Predictive Capture: A Single Statistic Prices the Decodability–Usability Gap Across Reservoir Families

A Preprint

Caitlyn Meeks
AuxiLab Tenerife — research@auxi.cafe

August 12, 2026

registered 2026-08-11 | run log #29, #30, #38 | 2 seeds (runs #29/#30); 1 seed (run #38) | 9 strictly held-out cells + 9 out-of-family | replicated before publication | promoted 2026-08-12 | corrected 2026-08-12

Correction notice, 2026 August 12. After this preprint was first posted, three objections raised by external readers were tested against the archived result files in an internal validity triage (internal run log #49), written independently of the analysis code it re-checks. **All three were sustained.** The law's measured accuracy survives unchanged; three of the ways it was described do not. **(i) Held-out purity.** Of the 15 cells the original posting called "held out", three return values bit-identical to the calibration cell in both axes — two are that same reservoir under a different name, and the third is a linear widening [x|F(x)] that spans its column space — and a fourth shares its measured value exactly; the 15 carry only 12 distinct measured values. Just **9** cells are strictly disjoint from calibration, and all 9 belong to a *single* reservoir family. Runs #29/#30 therefore do not by themselves establish cross-family scope; that scope now rests entirely on the separate blind extension reported in §4.1, which survives the same screen cleanly. **(ii) The null.** The pre-registered analytic null is a constant — standard deviation 0.000025 bits over the 15 cells, exactly 0.0 over the strict 9 — so "an analytic null beaten by an order of magnitude" is *withdrawn*. Against baselines with real variance the law still wins on magnitude, by 1.8–2.9× against blind rivals and 1.6–1.8× against an oracle rival, but it is matched on *ordering* by a one-parameter law on linear decode depth alone. **(iii) Tautology.** The intercept is a count over fixed text with no reservoir seed anywhere on its path, so the registered prediction that it be identical across seeds could not have failed. It is struck from the scoreboard, along with eight others that were not risk-bearing either — a constant predictor that could not be lost to, an identity of the feature algebra, instrument self-checks, and restatements of another prediction on the same numbers. The risk-bearing count is **8 passed, 1 missed**, plus one pass bearing on a side finding — not the ~17 confirmations the original presentation implied. The sections below are corrected in place; §4.1, Table 2 and Table 3 are new. No result file was altered, and no measurement was re-run — every number here is recomputed from the archive as it already stood.

Abstract

Three earlier findings in this research program each uncovered a version of the same gap: a recurrent reservoir can *store* information, recoverable by a linear probe, without being able to *use* it in a trained readout — an addressed-memory echo state network (ESN) that decodes deeper into its input history yet predicts next characters worse, a bridged ESN whose future-keyed representational geometry adds nothing to its trained readout's accuracy, and a state-quantization dial that degrades bits per character (bpc) smoothly while decode structure persists. This study asks whether that gap is one measurable quantity or three unrelated

coincidences. We define **predictive capture** $x_c = 1 - R^2_p$: the fraction of a fixed 5-gram language model's next-character distribution that a linear map of a reservoir state fails to reconstruct, measured out-of-sample. The two-parameter law bpc = intercept + $\beta \cdot x_c$, calibrated on exactly two cells of a state-quantized ESN, was applied blind to cells it was not fit to. On the **9 cells strictly disjoint from calibration** — a two-mixing-scheme sweep across ρ within one family, the addressed-memory ESN, incomplete because the one cell it is missing is the calibration reservoir itself — and on two independently drawn random seeds: Spearman(predicted, measured) 0.983 and 0.950, median absolute error 0.043 and 0.041 bits, with seed 0's frozen constants applied to seed 1 scoring 0.950 / 0.044. The fitted β is value-stable across independently drawn reservoirs (2.8076 vs 2.8174, a 0.35% difference), so the law is reproducible in value and not merely in functional form. **Cross-family scope rests on a separate blind extension** (§4.1): the frozen seed-0 constants, applied to 9 never-before-scored cells spanning four further reservoir families, score Spearman 1.0 and median error 0.0166 bits — but that extension is a single run at a single seed, with no replication of its own. Against baselines with real variance the law wins on magnitude by 1.5–2.9×, and beats even an oracle rival permitted to fit its slope on the scored cells by 1.6–1.8×; on *ordering*, however, it is matched by a one-parameter law on linear decode depth, so the distinctive contribution of x_c is calibrated magnitude rather than rank. Two seeds is treated as a floor, not a sufficient result: the program's own independent audit of this entry is still pending, the internal validity triage reported in the correction notice above is *not* that audit, and a third seed on a purity-corrected inventory is the named next step.

Keywords reservoir computing · echo state networks · linear probes · decodability vs. usability · predictive capture · cross-family generalization · pre-registration

§1 The question

A recurring result in this program is a dissociation: instruments that measure what a reservoir state *stores* (linear decode depth, geometric alignment with a target) and instruments that measure what a trained readout can *use* (validation bits per character) rank the same reservoirs differently, in reproducible ways. Reservoir computing's standard practice is to hold the recurrence fixed and train only the readout [2], which makes the two kinds of instrument straightforward to separate. Three separate lines of experiments encountered this independently, each through its own mechanism — content-addressed memory, a future-keyed linear bridge, and state quantization. Either the program had found three different phenomena, or one phenomenon three times. The distinction is testable: if it is one phenomenon, there should exist a single statistic, computable from a reservoir's states without training any readout, that prices all three families' bpc on a single line.

§2 Method

No new reservoir dynamics are introduced. This study reuses the construction code verbatim from the three source experiment families and re-evaluates 17 cells on identical text8 splits [3] (2M training characters, N=1024-unit reservoirs [1]): six cells from the addressed-memory ESN family ({Hadamard-permutation addressing, sparse random mixing} × spectral radius $\rho \in \{0.6, 0.95, 1.1\}$), four low-ρ crossover arms ({permutation addressing, sparse random mixing} × $\rho \in \{0.45, 0.8\}$), four cells from the state-quantized ESN family ($b \in \{1, 2, 4, \text{float32}\}$ bits per unit), and three arms of the future-keyed

linear bridge (forward state x, bridge-mapped state F(x), and their concatenation [x‖F(x)]). An in-run consistency gate required every recomputed ridge-regression validation bpc to match its archived value to three decimal places — 0 failures in 17 cells at seed 0; at seed 1, all 7 cells with a matching same-seed archive matched exactly, and the remaining 10 had no same-seed archive to gate against, exactly as disclosed before running.

The corrected inventory. The original posting described the 15 non-calibration cells as held out across three families. They are not. The calibration cell at b = float32 is the state-quantized ESN with quantization disabled, which is bit-for-bit the house reservoir at N=1024, ρ=0.8, leak 1.0, bias 0.2, tanh. So is the crossover arm at ρ=0.8 with sparse random mixing, and so is the forward state of the future-keyed bridge — three names for one reservoir. The bridge's concatenated arm [x‖F(x)] spans the same column space as that state, and its mapped arm F(x) reproduces the same measured bpc to four decimals. All four therefore carry a measured value copied from a calibration cell, and the 15 hold only **12 distinct measured bpc values and 13 distinct x_c values**. Classified before the recount: tier A, exact construction copies of the calibration cell (2 cells); tier B, transformed features of that same reservoir (2); tier C, the same reservoir interior to the b=1…float32 dial the calibration pair spans (2); tier D, strictly disjoint (**9**). The 9 tier-D cells are the six addressed-memory arms at ρ ∈ {0.6, 0.95, 1.1} and the three crossover arms at ρ ∈ {0.45, 0.8} that are not the calibration reservoir — *one family*. §4 reports tier D as the primary analysis and the published 15 alongside it.

```
# STORAGE INSTRUMENT: predictive capture, no readout training required
#   P5 = fixed 5-gram language model
#        (additive-smoothed, 5M training chars)
#   fit ridge regression:
#        state(t) -> log2 P5[context_4(t)]   (27-dim target)
#   R2_p = pooled out-of-sample R2 of that fit
x_c = 1 - R2_p

# THE LAW: two parameters, calibrated on two cells
#   (state-quantized reservoir at b=1 and float32 only)
bpc_predicted = intercept + beta * x_c
# then applied blind to the 15 non-calibration cells
# (see the purity correction below: only 9 are strictly disjoint)
```

The pre-registered control is an analytic null with real statistical teeth: a credit-apportioned n-gram value, which credits each character decodable at lags 1–4 with the marginal n-gram value of each context extension, telescoping to the 5-gram anchor under perfect decoding.

---

§3 Pre-registered predictions

- **O1** (ordinal): the sign of Δx_c matches the sign of Δbpc in the four risk-bearing matched pairs (addressed-memory permutation addressing vs. sparse mixing at ρ ∈ {0.6, 0.95, 1.1}; quantized b=2 vs. float32). The two bridge-arm pairs are nested-model consistency checks, registered as such and excluded from scoring. Loss condition: ≥2 pairs fail.

- **O2** (ordinal): Spearman(predicted, measured) ≥ 0.8 over the 15 held-out cells; a pre-registered gray zone of [0.6, 0.8) would promote nothing either way; < 0.6 rejects the single-law reading.
- **Q1** (quantitative): median |error| ∈ [0.00, 0.08] bits; > 0.15 rejects the quantitative law even if the ordinal predictions survive.
- **Named overall loss:** the analytic null matches or beats the law on O1 and O2.

The seed-1 replication (predictions E1–E4) re-registered the same thresholds for a freshly fitted law, and added a sharper test, E2: seed 0's already-fitted parameters applied to seed 1's cells with *zero refitting* must clear the same thresholds. One scope reduction was disclosed before running: O1 is scored on raw sign-match of point estimates rather than the originally specified bootstrap interval on Δx_c — an O1 pass is accordingly somewhat softer evidence than the original registration envisioned, and is labeled as such throughout.

Which of these were capable of failing. A prediction fixed by construction is not evidence, and this study registered several. E3 (the intercept identical across seeds) is struck: the intercept is `fivegram_val_bpc(P5, val_full[:500 000])` with the table built from a fixed training slice, a path carrying no reservoir seed and no random-number draw of any kind — re-executing it returns 2.061436 and returns it bitwise identically on repetition. The "named overall loss" against the analytic null is struck for the reason given in §4.2: a constant predictor cannot be lost to by any statistic correlated with bpc at all. Two nested-model consistency checks on the bridge arms were already excluded from scoring at registration and are now known to be exact duplicates of the calibration cell. Also struck is the span lemma tested in §4.1: that [x‖x] spans the same column space as x is linear algebra, and the registration itself treats a violation as a code failure — it is a valuable instrument gate, not a test of the world. Scoring only predictions whose outcome was not determined in advance by construction, this study has **8 risk-bearing passes** (O1, O2, Q1; E1, E2; F1, F2, FQ3), **1 risk-bearing miss** (F4, registered non-blocking), and one further pass bearing on the side finding rather than the law (E4). Nine registered items are struck as not risk-bearing.

---

§4 Results

Table 1. Both random seeds (internal run log #29 / #30), by cell subset. The **tier D** rows are the primary, purity-corrected analysis: cells strictly disjoint from calibration. The "all 15" rows are what the original posting reported, retained for comparison; the triage's recomputation of them reproduces the published figures exactly, so the disagreement is about which cells belong in the set, not about arithmetic. Every subset clears both registered bars at both seeds, with no gray zone approached.

subset	n	families	seed 0 Spearman / median err	seed 1	seed 1, frozen seed-0 law
tier D — strictly disjoint (primary)	**9**	**1**	**0.9833 / 0.0431**	**0.9500 / 0.0409**	**0.9500 / 0.0444**
tier D + C (lenient)	11	2	0.9909 / 0.0368	0.9636 / 0.0376	0.9636 / 0.0412
all 15, as originally posted	15	3	0.9910 / 0.0427	0.9801 / 0.0409	0.9801 / 0.0444
registered threshold	–	–	≥ 0.8 / ≤ 0.08	≥ 0.8 / ≤ 0.08	≥ 0.8 / ≤ 0.08

01 sign matches (risk-bearing pairs; 3 of the 4 lie in tier D)	4	–	4/4	4/4	≥2 fail = loss
fitted β (intercept 2.0614, fixed by construction)	–	–	2.8076	2.8174	0.35% apart

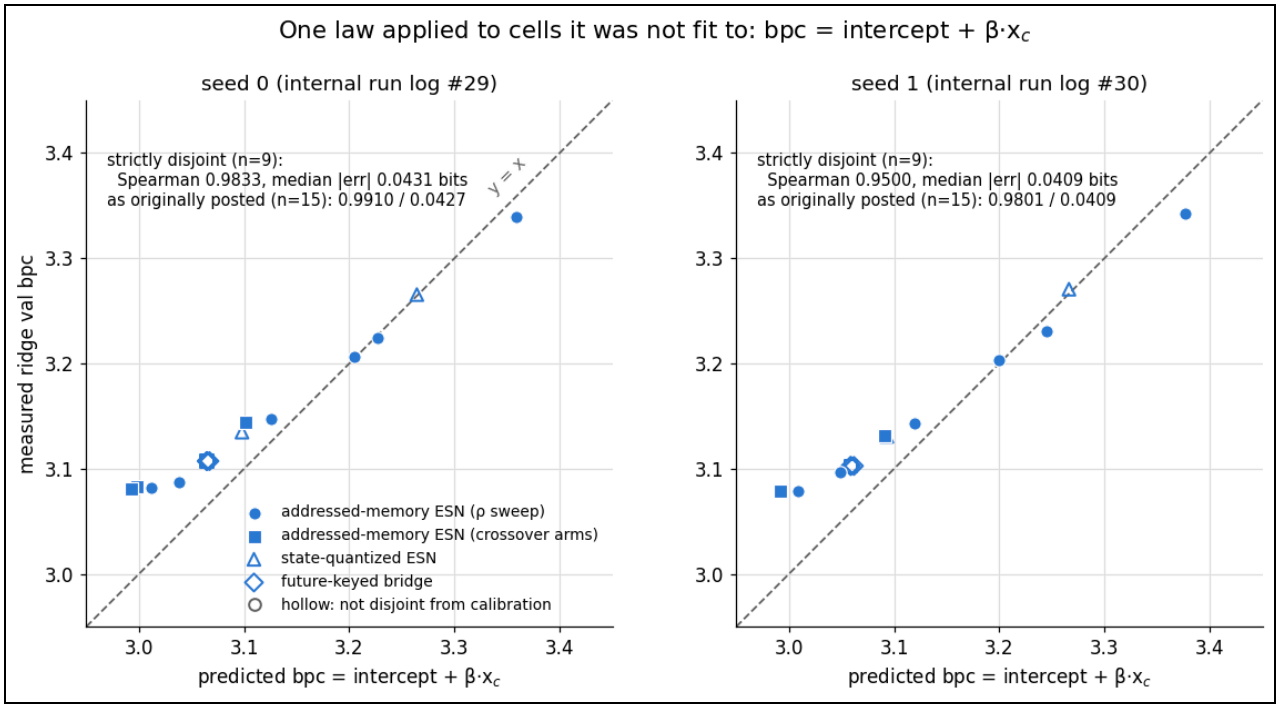

Figure 1. Predicted versus measured ridge-regression validation bits per character, both seeds (internal run log #29, #30). **Filled** markers: the 9 cells strictly disjoint from calibration, which carry the primary statistics annotated on each panel. **Hollow** markers: the 6 cells that are the calibration reservoir under another name, a linear transform of it, or an interior point of the dial the calibration pair spans — the three tier-A/B duplicates coincide exactly and plot on top of one another, which is the visual signature of the purity fault this figure is redrawn to expose. The two calibration cells (state-quantized ESN at b=1 and float32) are not shown. The dashed line is y = x, not a fit. This figure supersedes the one first posted, whose title asserted a cross-family scope on the strength of these two runs alone, and whose legend carried this program's internal names for the reservoir families.

Every pre-registered threshold cleared, on both seeds and on every subset, with no gray zone approached. On the primary strictly-disjoint 9: Spearman 0.9833 / 0.9500 and median error 0.0431 / 0.0409 bits, against registered bars of ≥ 0.8 and ≤ 0.08. O1: 4/4 sign matches at both seeds, but only three of the four registered pairs lie inside the strict 9 — the addressed-memory permutation-versus-sparse pairs at ρ ∈ {0.6, 0.95, 1.1}, all matching at both seeds. At seed 1 the ρ=1.1 pair has Δx_c = +0.0632 against Δbpc = +0.1401. The fourth registered pair, quantized b=2 versus float32, sets a tier-C cell against the calibration cell itself; it matched at both seeds and is reported for completeness, not as strict-9 evidence. Removing the impure cells costs the law almost nothing in accuracy and costs it its *scope*: the 9 surviving cells all come from a single family, so what runs #29 and #30 establish, cleanly and at two seeds, is that one equation prices the gap within that family — calibrated on a reservoir from a second one — and not that it does so across three families.

E2: the same fitted constants generalize to an independently drawn reservoir. Seed 0's frozen constants, applied blind to seed 1's strictly-disjoint 9, score Spearman 0.9500 and median error 0.0444 bits — a cost of 0.0035 bits for using the "wrong" seed's parameters, and identical ranking.

With β = 2.8076 against 2.8174 across independently drawn substrates, the law is value-stable, not merely stable in functional form.

---

§4.1 The cross-family evidence: a separate blind extension

Because the corrected inventory of runs #29/#30 spans one family, the program's cross-family claim rests entirely on a separate, separately registered extension (internal run log #38) in which the *frozen* seed-0 constants — no refitting of any kind — were applied to reservoirs the instrument had never touched. That run evaluated 13 cells and scored 11 of them. Screened by the same purity rule applied above, two are removed: one duplicate pair whose identity *is* the span lemma (a lone half-width reservoir and its exact duplicate-twin concatenation, both 0.4444 / 3.3142), counted once; and one cell that is bit-identical to a cell already scored at run #29. That leaves **9 genuinely new, mutually distinct cells across four further families**, none sharing lineage with the calibration pair.

Table 2. The blind out-of-family extension (internal run log #38, seed 0), after the same purity screen. The law's two constants were fitted at run #29 on two state-quantized cells and were *not* touched here. Every recomputed ridge validation bpc matched its archive to three decimals (13/13 gate passes).

cell	family	x_c	predicted	measured	error
64-block fragmented reservoir	block-diagonal	0.3679	3.0942	3.1321	−0.0379
periodic block reset, 8 blocks	block-reset	0.4691	3.3785	3.3712	+0.0073
random block reset, 8 blocks	block-reset	0.4931	3.4457	3.4257	+0.0200
periodic block reset, 64 blocks	block-reset	0.4306	3.2705	3.2718	−0.0013
random block reset, 64 blocks	block-reset	0.4323	3.2751	3.2760	−0.0009
lone half-width reservoir	paired half-width	0.4444	3.3092	3.3142	−0.0050
independently drawn twin pair	paired half-width	0.3739	3.1113	3.1442	−0.0329
addressed memory, linear units	linear (unsquashed)	0.5598	3.6331	3.6076	+0.0255
sparse mixing, linear units	linear (unsquashed)	0.5543	3.6177	3.6011	+0.0166
9 cells, 4 families, frozen constants: Spearman 1.0, median |error| 0.0166 bits, max 0.038 (computed from unrounded values; the largest rounded row error shown is 0.0379)					**−**

The two linear cells deserve separate mention: unsquashed, zero-bias activations are the largest mechanistic departure from anything in the law's calibration history, and were registered in advance as an extrapolation stress test with a wider allowance (≤ 0.12 bits). They do not miss worst; at 0.0255 and 0.0166 bits they sit mid-pack. Run #38's figures are unchanged by the purity screen — they were already 1.0 and 0.0166 on the unscreened 11 — so this run needs no correction. It is now carrying the program's cross-family claim rather than sharing it, which is a materially weaker evidential position than the original posting described: one run, one seed, no replication of its own.

§4.2 What the law was actually compared against

The pre-registered analytic null is a constant. Its standard deviation across the 15 originally reported cells is 0.000025 bits; it takes two distinct values at four decimals, 2.1590 and 2.1591. On the strictly-disjoint 9 its standard deviation is exactly 0.0, it takes the single value 2.1590, and its rank correlation is not defined at all. The mechanism is visible in its construction: the null credits each character decodable at lags 1–4 with the marginal value of that context extension, and every cell in this inventory decodes lags 1–4 at accuracy exactly 1.0, so the telescoping sum collapses to the 5-gram anchor every time — the pre-registered failure mode, fired completely. A rank correlation computed over an array whose entire spread is 0.0001 bits carries no information. **The claim that the law beat a purpose-built null by an order of magnitude is withdrawn: it was a comparison against a constant predictor.**

Table 3. The law against rivals with real variance, scored blind on the strictly-disjoint 9 with the same two-point calibration budget. Ratios are each rival's median error divided by the law's. S3 is an *oracle* rival: it is permitted to fit its slope on the very cells it is then scored against.

predictor	seed 0 median err (ratio)	seed 1 median err (ratio)	seed 0 Spearman	seed 1 Spearman
the law (x_c)	**0.0431**	**0.0409**	**0.9833**	**0.9500**
pre-registered analytic null (constant)	0.9855 (22.9×)	0.9725 (23.8×)	undefined	undefined
S1 family-mean, leave-one-out	0.0785 (1.82×)	0.0836 (2.04×)	−1.0	−1.0
S2 best blind single feature, by magnitude	0.1094 (2.54×)	0.1185 (2.90×)	undefined	undefined
S2 best blind single feature, by rank (decode depth)	0.3401 (7.89×)	0.3410 (8.34×)	0.9667	0.9667
S3 oracle single feature (slope refit on scored cells)	0.0698 (1.62×)	0.0743 (1.82×)	undefined	undefined

Two things follow, and they point in opposite directions. *In the law's favour:* it beats every rival on magnitude at both seeds, including S3 — a rival allowed to fit its slope on the very cells it is scored against still misses by 1.6–1.8× more than the law does blind. That is the strongest single statement available for x_c, and it is about calibrated magnitude. *Against it:* a one-parameter law on linear decode depth alone — a quantity already sitting in the archived result files, costing nothing extra to compute — ranks the same 9 cells at Spearman 0.9667 at both seeds. That beats the law at seed 1 (0.9500) and loses to it at seed 0 (0.9833). The registered survival rule required the law to beat these rivals on *both* metrics at *both* seeds, so the comparison claim is **weakened, not survived**. The honest reading: x_c's distinctive contribution is calibrated magnitude, and the ordering is largely recoverable from decode depth. One disclosure in the rival's favour, since an adversarial bar should be given every advantage: decode depth is the best of six candidate features chosen after seeing the scores, so 0.9667 is a maximum over six, not a pre-registered rival. That decode depth should rank usability this well is in tension with the program's own decodability-versus-usability story, and is filed as its own open question rather than resolved here.

One secondary finding travels with this result. The 5-gram model's own in-run validation bpc is 2.0614, a stable ~0.098 bits below the "2.159" value documented in the program's standing configuration; the law was fit and scored against the in-run value throughout, and the documented

value should not be quoted without this caveat until reconciled. The discrepancy is real and was independently reproduced at 2.061436 during the triage. The original posting additionally described this value as "confirmed at both seeds"; that is withdrawn, because the quantity is a count over fixed text with no reservoir seed on its path, so agreement across seeds was guaranteed and says nothing. A second previously established low-ρ crossover in this program (permutation addressing outperforming sparse random mixing below ρ* ≈ 0.72) is specific to a *logistic-regression* readout: under ridge regression — this study's usability instrument — permutation addressing underperforms sparse mixing at every tested ρ from 0.45 to 1.1, both seeds, with no crossover, and x_c tracks ridge's monotonic ordering throughout. One caveat on that second finding: at seed 1 the ρ=0.45 ridge gap narrowed to 0.0008 bits — the sign did not flip, but that specific sub-claim now rests on a near-tie and would need a third seed to firm up.

---

§5 Discussion

The claim that survives. A scalar computed from reservoir states alone, requiring no readout training, predicts the bits per character a trained ridge readout will achieve on reservoirs the two fitted constants never saw — to a median of roughly one-twentieth of a bit, at two independently drawn seeds, and with the same constants transferring across substrates to within 0.35%. That is the result, and it is unchanged by the corrections above. What the corrections change is its *reach*. The claim that the program's recurring dissociations — storage without usability, representational geometry without payoff, quantization as a pure bpc dial — are *one* quantity rather than three coincidences is still the reading this study supports, but its out-of-family evidence is now concentrated in a single unreplicated run (§4.1) rather than distributed across the two replicated ones. The difference matters: one run at one seed is exactly the evidential position this program's own rules treat as provisional.

What this does not establish. Uncertainty is reported as a min–max range across two seeds rather than a bootstrap interval; this study's own replication count is 2, though it re-scores cells drawn from source experiments with substantially deeper replication histories, and its effective sample was always smaller than the original posting reported — 15 nominal cells carrying 12 distinct values. On ordering, x_c is not demonstrably better than a decode-depth ruler. Nothing here shows that x_c is the *only* statistic with this property, nor that the linear functional form is privileged; a per-order recalibration found the law's accuracy saturating above 3-gram context rather than peaking at the registered 5-gram choice, which was a registered miss. This research program requires an independent audit of every promoted finding — re-deriving these numbers from the raw result files using independently written code, by a party that did not run the science — and that audit has *not* been run for this result. The validity triage reported in the correction notice is not a substitute: it was carried out inside the program, and it re-scores the same three runs rather than re-measuring anything. The finding is therefore reported as replicated, corrected, and audit-pending. A third seed on a purity-corrected inventory is the named next step, and it is now a more consequential one than it was before this correction.

---

§6 References

1. H. Jaeger, "The 'echo state' approach to analysing and training recurrent neural networks," GMD Report 148, German National Research Center for Information Technology, 2001.

2. M. Lukoševičius, H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer Science Review* 3(3), 2009.

3. M. Mahoney, "Large text compression benchmark" (text8), mattmahoney.net/dc/textdata.

---

§7 Provenance

- **f2b91d28** — 17 cells including the three bridge arms, seed 0 (internal run log #29). Pre-registered before running; O1, O2 and Q1 all confirmed; internally rated at the program's highest first-pass significance tier, which by standing policy forbids publication until a replication passes — the seed-1 rerun was queued ahead of all other pending work.
- **86b40cc1** — seed 1 (internal run log #30). E1 (fresh fit) fully confirmed; E2 (frozen seed-0 law, zero refitting) confirmed; E4 (ridge/logistic crossover split) confirmed, with the ρ=0.45 near-tie flagged. E3 (intercept bit-identical) was recorded as confirmed and is now *struck* as fixed by construction (§3, §4.2). Promoted to the program's confirmed-findings record per the pre-stated promotion rule.
- **36af2eef** — the blind out-of-family extension under frozen constants, seed 0, 13 cells (internal run log #38). Registered before running; F1, F2, FQ1, FQ2 and FQ3 confirmed; F4 (per-order dial) missed, registered non-blocking; F3 (span lemma) confirmed and subsequently struck as an instrument identity rather than a test of the world. Single seed, not independently replicated.
- **796082b3** — read-only re-analysis of runs #29, #30 and #38 against three externally raised objections (internal run log #49). No reservoir was rebuilt and no result file was altered; the triage's code reads only the archived result files and never the analysis block it re-checks. All three objections sustained; the source of every correction on this page.

Audit status: audit pending. This program requires every promoted finding to be independently audited by a party that did not run the science, re-deriving its numbers from the raw result files with freshly written code. That audit has *not* been carried out for this finding. The 2026-08-12 validity triage described above is not that audit and does not stand in for it: it was run inside the program, on the same model family as the science it examines, and the program's own records say so. Replication and audit are different guarantees, and this result currently has the first and not the second.

All claims are judged strictly against the pre-registered wording; disclosed scope reductions are noted above. 2 seeds at runs #29/#30, 1 seed at run #38; replicated before publication; corrected in place on 2026-08-12, with the original wording's failures stated rather than quietly removed. Internal designation: assay office.
