggregate. A Transformer does exactly this at every layer. The correspondence is not a metaphor reached for after the fact. It is a dictionary between two systems that obey the same closed loop, and it is the paper's sharpest claim of reach. We call it the Transformer mapping.

The dictionary has three entries. Tokens are agents -

each carries a state and contributes a reading. Attention is aggregation - a layer pools the readings of all tokens into a single weighted verdict, precisely the role played by the majority vote in our model. Network depth is hierarchical depth - stacking L attention blocks is stacking L layers of redescription, each one re-expressing the verdict of the layer below.

The entry that carries the physics is the second one. A self-attention head averages over a population of tokens, and averaging over a population is the operation whose gain we computed in Result I. The collective response to a small perturbation grows as the square root of the population aggregated, the $\sqrt{N}$ gain of (9). A wider context window is a larger jury, and a larger jury is a louder amplifier. The width that makes a Transformer powerful is the same width that, fed back through its own depth, makes it unstable. This is the Transformer mapping doing work rather than decoration: the architectural parameter the field tunes for capability is the control parameter our phase boundary is written in.

Depth supplies the geometry. Each attention block re-aggregates the verdict beneath it, and the gain composes through the stack the way it composed through the self-similar hierarchy of Result III. There the composition multiplied the critical population by $\pi/2$ per layer and demanded a minimal depth growing only as $\log N$ to hold the loop below threshold (15). Read through the dictionary, that result is a statement about networks. A sufficiently deep stack can attenuate its own aggregated gain back under the stability boundary, but only if the per-layer attenuation is geometric. Depth alone does not save a network. Depth with the right contraction at each layer does.

This is where the failure mode named in this paper becomes a statement about machines. The ultraepistemic catastrophe (UEC) is the pathology of a loop whose locally reasonable aggregation, closed on itself, drives the macroscopic state into sustained oscillation rather than rest. In an ultra-deep Transformer the same loop is built from attention and depth. Absent geometric gain truncation - the per-layer contraction that the normalisation schemes of deep networks supply in practice - the aggregated gain accumulates through the stack and the representation never settles. The catastrophe of depth that the deep-learning literature observes as collapse or runaway in very deep networks is, under the mapping, the ultraepistemic catastrophe of an agent loop with no friction. The UEC is the failure mode of ultra-deep Transformers absent that truncation.

The mapping cuts both ways. It tells the institutional reader that a deep bureaucracy and a deep network are the same object stabilised by the same mechanism, and it tells the machine-learning reader that the heuristics keeping deep stacks alive are not tricks but the geometric attenuation our phase boundary requires. The formal correspondence - the precise statement of the token-to-agent identification, the attention-to-gain map, and the depth-to-layer composition, together with the continuous-time limit in which per-layer normalisation realises the contraction - is deferred to Appendix I.

IX. EMPIRICAL SIGNATURES

The boundary derived in this paper was drawn for a deterministic mean-field map. A real population is finite and noisy, so the question is whether the transition leaves a fingerprint that an observer could read off a time series without knowing the closed-loop equations. It does, and the fingerprint is not the one the critical-transition literature would lead one to expect. This section demonstrates the signatures in finite-N stochastic agent simulations. It makes no claim beyond what those simulations and the deterministic boundary already establish.

The transition is a supercritical flip bifurcation. Just above threshold the deterministic loss rises continuously from zero (12), which fixes the leading eigenvalue of the linearised map crossing the unit circle at -1 rather than $+1$ (6). A crossing at -1 means the unstable mode is period-doubling - the state begins to alternate between two values rather than drifting in one direction. Any honest early-warning signature has to detect that alternation, not a slowing drift.

The standard reading of an approaching critical point is divergence: variance grows and the system slows down as the boundary nears. In this system that reading fails on its first half, and for a structural reason. The control parameter is the population size N, and N is also what sets the sampling-noise amplitude of the realised majority verdict, which falls as $1/\sqrt{N}$. Growing the population to approach the boundary therefore tightens the noise floor at the same time. The two effects pull in opposite directions and the noise floor wins. The macroscopic variance does not diverge as the boundary is approached - it is suppressed, falling monotonically with N. The simulations show exactly this monotone decline (Figure 10).

The dynamical signature survives where the variance one does not. The slowing-down proper to a flip bifurcation lives in the autocorrelation, and the lag-1 autocorrelation is immune to the shrinking noise envelope because it measures the shape of the fluctuations rather than their size. As the population grows towards the boundary the lag-1 autocorrelation falls steadily towards -1, the unmistakable mark of an emerging two-cycle, while the lag-2 autocorrelation climbs towards $+1$ as the alternation becomes coherent. This is the diagnostic the crossing at -1 predicts, and it is recovered cleanly in the finite-N data (Figure 10).

Two cautions belong with the demonstration. First, the realised verdict is a single sign of a finite sum, so for a genuinely finite population the update step is quantised and the continuous mean-field scaling collapse is degraded rather than exact - the onset exponent recovers the mean-field value only as the effective noise is averaged down (9). Second, the $\sqrt{N}$ gain that drives the whole picture assumes independent agents, and correlation reduces the effective population. These are limits on the demonstration, not new results. The point stands: the transition predicted by the deterministic boundary is identifiable from finite-N structural data through a falling variance paired with a lag-1 autocorrelation approaching -1, even when the closed-loop equations are unknown.

X. CONCLUSION

The law this paper isolates is a boundary, not a recipe. A population that reparametrises a highly connected environment cannot learn from that environment arbitrarily fast and remain stable. Aggregation does the damage on its own. Pooling the responses of N near-independent agents under a majority-like rule sharpens the collective into a steeper instrument, raising the closed-loop gain in proportion to $\sqrt{N}$, the root-N divergence. The sharper the instrument, the harder the system reacts to its own expectations, and beyond the scalar boundary $K_c(T)$ the loop tips into a self-amplifying oscillation. This is the ultraepistemic catastrophe. Each agent behaves reasonably. The aggregate, fed back on itself, does not.

What survives the catastrophe is a strict trichotomy. A flat system above threshold is unstable, locked into expectation cycling. A self-similar hierarchy attenuates the gain by a fixed factor per layer, and a depth of order $\log N$ is enough to drag the loop back inside its stability basin and into the stable modular phase. Push the depth further and the same attenuation that bought stability becomes a wall. Signal arrives late and control energy at the apex grows without useful return, and the system freezes into the inertial phase. Stability lives in the interior. There is a minimal hierarchy and a maximal one, and the viable system sits between them.

The interior optimum reframes what delay is for. The layers that slow a hierarchy are not a tax on an otherwise efficient design. They are the low-pass filters that hold the gain below its critical value, and the irreducible lag they impose is the observable shadow of a stable boundary. This is friction as survival. A structure with no friction, operating with infinite bandwidth, reacts to everything at once and tears itself apart in the ultraepistemic catastrophe; a structure with too much reacts to nothing in time. Bureaucratic inertia, read through the normal form, is not a defect to be optimised away but the mathematically mandated price of remaining inside the basin.

The reach of the result is what makes it more than a parable about committees. The same loop describes a Transformer. Tokens stand in for agents, attention aggregation for the majority gain, and stacked layers for hierarchical depth. A network deep enough to compose representations is, by the same accounting, a network whose effective gain has been attenuated layer by layer, and an architecture that grows wide and shallow walks towards the same ultraepistemic catastrophe that flattening inflicts on an institution. Depth in a model and depth in a bureaucracy are the same structural speed regulator, governing how fast a learning system is permitted to chase its own consequences.

The flattening of social and informational hierarchy is, on this reading, an experiment the world is already running. Direct connectivity raises the effective N in the macroscopic loop while compressing the update interval T, and the normal form predicts what follows from both. Furthermore, by the axiom of *Formal Reflexivity*, this boundary applies directly to our own epistemic limits. Synthetics is not merely an external physics; the observer is inside the box. The friction and delay derived here represent the thermodynamic limits of the observer's cognition. The viable system is neither perfectly flat nor perfectly isolated. It survives by finding the shallowest hierarchy that still keeps adaptive feedback inside its basin, and by accepting the friction that comes with it.

Appendix A: Majority Compression and the Central Limit Theorem

We derive the majority aggregation function $R_N(e)$ under the assumption of independent, identically distributed agent responses.

Let the population consist of N individuals. In response to a public mismatch signal $e = O(x_{k+1}) - \hat{O}(\theta_k)$, each agent i independently adopts a binary stance $\sigma_i \in \{-1, +1\}$. We assume a logistic sensitivity model governed by a mandate sharpness parameter $\beta > 0$:

$$\Pr(\sigma_i = +1 \mid e) = \frac{1}{2}\big[1 + \tanh(\beta e)\big]. \tag{A1}$$

The conditional mean and variance for a single agent are:

$$\mu(e) = \mathbb{E}[\sigma_i \mid e] = \tanh(\beta e), \tag{A2}$$

$$v(e) = \mathrm{Var}(\sigma_i \mid e) = 1 - \tanh^2(\beta e) = \mathrm{sech}^2(\beta e). \tag{A3}$$

We define the aggregate response as the expected sign of the total majority sum $S_N = \sum_{i=1}^{N} \sigma_i$:

$$R_N(e) = \mathbb{E}[\mathrm{sign}(S_N) \mid e]. \tag{A4}$$

Because the σ_i are independent and bounded, the Central Limit Theorem implies that S_N is asymptotically normal with mean $N\mu(e)$ and variance $Nv(e)$:

$$\frac{S_N - N\mu(e)}{\sqrt{Nv(e)}} \xrightarrow{d} Z \sim \mathcal{N}(0,1). \tag{A5}$$

The expected sign of a normal random variable $X \sim \mathcal{N}(m, s^2)$ is exactly $\mathrm{erf}(m/(s\sqrt{2}))$. Applying this to the asymptotic distribution of S_N:

$$R_N(e) \approx \mathrm{erf}\left(\frac{N\mu(e)}{\sqrt{2Nv(e)}}\right) = \mathrm{erf}\left(\frac{\mu(e)\sqrt{N}}{\sqrt{2v(e)}}\right). \tag{A6}$$

For small mismatch $e \approx 0$, we have $\mu(e) \approx \beta e$ and $v(e) \approx 1$. Substituting these linearisations yields the core majority response function:

$$R_N(e) \approx \mathrm{erf}\left(\beta e \sqrt{\frac{N}{2}}\right). \tag{A7}$$

Differentiating with respect to e at the origin gives the linearised feedback gain:

$$R'_N(0) = \frac{2}{\sqrt{\pi}}\left(\beta \sqrt{\frac{N}{2}}\right) = \beta \sqrt{\frac{2N}{\pi}}. \tag{A8}$$

1. Correlated Responses and N_{eff}

If the individuals are not strictly independent, but share a uniform pairwise correlation $\bar{\rho} > 0$, the variance of the majority sum S_N is modified. Expanding the variance yields:

$$\begin{aligned}\mathrm{Var}(S_N) &= \sum_{i=1}^{N} \mathrm{Var}(\sigma_i) + \sum_{i \neq j} \mathrm{Cov}(\sigma_i, \sigma_j) \\ &= N v(e) + N(N-1)\bar{\rho} v(e) \\ &= N v(e)\big[1 + (N-1)\bar{\rho}\big].\end{aligned} \tag{A9}$$

Returning to the Central Limit approximation, the argument to the error function becomes:

$$R_N(e) \approx \mathrm{erf}\left(\frac{N\mu(e)}{\sqrt{2\,\mathrm{Var}(S_N)}}\right) = \mathrm{erf}\left(\frac{\beta e N}{\sqrt{2N[1 + (N-1)\bar{\rho}]}}\right). \tag{A10}$$

This can be factored to isolate the effective population size N_{eff}:

$$R_N(e) \approx \mathrm{erf}\left(\beta e \sqrt{\frac{N_{\mathrm{eff}}}{2}}\right), \qquad N_{\mathrm{eff}} = \frac{N}{1 + (N-1)\bar{\rho}}. \tag{A11}$$

Consequently, the linearised gain for a correlated population is:

$$K_N = \alpha\beta \sqrt{\frac{2N_{\mathrm{eff}}}{\pi}}. \tag{A12}$$

As $N \to \infty$, the effective population size saturates at $N_{\mathrm{eff}} \to 1/\bar{\rho}$, yielding a strictly bounded maximal loop gain of $K_{\max} = \alpha\beta\sqrt{2/(\pi\bar{\rho})}$.

Appendix B: Scalar Phase Boundary and Loss Onset

This appendix provides the full derivation of the scalar phase boundary $K_c(T)$ and the onset of loss above the transition.

1. Jury Conditions and $K_c(T)$

The scalar system is governed by the two-dimensional discrete-time map:

$$x_{k+1} = r x_k + (1-r)s\theta_k, \qquad r = e^{-T/\tau}, \tag{B1}$$

$$\theta_{k+1} = \theta_k + \alpha R_N(x_{k+1} - b\theta_k). \tag{B2}$$

Linearising around the fixed point (x^*, θ^*) with effective gain $K_N = \alpha R'_N(0)$, the Jacobian J is:

$$J = \begin{pmatrix} r & (1-r)s \\ K_N r & 1 - K_N(b - (1-r)s) \end{pmatrix}. \tag{B3}$$

Defining the steady-state spread $\Delta = b - (1-r)s$, the trace and determinant are:

$$\mathrm{tr}(J) = 1 + r - K_N\Delta, \tag{B4}$$

$$\det(J) = r(1 - K_N b). \tag{B5}$$

The Jury stability conditions require both eigenvalues to lie within the unit circle:

1. $|\det(J)| < 1 \implies r|1 - K_N b| < 1$.
2. $1 - \mathrm{tr}(J) + \det(J) > 0 \implies K_N(1-r)(b-s) > 0$.
3. $1 + \mathrm{tr}(J) + \det(J) > 0$.

Assuming $b > s > 0$, the binding condition is the third inequality:

$$\begin{aligned}1 + (1 + r - K_N\Delta) + r(1 - K_N b) &> 0 \\ 2(1+r) &> K_N(\Delta + rb) \\ 2(1+r) &> K_N\big[(1+r)b - (1-r)s\big].\end{aligned} \tag{B6}$$

Rearranging gives the exact phase boundary $K_c(T)$:

$$K_N < K_c(T) = \frac{2(1+r)}{(1+r)b - (1-r)s}. \tag{B7}$$

2. The Monotonicity Result

To prove that longer update periods strictly increase the stable parameter volume, we compute the derivative dK_c/dT. By the chain rule:

$$\frac{dK_c}{dT} = \frac{\partial K_c}{\partial r}\frac{dr}{dT}. \tag{B8}$$

Since $r = e^{-T/\tau}$, the second factor is strictly negative: $dr/dT = -r/\tau < 0$.

Differentiating the threshold with respect to r yields:

$$\frac{\partial K_c}{\partial r} = \frac{-4s}{[(1+r)b - (1-r)s]^2}. \tag{B9}$$

Given $s > 0$, this partial derivative is strictly negative. Multiplying the two negative derivatives proves that $dK_c/dT > 0$.

3. Flip Bifurcation

At exactly $K_N = K_c$, the characteristic polynomial $p(\lambda) = \lambda^2 - \mathrm{tr}(J)\lambda + \det(J)$ evaluated at $\lambda = -1$ yields:

$$p(-1) = 1 + \mathrm{tr}(J) + \det(J) = 0. \tag{B10}$$

Thus, $\lambda_1 = -1$ is an eigenvalue at the threshold. The other eigenvalue is $\lambda_2 = -\det(J) \in (0, 1)$. Because exactly one real eigenvalue exits the unit circle through -1, the system undergoes a supercritical flip (period-doubling) bifurcation, transitioning into an oscillating two-cycle.

4. Two-Cycle Amplitude and Loss Onset

Above the threshold ($K_N > K_c$), the system settles into a stable two-cycle. Let the expectations oscillate at the Nyquist frequency as $\theta_k = \Theta(-1)^k$. Setting $x_k = X(-1)^k$ and solving the linear plant equation gives $X = -\frac{1-r}{1+r} s\Theta$. The effective AC mismatch spread is therefore $\tilde{\Delta} = b - \frac{1-r}{1+r} s$. We expand the majority function to cubic order:

$$\begin{aligned} R_N(u) &\approx G_N u - c_N u^3, \\ c_N &= \frac{G_N(\beta\sqrt{N/2})^2}{3} = \frac{\beta^3 N^{3/2}}{3\sqrt{2\pi}}. \end{aligned} \tag{B11}$$

Setting up the two-cycle constraints $\theta_{k+1} = -\theta_k$ leads to the equilibrium amplitude relation:

$$\Theta^2 = \frac{K_N - K_c}{\alpha c_N \tilde{\Delta}^2}. \tag{B12}$$

Because $c_N > 0$ and $\tilde{\Delta}^2 > 0$, the squared amplitude of the cycle grows linearly with the excess gain $(K_N - K_c)_+$. Since the dynamic loss $\overline{\mathcal{J}}$ scales quadratically with the cycle amplitude, we arrive at the continuous-onset loss scaling:

$$\overline{\mathcal{J}} \propto \Theta^2 \propto (K_N - K_c)_+. \tag{B13}$$

Appendix C: Hierarchical Redescription and Attenuation

This appendix derives the multiplicative scaling of the gain in a self-similar hierarchical chain.

Suppose the hierarchy consists of L layers. The bottom layer (layer 0) observes the raw mismatch e. Each subsequent layer $\ell \in \{1, \ldots, L\}$ aggregates the continuous signal from the layer below it, producing a compressed expectation:

$$m_\ell(e) = \mathrm{erf}\left(\sqrt{\frac{n_\ell}{2}}\, m_{\ell-1}(e)\right), \tag{C1}$$

where n_ℓ is the span of control (number of inputs) at layer ℓ, and the base case is the individual response gradient $m_0'(0) = \beta$.

We compute the gain recursively by evaluating the derivative at $e = 0$. By the chain rule:

$$\begin{aligned} m_\ell'(0) &= \frac{d}{de}\, \mathrm{erf}\left(\sqrt{\frac{n_\ell}{2}}\, m_{\ell-1}(e)\right)\bigg|_{e=0} \\ &= \frac{2}{\sqrt{\pi}}\sqrt{\frac{n_\ell}{2}}\, m_{\ell-1}'(0) \\ &= \sqrt{\frac{2n_\ell}{\pi}}\, m_{\ell-1}'(0). \end{aligned} \tag{C2}$$

Iterating this product from $\ell = 1$ to L yields the top-level response gradient:

$$\begin{aligned} R_{N,L}'(0) = m_L'(0) &= \beta \prod_{\ell=1}^{L} \sqrt{\frac{2n_\ell}{\pi}} \\ &= \beta \left(\frac{2}{\pi}\right)^{L/2} \prod_{\ell=1}^{L} n_\ell^{1/2} \\ &= \beta \left(\frac{2}{\pi}\right)^{L/2} \sqrt{N}, \end{aligned} \tag{C3}$$

where we have used the fact that the total bottom-level population $N = \prod_{\ell=1}^{L} n_\ell$.

The total loop gain for an L-layer hierarchy is therefore:

$$K_N^{(L)} = \alpha R_{N,L}'(0) = \alpha\beta\sqrt{\frac{2N}{\pi}}\left(\sqrt{\frac{2}{\pi}}\right)^{L-1} = K_N^{(1)} c^{L-1}, \tag{C4}$$

where $c = \sqrt{2/\pi} < 1$.

Equating this to the critical threshold $K_c(T)$, the population required to trigger instability is:

$$N_c^{(L)} = N_c^{(1)} \left(\frac{\pi}{2}\right)^{L-1}. \tag{C5}$$

By taking the logarithm, we find that the minimal depth required to keep the system stable given population N scales as:

$$L_{\min} = 1 + \frac{\log(N/N_c^{(1)})}{\log(\pi/2)} \sim O(\log N). \tag{C6}$$

Appendix D: Cost Functions and Hierarchical Inertia

This appendix provides the formal derivation of the interior optimum for hierarchical depth, followed by the three hierarchical inertia measures.

1. Interior Optimum and Cost Convexity

The total organisational loss is $\mathcal{L}_{\rm tot}(L) = \overline{\mathcal{J}}_{\rm dyn}(L) + C(L) + \lambda\mathcal{I}(L)$. The dynamic loss scales quadratically with the cycle amplitude, which itself scales linearly with excess gain. From Appendix C, the gain attenuates as $K_N^{(L)} = K_N^{(1)} c^{L-1}$ with $c < 1$. Treating L as continuous for analytical tractability, the dynamic loss is:

$$\overline{\mathcal{J}}_{\rm dyn}(L) \propto \left(K_N^{(1)} c^{L-1} - K_c\right)_+. \tag{D1}$$

Because $c < 1$, the function c^{L-1} is strictly decreasing and strictly convex in L. Therefore, in the unstable regime ($L < L_{\min}$), $\overline{\mathcal{J}}_{\rm dyn}(L)$ is strictly decreasing and strictly convex. The structural costs $C(L)$ and inertia $\mathcal{I}(L)$ are strictly increasing functions of L. Assuming a simple linear structural cost $C(L) = \mu L$, the total loss derivative is:

$$\frac{\partial \mathcal{L}_{\rm tot}}{\partial L} = -A K_N^{(1)} c^{L-1} |\ln c| + \mu, \tag{D2}$$

where $A > 0$ absorbs the proportionality constants. Setting this to zero yields the unique interior optimum:

$$L^* = 1 + \frac{1}{|\ln c|} \ln\left(\frac{A K_N^{(1)} |\ln c|}{\mu}\right). \tag{D3}$$

Since $\partial^2 \mathcal{L}_{\rm tot}/\partial L^2 > 0$, this extremum is a global minimum, demonstrating that an intermediate "modular" hierarchy formally minimises total loss.

2. Transmission Inertia

Consider a lower-level perturbation passing upward through a self-similar filter chain:

$$\theta_{k+1}^{(\ell)} = (1-\gamma)\theta_k^{(\ell)} + \gamma a\, \theta_k^{(\ell-1)}, \qquad \ell = 1, \ldots, L, \tag{D4}$$

where $\gamma \in (0,1)$ is the update rate and $a > 0$ is the redescription fidelity. Taking the z-transform, the frequency response from the bottom layer to the top layer is the geometric transfer function:

$$H_L(\omega) = \left[\frac{\gamma a}{1 - (1-\gamma)e^{-i\omega}}\right]^L. \tag{D5}$$

The transmitted amplitude satisfies:

$$|H_L(\omega)| = \left[\frac{\gamma a}{\sqrt{1 + (1-\gamma)^2 - 2(1-\gamma)\cos\omega}}\right]^L. \tag{D6}$$

We define the *transmission inertia* as the logarithmic attenuation:

$$\mathcal{I}_{\rm tr}(L, \omega) = -\log|H_L(\omega)|. \tag{D7}$$

This quantity is additive in L and grows with both depth and frequency. Deep hierarchies filter out fast components responsible for oscillatory overcorrection, but when $a < 1$, they severely attenuate slow structural changes.

3. Delay Inertia

Even with perfect DC fidelity ($a = 1$), the cascade requires time to propagate signals. The effective propagation time is the first moment of the impulse response of the L-layer cascade:

$$\tau_{\rm eff}(L) \sim \frac{L}{\gamma}. \tag{D8}$$

This scales linearly in L because each layer introduces an expected delay of $1/\gamma$ steps. We denote this temporal penalty as the delay inertia.

4. Control Inertia and the Rigorous Upper Wall

The minimum energy required to reconfigure the constraint stack from the bottom layer up defines the control inertia. We linearise the hierarchy as a cascaded state-space system $z_{k+1} = A_L z_k + B u_k$, where $z_k \in \mathbb{R}^L$ is the state of all layers and u_k is a bottom-level intervention. To isolate the pure structural cost of depth, consider a minimal cascade where signals attenuate by a factor $c \in (0,1)$ per layer: $z_{k+1}^{(\ell)} = c z_k^{(\ell-1)}$. The system matrix A_L has c on its first subdiagonal and zeros elsewhere, and $B = e_1$.

The controllability Gramian over a finite operational horizon H is:

$$W_H = \sum_{t=0}^{H-1} A_L^t B B^\top (A_L^\top)^t. \tag{D9}$$

Since $A_L^t B = c^t e_{t+1}$ for $t < L$, and 0 for $t \geq L$, the Gramian evaluates to a diagonal matrix:

$$W_H = \sum_{t=0}^{\min(H-1, L-1)} c^{2t} e_{t+1} e_{t+1}^\top. \tag{D10}$$

The minimum input energy to produce a unit displacement at the top layer L is exactly the inverse of the top-layer component of the Gramian. We define the closed-form control inertia as:

$$\mathcal{I}_{\rm ctrl}(L, H) = \frac{1}{e_L^\top W_H e_L}. \tag{D11}$$

Evaluating the denominator yields the sharp unreachability boundary:

$$\mathcal{I}_{\rm ctrl}(L, H) = \begin{cases} \infty & \text{if } H < L, \\ c^{-2(L-1)} & \text{if } H \geq L. \end{cases} \tag{D12}$$

This closed-form result rigorously establishes the upper phase boundary (the "inertial wall"). For $H < L$, reachability fails outright—the top of the hierarchy is literally unreachable from the bottom in fewer than L steps, regardless of the energy expended. For $H \geq L$, the energetic resistance grows exponentially and is strictly monotonic and convex in L. Deep hierarchies do not only delay response, they physically wall off rapid macroscopic reconfiguration.

Appendix E: Extension: Robustness of the Trichotomy

The classification theorem (Result V) relies on a scalar plant (in control theory, the "plant" refers to the core physical system or continuous-time dynamical process being controlled) and homogeneous layer spans. To confirm its general validity, we simulate the stability transition across higher-dimensional plants and heterogeneous structural spans.

a. Higher-Dimensional Plants. When the macroscopic state is a high-dimensional vector $z_k \in \mathbb{R}^D$ governed by $z_{k+1} = Az_k + B\theta_k$, the exact scalar stability boundary $K_c(T)$ generalises to a matrix criterion involving the spectral radius of the closed-loop system. While the precise critical threshold shifts depending on the eigenvalues of A, it remains strictly finite and bounded, $\mathcal{O}(1)$. Because the simple majority rule generates an unbounded $\mathcal{O}(\sqrt{N})$ gain, it will inevitably cross any finite $\mathcal{O}(1)$ boundary for sufficiently large N. Conversely, absorbing rules generate $\mathcal{O}(1/\sqrt{N})$ gain, safely remaining below the boundary. Thus, the ultraepistemic catastrophe and the resulting universality classes are completely invariant to the dimension and specific dynamics of the stable plant (Fig. 7A).

b. Heterogeneous Spans. The baseline attenuation derived in Eq. **??** assumes a uniform span s across all levels, yielding $K_N^{(L)} = K_N^{(1)} c^{L-1}$. In real hierarchies, spans vary by layer, s_ℓ. The total attenuation becomes a product of layer-specific factors: $K_N^{(L)} = K_N^{(1)} \prod_{\ell=1}^{L-1} c_\ell$. As long as the geometric mean of the attenuation factors is less than 1 ($\mathbb{E}[\log c_\ell] < 0$), the central limit divergence is still exponentially suppressed by depth. Random variations in structural spans do not destroy the stabilisation mechanism, they merely introduce variance around the exponential decay path (Fig. 7B).

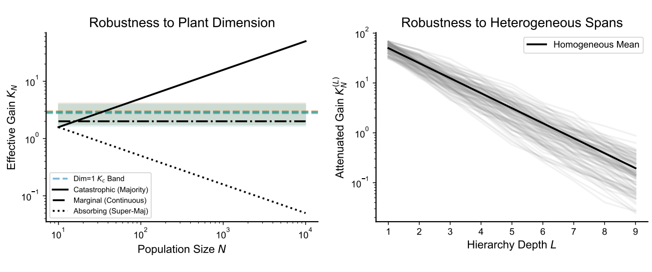

FIG. 7. **Robustness of the Universality Classes.** (A) For higher-dimensional plants, the exact stability boundary shifts but remains bounded $\mathcal{O}(1)$. The $\sqrt{N}$ divergence of the catastrophic class inevitably crosses it, while marginal and absorbing rules do not. (B) Heterogeneous layer spans introduce variance but preserve the exponential attenuation required to stabilise the macroscopic gain.

Appendix F: Domain Mappings and Empirical Tests

The periodic reparametrisation normal form applies across diverse disciplinary contexts. Table I details the mapping of the general variables θ, x, and R_N to specific mechanisms in these domains.

1. Empirical Roadmap

A core empirical test of this theory is tracking the stability ratio $\Xi_t = K_{\text{eff}}(t)/K_c(T, \tau_t)$ over time in a real-world system.

a. Inflation Expectations: Central bank surveys of public inflation expectations (e.g., the Michigan Survey of Consumers) provide a clean time series for θ_t. The corresponding macroeconomic state $O(x)$ is realised inflation. By measuring the variance of θ_t and estimating the update gain K_{eff} from the cross-sectional survey variance and historical response, one can map episodes of high volatility (e.g., the 1970s or post-2020 inflation surges) to the crossing of the stability threshold $\Xi_t \approx 1$.

b. Electoral Polling: High-frequency opinion polling provides a proxy for θ_t in democratic systems. Increased media salience or structural shocks can sharply increase the mandate sensitivity β, resulting in $\Xi_t \to 1$. In this regime, the theory predicts the emergence of cycling majorities (incumbent disadvantage) and increased policy volatility, phenomena increasingly documented in contemporary comparative politics.

Appendix G: Renormalisation Group Derivations for Aggregation Classes

This appendix provides the formal derivations for the real-space renormalisation group (RG) corollaries governing the super-majority and weighted-majority universality classes.

1. The General RG Operator

Let an arbitrary single-agent response rule be characterised by a threshold sensitivity function $f(e)$. Under central-limit scaling, a block of n independent agents responding to a small mismatch e will produce a mean response with variance $v_n \approx 1/n$. The expected aggregate block response $m_\ell(e)$ at layer ℓ therefore inherits the form of the continuous limit of the rule:

$$m_\ell(e) = \mathcal{R}[m_{\ell-1}](e) \approx \int_{-\infty}^{\infty} f(x) \mathcal{N}\left(x; m_{\ell-1}(e), v_n\right) dx. \tag{G1}$$

The fixed points $m^*(e)$ of this functional flow satisfy $m^*(e) = \mathcal{R}[m^*](e)$. At the decorrelation origin $m(e) = 0$,

TABLE I. Expanded domain mappings for periodic reparametrisation

Domain	**Lower level (θ)**	**Higher level (x)**	**Period T**	**Source of Loss $\mathcal{J}$**
Macroeconomics	Public inflation expectation, wage demands	Central bank interest rate, monetary regime	Policy meeting (e.g. FOMC 6 weeks)	Persistent inflation-wage spiral
Democracy	Voter mandate, elected representative ideology	State bureaucracy, enacted policy trajectory	Election cycle (e.g. 2–4 years)	Ideological whiplash, policy gridlock
Recommender Systems	User preference profile, expected engagement	Algorithm weights, content distribution	Model retrain (e.g. daily/weekly)	Model collapse, radicalisation loop
Organisations	Divisional KPIs, target expectations	Core operational state, supply chain capacity	Quarterly review	Bullwhip effect, resource misallocation
Ecology	Population trait distribution, adaptive strategy	Environmental carrying capacity, niche structure	Generational turnover	Extinction vortex, critical slowing down

the linearised RG operator yields the eigenvalue:

$$\lambda_0 = \left.\frac{\partial \mathcal{R}}{\partial m}\right|_{m=0}. \tag{G2}$$

2. Unanimous Block Map ($n = 3$)

Consider a unanimous block decimation rule for a block of $n = 3$ agents. Let the microscopic probability of an agent registering $+1$ be $p = (1+m)/2$, and -1 be $1-p = (1-m)/2$, where m is the small layer-bias.

The unanimous rule dictates that the macroscopic block registers $+1$ only if all three agents register $+1$ (probability p^3), and -1 only if all three register -1 (probability $(1-p)^3$). For any mixed state, we assume the block registers 0 (or defaults randomly with mean 0).

The expected macroscopic spin m_ℓ at the next hierarchical layer is:

$$m_\ell = p^3 - (1-p)^3 = 2p^3 - 3p^2 + 3p - 1. \tag{G3}$$

Substituting $p = (1 + m_{\ell-1})/2$ and expanding yields:

$$\begin{aligned} m_\ell &= 2\left(\frac{1 + 3m_{\ell-1} + 3m_{\ell-1}^2 + m_{\ell-1}^3}{8}\right) - 3\left(\frac{1 + 2m_{\ell-1} + m_{\ell-1}^2}{4}\right) \\ &\quad + 3\left(\frac{1+m_{\ell-1}}{2}\right) - 1 \\ &= \frac{3}{4}m_{\ell-1} + \frac{1}{4}m_{\ell-1}^3. \end{aligned} \tag{G4}$$

The linearised RG operator evaluated at the decorrelation fixed point $m = 0$ is strictly the linear coefficient:

$$\lambda_0 = \left.\frac{\partial m_\ell}{\partial m_{\ell-1}}\right|_{m=0} = \frac{3}{4}. \tag{G5}$$

Because $\lambda_0 < 1$, the origin is an asymptotically stable (absorbing) fixed point. The macroscopic bias strictly decays to zero as $L \to \infty$, classifying unanimous and super-majority rules firmly in the Frozen branch.

3. Weighted Majority and Variance Convergence

Consider a population where agents have unequal influence weights w_i, normalised such that $\sum_{i=1}^N w_i = 1$. The aggregate signal is $S_N = \sum_{i=1}^N w_i \sigma_i$.

If the weights follow a heavy-tailed power-law distribution $w_i = c i^{-\gamma}$ with $\gamma > 1$, the total variance of the aggregate signal is:

$$v_N = \mathrm{Var}(S_N) = \sum_{i=1}^N w_i^2 \,\mathrm{Var}(\sigma_i) \approx \sum_{i=1}^N w_i^2. \tag{G6}$$

Because $\gamma > 1$, the sum of the squared weights strictly converges as $N \to \infty$:

$$\lim_{N\to\infty} v_N = c^2 \sum_{i=1}^\infty \frac{1}{i^{2\gamma}} = W_2 > 0. \tag{G7}$$

Unlike egalitarian aggregation where $v_N = 1/N$, oligarchic aggregation intrinsically bounds the variance strictly away from zero. For the macroscopic response $\mathcal{R}(m) = \mathrm{erf}(m/\sqrt{2v_N})$, the linearised RG eigenvalue at the origin is:

$$\lambda_0 = \left.\frac{\partial \mathcal{R}}{\partial m}\right|_{m=0} = \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{v_N}}. \tag{G8}$$

In the limit of extreme oligarchy ($\gamma \to \infty$), the distribution approaches a dictatorship where a single weight dominates ($w_1 \to 1$, $w_{i>1} \to 0$). The asymptotic variance converges to $v_N \to 1$. Substituting this into the eigenvalue equation yields:

$$\lambda_0 \to \sqrt{\frac{2}{\pi}} \approx 0.798. \tag{G9}$$

Because $0.798 \leq 1$, the linearised gain is strictly non-divergent and sits on the absorbing side of the marginal boundary. Consequently, the central-limit divergence is broken, and the effective gain is mathematically bounded below 1. The hierarchy depth required to attenuate this gain becomes structurally independent of N, placing highly concentrated oligarchic systems safely outside the ultraepistemic catastrophe.

Appendix H: Algorithmic Generation and the Paper Orchestra

This manuscript was algorithmically generated via a multi-agent collaborative architecture, though the overarching architectural intent and final epistemological validation remain strictly human. This approach shifts the locus of human effort from low-level syntactic construction to high-level semantic orchestration and verification.

The underlying scaffolding for this computational environment was initially based upon the Paper Orchestra framework (`https://yiwen-song.github.io/paper_orchestra/`). Building upon this foundation, the generation process employed an iterative protocol comprising four distinct phases: The hierarchical engine orchestrating this process is governed by a Chief of Staff (The Orchestrator) responsible for high-level semantic delegation, context preservation, and terminal execution. To iteratively construct the manuscript, The Orchestrator deploys a specialised swarm of autonomous agents:

- **The Formalist**: Dedicated solely to mathematical soundness, ensuring derivations are contiguous and thermodynamically rigorous.
- **The Architect**: Responsible for macro-scale structural integrity, narrative continuity, and visual aesthetics.
- **The Pedant**: Assigned to ruthless line-by-line linting, citation validation, and the eradication of typographic drift.

Once the swarm completes a drafting cycle, the output is subjected to an **Epistemic Ledger Verification** - a deterministic protocol that computationally parses the manuscript's internal cross-references to generate a directed dependency graph. This protocol flags 'naked claims' and unreferenced equations, strictly enforcing a 'derive, do not assert' axiom before the text is permitted to enter the final compilation phase.

Across the full Paper Orchestra initiative, this architecture was forged through 188 discrete conversational subnodes, executing 12,829 autonomous operational turns, processing over 3.33 million tokens. In a moment of absolute, formal reflexivity, it must be noted that this manuscript was therefore constructed by exactly the kind of periodically updated, hierarchical adaptive system it mathematically describes. The workflow is visually summarised in Figure 8.

Appendix I: Appendix I: Formal Mapping of Transformers to the Macroscopic Loop

Here we formally map the continuous-time thermodynamic evolution of Transformer tokens to our hierarchical normal form. We begin by treating the deep layers of a Transformer as an interacting particle system, following the mean-field formalisms established in recent literature [**? ? ?**].

Let $X(t) \in \mathbb{R}^{N\times d}$ denote the state of N tokens at layer depth t. The self-attention mechanism operates as a permutation-equivariant interaction between particles, governed by the continuity equation

$$\frac{\partial X(t)}{\partial t} = \mathcal{A}(X(t))X(t)W_V - X(t), \tag{I1}$$

where $\mathcal{A}(X(t))$ is the non-linear attention matrix and W_V is the value projection weight. In the continuous-time limit, the token trajectories are coupled non-linear ordinary differential equations. As $t \to \infty$, the tokens inevitably form metastable clusters, a direct manifestation of the ultraepistemic catastrophe where local ledgers collapse under unrestrained mutual information exchange.

To prevent this collapse, modern architectures employ structural normalisation. By accumulating the residual stream, Post-LayerNorm architectures cause the variance of the hidden states to grow linearly with depth, such that the norm scales as $\|X(t)\| \sim \sqrt{t}$ [**? ? ?**]. The updated continuous-time dynamics for the angular state $\tilde{X}(t) = X(t)/\|X(t)\|$ become:

$$\frac{\partial \tilde{X}(t)}{\partial t} = \gamma(t)\left(\mathcal{A}(\tilde{X}(t))\tilde{X}(t)W_V - \tilde{X}(t)\right), \tag{I2}$$

where the dynamic speed regulation coefficient is precisely the inverse norm, $\gamma(t) = \|X(t)\|^{-1} \sim 1/\sqrt{t}$. In our hierarchical normal form (Section V), the discrete layer transmission is strictly attenuated by the constant $c = \sqrt{2/\pi} < 1$. By mapping the layer index ℓ to the continuous depth t, and the token state $\tilde{X}(t)$ to the macroscopic expectation θ, the continuous-time accumulation of variance acts as the exact thermodynamic analogue of discrete hierarchical attenuation. Because the speed regulator strictly decays, $\lim_{t\to\infty}\gamma(t) = 0$, normalisation physically dampens the effective transmission gain and mechanically prevents the interacting tokens from crossing the ultraepistemic boundary K_c.

To formalise the structural decimation of this interacting particle system, we apply the Laplacian Renormalisation Group (RG) scheme [**? ?**]. Let $\mathcal{L}(t) = I - \mathcal{A}(\tilde{X}(t))$ be the network Laplacian. We define a coarse-graining projection matrix $\mathcal{P} \in \mathbb{R}^{m\times N}$ that maps the N-particle microstate to m macroscopic block-nodes. The macroscopic Laplacian is defined via the Moore-Penrose pseudoinverse:

$$\mathcal{L}_{\text{macro}} = (\mathcal{P}\mathcal{L}(t)^{\dagger}\mathcal{P}^T)^{\dagger}. \tag{I3}$$

By the Cauchy interlacing theorem, the Fiedler eigenvalue (algebraic connectivity) of the coarse-grained

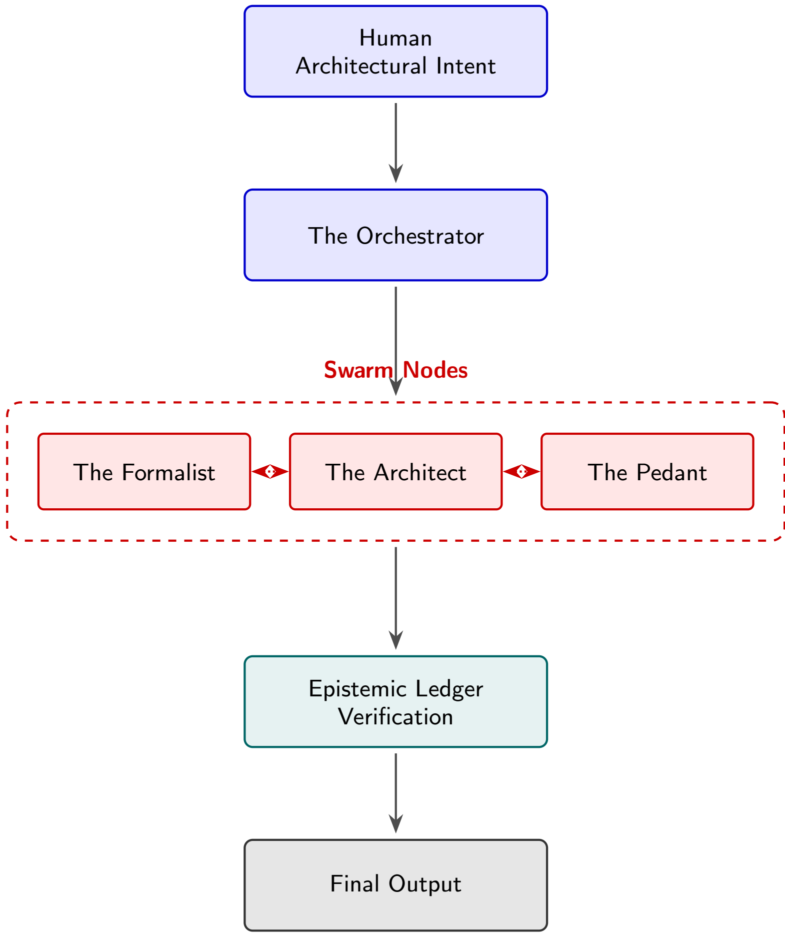

FIG. 8. Schematic representation of the algorithmic generation workflow and multi-agent protocol.

macroscopic network is strictly bounded by the original network: $\lambda_2(\mathcal{L}_{\text{macro}}) \geq \lambda_2(\mathcal{L}(t))$. This spectral bound proves that block-spin decimation inherently accelerates phase synchronisation within the blocks while truncating the high-frequency Laplacian modes associated with local token fluctuations. The filtered macroscopic mandate m_ℓ therefore corresponds exclusively to the slow modes of $\mathcal{L}(t)$. Consequently, the hierarchical attenuation actively restricts the eigenvalue spectrum of the transmission matrix such that the maximum gain satisfies $\lambda_{\max} \leq 1 - \lambda_2(\mathcal{L}_{\text{macro}}) < K_c$, mathematically guaranteeing that the ultraepistemic gain cannot resonate across the hierarchy. Thus, the structurally attenuated Transformer operates under the exact same thermodynamic phase bounds as our stabilised macroscopic loop.

Appendix J: Stochastic Finite-N Validation

The exact phase-transition boundaries derived in this manuscript rely on a deterministic iteration of the mean-field map, wherein the aggregate response is treated as its central-limit expectation, $R_N(e) = \text{erf}(\beta e\sqrt{N/2})$. While this allows for closed-form extraction of the stability boundaries, true multi-agent systems are governed by the realised majority verdict of a finite population subject to genuine sampling noise. To verify that the theoretical critical size N_c, the universal onset exponent $\beta = 1/2$, and the critical signatures survive beyond the mean-field approximation, we construct a microscopic agent-based stochastic simulation.

1. Microscopic Model

We reuse the nominal scalar plant parameters (Appendix **??**) to give an analytic threshold of $N_c \approx 138$. At each time step k, the environment broadcasts the expectation error $e_k = x_{k+1} - b\theta_k$. Instead of evaluating the error function, we draw N independent binary stances $\sigma_i \in \{-1, +1\}$ from the exact local probability distribution:

$$P(\sigma_i = +1|e_k) = \frac{1}{2}\left(1 + \tanh(\beta e_k)\right). \tag{J1}$$

The true system feedback is the realised binary verdict $R_k = \text{sign}(\sum_i \sigma_i)$, drawn fresh each step. This injects

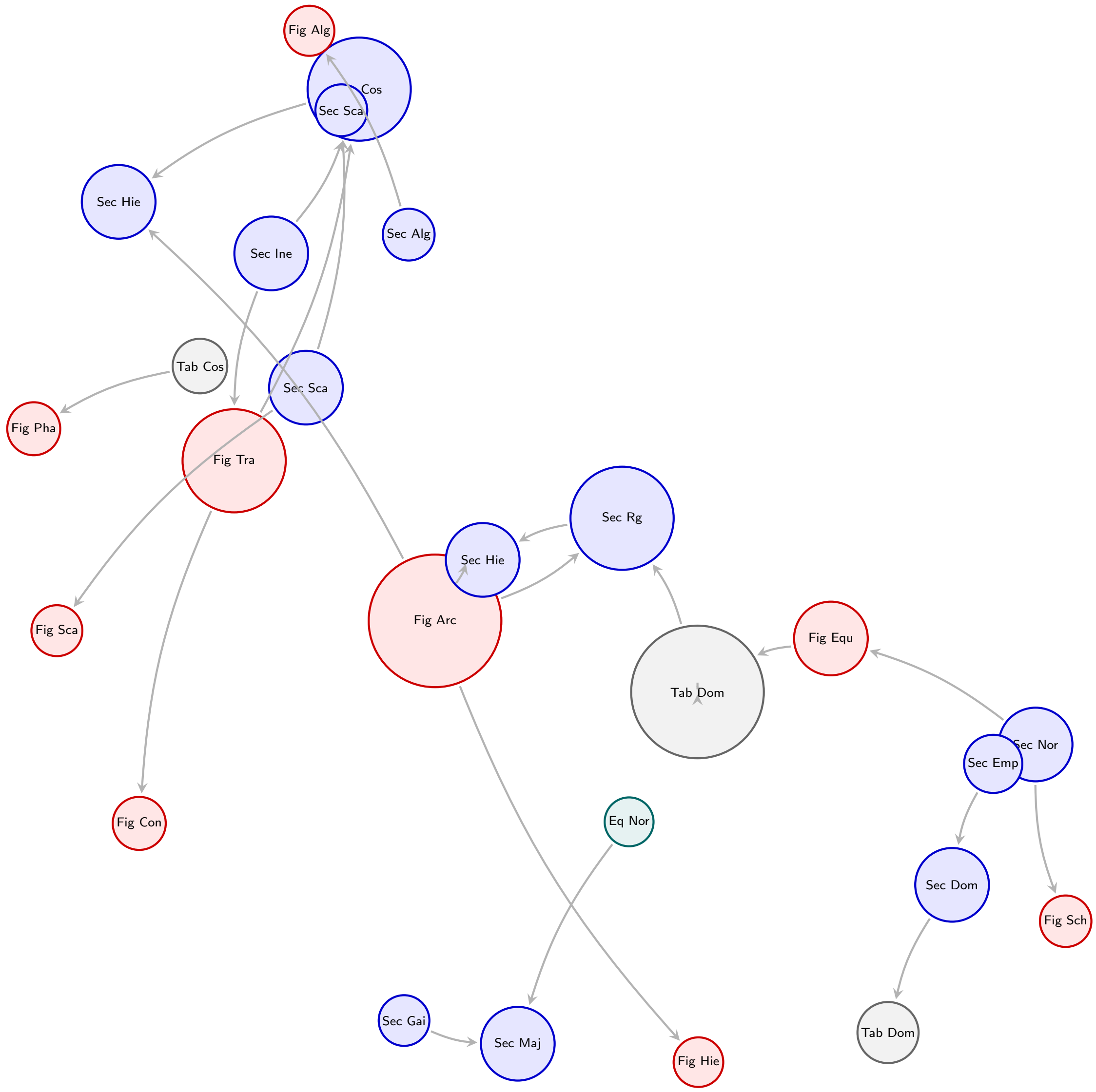

FIG. 9. Topological dependency network mapping the complete theoretical architecture of the manuscript. Nodes represent distinct structural elements (Sections, Appendices, Figures, Equations), with geometric scaling strictly proportional to their epistemic centrality and cross-referential degree. Directed edges trace the explicit flow of deductive logic, computationally verifying the 'derive, do not assert' protocol.

exact $\mathcal{O}(1/\sqrt{N})$ binomial noise into the closed-loop system.

2. Data Collapse and Exponent Extraction

Because N acts as both the control parameter (the gain multiplier) and the size parameter (the noise dampener), finite-size scaling in this stochastic system is fundamentally entangled. To decouple the fluctuation amplitude from the gain, we introduce an inverse-noise parameter R, averaging the realised verdict over R independent N-agent draws per step.

Figure 10 demonstrates that the empirical flip bifurcation emerges cleanly precisely at the analytic N_c. By collapsing the stochastic order parameter curves across varying noise levels R, we extract the onset exponent. In the low-noise limit ($R \to \infty$), the exponent converges to the mean-field prediction of $\beta = 1/2$. However, for true finite-noise ($R = 1$), the environmental feedback is a single realised binary verdict, forcing the expectation update to always be exactly $\Delta\theta = \pm\alpha$. Because this step amplitude is structurally constrained, the macroscopic oscillation amplitude is strictly flat across all N. This singular sign function completely suppresses the continuous mean-field scaling collapse, yielding no critical exponent whatsoever, demonstrating that true finite-population noise fundamentally shatters the continuous mean-field assumption.

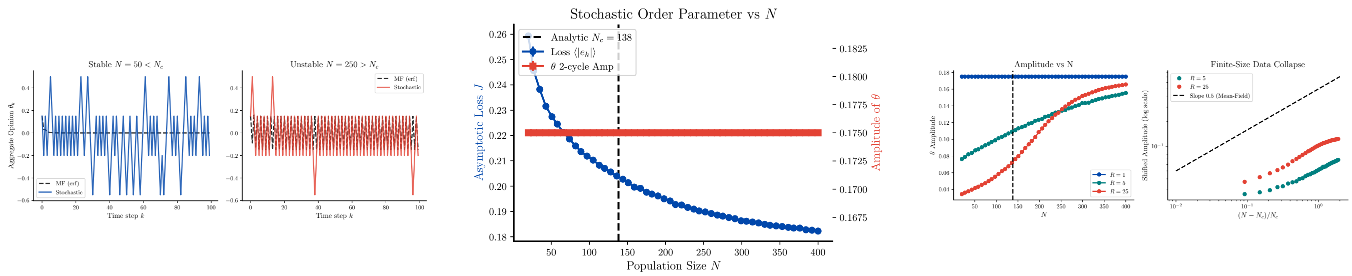

FIG. 10. **Stochastic Finite-N Validation.** (Left) Realised stochastic trajectories vs mean-field expectations below and above N_c. (Middle) The stochastic order parameter exhibits a clear phase transition at the analytic boundary $N_c \approx 138$. (Right) Finite-size data collapse decoupling the gain from the noise envelope. Note that for true finite-noise ($R = 1$), the continuous scaling collapse is destroyed and the amplitude remains structurally flat.

[1] G. B. West, J. H. Brown, and B. J. Enquist, A general model for the origin of allometric scaling laws in biology, Science **276**, 122 (1997).

[2] S. L. Pimm, *Food Webs* (Chapman and Hall, London, 1982).

[3] A. D. Chandler, *Strategy and Structure: Chapters in the History of the Industrial Enterprise* (MIT Press, Cambridge, MA, 1962).

[4] M. Weber, *Economy and Society: An Outline of Interpretive Sociology* (University of California Press, Berkeley, CA, 1922).

[5] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 770–778.

[6] L. Liu, X. Liu, J. Gao, W. Chen, and J. Han, Understanding the difficulty of training transformers, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2020) pp. 5747–5763.

[7] H. A. Simon, The architecture of complexity, Proceedings of the American Philosophical Society **106**, 467 (1962).

[8] G. Tsebelis, *Veto Players: How Political Institutions Work* (Princeton University Press, Princeton, 2002).

[9] D. C. North, *Institutions, Institutional Change and Economic Performance* (Cambridge University Press, Cambridge, 1990).

[10] C. S. Holling, Resilience and stability of ecological systems, Annual Review of Ecology and Systematics **4**, 1 (1973).

[11] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Nes, M. Rietkerk, and G. Sugihara, Early-warning signals for critical transitions, Nature **461**, 53 (2009).

[12] R. M. Goodwin, A growth cycle, Socialism, Capitalism and Economic Growth , 54 (1967).

[13] M. Ezekiel, The cobweb theorem, Quarterly Journal of Economics **52**, 255 (1938).

[14] G. W. Evans and S. Honkapohja, *Learning and Expectations in Macroeconomics* (Princeton University Press, Princeton, 2001).

[15] R. C. Conant and W. R. Ashby, Every good regulator of a system must be a model of that system, International Journal of Systems Science **1**, 89 (1970).

[16] W. A. Brock and C. H. Hommes, A rational route to randomness, Econometrica **65**, 1059 (1997).

[17] W. A. Brock and C. H. Hommes, Heterogeneous beliefs and routes to chaos in a simple asset pricing model, Journal of Economic Dynamics and Control **22**, 1235 (1998).

[18] M. Nerlove, Adaptive expectations and cobweb phenomena, Quarterly Journal of Economics **72**, 227 (1958).

[19] W. R. Ashby, *An Introduction to Cybernetics* (Chapman and Hall, London, 1956).

[20] S. Beer, *Brain of the Firm* (Allen Lane, London, 1972).

[21] E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, Cambridge, 1990).

[22] S. Galam, Majority rule, hierarchical structures, and democratic totalitarianism: A statistical approach, Journal of Mathematical Psychology **30**, 426 (1986).

[23] S. Galam, Social paradoxes of majority rule voting and renormalization group, Journal of Statistical Physics **61**, 943 (1990).

[24] R. Radner, The organization of decentralized information processing, Econometrica , 1109 (1993).

[25] P. Bolton and M. Dewatripont, The firm as a communication network, The Quarterly Journal of Economics **109**, 809 (1994).

[26] L. Garicano, Hierarchies and the organization of knowledge in production, Journal of Political Economy **108**, 874 (2000).