The Metrology of Latent Hierarchy

Evan Tabak Atlas — ORCID 0009-0007-7374-2338

Essay — body complete, 2026-09-15. The abstract below is canon (CONSTITUTION §1 carries its short form) and stands verbatim; the body is written to it. The gate this draft carried until 2026-08-31 — that the essay ships behind a supplemental provisional — is gone: that provisional was dissolved on 2026-08-31, and the essay now ships on owner sign-off. Its Figure 1 is the verdict map (CONSTITUTION §8 — the sealed-fit inventory generated by Fenopan x3, re-cut by x9, by the amendment-consumption pass, by the audit-corrections pass and by the x13 consumption pass; §8's source of truth is record v21 (28 rows; canon 1.25.0), and the figure is cut from whichever inventory record §8 currently points at — rendered deterministically by `make_figures_verdict.py`). Dated note (2026-09-15): this essay moved from the `out/v18` cut to the `out/v21` cut the same day the re-cut landed; `out/v18` (27 rows) stays sealed beside it and keeps its own DOI, and the sentences the re-cut restated are restated here. Epistemic tags are the canon's (§0): [REPO] measured in a sealed record, [VERIFIED] an external source fetched and quoted, [DERIVED] our own arithmetic or structural claim. Every measured number below names the repository and the record it was measured in; the appendix gives the DOIs. This essay's own deposit: Zenodo `10.5281/zenodo.22759978` (preprint; DOI reserved 2026-09-15; published 2026-09-15 on the owner's sign-off).

Figure 1 — the verdict map (CONSTITUTION §8)
one estimator mirror, one claim discipline, 28 sealed cells (cut v21: CITED 15, CONTROL 3, DERIVED 1, VENDORED 9)

Generated by make_figures_verdict.py from the sealed inventory cut §8 points at (fenopan.inventory.CURRENT_CUT). A view of the sealed numbers; no re-fit.

Abstract

Boethius divided music into three: the music of the spheres, the music of the body, and the music of instruments. This essay reports what happened when one mathematical object was made to play all three — and a fourth he did not name, the music of the body politic.

The object is a multiplicative cascade: a hierarchy of hidden states, slow over fast, each multiplying the next, with a property most beautiful ideas lack — an exact likelihood, so it can be fitted, compared, and above all *refuted*. I built it first as an instrument, with dials for its parameters and a gesture for holding its layers still. Then I pointed the same object, running in reverse as an inference engine, at matter falling into black holes, at the human heartbeat, and at the recorded catastrophes of civilization — wars, defaults, outbreaks — asking each time not "does the pattern appear?" but "does the mechanism survive a fair fight against its rivals, at a sample size where the fight can be decided at all?"

The answer is a map, and the map is the finding. The cascade wins where a system genuinely re-occupies discrete regimes across separated timescales — performed music (as read by the *discrete* corpus estimator; a continuous re-estimation of the same performances reads several times lower, so the piano rung is a discrete-crosswalk value and is quoted as one), the multi-state X-ray binaries — and loses, decisively and instructively, where dynamics are smooth or rough, or where the apparent structure belongs to the telescope schedule, the reporting baseline, or the statistic itself. Along the way the program kept finding the same buried flaw in four literatures: mechanism claims made with instruments whose power to detect the mechanism had never been measured — including the field-standard test of multifractality, which turns out to be nearly blind to its own canonical example. Every negative result here therefore travels with a certificate: proof, obtained before the data was read, that the instrument could have seen the thing it did not see.

What emerges is not a theory of everything but something rarer: a calibrated account of where hidden hierarchy is real, where it is an artifact of our instruments, and where — at the boundary where self-excitation becomes indistinguishable from fate — the honest answer is that no one can yet tell. Boethius, it turns out, was proposing a research program. It just needed an estimator.

1. The instrument

The object came first as a thing to play. The Markov-Switching Multifractal is volatility — or intensity — written as a product of *k* independent two-state multipliers on a geometric ladder of switching rates, and because its latent state space is a finite Markov chain of 2^k states, its likelihood is computable exactly by a Hamilton-style forward recursion [VERIFIED — Calvet & Fisher 2001/2004]. That exactness is the whole reason this object and not another one: a descriptive multifractal formalism can be computed on anything and refuted by nothing, while a cascade with an exact likelihood can enter a model tournament and *lose*.

Built as an instrument, the cascade's parameters become dials. The multiplier `m_hi` sets how hard the layers pull; the depth *k* sets how many of them there are; the base *b* and the fastest rate set how far apart their clocks sit. The *pin* is the gesture that holds one layer still while the others run — and on a loop-free graph it acquires exact counterfactual semantics, because clamping a node's latent state and re-running the downward pass of an upward–downward message schedule is a well-defined operation rather than a metaphor [DERIVED]. An instrument you can play is an instrument whose dials you can calibrate, and calibration is where this program's real subject begins.

The first calibration was a translation problem. The instrument is parameterised discretely (a binary multiplier on a *b*-ary ladder); the inference arm measures continuously (the intermittency coefficient λ^2 of a log-correlated field). The founding experiment asked whether the crosswalk between them — λ^2_equivalent $\approx \beta(k,b)\cdot\mathrm{var}(\ln M)/\ln b$ — is real, by simulating the discrete cascade and fitting it with the *same* vendored continuous estimator the inference arm uses. It is: the formula is the asymptotically exact intermittency coefficient of the field, and the estimator recovers it up to a characterised, multiplier-independent finite-*k* factor $\beta(k,b)$ [REPO — Fenopan x1, record v0]. The cells that matter here are the controls, carried in the sealed-fit inventory: a known-truth cascade reads $\hat{\lambda}^2 = 0.76$ against a predicted 0.755; a vanishing multiplier reads 1.5×10^{-6}; a rough monofractal of true $H = 0.30$ reads $\hat{H} = 0.30$ at $\lambda^2 = 0.020$ — roughness is separable from intermittency, and the columns behave before any real number is read [REPO — Fenopan x3, record v21].

Two things that calibration then did to the instrument's own headline are the reason this essay is a metrology and not a tour. The program's most-quoted number is the performed-piano multiplier: $\hat{m}$_hi = 1.619 over 1,264 MAESTRO performances, with min-BIC selecting MSM-Cox on 84 % (101/120) of a seeded subsample fit by a *different* estimator — two analyses, stated as two, and still insufficient as a generator at classifier resolution (0.758) [REPO — Fenophone `msm-corpus-fit` + `msm-cox-hawkes`]. First, a continuous re-estimation of the same performances reads five to six times lower than the crosswalk band — not because either instrument is broken, but because the corpus fit runs through a Poisson counting channel, and the identical synthetic cascade pushed through *that* channel attenuates onto the real band [REPO — Fenopan x12, records v12 and v14]. Second, and harder: the discrete estimator at the corpus operating point carries a large zero-correlation floor, and once that floor is re-cut at the corpus's own marginal mean it reaches the sealed multiplier at a Fano factor of ≈ 1.58 against the corpus's own 1.552 — **the sealed rung sits at its own zero-correlation floor, within 0.014** [REPO — Fenophone `msm-corpus-fit`, null-floor re-cut of 2026-08-31]. The earlier headline that "0 % survives decorrelation" was a clamp of a negative raw difference; published properly, on the n = 56 admissible performances the nulls were matched to, it is *no detectable difference* (paired real − shuffle: median $\Delta = -0.0072$, Wilcoxon $p = 0.114$; real − negative-binomial: −0.0078, $p = 0.714$). The conclusion survived the proper test. The point estimate did not, and the canon no longer prints it.

No sealed number moved in any of that. What moved was what the number *means*: the piano rung is quoted only as a discrete crosswalk of a latent model parameter, flagged as sitting at a calibrated null floor, and never in a fine cross-domain ordering. That is the shape of almost every result in this essay. The

measurements are stable; the licences attached to them are what the program spends its time earning, losing, and re-earning.

2. Frozen and running music

A cascade has two phases. The running phase is a process in time — music, a flare record, an order book. The frozen phase is a quenched, bounded realization of the same multiplicative logic: a designed artifact, a relief, a façade, a tree. A city is the time-integral of a running intensity; a building is a cascade that stopped. The frozen phase has its own exact machinery, because on a chain *or a tree* the likelihood is an upward–downward message pass [VERIFIED — Crouse, Nowak & Baraniuk 1998] — machinery never, as far as the program's own targeted search could establish, applied to measured anatomical or architectural trees.

Where the design doctrine is recursive and multiplicative, the frozen thesis holds. Gothic ornament is generated by a doctrine of that kind, and the built-form arm finds real persistence in it — 49/49 transects beat their shuffles — while two of its pre-registered structural predictions are refuted outright and the diffuser cascade it hoped to find had to be *designed in* rather than discovered [REPO — Fenophone research, carried as a cited cell in Fenopan x3, record v21].

Then the arm went looking for the opposite. The canonical unicursal labyrinths — the Chartres eleven-circuit, the Cretan seven — are the **anti-cascade**: they maximise path length at *zero measure heterogeneity*, the combinatorial-topological dual of a recursive multiplicative design [REPO — Fenopan x7, record v6]. Read as one-dimensional series against the program's own estimators, they carry no cascade on any axis it can read. Detrended fluctuation analysis is *scalebound* — $\alpha = 1.76$ with a strong crossover ($\Delta = 0.49$; α_{small} 1.98 → α_{large} 1.49), outside the long-range-correlated band $\alpha \approx 1$ that the ornament positive control occupies ($\alpha = 1.01$, detected). The two-point multifractal width shows **no** excess over its surrogates (width $z = -6.3$, *below* the band). The topology is exactly self-dual: the complement of the level sequence is its reversal, the same $\mathbb{Z}_2$ retrograde symmetry a musical retrograde is, and 72 of the 1,828 depth-12 designs share it.

Two features of that record are worth more than the verdict. The *scattering* leverage does read a positive excess ($z = +7.8$) — and it is an artifact, the isolated-singularity high side of the statistic firing on the turns, generic to **100 %** of the enumerable null, with Chartres at the null median. And "generic to 100 % of the null" is sayable because **for once the true null is enumerable**: Phillips' three conditions on valid unicursal designs give $M(12) = 1{,}828$, reproduced byte-exactly, and against that exact null Chartres is not special on α (22nd percentile), on width (40th), or on scattering z (52nd). Nearly every negative in the literature this essay engages is a negative against a *surrogate* approximation to a null; here the null could be counted, and counting it turned a "significant" statistic into a measured property of the design class.

The record also carries its own correction: an earlier reading — the level string's Lempel–Ziv complexity at 0.004 against an ornament cascade's 0.71 — was a sampling-density artifact, replaced by the resolution-invariant statement that the string's description length is set by its circuit count. Saying correctly that the

labyrinth is a minimal-complexity algebraic object required withdrawing the more dramatic number.

What the negative buys is **scope**. Freezing applies where the design doctrine is recursive and multiplicative; it does not apply where the doctrine is combinatorial and topological. That is a sharper thesis than the one the arm set out with, and it was bought by a refutation.

3. The spheres

Pointed at the sky, the same object stops being an instrument and becomes a defendant. The question is never "does the pattern appear?" — scale-free patterns appear everywhere and mean almost nothing — but "does the mechanism survive a fair fight against its rivals, at a sample size where the fight can be decided at all?"

It wins where a system genuinely re-occupies discrete regimes across separated timescales. GRS 1915+105, the multi-state X-ray binary, beats the best linear rival out-of-sample (+63.2 multi-fold out-of-sample log-likelihood; +840 BIC over CARMA), and the wins **scale with multi-stateness** — 7 of 17 X-ray binaries at $k \leq 6$ [REPO — Fenocosm v0–v15, cited in Fenopan x3, record v21]. That is the boundary stated as a gradient rather than a claim, which is what a real mechanism looks like and an artifact does not.

It loses, instructively, everywhere the gradient runs out. The daily edge **reverses** at sub-daily cadence: the same source carries two clocks about 5.6 decades apart, at the disc's two dynamical edges [REPO — Fenocosm v9/v10/v15]. On active galactic nuclei — the damped random walk's home turf — the cascade loses on **0/4** well-sampled real light curves at a median ΔBIC of −311, with the emission-line validation passing 5/5 and the synthetic prediction reproduced: smooth, not regime-switching [REPO — Fenocosm v24]. Twice more the flat multi-state incumbent holds: a freely-fit quantum-dot blinking ladder does **not** collapse onto a two-parameter geometric one (0/30 false cascade-wins above the floor, a genuine cascade recovered above ≈ 52 transitions against the flat incumbent's ~10^4-event onset) [REPO — Fenocosm v30], and on M-dwarf flare photometry the published three-state HMM is **not out-raced** above a floor of ≈ 1,600 flares (0/12 false cascade-wins, the genuine cascade recovered from ≈ 60) [REPO — Fenocosm v31]. And on the one real repeater FRB in the roster the verdict is an estimability statement rather than a world statement: 60 real CHIME bursts against a selection floor of $N^* \approx 800$, with the fitted span $S = 1.00$ a ladder-degeneracy boundary solution and not a measured absence of hierarchy [REPO — Fenocosm v27].

The forecast altitude

Through 2026 the spheres arm acquired something the rest of the program did not have: a *product face*, a causal 24-hour M-class flare now-cast scored against a public skill ledger. A forecast is a different altitude from a likelihood tournament, and the last two phases of work were spent measuring what the forecast's skill is actually made of. The answer is an ordering, and it is not the flattering one: **calibration first, then structure, then depth.**

Calibration first. The sealed forecaster issues a mean 0.195 on a tail where the observed M-day frequency is 0.144. A two-parameter Platt map fitted on the training block alone and never refitted moves the tail Brier skill score from +0.3385 to +0.3645 — a paired ΔBSS of **+0.026** [+0.018, +0.035] — which is **larger than the cascade's entire margin over persistence**, +0.021 [+0.012, +0.030] on the same resamples [REPO — Fenocosm v45]. The pre-registered reliability bearing *failed* and is reported as failed — the worst calibrated bin still misses the diagonal by 0.075. Calibration bought skill, not a flat reliability diagram.

Structure second, where structure means "any model at all" rather than "the right model". An eleven-parameter ridge logistic regression on the same daily counts scores +0.3734 where the Platt-calibrated 64-state cascade scores +0.3645; the paired interval is **−0.0089 [−0.0165, −0.0019]**, excluding zero, and stacking the cascade's posterior onto the logistic adds **+0.0015 [−0.0022, +0.0051]** — nothing [REPO — Fenocosm v46]. Against the *uncalibrated* incumbent the gap is four times larger, so most of the published forecaster's distance from a regression was miscalibration rather than mechanism.

Depth last, and depth is where the hierarchy was supposed to live. Raw, k = 6 over k = 2 is +0.0729 [+0.0527, +0.1069]; calibrate every entrant and the cheapest depth indistinguishable from k = 6 is **k = 4** (+0.0032, covering zero). Against a free flat Poisson HMM chosen on training BIC the calibrated difference is **+0.0009 [−0.0090, +0.0097]**: a five-state model with no hierarchy at all matches the geometric six-layer cascade at forecast altitude — *while the same cascade still wins the likelihood race on the same training block by ΔBIC 97, on four parameters against twenty-five* [REPO — Fenocosm v47]. A parsimony verdict and a forecast verdict are different questions about the same data, and the program had answered only the first.

Three further readings sharpen rather than soften this. The cascade's **margin over persistence widens with lead time** — +0.0210 [+0.0124, +0.0294] at one day to **+0.0861 [+0.0515, +0.1230]** at fourteen — because h-matched persistence decays faster than the filter does, so the original seal's "modest edge" was a one-day statement; the count-only logistic nonetheless beats the cascade at *every* horizon [REPO — Fenocosm v49]. The X-class tile is real — 568 real ≥ X1.0 flares on 122 X-days against a climatology of 0.016701, the frozen ladder's own 64-state belief beating X persistence by +0.0718 [+0.0256, +0.1274] [REPO — Fenocosm v50] — but it is **a rate effect**: a pure size lottery with one constant flare-level X fraction scores +0.0852 calibrated, and the paired difference against the per-latent-state route is +0.0170 [−0.0097, +0.0514], covering zero [REPO — Fenocosm v52]. The cascade sets *when*, not *how big*, and even that is scoped as an estimability statement.

What *does* buy skill is exogenous information. Two public, issue-time-available activity covariates in the exposure gain **+0.0507 [+0.0382, +0.0665]**, repair the quiet-year floor (mean issued 0.0533 → 0.0308 against an observed 0.0137), and — the diagnostic sentence of the whole arc — **collapse the fitted ladder span from S = 146.8 to S = 4.16**: the cascade's slow rungs had been tracking the solar cycle, and once the covariate supplies it directly the hierarchy has nothing left to carry [REPO — Fenocosm v48, v51].

A second domain was then opened to see whether the discipline generalises. Geomagnetic activity, read as the daily count of storm-level three-hour Kp intervals, is a good test because it carries a *known* recurrent structure — Bartels' 27-day solar rotation — so any claimed cascade skill must beat a baseline that exists by construction. The synthetic two-world atlas, run before any real claim, returned **UNDECIDABLE AT THIS N** on three of its four arms: a no-hierarchy world carrying only the season, the recurrence and a two-state switch selects **k* = 4 on 3/3 seeds** with a compact ladder, so the depth gain is not a hierarchy detector here [REPO — Fenocosm v54]. On the real record the verdict is the registered negative: the calibrated cascade beats persistence at one day (+0.0462 [+0.0345, +0.0563]) and **falls below the flat 27-day recurrence baseline at two days** (−0.0412 [−0.0560, −0.0238]); the selected depth is k* = 3 with a ladder flagged degenerate at S = 1.00, and a free three-state HMM describes the training block better than any cascade depth by 445 BIC. Beyond two days, count-only Kp forecasting reduces to recurrence plus a trailing activity level [REPO — Fenocosm v55]. The second domain carries no serviceable cascade now-cast, and the arc's deliverable is saying so with intervals. None of this arc is a row of Figure 1, and deliberately so: a Brier skill score is not a verdict word and the map is a mechanism map, so the geomagnetic reading is carried as a dated note beside the map's solar-flare row instead [REPO — Fenopan x3, record v21].

4. The body

The physiological arm produced the program's best negatives, and it produced them in the order a metrology demands: the instrument first, the verdict second.

The heartbeat is where the cascade dies most cleanly. On interbeat-interval *durations*, raced by exact likelihood against a roster that included a tempered-Lévy renewal alternative, the cascade wins **0 of 83** records at a median of −27,246 nats while the tempered-Lévy model wins 79/83 [REPO — Fenosoma v0]. On heartbeat **log-volatility** — the field the financial rungs are measured on — the healthy cohort reads *rough, not cascading*: $\hat{H} = 0.111$ [0.088, 0.131], with both continuous members rejected 54/54 and 29/29 [REPO — Fenosoma v2]. Placed on the program's cross-domain amplitude coordinate this is the sentence that makes the ladder worth having: healthy heartbeat log-volatility is rough **and nearly non-intermittent** — index-like roughness with about a hundred times less intermittency than any financial asset re-estimated with the same estimator (heart λ^2_sfbm $\approx 4\times10^{-4}$ against stocks ≈ 0.058, indices ≈ 0.038) [REPO — Fenopan x2, record v1]. The coordinate is the identifiable combination $\lambda^2 = H(1-2H)v^2$ read at matched scale [VERIFIED — Wu, Muzy & Bacry 2022]; $\hat{H}$ is never compared across domains, because H and v^2 sit on a measured identifiability ridge. "It's fractal" would have grouped these objects. Measured on one identifiable coordinate at matched scale, they are two orders of magnitude apart.

Sleep staging is the negative with the most transferable methods asset attached. Nothing beats climatology on any clock or any coarsening [REPO — Fenosoma v1, v4] — and the reason is visible in the layers. With clock, power and readout all proven sufficient, **every layer of the fitted regime filter tracked respiration** (Spearman correlations up to 0.43) and nothing tracked sleep stage (≤ 0.12) [REPO — Fenosoma v4]. The

filter converged cleanly onto the wrong physiological signal. The standing rule taken from that measurement now binds every physiological or neural fit the program will run: no cascade claim ships without the per-layer readout, each layer's smoothed marginal correlated against a pre-declared nuisance panel and reported beside the headline whichever way it lands. A nuisance oscillator that is not looked for is a mechanism claim waiting to happen.

One positive sits inside the physiological negative, and it has to be stated the way the map now states it. All **54/54** healthy-RR records **depart the reversible-linear-Gaussian (IAAFT) null** under a wavelet scattering omnibus, with the time-asymmetry family relatively dominant on 28/54 [REPO — Fenosoma v3, scoped by Fenopan v20 and restated in the map's v21 cut]. The word *nonlinear* has left that cell: what survives is *not reversible-linear-Gaussian*, plus the features as biomarkers (§7 says why). It was never a cascade either.

The arm's most recent work, though, is not a verdict at all. It is a measurement of what an event-count licence is worth. An earlier record had exported a constant — a recovery floor of roughly 800 events, above which a fitted mechanism could be called recovered [REPO — Fenosoma v6] — and that constant travelled into two clauses of the program's canon. It was withdrawn on 2026-08-31, because the battery that produced it simulated its truth on exactly the grid it was then fitted on; re-run off-grid at its own pinned truth point, recovery collapsed to 0/6, 2/6 and 1/6 at N = 200/800/3,200 against 1/6, 6/6, 6/6 on-grid. The floor did not move — it **disappeared** [REPO — Fenosoma v9].

The re-measurement, run off-grid as the withdrawal required, returned something more useful than a number. On a seven-point panel at twelve seeds, recovery reads **2/12, 5/12, 9/12, 5/12, 2/12, 1/12, 0/12** at N = 200, 400, 800, 1,600, 3,200, 6,400, 12,800 — **non-monotone in N**, peaking at 800 and reaching 0/12 (Clopper–Pearson [0.00, 0.26]) at 12,800, as events per fit tick sweep 0.47 → 32.77. *No* N in panel clears the gate. Re-fitting the **identical streams** at a finer intensity resolution takes N = 12,800 from **0/12 → 4/12 → 11/12** ([0.62, 1.00]) with not one event added [REPO — Fenosoma v10]. So the clause the canon now exports is not a floor but a pair:

> An event-count recovery floor is licensing-bearing only as the pair (N, fit grid): "N ≥ n" with no declared intensity resolution is not a floor, and the canon should export the pair or export nothing.

Read as one unit — v6 with v8, v9 and v10 — this is the map's newest row and its twenty-eighth: excitation-on-trend beats trend-alone (23/23 by BIC; the branching ratio surviving detrending at **0.83 [0.75, 0.91]**, never v6's 0.91), scoped by v9 to a comparison *inside a declared basis* rather than a mechanism attribution, licensed by v10's pair, with v6's withdrawn floor deliberately absent [REPO — Fenopan x3, record v21]. Two riders travel with it. The failure is one-sided — the Hawkes diagonal is 12/12 at every N, and only the cascade is not recovered as itself. And the fitted branching ratio on **zero-excitation cascade truth**, a true self-excitation of exactly zero, **rises monotonically with N**, from 0.484 at N = 200 to 0.856 at N = 12,800. More events make the mimicry worse.

5. The body politic

The fourth music is the one Boethius did not name: wars, defaults, outbreaks — the recorded catastrophes of civilization, read as event streams. The arm that reads them began by refusing to read them. Its founding record is not a verdict about history but a **decidability atlas**: at what sample size can a given mechanism be told from its rivals at all? Known onsets, exact likelihoods, seeded panels: flat Poisson is identifiable from N ≈ 100, MSM-Cox from ≈ 300, Hawkes from ≈ 1,000, a flat hidden Markov model only from ≈ 10,000 [REPO — Fenocrisis v0]. Below the relevant floor, the committed result is the *undecidability statement*, not a world statement. The same record supplies the program's most quoted trap in its purest form: a process with true self-excitation of exactly zero, fitted by a Hawkes model with a constant background, reads $\hat{\eta} = 0.80$.

Panel power lives in the cross-section, and the licence is **coupling-conditioned rather than a bare domain count**: moderately coupled domains reach the recovery threshold at D = 2, weakly coupled ones nowhere in the grid [REPO — Fenocrisis v4]. "Three domains" is not a design; three *coupled* domains is.

With the floors in place, the entanglement claim could be raced. On onset timing, the cross-domain coupling **survives raw nulls and dies under the declared background family**, and the mechanism is undecidable at the corpus's own N [REPO — Fenocrisis v1 + v2]. On the consequence channel — severity rather than timing — a converged multi-seed re-run finds a seed-robust **within-conflict severity dependence on recent history** (rank p = 0.010, primary ΔBIC excess ≈ 7,654), a far fainter within-disaster one (≈ 34 BIC), and **no cross-domain shared severity regime** at all (contemporaneous correlation −0.115, one-sided p 0.949) [REPO — Fenocrisis v9]. Coupling fails on the consequence channel as it failed on onset; the within-domain clustering is a within-domain positive. And the program declines the obvious word for it: the sealed record stores only $|\hat{\varphi}|$, so the **sign is unrecoverable**, and until a signed record seals, no repository in the program prints a direction for that dependence. Severity is the open door in this arm — the one live channel — and it is being held open with the honest label rather than the interesting one.

Two apparatus results from this arm are load-bearing downstream. The first is the **span mirror**: an MSM-Cox ladder fitted to single-mechanism near-critical power-law-Hawkes truth reports a broad, genuinely loaded span — 1.14 decades, 87 % of a true multi-scale cascade's 1.31 — on data with *no ladder hierarchy at all* [REPO — Fenocrisis v2]. The ladder has one vocabulary for long memory, scales, exactly as Hawkes has one for elevated rates; a fitted span is never evidence of hierarchy until self-excitation has been raced.

The second landed this month and is the cleanest piece of self-auditing in the constellation. The panel likelihood is **exactly invariant under $\kappa \to -\kappa$**: over three values on a fitted stream the worst $|\text{loglik}(+\kappa) - \text{loglik}(-\kappa)|$ is $\mathbf{0.000 \times 10^{0}}$, bit-identical, and the mechanism is structural — flipping the loading's sign is a pure relabelling of a latent state space symmetric about zero, so it does not go away with more data [REPO — Fenocrisis v13]. Every signed-$\hat{\kappa}$ gate in the constellation is therefore a gate on a non-identified sign. The

honest response was neither to hide it nor to overstate it: re-scoring the sealed atlas on $|\hat{\kappa}|$, from that record's own stored values, moves it from 53 loaded cells at S = 1.1420 to **54 at S = 1.1516** — it moves **no verdict**, and it is carried forward rather than retro-fitted into sealed code.

The same record extended the degenerate-cell map into the corner where every real seismicity candidate operates — an Omori–Utsu decay of p ≈ 1.1, a power-law Hawkes tail at α ≈ 0.1, the opposite corner from the rough-volatility band mapped earlier. The span mirror **reproduces there**: on single-mechanism truth with no hierarchy, fits read mean $|\hat{\kappa}|$ = 0.273 with 0.40 of reportable replications above the loading gate (Wilson [0.23, 0.59]) and a span of **S = 1.53 decades** on the loaded ones ([1.17, 1.89]; ladder-degenerate replications excluded from both numerator and denominator — 11 of 36 — flagged-included, 1.27), against a re-run reference cascade's 1.15 and with a matched positive control certified at the same tiers. And the record insists on a calibration the consuming canon had almost dropped: a no-hierarchy **flat-Poisson null fits S = 1.39 at $\hat{\kappa} = -0.03$** — an *inert* ladder whose noise span sits **above** the reference cascade's. So "broad" means *a substantial fraction of a genuine cascade's span*, and **it is the loading gate, not the span magnitude, that separates a loaded ladder from an inert one**. The record also reports that its registered prose ("a non-trivial fraction of replications") and the predicate actually frozen in its code (*any* non-zero loaded fraction) are not the same bar, that the measurement cleared the frozen one at 0.40, and that it is not being presented as having cleared the prose one. That sentence costs the record something and it is in the record anyway.

The crossover that corner was built to measure came back **N-dependent and read tier by tier, never as one number**: no crossover at all at N = 300 (13/36 = 0.36 [0.22, 0.52]); an isotonic crossing at η = 0.850 at N = 1,000 that a fixed-scale-band control shows to be **band-confounded**, and which is therefore refused as a selection onset in N alone; and at N = 10,000 a crossover **left-censored at ≤ 0.8** — an upper bound, not a measurement [REPO — Fenocrisis v13]. The pooled figure across tiers exists and is explicitly not the claim-bearing quantity. That corner is now part of the map: the v21 cut extends the crisis-coupling row's degenerate-cell sentence into it, carrying the tier-by-tier band-axis reading the record requires — **and the row's verdict word does not change**, because v13 is a known-truth apparatus record that reads no catalog [REPO — Fenopan x3, record v21].

6. The traps

A trap, in this program's sense, is not a mistake. It is a place where a likelihood does its job correctly and the model class does the lying. The registry is kept because each trap in it has been walked into, deliberately, in more than one arena the program owns.

The branching-ratio trap. A Hawkes process has exactly one vocabulary for an elevated rate — past events — so any unmodelled background variation is absorbed as apparent self-excitation. Reproduced in-house five times, independently: true η = 0 read as $\hat{\eta}$ = 0.80 on synthetic onsets [REPO — Fenocrisis v0]; a provably zero-self-excitation smooth solar cycle reproducing η = 0.956 against a committed 0.957, retiring the program's own number [REPO — Fenocosm v21]; on real ant-colony data, a naive $\hat{\eta}$ of 0.85 reading

higher still (0.95) on a self-excitation-free intensity-preserving surrogate, strictly higher on 15/16 replicates [REPO — Fenopan x6, record v5]; on the real GOES flare record, $\eta = 0.999$ at both thresholds reproduced exactly by a zero-self-excitation null → CONFOUNDED [REPO — Fenocosm v29]; and, as §4 records, an $\hat{\eta}$ on zero-excitation cascade truth climbing to 0.856 as events are added [REPO — Fenosoma v10]. The standing rule is that a branching ratio is never evidence of contagion until a regime or background alternative has been raced and a bidirectional recovery gate has passed at the actual N.

That rule's *novelty* claim did not survive the program's own audit, and the correction is part of the finding. The trap is published and quantified — calibrations of Hawkes models on regime-switching Poisson mixtures with true $n = 0$ return estimates hovering around unity that Poisson residual tests cannot reject [VERIFIED — Filimonov & Sornette 2015]; on real tick data the average branching ratio falls monotonically from 0.74 under a constant background to 0.41 under a flexible one [VERIFIED — Omi, Hirata & Aihara 2017]; and the identifiability failure is proved in the fully nonparametric case [VERIFIED — Chen & Hall 2013, 2016]. The earlier in-house wording — that the rule did not exist in print — is **struck as false**. What the program contributes is not the rule but the cross-domain *measured confusion/recovery framing*: one trap, four arenas, a bidirectional recovery gate at the actual N in each. Striking one's own novelty claim is the same act as withdrawing the 800-event floor, and it is why this essay can be trusted about anyone else's instruments.

The width pincer. The field-standard test of multifractality — a two-point multifractal spectrum width read against IAAFT surrogates — has near-zero detection power against its *own canonical positive control*. On the binomial cascade, the width flags 0.125 where a wavelet scattering omnibus flags 1.000 on identical nulls; it is not rescued by any test rule and it is not a length artifact [REPO — Fenocosm v22 and Fenosoma x7, record v3 — found independently in two arenas in one day]. The mechanism is structural: width is a symmetric two-point functional with zero power against time-asymmetry *by construction*, IAAFT hands the marginal-encoded part of cascade multifractality back to the surrogate, and what IAAFT destroys is fourth-order — which width does not measure. Hence the **pincer**: the width reads too high when nothing is there (trends, finite size, isolated singularities) and too low when something is (marginal-encoded, fourth-order, time-asymmetric). What dies is every "the width did not exceed its surrogates, therefore linear" inference — void as *uninformative*, not as wrong; what survives is every likelihood-based result, and width as a biomarker. The scattering audit does not vindicate multifractality either: the specifically multiplicative cross-scale family dominates in only 3 of 22 sources [REPO — Fenocosm v22]. Since 2026-09-15 the pincer has a **third jaw**, and §7 is where it is paid for: *the scattering vector — τ first, and on matched pulse trains every family — reads* nonlinear *when a* linear *irreversible process would do; an IAAFT excess is evidence of nonlinearity only after an irreversible linear null has been raced* [REPO — Fenopan x13, record v20]. A second saturation regime is priced beside it on a counterparty's own data: any Gaussian process rank-mapped onto a marginal that is 59 % exact zeros fires the omnibus 8/8, so a flag on such a marginal is uninformative by construction [REPO — Fenocosm v60].

The observation process, and the nuisance oscillator. Exposure masks manufacture Weibull clustering (shape ≈ 0.46–0.54 out of a homogeneous Poisson through a real mask); reporting baselines absorb or manufacture coupling; smooth periodicity inverts width signs; detection floors read as smoothness [REPO — Fenocosm v21, v22]. No event-time claim ships without its mask-aware surrogate band, and no physiological fit without its per-layer nuisance readout (§4).

The audit campaign

If four traps are real, they are not the program's private property. From 2026-09-01 the constellation ran the obvious and uncomfortable experiment: point the calibrated instruments at the literature that uses the uncalibrated ones, under one protocol applied unchanged to every target. The pipeline is five stages — reconstruct the published pipeline with every unstated choice recorded as unstated; reproduce it on the paper's own data or a named public equivalent; measure **positive-control power at the paper's own operating point** before any null is read; run seven pre-committed calibrated null families; re-analyse with the modern stack — and one of four verdict words: SURVIVES, ATTENUATED, INDETERMINATE, REVERSES. The register is **allied**: every card is published including survivals, the stratum is always the named papers and pipelines rather than "the field".

Eight cards are executed. Movahed *et al.* 2006 on sunspot multifractality **REVERSES**: the published shuffle rule flags spectrum-matched *linear* nulls carrying no multifractality at all (an IAAFT surrogate of the target itself 24/24, phase-randomised 23/24, AR(2) 22/24) while the modern IAAFT rule clears every one of them (≤ 1/24), and Fourier-truncating the 11-year band — 57 % of the variance — drops the width by 70 % [REPO — Fenocosm v39]; the paper's own facts reproduce, and its adaptively detrended residual is genuinely multifractal on all three modern instruments — allied nuance, reported with the same prominence [REPO — Fenocosm v43]. The Ihlen 2012 MFDFA tutorial, audited *as the instrument*, is **ATTENUATED**: its shipped raw reading is tripped 24/24 by a linear surrogate of the canonical cascade, while the modern IAAFT rule clears that null to 1/24 but is itself **blind to the deterministic binomial — 0/24** at N = 1,024 and 4,096 — where the scattering omnibus and a Chhabra–Jensen statistic detect it 24/24 [REPO — Fenocosm v40]. That card also carries an erratum retracting its own allied centrepiece: the tutorial contains no surrogate guidance at all.

The gait line splits by target: **Hausdorff SURVIVES** — stride intervals are long-range correlated (control $\alpha = 0.905$, bootstrap [0.850, 0.958]) and the low-order exponent discriminates Huntington's disease at AUC 0.85, $p = 0.001$ — while the **Dutta** MFDFA width successor **REVERSES**: under Dutta's *own* stated conventions the widths read 0.88 / 0.88 / 0.84 against a published 3.7 / 2.2 / 2.15, permutation $F = 0.09$, $p = 0.918$, no cell of four carrying the published direction [REPO — Fenocosm v41, v42]. At gait's N ≈ 256 the card measures the regime that makes this inevitable: the width detects a genuine cascade 0/24 while the one powered instrument fires on a drift trend 16/24. At that length "unaffirmable either way" became a pre-registered, publishable verdict.

Ivanov *et al.* 1999 — the founding positive of the biosignal multifractality literature — is **ATTENUATED**, and the allied half is the larger one: the healthy heartbeat's multifractality is **genuine under the marginal-and-spectrum-preserving null** (16/18 records clear their calibrated budget), so the 1999 conclusion survives the modern instrument; what cannot certify it is the paper's own attribution step, which a heavy-tailed monofractal reproduces 16/24. The congestive-heart-failure "loss of multifractality" holds as a **loss of nonlinear width** (3/14 against 16/18 healthy), not as a narrower raw spectrum (AUC 0.50) [REPO — Fenocosm v56]. Ihlen & Vereijken 2010 is **ATTENUATED**, and the finding is a convention rather than a statistic: the paper's rule flags 51 of 987 series under its primary reading and 63 under the rank reading — the nominal rate — while reading "the 95 % confidence interval of the surrogate widths" as a *standard error* flags **478 of 987**, and the successor instrument's published rule 495; that convention fires on about a third of monofractal nulls where the rank reading of the same ensembles fires ≤ 1/24 at full power [REPO — Fenocosm v57].

The neural card is **ATTENUATED** with one reversal inside it: He 2011's exponent and its coupling to the Hurst exponent survive (grand-average $\beta = 0.63$, $r = 0.91$ across 280 series) while the *scale-free description* fails the paper's own goodness-of-fit on 143/280 series — a linear crossover — and the network ordering inverts; Zorick & Mandelkern 2013's published width is the width of a linear process with the EEG's spectrum and marginal; Rácz *et al.* 2019's per-band nonlinearity pattern reproduces, and so does a linear Gaussian EEG on 88 % of channels [REPO — Fenocosm v58]. The home-domain anchor — "musical rhythm obeys a 1/f power law" — is **ATTENUATED** with the **1/f form REVERSING**: the slope exists, the paper's own onset-shuffle null controls it, its genre ordering reproduces ($r = 0.77$), and the published population numbers even return under a tempo mapping the text never states — but **β moves +0.17 per tempo octave**, the two decades disagree, and the DFA reading of the same raster sits 0.59 below the spectral one. The slope is the envelope of a metric hierarchy, not a power law [REPO — Fenocosm v59].

The eighth card matters most for the program's credibility, because the result it audits is the one the constellation's own papers cite as expressiveness evidence. Roeske, Kelty-Stephen & Wallot 2018 on birdsong **REVERSES**. The positive reproduces — 21 of 23 songs beyond their IAAFT band at 8.1 SD, exceeding every calibrated *linear* budget — but the same songs with their **notes permuted as units** pass the paper's rule **184/184 times at 11.6 SD**, *farther* than the songs themselves; against its own note-shuffled ensemble the original is wider on **0/23** and narrower on 15/23; and a single cusp in a Gaussian monofractal passes 24/24 at 25 SD. The width is the note envelopes — the isolated-singularity high side of the program's own trap registry — not structure across timescales [REPO — Fenocosm v60].

What the eight cards found. The modal outcome is not a reversal. It is: *the conclusion survives, but the step cannot certify it.* Six of eight bundle words are ATTENUATED, and inside them the paper's substantive claim usually stands under a better null while the published pipeline is shown unable to license it. Two mechanisms carry almost all of it. The first is a **null/statistic mismatch**: the shuffle rule has no specificity against linear correlation, the IAAFT width rule has no power against marginal-encoded structure, the powered short-N statistic has no specificity, and the scattering omnibus has a priced high side on isolated

singularities. The second is **unstated conventions that carry headlines**: a confidence interval read as a standard error, a tempo mapping, an outlier cut, a scale range, a decibel transform. Each is invisible in the published text and each moves a headline by more than the effect it reports.

The campaign's most citable object is therefore not a verdict word at all. It is the **cross-card instrument table**: nine instruments, scored on five data forms — index and count series, interval series, response-time series, continuous neural signals, and music as both symbolic raster and audio envelope — at lengths from N = 256 to N = 120,000, each cell carrying a measured power on a genuine cascade and a measured false-positive rate against calibrated nulls [DERIVED — a tally across the eight cards' sealed records]. Its published core, sealed across three cards and two campaign-level batteries, reads:

instrument (rule)	power on a genuine cascade	specificity on calibrated nulls
width vs **shuffle** (the practiced rule)	binomial 22/24 at N = 4,096	**none against linear correlation**: IAAFT-of-target 24/24, phase 23/24, AR(2) 22/24
width, raw, no null	wide on the cascade ($\Delta\alpha \approx 1.22$)	**none**: IAAFT-of-cascade 24/24, isolated singularity 24/24
width vs **IAAFT** (the modern rule)	**blind** to the deterministic binomial: 0/24 at 1,024 / 4,096 / 3,120 / 256	clears every linear null (≤ 2/24); trips isolated singularity 24/24
wavelet-leader c_2 vs IAAFT	shares the blind spot (0/24 at 4,096)	clears Gaussian monofractals and heavy tails
Chhabra–Jensen $\Delta\alpha$ + t_MF	24/24 at 4,096 and at 256	**none at N = 256**: drift 16/24, AR(2) 13/24, IAAFT 11/24
scattering omnibus	24/24 at N ≥ 3,120; 14/24 at 256	clears every linear null at 256 (≤ 3/24); isolated singularity 24/24

Sealed across Fenocosm records v37, v38, v39, v40 and v41. Read down the columns and the field's standing disagreement dissolves into a measurement: there is no single instrument that is both powered and specific at every length, and which one a study used predicts its conclusion better than the data does.

7. The boundary where dominoes become weather

Everything above converges on one question, and it is the question the program was built to answer: given a clustered event stream, is the clustering **exogenous regimes** (weather), **endogenous self-excitation** (dominoes), or **smooth-or-rough drift**? It is the standing argument in seismology, market microstructure, epidemiology, neuroscience and crisis studies, argued everywhere with incompatible apparatus and no power curve.

Posed as classification, it is ill-posed near the boundary, and the reason is a theorem rather than a failure of effort. A nearly-unstable Hawkes process with power-law kernel tail $\alpha \in (½, 1)$ converges to rough volatility with $\mathbf{H = \alpha - ½}$ [VERIFIED — Jaisson & Rosenbaum 2015], and as $\alpha \to ½^+$ the endogenous

limit *is* the log-correlated cascade. Inverting against the physiological measurement, $\hat{H} = 0.111$ implies a kernel exponent $\alpha \approx 0.611$ [0.588, 0.631] [DERIVED bridge]. Rough drift may simply be the scaling limit of dominoes, and the trichotomy a dichotomy plus an observational limit.

The degeneracy is now **bidirectional on known truth**: a synthetic rough world reads as self-excitation, and near-critical power-law Hawkes truth reads as a cascade in 0.81 of head-to-head replications at $\eta \geq 0.99$, the crossover setting in from $\eta \approx 0.8$ [REPO — Fenocrisis v2]. And it has a real-data instance. Nine *Leptothorax* colonies — the only public stream above all four decidability floors at once — land exactly on the boundary: the held-out tournament **decisively rejects** the smooth two-clock null (16/16), and then the colonies **scatter along the identifiability ridge**, $\mathrm{corr}(\hat{H}, \ln \lambda^2) = \mathbf{-0.975}$, reproducing on real ants the −0.98 the estimator's own coordinate system predicts. Rough colonies read as drift, low-$\hat{H}$ colonies at the amplitude floor as shallow cascades, and the tournament prefers the cascade corner by only ~0.06 nats per bin over a generic hidden Markov model [REPO — Fenopan x6, record v5]. The recovery gate makes the reading trustworthy: the same tournament separates the corners when they are strong, so their merging on the near-floor colonies is a property of the *data*, not the instrument.

Which brings the essay to its sharpest conjecture, and to its refutation. **T4** held that **time-irreversibility is the trichotomy's actual discriminator**: the cascade's latent chain is reversible by detailed balance, Hawkes is not, and both inference arms had already built the proven-odd-under-reversal statistic without noticing what it might discriminate. T4 carried a stated risk to the program's own flagship win, since all 22 Swift/BAT sources — GRS 1915+105 among them — **depart the IAAFT null**, with the time-asymmetry family dominant in six [REPO — Fenocosm v22] — which is exactly why this program had to run it. That cross-reference is now scoped in the map itself: the v21 cut reads the X-ray-binary row's τ excess as a departure from the reversible-linear-Gaussian null **and nothing stronger** — no "nonlinear", "endogenous" or "multiplicative" word rides on the scattering evidence — while leaving the row's verdict untouched, because that verdict is an out-of-sample *likelihood* verdict and x13 raced no tournament [REPO — Fenopan x3, record v21].

It has now been run in three arenas, and it fails in all three. In a **coarse counts arena** the reversibility *proof* holds — $E[\tau] = 0$ is an exact null, not a surrogate — but τ's sign is a near coin-flip on self-exciting generators [REPO — Fenocrisis v2]. At **fine resolution**, on sixteen sealed ant replicates at 0.5-s bins with a nine-width sweep, the reading is **uninterpretable by rules frozen in code and printed before the run**, and both failing predicates are measurements rather than shrugs: the panel is null-ish (1/16 significant after correction; median $|z| = 1.64$; $\bar{\tau} = -1.9 \times 10^{-4}$, negative on 15/16) and an **off-grid** power battery certifies power 1.000 (8/8) at $\eta = 0.6$ — so a certified negative was in reach — but the instrument's own false-positive rate against the *rough* world at that width is **0.333**, above the pre-registered 0.15 ceiling, and rough drift is precisely the corner these colonies occupy; while a rate-matched control reproduces the ridge slope on worlds known to be null, voiding the second reading. The wash-out scale is measured: power falls below one-half by a bin width of 5–10 s for a ~1.7-s excitation [REPO — Fenopan x10, record v19]. In the **seismicity corner**, τ does not discriminate on the band the statistic actually sums [REPO — Fenocrisis v13].

And then the null that should have existed from the beginning was built. **Every standard surrogate family the program owned — shuffle, Fourier, IAAFT, autoregressive — is time-reversible by construction**, so no control it had could ever *fail* on a time-asymmetry statistic. The missing control is Poisson shot noise with an asymmetric two-exponential kernel: linear, irreversible, matched per target on the marginal by exact rank remap and on the spectral density by a zero-phase shaping filter, read by the siblings' *own* vendored instruments so that irreversibility is the only property the ensemble adds. With 39 matched members per cell and a coverage rule frozen before the real pass, **22/22** Swift/BAT sources are covered at every rise:decay ratio (Clopper–Pearson [0.846, 1.000]) and **54/54** healthy-RR records at at least one ratio ([0.934, 1.000]), with median coverage probability 1.000 at every ratio in both domains: the matched linear pulse trains are *more* time-asymmetric than any real target — by an order of magnitude on the cardiac cohort [REPO — Fenopan x13, record v20]. The same controls catch the omnibus in the pincer's **third jaw**: on matched pulse trains the scattering omnibus flags 1.000 and *every* family flags ≥ 0.935, against sealed budgets of 2/51 and 0.010, so the program's own 22/22 and 54/54 scattering departures sit inside what a matched linear pulse train produces.

So T4 now carries the grade **REFUTED AS POSED**, in all three arenas, and the scope clause is sharpened with it: irreversibility is evidence that a *pure reversible-ladder description is incomplete* — and "incomplete" can be completed by a **linear irreversible driver**; no nonlinear, multiplicative or endogenous component is implied by it. A τ excess separates *reversible-linear* from *everything else*, which is not the trichotomy's boundary. The grade travels with a **re-entry condition**, so it can be earned back rather than argued back: a time-asymmetry statistic with certified specificity against **both** rough drift **and** matched irreversible-linear pulse trains, at the arena's own resolution and N. Until such a statistic exists, no verdict in this program rides on τ. And the trichotomy this section opened with gains a **fourth corner** — *linear superposition of asymmetric pulses*: irreversible, linear, carrying no hierarchy at all, and the corner every τ-based discriminator was silently reading. What x13 does *not* say is that any source or heart is shot noise, and it touches none of the likelihood tournaments. The honest position at the boundary is the one the abstract already promised: where self-excitation becomes indistinguishable from fate, no one can yet tell — and now the program can say *how far* from telling, at what resolution, against which null.

8. The metrology

The thesis of this essay is that the map in Figure 1 is not the finding. The finding is the **instrument that produced it, measured**.

Consider what the program can now state about its own limits. There is an **atlas** of identifiability floors on known onsets — flat Poisson from N ≈ 100, MSM-Cox ≈ 300, Hawkes ≈ 1,000, a flat hidden Markov model ≈ 10,000 [REPO — Fenocrisis v0]. There is an **event-altitude selection floor**, N*_selection ≈ 800 at a multiplier of 1.8, travelling with its uncertainty rather than as a constant: a six-seed panel crossing on a non-monotone rate curve, Clopper–Pearson 95 % on 5/6 = [0.36, 1.00], grid bracket [400, 1,600] [REPO — Fenocosm v32 + v35]. There is a hard distinction between that **selection** floor and a **recovery** floor: the

corrected reading puts the multiplier's recovery *below* the selection floor while a fitted **span is not recovered anywhere in the tested range**, so a span remains a model-selection quantity until shown recovered at the actual N [REPO — Fenocosm v35]. There is a **decidability surface** in (dwell count, events per dwell) that refines the program's own "about four dwells" reading into an absolute **event-count floor of order 10^3 events** — certified specific only against a fast autoregressive null, because a matched trend or periodic component is called a slow clock essentially all the time [REPO — Fenopan x5, record v4]. There is the pair clause of section 4 above [REPO — Fenosoma v10]. And since 2026-09-15 the floor clause has one more condition: a **ladder-degeneracy-flagged fit supplies no amplitude at all** — the real Kp fit reads $\hat{m}_{hi} = 3.372$, far above the floor, on a ladder flagged S = 1.00 whose parameters the flag declines to report [REPO — Fenocosm v55].

Every one of those is a statement about what *cannot* be decided, obtained at the same standard as a positive result. That is what "metrology" means here, and it has a price list. Negative results are deliverables, published in the same register as positives — but only with the certificate that the instrument could have seen a positive. The sunspot reversal needed its power certificate *before* any null was read. The ant irreversibility record had a certified negative within reach and **refused it**, because its own specificity world failed at the reference width. The cross-domain audit pre-registers INDETERMINATE as a publishable verdict for the short-N regime, and used it. A field that cannot say "unaffirmable either way" will say something else instead.

The same discipline has been applied inward, repeatedly, at cost. The 800-event recovery floor the program exported was withdrawn as a grid-matching artifact and came back as a pair, not a number. The most-quoted parameter in the canon was shown to sit at its own calibrated zero-correlation floor, and the clamped point estimate that had made that finding look stronger was withdrawn while the finding was strengthened. A novelty claim was struck as false against prior art the program had not found the first time. An aggregation result's real-data half was downgraded from "closes the gap" to *consistent with, but not powered to exclude* [REPO — Fenopan x8, records v7 and v17]. The sharpest conjecture in the canon was refuted as posed, in three arenas, by the program that proposed it — and the time-asymmetry departures the flagship arm had counted as evidence were shown to be covered by a linear null that no one, this program included, had built.

None of those is a scandal and all of them are the result. A map of where hidden hierarchy is real is worth exactly as much as the calibration of the instrument that drew it, and the only way to establish that calibration is to keep measuring the instrument against worlds whose truth is known, and to publish what comes back. Twenty-eight sealed cells, four provenance grades, one estimator mirror under one claim discipline — and, beside them, a growing register of the things this apparatus provably cannot tell apart at the sample sizes the world supplies.

What it means for a field to know what it cannot decide is simply this: the argument stops being about who is right and becomes about which cell of the decidability surface the data sits in. That is a smaller kind of claim and a much larger kind of progress.

9. Coda: the four musics as one program

Boethius divided music into the spheres, the body, and instruments, and the division looks like poetry until you try to test it. One object, played in all three and in the fourth he did not name, sorts them: it wins on performed music and multi-state X-ray binaries, where a system genuinely re-occupies discrete regimes across separated timescales; it loses on hearts, on smooth galactic nuclei, on labyrinths, on crisis onsets, and on geomagnetic storms, and each loss says something specific about *why*. The sorting is law-shaped, and the law has four clauses: recurrence rather than mere hierarchy; amplitude above the estimator's floor at the available N; no competing modulation of comparable variance in the observation process; and a measured object that is the unit where the multiplicative feedback closes, not an aggregate over many such units.

But the essay's real claim is smaller and more durable than the sorting. Across the constellation's record the program found the same buried flaw in four literatures — mechanism claims made with instruments whose power to detect the mechanism had never been measured — and then found it at home: in its own branching ratios, its own spans, its own most-quoted parameter, its own sharpest conjecture, its own citations. The instrument that can find that flaw abroad is the one that has been made to fail at home, on purpose, with the failures published.

Boethius was proposing a research program. It needed an estimator, and it turned out to need something the estimator alone could not supply: a metrology — a standing account of what the estimator can see, at what length, against which null, and a habit of saying so when the answer is that it cannot see anything at all.

Reproducibility

Figure 1 is rendered deterministically from the sealed inventory cut the canon's §8 points at by `make_figures_verdict.py`, which reads the committed record and re-fits nothing: the figure is a view of sealed numbers, not a new computation. Every record quoted here is regenerable from its own deposited object — each carries its JSON, its findings file with the epistemic tags, its pinned seeds and its regeneration command line — and sealed records are append-only, so a correction is a dated append or a new record beside the old one, never an edit in place — the inventory's four prior cuts all still reproduce beside v21. Three gates run over the apparatus: an integrity and anchor gate on the vendored estimators, a citation gate binding every vendored headline to the record it is read from and the canon value it quotes, and a prose gate binding each record's narrative to its JSON.

Data availability

No raw third-party data is redistributed. The measurements quoted here read: Swift/BAT transient-monitor daily light curves (Krimm *et al.* 2013); PhysioNet beat-annotation and polysomnographic databases (Goldberger *et al.* 2000; ODC-By); the Richardson *et al.* 2017 *Leptothorax* interaction data (Dryad, CC0);

WDC-SILSO sunspot numbers (Royal Observatory of Belgium); the Penticton/DRAO F10.7 radio flux (NRC Canada); GFZ Potsdam Kp (CC BY 4.0) and WDC Kyoto Dst (scientific use, no commercial application); public lexical-decision, EEG, fMRI and gait corpora named in the cards that read them; and a committed, license-clean derived realized-variance panel for 180 equities. Each sealed record carries its own per-file provenance, licence basis and pinned checksums, and is deposited as a standalone object with its own DOI (appendix); the few that carry no DOI yet are marked *deposit pending* there, and the essay quotes from them only what the canon already exports as a clause. The essay itself is deposited as a standalone Zenodo preprint at `10.5281/zenodo.22759978` (Markdown, PDF and Figure 1; DOI reserved 2026-09-15; published 2026-09-15 on the owner's sign-off), citing every deposited record in the appendix by DOI in its metadata. **Dated note (2026-09-15):** Figure 1 and every inventory citation moved from the `out/v18` cut to the `out/v21` cut on the day the re-cut landed; both cuts are deposited, and this essay quotes only what `out/v21` carries.

Appendix — provenance table

Every DOI carries the prefix `10.5281/zenodo.`; the suffix is given. Records marked pending are sealed and regenerable but not yet deposited; their numbers travel here as canon clauses with their record id, never as an undocumented figure.

repository	record	what it carries here	DOI suffix
Fenopan	v0, v12, v14	x1 crosswalk; x12 piano rung and its null calibration	pending
Fenopan	v1	x2 — the matched-scale amplitude ladder	22183540
Fenopan	v4	x5 — the decidability surface; the event-count floor	pending
Fenopan	v5	x6 — the ant-colony trichotomy tournament	22180448
Fenopan	v6	x7 — the labyrinth, the anti-cascade	22180453
Fenopan	v7, v17	x8 — the aggregation law, synthetic and real halves	22183544, 22180457
Fenopan	v21	x3 — the sealed-fit inventory, x13 consumption cut (28 rows); **Figure 1**	22760255
Fenopan	v18	x3 — the prior inventory cut (27 rows), sealed beside v21	22759719
Fenopan	v19	x10 — T4 at fine resolution	22759721
Fenopan	v20	x13 — the irreversible-but-linear null	22759725
Fenocosm	v0–v15, v24, v27, v30, v31	X-ray binaries, AGN, repeater FRB, blinking, stellar flares	via v21

repository	record	what it carries here	DOI suffix
Fenocosm	v21, v22	event-time surrogates; the scattering audit and the width pincer	22180488, 22180490
Fenocosm	v29	the real-stream branching-ratio confound	22180496
Fenocosm	v32, v35	the selection/recovery split and its corrected reading	pending, 22183546
Fenocosm	v37, v38	the attribution-power and t_MF batteries	22759648, 22759652
Fenocosm	v39, v43	audit card 05 — sunspots — and its addendum	22759656, 22759667
Fenocosm	v40	audit card 01 — the tutorial as instrument	22759659
Fenocosm	v41, v42	audit card 02 — the gait line — and its addendum	22759663, 22759665
Fenocosm	v56, v57	audit cards 03 (heartbeat) and 04 (cognition)	22759699, 22759701
Fenocosm	v58, v59, v60	audit cards 06 (neural), 07 (rhythm), 08 (birdsong)	22759703, 22759705, 22759707
Fenocosm	v45, v46, v47	calibration; cheap challengers; the depth ablation	22759673, 22759675, 22759677
Fenocosm	v48, v51	covariate augmentation and the live supply	22759679, 22759687
Fenocosm	v49, v50, v52	horizons; the real-label X class; the marked channel	22759683, 22759685, 22759689
Fenocosm	v53, v54, v55	the geomagnetic arc: data, atlas, tournament	22759693, 22759695, 22759697
Fenosoma	v0	heartbeat durations by exact likelihood	22180536
Fenosoma	v1, v4	sleep staging; the per-layer nuisance readout	22180540, 22180547
Fenosoma	v2, v3	log-volatility roughness; RR nonlinearity under scattering	22180542, 22180545
Fenosoma	v6, v9, v10	the exported floor, its withdrawal, its re-measurement as a pair	22180549, 22183551, 22759711
Fenocrisis	v0	the decidability atlas; the branching-ratio trap	22180514
Fenocrisis	v1, v4	the onset-coupling tournament; panel power	pending
Fenocrisis	v2	the degenerate cell; the span mirror; T4 in counts	22180516
Fenocrisis	v9	the severity channel	22180520
Fenocrisis	v13	the κ-sign non-identifiability; the Omori corner	22759715
Fenophone	`msm-corpus-fit`, `msm-cox-hawkes`	the piano corpus fit, its null floor, the selection share	22180442, 22180446

References

- Calvet, L. & Fisher, A. (2001, 2004). The Markov-Switching Multifractal. [VERIFIED]
- Chen, F. & Hall, P. (2013). *J. Appl. Prob.* 50(4):1006; (2016). *JCGS* 25(1):209. [VERIFIED]
- Crouse, M., Nowak, R. & Baraniuk, R. (1998). Wavelet-based hidden Markov trees. [VERIFIED]
- Filimonov, V. & Sornette, D. (2015). *Quantitative Finance* 15(8):1293. [VERIFIED]
- Jaisson, T. & Rosenbaum, M. (2015). Nearly unstable Hawkes processes. [VERIFIED]
- Omi, T., Hirata, Y. & Aihara, K. (2017). *Phys. Rev. E* 96:012303. [VERIFIED]
- Wu, P., Muzy, J.-F. & Bacry, E. (2022). *Physica A* 604:127919. [VERIFIED]

The eight audited papers are cited in full, with quoted claims and locators, in each card's sealed record.