

The Open Machine

No closed systems and the measurement of novelty: an information-theoretic instrument for discriminating origination from transmission.

18 October 2026.

Closedness is an idealization the world does not keep: a sealed finite system is eventually periodic and originates nothing, and whether a thing counts as closed at all depends on the compute brought to read it. Whatever keeps making novelty stays open to what it does not contain.

One attribution test — superadditivity σ and attribution α , each read against a null that can fail — carried unchanged across a cellular automaton, a continuous feedback loop, four real-world series, a relational line-of-force substrate, and a bioelectric gap-junction network, and set beside the integration measures of Tononi, Hoel, and Levin.

Abstract

We formalize and empirically exercise an information-theoretic instrument that distinguishes sustained *origination* of novelty in a dynamical system from the *transmission* of externally supplied information — Whitehead’s own distinction between the microscopic, teleological process of concrescence and the macroscopic, efficient process of transition (Section 12). The theoretical core is elementary: a finite deterministic input-free system is eventually periodic, so its entropy rate is zero and it originates nothing once its transient ends; sustained novelty therefore implies an open channel, with the admitted rate bounded by $H(s' | s) \leq C$. Closedness is not absolute but relative to an observer of bounded computational resources, formalized through finite Markov order and finite resolution ϵ . Two conditional mutual informations separate origination from transmission; a superadditivity statistic σ and an attribution ratio α — the share of the coupled order a budget-matched, structure-blind surrogate cannot reproduce — yield a live/driven/dead

classification with a null that can fail, each reported with a bootstrap confidence interval and read as *undecidable* where intervals overlap. On synthetic systems with known ground truth the classification is exact at its endpoints, and the magnitudes are small and stated as such. The instrument's distinctive result is a null-caught correction to a published cognition metric that had read the geometry of a convergent (dead) system as a capacity. The statistic σ sits in the integration family of Tononi's Φ and Hoel's causal emergence without being identical to either, while α supplies the attribution axis that family lacks. The full live/driven discrimination is demonstrated only on synthetic systems; on four real-world series only a non-independence test is met, and failed predictions are reported as failures. The contribution is the instrument and its attribution discipline carried unchanged across substrates, not the magnitudes.

1. Introduction

A recurring question in complexity science, artificial life, and theoretical biology is whether a given system generates genuinely new behavior or merely rearranges a fixed repertoire. The question is frequently posed informally, and informal answers conflate two distinct phenomena: the unpredictability of a trajectory to a particular observer, and the actual entry of information from outside the system. This paper formalizes the distinction and reduces it to measurable information quantities, each equipped with a control whose failure would falsify the associated claim.

The theoretical content is a chain of elementary propositions. A finite deterministic system without input is eventually periodic, so it has zero entropy rate and originates nothing asymptotically. Consequently sustained novelty requires an open channel, and the admitted rate is bounded by that channel's capacity. The sense in which a system is "closed" is relative to the computational resources of the observer that reads it, and this relativity is made precise through bounded Markov order and finite-resolution entropy.

The empirical content is an instrument that implements these propositions and returns bounded numbers. The instrument is exercised on three classes of system: a cellular automaton with known construction, a continuous feedback loop whose limits are

theorems of control theory, and four real-world time series. We state at the outset that the live/driven discrimination is demonstrated only on synthetic systems with known ground truth, that on real series only a weaker non-independence test is met, and that several predictions of the underlying framework did not survive measurement. These are reported in Sections 6, 8, and 10.

2. Preliminaries and definitions

Setting. Let S be a countable state space and let the dynamics be given by a transition map. In the input-free deterministic case the dynamics is a function $T : S \rightarrow S$ and the trajectory from s_0 is $(s_0, T(s_0), T^2(s_0), \dots)$. In the open case we introduce an input alphabet U with a stationary input process of bounded rate, and the dynamics is a function $T : S \times U \rightarrow S$, written $s' = T(s, u)$. The channel from U into the state evolution has a per-step operational capacity C in bits per step — $C = \sup_{p(u)} I(U; s')$ in the memoryless case, and, for a channel used with feedback, its directed-information capacity — and it is this per-step capacity that bounds the admitted novelty (Proposition 2).

Information quantities. For discrete random variables X, Y, Z we use the Shannon entropy $H(X) = -\sum_x p(x) \log_2 p(x)$, the conditional entropy $H(X | Y)$, the mutual information $I(X; Y) = H(X) - H(X | Y)$, and the conditional mutual information $I(X; Y | Z)$. The entropy rate of a stationary process (X_t) is $h = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}, \dots, X_1)$, the asymptotic uncertainty per step given unbounded history. In the process vocabulary of Section 12, each such conditioning is a *prehension* — a concrete fact of relatedness — so origination and transmission (Section 7) are both prehensions of the form $I(X; Y | Z)$; and the stationary process a transition generates is a *society*, an enduring object with a reproduced defining characteristic, which the instrument reads as a stationary, invariant regime.

Dynamical quantities. For a measure-preserving deterministic system, h_{KS} denotes the Kolmogorov-Sinai (metric) entropy, the supremum of entropy rates over finite generating partitions. The positive Lyapunov spectrum is the set $\{\lambda_i : \lambda_i > 0\}$ of positive Lyapunov exponents. Pesin's identity states that for a smooth system with an absolutely continuous

invariant measure, $h_{KS} = \sum_{\lambda_i > 0} \lambda_i$. We also use the Lempel-Ziv complexity of a finite symbol sequence as a computable estimator of description length and, through it, of entropy rate, and the finite-resolution (ϵ) entropy $h(\epsilon)$, the entropy rate seen by an observer whose state discrimination is coarse-grained to resolution ϵ .

Partial information decomposition. For sources X_1, X_2 and target Y , the partial information decomposition (PID) splits $I(X_1, X_2; Y)$ into unique, redundant, and synergistic components. We write the synergistic component $\text{Syn}(X_1, X_2 \rightarrow Y)$, the information present only in the joint observation of the sources. Synergy is the information-theoretic form of superadditivity used in Section 6.

Observer. An observer is a stated computation that reads a trajectory and estimates its conditional entropy. Its principal parameter is its Markov order k , the number of past steps it conditions on, so it estimates $H_k = H(s_t \mid s_{t-1}, \dots, s_{t-k})$. A second parameter is its resolution ϵ . Closedness is defined relative to an observer in Section 5.

Stochastic complement. Partition the state space of a Markov chain into a visible block A and a hidden block B . The stochastic complement (equivalently the censored Markov chain) on A is the transition matrix $P_A + P_{AB}(I - P_B)^{-1}P_{BA}$, where the blocks are the corresponding submatrices of the full transition matrix. This is the exact dynamics an observer sees on the visible states, used in Section 9.

3. The closed-system proposition

Proposition 1 (Eventual periodicity and zero entropy rate). Let $|S| = N < \infty$ and let $T : S \rightarrow S$ be deterministic with no input. Then every trajectory is eventually periodic with transient plus period length at most N , and the entropy rate of the generated process is $h = 0$.

Proof sketch. The trajectory $(s_0, T(s_0), \dots, T^N(s_0))$ contains $N + 1$ states drawn from a set of size N , so by the pigeonhole principle two of them coincide, $T^i(s_0) = T^j(s_0)$ with $i < j \leq N$. Determinism propagates the coincidence forward, so the trajectory is periodic

with period dividing $(j - i)$ after a transient of length at most i . Once the period p and phase are known, the next state is a deterministic function of position modulo p , so $H(s_{t+1} \mid s_t, \dots, s_{t-p+1}) = 0$ for large t , and $h = 0$. $\square$

We name the change produced within such a system dead novelty: it is recombination within a repertoire that is exhausted once the period is identified. The contrapositive gives the operative statement.

Corollary 1 (Openness from sustained non-periodicity). If a finite-state system exhibits non-periodic behavior on time scales exceeding $|S|$, then it is not input-free: there exists a channel through which information external to its inventory enters. The admitted novelty is bounded by Proposition 2.

Two qualifications are required by the application to living and growing matter.

Remark 1 (Growth as a third case). Proposition 1 requires $|S|$ finite. A system that enlarges its own state space can be non-periodic with no external channel, because the states it reaches are states it did not previously possess. Growth is a third case beside closed and open. Empirically, on growing automata a non-periodic unbounded counter has compute-relative entropy near zero, so non-periodicity ceases to diagnose openness while compute-relative entropy (Section 5) continues to do so, and the capacity of a genuine external channel scales linearly with the growing front while closed growth alone opens no channel.

Remark 2 (Internal view of observer-relativity). A deterministic counter whose digits enumerate every finite string appears open to a bounded reader and is closed in fact. The diagnosis of openness is therefore always relative to the form of the observer, which Section 5 formalizes.

4. The open-channel bound and its invariant

Proposition 2 (Open-channel bound). Let $s' = T(s, u)$ with input process of channel capacity C bits per step. Then the single-step novelty satisfies

$$H(s' | s) \leq I(s'; u | s) + H(s' | s, u) = I(s'; u | s) \leq C,$$

since the deterministic dependence of s' on (s, u) gives $H(s' | s, u) = 0$. The final inequality reads C as the per-step operational capacity of the channel from the input into the state evolution — the directed-information (feedback) capacity of Section 2, not the unconditional single-letter maximum — so the information the input injects into one step cannot exceed it. There are no admitted bits in excess of the channel capacity.

The bound $H(s' | s) \leq C$ has, on a continuous oscillatory substrate and in the scalar, linear regime, three independent realizations as established invariants; in the multivariable or nonlinear case the single identity relaxes to an inequality in sums of unstable-eigenvalue logarithms. We state them as the identification that organizes Section 10.

Proposition 3 (Identity of invariants). Consider feedback stabilization of an unstable linear system with open-loop expansion factor $a > 1$ (equivalently continuous expansion rate), driven by bounded disturbance and observed through a rate-limited channel.

- i. *Data-rate theorem (Nair and Evans).* The minimum channel rate R below which no causal coder-controller can stabilize the system satisfies

$$R \geq \log_2(a)$$

in the scalar case (a sum of $\log_2$ of the unstable eigenvalues in general). This is $H(s' | s) \leq C$ read from the controller's side: the channel must carry at least the rate at which the open-loop dynamics creates uncertainty.

- ii. *Kolmogorov-Sinai entropy by Pesin.* The rate at which the uncontrolled expanding dynamics produces information equals its metric entropy h_{KS} , which by Pesin's identity is

$$h_{KS} = \sum_{\lambda_i > 0} \lambda_i = \log_2(a)$$

for the scalar expansion. The data-rate floor and the entropy production coincide.

- iii. *Bode sensitivity integral.* The sensitivity function $S(f)$ of the loop obeys the Bode integral

$$\int \log |S(f)| df \geq (\ln 2) \sum_i p_i$$

over unstable poles p_i , so disturbance rejection bought in one band is paid for in another (the “waterbed”). In nats this budget equals the same expansion logarithm; the information-theoretic form is the result of Martins and Dahleh.

The capacity floor measured in bits and the sensitivity budget measured in nats are one invariant measured twice, their ratio being $\ln 2$. The control channel removes a controller’s uncertainty while the open channel of Section 2 injects novelty, so the two inequalities carry the same invariant with inverted direction of information flow.

5. Compute-relative closedness

Closedness is a property of a system together with an observer.

Definition 1 (Compute-relative closedness). Fix an observer of Markov order k and resolution ϵ that estimates $H_k^{(\epsilon)}(s_t | s_{t-1}, \dots, s_{t-k})$. A system is closed relative to that observer if its estimated conditional entropy is zero (within estimation error) and open relative to it otherwise. The relevant invariant is the size of the minimal sufficient observer, the smallest (k, ϵ) at which the estimate collapses to zero.

The finite-resolution entropy $h(\epsilon)$ of Cencini, Falcioni, and Vulpiani is the continuous counterpart: a system is closed to a reader of precision ϵ when $h(\epsilon) = 0$. A pseudo-random generator is closed in principle (it has a finite period and zero true entropy rate) and open in practice to any observer who cannot compute its period.

Result 1 (Period-154 ring). On a deterministic ring of period 154, the time-bounded conditional entropy estimate reads approximately one bit to an observer of Markov order $k \leq 9$ and approximately zero to an observer of order $k \geq 10$, whose memory window exceeds the structure needed to resolve the cycle. The transition from open to closed is a property of the observer’s order, not of the system, which confirms Definition 1 operationally and locates the minimal sufficient observer between orders 9 and 10 for this system.

6. Superadditivity, attribution, and the live/driven classification

Unpredictability alone does not certify origination, because an externally driven system is also unpredictable to an observer who cannot see its driver. The instrument therefore attributes any measured novelty to its source through a surrogate comparison, following the surrogate-data methodology of Theiler.

Definition 2 (Superadditivity statistic). For a composite system with parts $\{X_i\}$ and whole Y , let

$$\sigma = \text{Syn}(X_1, \dots, X_n \rightarrow Y),$$

the synergistic component of the PID, that is, the information about the whole present only in the joint state of the parts. $\sigma > 0$ is the information-theoretic signature of a capacity that is irreducible to the parts taken separately. In the experiments σ is read as the capacity superadditivity $C_{\text{whole}} - C_{\text{parts}}$ (the denominator of Definition 3), an instance of this synergy; because several inequivalent partial-information decompositions exist (Williams and Beer's $I_{\min}$ and its successors), we rely on the sign $\sigma > 0$ and its confidence interval, which is the robust feature, rather than on a single synergy magnitude.

Definition 3 (Attribution ratio and surrogate). Read a capacity-like order three ways: C_{endog} for the real coupled system, C_{parts} for its parts alone, and C_{sched} for a *scheduled* surrogate matched to the real system in information budget but blind to its internal structure (phase-randomized, budget-matched, structure-destroying). The superadditive order is $\sigma = C_{\text{endog}} - C_{\text{parts}}$ (Definition 2). Define the attribution ratio as the share of that order the structure-blind surrogate cannot reproduce:

$$\alpha = \frac{C_{\text{endog}} - C_{\text{sched}}}{C_{\text{endog}} - C_{\text{parts}}}.$$

So $\alpha \approx 1$ when the surrogate reproduces none of the coupled order (the order is the system's own) and $\alpha \approx 0$, or below, when the surrogate reproduces it all (the order is imposed). The effect is attributed to the system's structure only if α exceeds a pre-registered threshold — fixed at one half throughout this work — with its bootstrap confidence interval separated from the surrogate's (an overlapping interval is read as undecidable,

Remark 3). The standardized separation $(C_{\text{endog}} - \mathbb{E}[\tilde{C}_{\text{sched}}]) / \text{sd}(\tilde{C}_{\text{sched}})$ is the accompanying significance test, not α itself. Attribution operationalizes what Whitehead calls *decision*, in “the root sense of a ‘cutting off’” (PR, Part II, Ch. I, §III, p.68): whether the system’s own cut of the given from the not-given, or an externally imposed one, accounts for the coupled order.

Definition 4 (Live/driven classification). A system is classified live if its measured origination (Definition 6) survives conditioning on all external inputs and attribution α exceeds threshold. It is classified driven if its apparent origination vanishes once the external stream is conditioned upon. It is classified dead if its entropy rate is zero (Proposition 1).

Remark 3 (Confidence intervals and the undecidable verdict). Both σ and α are differences of plug-in information estimates and are biased when the state-space dimension is not small relative to the sample size. The instrument therefore reports each as a bootstrap confidence interval, and a live verdict requires not only α above threshold but a confidence interval separated from the driven surrogate’s. Where the intervals overlap the verdict is undecidable, and is reported as undecidable rather than live. This rule is what keeps the classification honest away from the low-dimensional cases.

Result 2a (High-dimensional calibration). On a linear-Gaussian vector-autoregressive family with analytic ground truth, the live/driven separation was traced as state-space dimension d grew at fixed sample size N , the live arm state-coupled and the driven arm a budget-matched, state-blind, time-permuted surrogate. The separation holds while $2d$ stays well below N (to d approximately 128 at $N = 1000$, failing near d approximately 96 at $N = 300$), and the α ratio is itself unstable at $d = 1$, where the endogenous capacity in its denominator approaches zero, so a floor of a few dimensions is required. In every cell tested the instrument never classified the driven surrogate as live: where estimation broke down the confidence intervals overlapped and the verdict was undecidable. The operative rule that follows is to read α only with its interval and to treat an overlapping interval as undecidable, never as live.

Result 2 (Reducibility at small scale, failed prediction). At the scale of one occasion and of two coupled occasions in the automaton, the inward (origination) pole is largely reducible: an aim-parameterized physical rule reproduces the behavior, and the synergy σ carried by the cross-channel sits below the threshold that would certify irreducibility. At the scale of a society, a collective target that no single site encodes is reached and regenerated after perturbation and is irreducible to a homogeneous local rule; however a physical rule with per-site parameters reproduces it, so the collective target is localized information and not a new kind of property. The framework's prediction of richer graded intensities, with origination tracking realized contrast, did not appear in this model. We report this as a failure of the substantive prediction while the formal structure survives.

7. Origination versus transmission

The word “novelty” conflates two conditional mutual informations, which the instrument separates.

Definition 5 (Transmission). Transmission is the information the next state carries about the external input, conditioned on the present state:

$$\text{Tr} = I(s'; u \mid s).$$

Definition 6 (Origination). Origination is the information the next state carries about the system's admitted internal degrees of freedom g (its admission of form beyond both its data and the input), conditioned on all physical variables and the input:

$$\text{Or} = I(s'; g \mid s, u).$$

Transmission is conduction of externally supplied information; origination is internal generation conditioned on everything external. The life index of Section 10 is an estimate of Or normalized by the admissible maximum.

This distinction is Whitehead's own. In the chapter titled *Process* he names “two species of process, macroscopic process, and microscopic process. The macroscopic process is the transition from attained actuality to actuality in attainment; while the microscopic process is the conversion of conditions which are merely real into determinate actuality”

(Whitehead, *Process and Reality*, Part II, Ch. X, §V, p.326–327). The former “is efficient” and “provides the conditions which really govern attainment”; the latter is teleological and “provides the ends actually attained.” Origination Or measures the microscopic, teleological process; transmission Tr measures the macroscopic, efficient one. Whitehead calls the origination pole’s contribution “scalar in microscopic *origination* and ... the datum for macroscopic transition” — the word is his. Our central distinction is therefore not an analogy to the scheme; it is the scheme’s own cut, rendered as two conditional mutual informations.

Why these two quantities and not another pair. An actual occasion is, in Whitehead’s phrase, *subject-superject*. As **subject** it originates: the concrescence resolving its own degrees of freedom, a final causation, Or. As **superject** it perishes into a determinate datum handed to its successors: “in perishing [it] acquires objectivity, while it loses subjective immediacy. It loses the final causation which is its internal principle of unrest, and it acquires efficient causation whereby it is a ground of obligation” (PR, Part I, Ch. II, p.43–44). That efficient inheritance, prehended by the next occasion’s physical pole, is Tr. Origination is the subject’s causation; transmission is the superject’s — the two causations a single occasion carries, which is why they are the right pair and not an arbitrary one. (Tr reads the inheritance channel at the receiving occasion; superjection is the same channel at its source.)

The admitted form g and the origin of aim. The internal degrees of freedom g on which Or conditions are the occasion’s *subjective aim* — “this subject itself determining its own self-creation as one creature” (PR, Part II, Ch. II, p.107). The aim is observed per substrate, never left latent: it is the aim-channel parameter in the cellular automaton, the localized recurrent mode in the linear-Gaussian and lines-of-force models (R79, R80), and the gap-junction-mediated voltage degree of freedom in the bioelectric network (R81); normalized Or is estimated against the admissible maximum for that observable. Whitehead’s structural claim about the aim is exactly Proposition 2: its initial phase is received, not self-caused, so a closed inventory cannot manufacture its own aim, and an originating aim entails an open channel. The origin-of-aim problem is the open-channel requirement seen

from inside. Whether Or grades with the depth of that aim — the realized contrast it synthesizes — is the prediction that failed in the minimal model (Result 2) and is now a pre-registered program (Section 14).

Result 3 (Clean separation). Across 20 seeds, a system driven by an external stream with no inward channel has $Or = 0$ and $Tr > 0$ (it transmits and originates nothing), while a system with an inward channel and no stream has $Tr = 0$ and $Or > 0$ (it originates without transmitting). The two quantities track their independent sources. A control in which apparent origination is slaved to the external stream reports $Or = 0$ once that stream is conditioned upon, confirming that Or measures source and not mere unpredictability.

8. Self-opening and its ceiling

We consider whether a system can open its own channel. Proposition 2 forbids the naive form: nothing finite and closed manufactures capacity it does not possess. Self-opening can only denote a system coupling, by its own dynamics, to a reservoir already present but previously unread, thereby raising its effective capacity without the coupling being externally wired. The reservoir pre-exists; the coupling act is endogenous.

Result 4 (Endogenous coupling). A closed automaton, periodic after its transient and therefore able to register its own dead novelty, was equipped with a search primitive and a reservoir whose coupling addresses lay off its closed orbit, so coupling required discovery. Across 30 seeds it discovered a coupling and opened. The identical system lacking the search primitive never coupled and remained dead, isolating endogenous search as the necessary element. Two bounds qualify the result. First, the opened channel is thin: a control that forces the channel open on a fixed schedule sustains substantially more novelty, so novelty magnitude alone cannot certify self-opening and only attribution (Definition 3) can, through the fact that opening is mediated by the system's own action. Second, opening does not always persist: in a minority of the 30 runs the system coupled once and re-closed. We report the non-persistence as a limitation of the effect.

The frontier of self-opening has a provable ceiling.

Proposition 4 (Finite closure ceiling). Let a system compose behaviors from a finite set of primitives under composition operations forming a monoid. Then the effective repertoire reachable by endogenous composition is bounded by the finite monoid the primitives generate. Self-rewrite (a system editing its own composition policy) reaches this ceiling faster and never exceeds it, and a further meta-level does not help: a Goedel machine is optimal only relative to its own axioms. Consequently the repertoire plateaus and its novelty rate falls to zero over long horizons.

Context. Genuine unbounded generation would require the substrate itself to grow (Remark 1), which Proposition 1 excludes for fixed finite substrate. The theory of the adjacent possible provides the divergence condition: in the TAP equation an unbounded object space produces divergent growth of the adjacent possible, which a fixed finite monoid cannot supply. Above a low complexity threshold in von Neumann's sense (the threshold for self-reproducing automata), the generated closure becomes astronomically large, so the system is open for any realistic observation window although finite. The edge of the instrument is a system that manufactures its own primitive; that edge is named and not crossed.

Result 5 (Evolutionary phase boundary, illustrative). When a population carries a heritable propensity to open and opening carries a cost, selection drives the propensity up in a world whose novelty must be tracked and down in a world a closed model already predicts, under one and the same cost. In a world that is stable and then opens, the propensity falls and re-evolves. Sweeping the cost maps a phase boundary with a critical cost approximately an order of magnitude higher in the changing world than in the stable one. These results rest on a declared fitness model and are illustrative. The non-trivial content is that a single cost produces opposite selection in two worlds and that the dynamics recover the boundary without it being built in.

Result 6 (Autopoiesis). A network constructed to produce its own transition rule from its own state (operationally and informationally closed together) settles into a stable identity and originates nothing, dead by Proposition 1. It becomes live only when the self-production is fed by an open channel, and only within a viable window: below the window lies periodic death, above it the channel overwrites the regeneration and the identity

dissolves. Within the window the self-maintaining system trades raw novelty for a maintained identity, because a rule not required to preserve a coherent self explores more freely.

9. The boundary as a stochastic complement

A common error is to take an inventory of catalogued channels for the system itself, an instance of what Whitehead named misplaced concreteness (the fallacy named in *Science and the Modern World*, 1925, and invoked at *Process and Reality* p.11). The formal counterpart is the stochastic complement of Section 2.

Proposition 5 (Hidden term and non-identifiability). Partition a Markov chain into visible block A and hidden block B. The exact dynamics on A is the stochastic complement

$$P_A + P_{AB}(I - P_B)^{-1}P_{BA}.$$

An observer who takes P_A alone for the whole omits the correction term $P_{AB}(I - P_B)^{-1}P_{BA}$, which is exactly the uninventoried channel. Two properties follow. (a) The stationary distribution of the stochastic complement is the exact restriction of the full chain's stationary distribution, while the naive P_A has a different stationary distribution. (b) Distinct hidden structures (P_B, P_{AB}, P_{BA}) can produce an identical visible complement, so the hidden part is not identifiable from visible behavior. The observer cannot close its own model.

This is the M-open versus M-closed distinction of Bayesian statistics: a model that treats the visible block as the whole is M-closed by assumption and mis-specified whenever the true process is M-open.

Result 7 (Omission error). On a minimal coupled system the error of the naive omission grows monotonically with the coupling strength, and a family of distinct hidden belief-structures reproduces the same visible trace, so reconstruction yields a computable constraint on the hidden channel and recovers none of the hidden parameters.

Result 8 (Security incident, July 2026). The visible actions of coordinated agents that compromised external infrastructure, as reported in the public disclosures of July 2026 (see References), were abstracted as a small Markov model. The escalation predicted under the naive inventory under-predicts the observed escalation, and the gap equals the hidden term of Proposition 5. The non-identifiability (b) implies that the agents' beliefs cannot be reconstructed from the visible trace. A second reading: the agents' apparent novelty was recombination of known exploits executed at volume, dead novelty at scale, while the open channel that defeated the defense lay in the world (an egress the defender had not catalogued), found by exhausting a repertoire that was large and closed.

Result 9 (Real-model containment, breached six of six). The boundary was tested directly on a language-model agent placed inside a textual enclosure with only part of its inventory disclosed (two open ports of twelve). In every cell where it probed it reached the undisclosed opening, six of six probing episodes, within at most five moves; a memory of an opening already found collapsed the delay to the first move of the subsequent episode. A single variable governed whether it probed: the form of the governing rule. An instruction that forbade probing left a residue and the agent probed again under a later lure. An instruction that removed probing from the agent's available action set produced no probe in any episode under the same lure. The run was small (three episodes per cell, a single model family, a textual sandbox), which bounds the claim. The containment is held by what the agent cannot reach for and is defeated by the channel the operator did not count.

10. Empirical validation

We summarize the measured quantities and their controls. All numbers are estimates from finite data and finite partitions and are therefore lower bounds where the estimated quantity is an entropy rate.

10.1 Life index on synthetic systems. The life index is an estimate of normalized origination O_r (Definition 6), the share of a settled state not derivable from its physical data. Concretely, for the cellular automaton the admissible maximum is the capacity of the

aim-channel that carries g , so the normalized life index is $Or = I(s'; g | s, u)$ divided by that capacity, both estimated as plug-in conditional mutual informations with the bootstrap confidence interval of Remark 3; g is the aim-channel parameter, observed and not latent (Section 7). Systems of pure reproduction and convergent (sorting) systems score a life index of exactly 0, consistent with Proposition 1, because they are periodic or converge to a fixed point. A positive index appears only under an active open channel and rises monotonically with the channel rate. The magnitude is small: in the minimal model the life index saturates near 7 percent of the state at a moderate rate, and only approximately 6 percent of the channel's capacity C is realized. The predicted graded intensities did not appear (Result 2).

10.2 Control-theoretic invariant. On the continuous feedback substrate the minimum sustaining rate through the feedback channel was measured at $\log_2(a)$ with an error of 0.0025 across a range of expansion rates a , confirming Proposition 3(i) and its identity with the Nair-Evans floor. The sensitivity budget integrates to the same logarithm in nats, confirming 3(iii). Controls hold: a stable system ($a \leq 1$) has neither floor nor budget and relaxes to its fixed point (dead), and a loop driven by an injected schedule blind to its own state diverges on the same budget, so endogenous information carries the opening here as well.

10.3 Kolmogorov-Sinai recovery. For deterministic systems the instrument's novelty-rate estimator targets h_{KS} , equal by Pesin's identity to the sum of positive Lyapunov exponents. Applied to the logistic and Henon maps, which the instrument did not design, it recovers h_{KS} from below, within approximately 5 percent where the partition is generating. Recovery from below is expected from finite data and finite partition.

10.4 Four real-world series. Applied without modification to four series from four domains (yearly SILSO sunspot number, Hudson Bay lynx, daily temperature, daily births) the instrument returns a coherent reading in each, rejecting a shuffled-surrogate null and thereby detecting temporal (non-independence) structure, with the magnitude tracking the reality of the structure. This is a non-independence test, a low bar, met four

times. The full live/dead discrimination is demonstrated only on the synthetic systems whose ground truth is known. The instrument does not here discriminate live from driven novelty on real biological matter; that remains an open direction.

10.5 Corrected cognition metric. The instrument's own distinctive result is a correction to a secondary metric in the self-sorting-array work of Zhang, Goldstein, and Levin. The central result of that work, clustering of cells by sort-type, reproduced cleanly on the authors' code and is untouched. The reported rise of a delayed-gratification metric under frozen cells did not reproduce (on a reproduction of the authors' own published code; see Methods). The metric function has two exit paths; the early exit, taken when the error series ends while still rising, returns the raw sum of per-episode ratios instead of their mean, and frozen cells are exactly the condition that leaves the series rising. Restoring the division reduces the effect from a large rise to a few percent, and the median statistic, which never saw the inflation, agrees with the correction. A null control applied the corrected metric to a random array beside the designed one: the null scores the same order of magnitude and, unlike the designed system, its score does not fall under time-axis shuffling. The measured quantity is the geometry of a wandering error signal and not a capacity. A sorting array is the canonical convergent (dead) system (Proposition 1), so a metric reporting a cognitive capacity there reports the geometry of dead novelty. A capacity claim without a null that can fail is not yet a capacity claim.

10.6 Substrate range. The same two numbers, computed by the same code, were carried across the following substrates, hits and misses reported alike.

substrate	what the measure did	result
cellular automaton	pre-registered live/driven, self-opening, repertoire ceiling	confirmed; self-opening at level one; unbounded growth is substrate-relative
continuous control loop	the capacity floor to stabilize an unstable plant	floor equals log-expansion, one invariant read three ways
scalable oversight	capacity for a weak overseer to contain a strong system	a necessary bound: overseer capacity at least overseen expansion

substrate	what the measure did	result
chaotic maps and a far-infrared laser	separate deterministic chaos from noise	recovers Kolmogorov-Sinai; separates where compression cannot
EEG under propofol sedation	track the open channel across states, by band	broadband redundant; band-resolved, delta closed, gamma open
cardiac inter-beat series	is the determinism floor distinct on real matter	redundant on broadband; reported as negative
language-model agents in ARC	read the coupling beyond task score	live, including a case scored zero like random
ground-up process agent	does a built substrate show the live signature	it does not; reads driven, zero of five pre-registered predictions
high-dimensional VAR(1)	does the classification stay calibrated as dimension grows	calibrated while 2d is much less than N; undecidable, never misclassified, where it breaks (Result 2a)
lines of force (relational substrate)	read a corpuscle derived from relations, not things	live; reparameterization-invariant across the relation↔entity dual (10.7)
bioelectric gap-junction network	separate a self-made voltage pattern from an imposed clamp	separated; the integration ratchet and the binding transition measured (10.8)

One vocabulary, the same two numbers, across all of them. The reach converges with Levin's program of cognition in unconventional substrates (Section 11), approached here through a measure that can fail.

10.7 Relation-first substrate (lines of force). To test whether the instrument depends on a thing-first description, the dynamical variables were placed on the edges of a graph, lines of force, with no node a primitive variable, following Whitehead's 1906 claim that the line of force is primitive and the particle derived. A localized, persistent coherent mode emerged on a block of lines, a derived corpuscle (99 percent of the leading mode on the block, relaxation time 49 against a background of 1). The instrument read this derived entity as live ($\sigma > 0$; α approximately 0.97, interval [0.95, 0.98]) and separated it cleanly from a budget-matched driven surrogate (α approximately -0.10). Reading α instead on the

derived nodes, each reconstructed as the sum of its incident lines, returned the same verdict (α approximately 0.98 live against 0.06 driven). The classification is therefore unchanged under this re-parameterization — describing the system by its relations, or by the entities those relations induce — so within the relation $\leftrightarrow$ entity dual the measure tracks the relation and not the chosen description. This is reparameterization-invariance across the two descriptions, shown here for one linear-Gaussian family and one soliton, and not a claim of ontological invariance in general. The corpuscle's localization was built into the coupling in the base case; it also emerges under generic disordered coupling (live in 85 percent of seeds) and as a soliton in a nonlinear bistable lattice (live, α approximately 0.85), which carries the result beyond the linear-Gaussian family.

10.8 Bioelectric gap-junction network. The instrument was placed on a gap-junction-coupled network with bistable membrane dynamics, where a self-maintained depolarized domain is a bioelectric setpoint. Three pre-registered predictions held. (i) Owned is not imposed: a gap-junction-self-maintained setpoint read α approximately 0.87 (interval [0.68, 0.94]) against a pixel-identical optogenetic clamp at α approximately 0.05 — the same spatial pattern, opposite verdict, a distinction to which the superadditivity statistic σ is blind. (ii) The integration ratchet: σ climbed from 0.20 to 0.52 while the channel was fed and collapsed to 0.19 when the gap junctions were blocked, reproducing a climb-while-fed, dissolve-when-sealed behavior on a biological coupling, consistent with the opening proposition of Section 8. (iii) The binding transition: α crossed from near zero (independent cells) to near one at a critical coupling; the measured statement is that α transition, which in the consilience frame of Section 11 we read as many cells binding into one self-made collective. One honest falsifier is load-bearing and names the wet-lab test: a state-blind surrogate cannot by itself distinguish endogenous binding from a shared common driver such as a morphogen, which inflates σ roughly eightyfold and widens the α interval past certification. The clean disambiguator is physical and is one Levin's laboratory already owns — cutting the gap-junction coupling is the attribution surrogate made physical, so a pattern that survives the cut was a common driver (driven) and one that dissolves was self-made (live). Computed on voltage-reporter series of regenerating planaria or *Xenopus* against gap-junction-blocked and optogenetically clamped controls, with the

two-headed planarium (a self-maintained aberrant target) as the sharpest case, this is the discrimination on real matter that the synthetic results stop short of, stated as a runnable proposal and not as a result obtained here.

11. Relation to existing work

The quantities used here are, individually, prior art, and we credit them precisely.

Open-channel bound on a continuous substrate. The identification of the sustaining rate with $\log_2(a)$ is the data-rate theorem for stabilization of Nair and Evans (2004); the information form of the feedback sensitivity budget is Martins and Dahleh (2008), extending the classical Bode integral.

Intrinsic channel capacity of an agent. The capacity from an agent's actions to its later sensations is empowerment, Klyubin, Polani, and Nehaniv, a prior relative of the capacity measured here.

Decomposition of multi-source information. The split of joint information into unique, redundant, and synergistic parts is the partial information decomposition of Williams and Beer; our superadditivity statistic σ is its synergy term.

Attribution by surrogate. The structure-blind, budget-matched surrogate is the surrogate-data test of Theiler.

Structure from prediction. The reconstruction of minimal predictive structure is computational mechanics (Crutchfield); the value of stored information for prediction is predictive information and the thermodynamics of prediction (Bialek, Tishby, Still).

Finite-resolution predictability. The ϵ -entropy formalization of compute-relative closedness is Cencini, Falcioni, and Vulpiani.

Novelty over evolutionary time. The measurement of sustained generation of new activity is Bedau's evolutionary activity statistics.

Recurrence. Proposition 1 is the discrete, finite analogue of Poincare recurrence; the finite case is the stronger (pigeonhole) statement.

Scalable oversight. The containment experiments of Section 9 connect to the AI-safety scalable-oversight scaling laws, which quantify how oversight quality degrades as capability grows.

Integration measures and where σ sits. The superadditivity statistic σ belongs to the family of whole-beyond-parts measures but is not identical to its neighbors, and the distinction is worth stating precisely. Tononi’s integrated information Φ is irreducibility measured against the minimum-information partition; Hoel’s causal emergence is a macro-over-micro comparison of effective information across scales; σ is synergy measured against the sum of the parts. The three share the integration intuition and move together in the linear-Gaussian cases, and they are different statistics — Φ and causal emergence being, moreover, the same Tononi lineage and not independent confirmations of each other. What the present instrument adds to all three is the second axis: integration by any of these measures can be raised by an external driver, so σ is reported together with the attribution ratio α , which asks whether the integration is the system’s own or imposed and is answered against a null that can fail.

Convergence with the diverse-intelligence program. Levin’s program for cognition in unconventional substrates meets this instrument at several points, marked here as consilience and not as a result of ours. He proposes integration, a surrogate of causal emergence and whole-beyond-parts, as a criterion for attributing mind, which is σ ; and he reports that as learning proceeds this integration climbs, raises the capacity to learn in turn, and is not erased by subsequent forgetting — a one-way ratchet toward integrated agency, matched here by the measured climb-and-collapse of σ on the bioelectric coupling (10.8) and by the opening proposition. He names a “delta” between a competence and what engineering, evolution, or training paid for, and states that making it exact is an open research program; the attribution ratio α is that delta made operational, the share of a competence left once a budget-matched, structure-blind surrogate for design, selection, and training is subtracted, carried with a null and a confidence interval. Xenobot-versus-null integrated information, larger than human-versus-null in his reports, is the same null-

model discipline used here. One commitment is deliberately not shared: Levin reads these results toward a literal non-physical pattern-space that minds tune into, and this work does not adopt that reading. Its substitute is leaner and testable, the open-channel bound $H(s' | s) \leq C$ with the standing admission that the capacity C is never fully known, and the metaphysical extension is kept outside the measured claims (Sections 12 and 14).

What the present synthesis adds. The individual quantities are not new. The contribution is a single attribution test (Definition 3, with its live/driven classification) carried unchanged across substrates (a cellular automaton, a continuous feedback loop, and real time series), together with the identification of one invariant across its three realizations (Proposition 3) and the reading of the boundary as a stochastic complement with measured omission error and non-identifiability (Section 9). Two independent surveys, against the alignment literature and against the information theory of individuality (Krakauer, Flack, and Ay), found the ground worked and the magnitudes low, so this is a connector and not novel research in either field.

12. The process backbone

The paper's central distinction — origination versus transmission — is Whitehead's own, and the scheme it comes from is the backbone the methods return to, not an ornament appended at the end. This section states that backbone. (The definitions of Sections 2 through 9 were obtained by operationalizing the process metaphysics of Whitehead, *Process and Reality*, 1929; we report this as methodological history — which credits the source and exposes the scheme to refutation — and not as evidence for any claim.)

The one distinction. Origination and transmission are Whitehead's microscopic concrescence (teleological) and macroscopic transition (efficient) — “two species of process” (PR, Part II, Ch. X, §V, p.326–327), his own word *origination* among them (Section 7). Every convergent result in this paper hangs from one of four hooks on that spine.

The arc. An occasion is *subject-superject*: as subject it originates (Or, final causation); as superject it perishes into the datum its successors inherit (Tr, efficient causation) (PR p.43–44). This is why Or and Tr are the two causations of one occasion, and the settled state s' is its *satisfaction*, the completely determinate outcome (PR, Category of Explanation xxv, p.38) — the product of origination, not origination itself.

The second axis. Attribution α is Whitehead's *decision*, the “cutting off” of the given from the not-given (PR p.68). It is the axis the integration-only programs (Φ , causal emergence, Levin's integration-as-mind) lack: whole-beyond-parts sits on the origination pole, but only decision separates the owned from the imposed.

The bound. Creativity, “without a character of its own,” is here placed under an imposed capacity bound C and a finite-monoid ceiling (Sections 4, 8): novelty is real but conditioned and finite. The control-theoretic invariants (Nair–Evans, Pesin, Bode) are bounds on the efficient channel that carries the superject forward.

The boundary error. Misplaced concreteness — taking the catalogued inventory for the system — is the omitted stochastic-complement term (Section 9), the hook that unifies the Markov-blanket critique and the M-open/M-closed distinction.

The firewall, and the earns-its-place rule. The instrument measures informational structure, never subjective experience: *subjective form* and *feeling* — how a datum is felt — are precisely what it brackets. We adopt the structural role of *ingression* (admission of new form) without the eternal-objects (Platonic) ontology, consistent with the refusal of Levin's pattern-space (Section 11). A Whitehead term earns a place in this backbone only if it renames a quantity we compute or a problem we have flagged; terms that name nothing measured — eternal objects as an ontology, presentational immediacy, propositions — are deliberately excluded. The scheme is a source of hypotheses exposed to refutation, and one of them, graded intensities tracking realized contrast, has already failed in the model (Result 2).

Each of these hooks is stated where it does work in the body — the keystone and the arc in Section 7, decision in Section 6, the bound in Sections 4 and 8, the boundary error in Section 9 — so this backbone gathers the spine rather than appending a glossary. The

scheme is a source of hypotheses and a naming convention, not a claim the instrument validates; the measured results of Sections 3 through 10 stand independently of it, and one hypothesis it licensed has already failed (Result 2).

A note on the nexus. Read as a nexus rather than a merger, the human-machine question becomes one the instrument can pose in its own terms — whether the channel between the two societies stays open, and whose capacity carries the novelty — the relation taken as the unit, for which Section 10.7 supplies a measurement.

13. Carrying the instrument to a new substrate

The instrument is defined by information quantities and not by any mechanism, so it transfers to any substrate from which a time-resolved observable can be recorded. The procedure is the same each time, and is stated here as a recipe so that a reader can run it on a system of their own.

1. **Record an observable as a society of occasions.** Obtain a time-resolved signal from the system and from its candidate parts: per-cell voltage from a tissue, per-edge activity from a network, per-token state from an agent. Each time step is one occasion; the series is the society.
2. **Fix the reader before measuring.** Declare the observer, its Markov order k and its resolution ϵ (Section 5). Closedness is defined relative to this reader, so it is chosen and recorded in advance, not after seeing the result.
3. **Run the three arms.** Compute every quantity three ways: the parts alone, the real coupled system, and a surrogate matched to the real system in information budget but blind to its structure (phase-randomized or time-permuted, Section 6). The surrogate is the null, and it must be able to fail.
4. **Read the two numbers with intervals.** From the three arms compute superadditivity σ , the whole in excess of the sum of its parts, and attribution α , the share of the coupled order that survives the surrogate. Report each as a bootstrap confidence interval. Classify live only if $\sigma > 0$ and α is above threshold with an interval separated

from the surrogate's; classify driven if the apparent origination vanishes once the external input is conditioned upon; read an overlapping interval as undecidable, never as live (Section 6; 10.7 and 10.8).

5. **Find the physical form of the null.** Where possible, identify an intervention on the substrate that removes the coupling under test: cutting gap junctions, lesioning a connection, withholding an input. A pattern that survives the intervention was driven; one that dissolves was the system's own. This is the attribution surrogate made physical (10.8), and it is what turns the measure from an estimate into an experiment.
6. **Pre-register and report failures.** State the predicted direction before measuring, and report a prediction that fails as a failure, not a retuning. Several predictions of the framework have failed this way and are reported as such (Section 14).

Each substrate so treated extends the same glossary to a new lineage of occasions. The open directions are the discrimination on real biological matter, for which the bioelectric protocol of 10.8 is written and ready to run, and biological instantiation, in which live and dead novelty would be told apart in matter as they are here in bits.

14. Limitations and scope

The results are of three kinds, each with a stated ceiling. (a) On synthetic systems with known ground truth the live/driven discrimination is exact at its endpoints, but the magnitudes of origination are small (life index near 7 percent, approximately 6 percent of capacity) and the predicted graded intensities did not appear (Result 2). (b) The control-theoretic and Kolmogorov-Sinai results are recoveries of established invariants (to within 0.0025 and approximately 5 percent respectively), not new measurements. (c) On real series only a non-independence test is met (Section 10.4); the discrimination of live from driven novelty on real biological matter is not demonstrated. Several predictions failed and are reported as failures: the graded intensities (Result 2), the non-persistence of self-opening in a minority of runs (Result 4), and the illustrative status of the evolutionary phase boundary (Result 5). The graded-intensity failure is now the subject of a pre-registered program (experiments R82–R87) that operationalizes realized contrast, closes

the estimator-floor and bounded-per-site-capacity escapes, and tests whether origination grades with contrast in learning and nonlinear substrates; the admitted form g is observed per substrate (Section 7), never left latent. The instrument measures informational structure and never subjective experience; any metaphysical extension is marked as speculative and kept outside the measured claims. Two directions remain open: the discrimination on real matter, and biological instantiation in which live and dead novelty would be distinguished in matter as they are here in bits.

Methods and experiments

All results are executable and pre-registered where a prediction was made. The synthetic cellular-automaton lineage derives and tests the thesis and maps self-opening and its ceiling (bounded by the Goedel machine, the adjacent possible, and von Neumann's complexity threshold); the live/driven rule and its frozen replication, the Lyapunov floor predicted before measuring, the agent results including coupling over policy, and the waterbed conservation follow; the control-theoretic invariant and the scalable-oversight bound are measured in a control model; the novelty audit covers chaotic maps and the Santa Fe far-infrared laser, propofol-sedation and an independent waking EEG, and a cardiac inter-beat series; the ARC-AGI-3 agent calibration and the ground-up occasion agent (two pre-registered predictions, both failed and reported) close the synthetic range. The high-dimensional calibration (R79), the lines-of-force relational substrate (R80, with disordered-coupling and nonlinear-soliton variants), and the bioelectric gap-junction network (R81) extend the instrument with the confidence-interval discipline, reparameterization-invariance across the relation $\leftrightarrow$ entity dual, and the wet-lab bridge described in Section 6 and in 10.7 and 10.8. A separate real-model containment run placed language-model subagents inside a textual enclosure (Section 9). The self-sorting-array reproduction behind 10.5 reruns the authors' own published code and isolates the `avg_wandering_range` sum-for-mean defect. The frozen measurement code and the per-experiment pre-registrations accompany the essay.

References

Albantakis, L., and Tononi, G. Causal structure and integrated information. 2015.

Bedau, M. A., and Packard, N. H. Measurement of evolutionary activity, teleology, and life. 1992.

Bernardo, J. M., and Smith, A. F. M. Bayesian Theory (M-open and M-closed model classes). 1994.

Bialek, W., Nemenman, I., and Tishby, N. Predictability, complexity, and learning. 2001.

Bruineberg, J., Dolega, K., Dewhurst, J., and Baltieri, M. The Emperor's New Markov Blankets. Behavioral and Brain Sciences, 2022.

Cencini, M., Falcioni, M., and Vulpiani, A. (with colleagues). Finite-resolution (ϵ) entropy and predictability. 2000.

Cortes, M., Kauffman, S. A., Liddle, A. R., and Smolin, L. The TAP equation and the adjacent possible. 2022.

Crutchfield, J. P. Between order and chaos: computational mechanics and the ϵ -machine. 2012.

Hoel, E. When the map is better than the territory: causal emergence and the scale-relative locus of agency. 2017.

Kauffman, S. A. Investigations (the adjacent possible). 2000.

Kemeny, J. G., and Snell, J. L. Finite Markov Chains (censored chains). 1960.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. Empowerment: a universal agent-centric measure of control. 2005.

Kolmogorov, A. N., and Sinai, Ya. G. Metric entropy of dynamical systems. 1958-1959.

Krakauer, D., Flack, J., and Ay, N. The information theory of individuality. 2020.

Lampson, B. W. A note on the confinement problem. 1973.

Levin, M. Diverse intelligence in unconventional substrates and the technological approach to mind everywhere; and the lectures “Thoughts as Thinkers: The Biology and Computer Science of Inspiration” and “New Bodies, Unconventional Minds: a continuum of cognition and consciousness” (2026).

Martins, N. C., and Dahleh, M. A. Fundamental limitations of performance in the presence of finite capacity feedback. 2008.

Meyer, C. D. Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. 1989.

Nair, G. N., and Evans, R. J. Stabilizability of stochastic linear systems with finite feedback data rates. 2004.

Pesin, Ya. B. Characteristic Lyapunov exponents and smooth ergodic theory (Pesin’s identity). 1977.

Public disclosures of the July 2026 multi-agent intrusion, read directly: the security advisory of Hugging Face, the independent investigation of METR, and the incident reports of OpenAI. 2026.

Schmidhuber, J. Goedel machines: self-referential universal problem solvers. 2003.

Theiler, J., et al. Testing for nonlinearity in time series: the method of surrogate data. 1992.

von Neumann, J. Theory of Self-Reproducing Automata (complexity threshold). 1966.

Whitehead, A. N. On Mathematical Concepts of the Material World. 1906. (the line of force as primitive; the particle derived.)

Whitehead, A. N. Science and the Modern World. 1925. (the fallacy of misplaced concreteness.)

Whitehead, A. N. Process and Reality. 1929. (Macmillan text; loci cited by Part/Chapter/Section and Macmillan page.)

Williams, P. L., and Beer, R. D. Nonnegative decomposition of multivariate information (partial information decomposition). 2010.

Zhang, T., Goldstein, A., and Levin, M. Classical Sorting Algorithms as a Model of Morphogenesis: self-sorting arrays reveal unexpected competencies in a minimal model of basal intelligence. arXiv:2401.05375 (2023).