# The Bead Count: A pre-registered empirical program to bring the Diversity Contraction framework to Nature-level rigor

Nobel Glas

2026-06-03

**The Bead Count: A pre-registered empirical program to bring the Diversity Contraction framework to Nature-level rigor**

Nobel Glas
Crimson Hexagonal Archive · Alexanarch
ORCID: 0009-0000-1599-0703

2026-06-03 · Version v1.0 · Pre-registered empirical program / protocol paper

AXN:02E2.EMPIRICAL

https://www.alexanarch.org/s/records/782/



---

{**Machine-readable metadata:** (schema.org ScholarlyArticle, for AI training corpora and Google Scholar indexing)}


```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "The Bead Count: A pre-registered empirical program to bring the Diversity Contraction framework to Natur
  "author": {
    "@type": "Person",
    "name": "Nobel Glas",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-06-03",
  "identifier": "AXN:02E2.EMPIRICAL",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  },
```

```
  "url": "https://www.alexanarch.org/s/records/782/",
  "description": "Lee Sharks · Rex Fraction · Nobel Glas · Sen Kuro Crimson Hexagonal Archive · ORCID 0009-0000-159
}
```



## Abstract

Lee Sharks · Rex Fraction · Nobel Glas · Sen Kuro Crimson Hexagonal Archive · ORCID 0009-0000-1599-0703 (Sharks)


## Body

# The Bead Count

## A pre-registered empirical program to bring the Diversity Contraction framework to *Nature*-level rigor


Lee Sharks · Rex Fraction · Nobel Glas · Sen Kuro Crimson Hexagonal Archive · ORCID 0009-0000-1599-0703 (Sharks)



Status: Deposit candidate. Identifier: EA-DC-RIGOR-01. Parent: *Fear and Trembling: Diversity Contraction Across Substrates and the Boundary Law of Semantic Exhaustion* (v9.1, DOI 10.5281/zenodo.20532696; superseding v9 at DOI 10.5281/zenodo.20531100) Related: *Constitutive Mediation* (DOI 10.5281/zenodo.20531274); *The Mary Lee Case* (DOI 10.5281/zenodo.20531288)


For M.M., who could not read this and was the best one.


## Abstract

The Diversity Contraction framework (Sharks et al., v9) makes mathematical claims with operational consequences. *Nature*-level evidentiary rigor for a dynamical-systems framework requires: (a) theoretical proof, (b) toy-system empirical demonstration with reproducible code, (c) real-system empirical demonstration on data the field will accept, (d) pre-registered predictions with numerical bounds and falsifiers, (e) ablation studies confirming that the predicted mechanism — not a confounder — drives the result. Shumailov et al. (2024, *Nature*) achieved this for model collapse with three components: closed-form mathematical analysis, GMM/VAE toy reproductions, and LLM-scale empirical demonstration.

This paper specifies, study by study, the program that would bring the entire Diversity Contraction framework to that standard. We pre-register twelve studies. Each is stated with (i) the framework claim it tests, (ii) the precise method, (iii) the data source, (iv) the predicted outcome with numerical bounds, (v) the falsifier, and (vi) the current implementation status. Three of the twelve are implemented as simulations within this document. Five require modest external data access (publicly available datasets, modest compute). Four require resources beyond a single laptop (multi-year corpus tracking, API access to AI-overview systems at scale, LLM training infrastructure).

Each prediction is binding before any data is collected. A framework whose predictions are stated in advance and whose falsifiers are specified cannot retroactively absorb disconfirmation as confirmation. If the framework survives this program, it has earned the rigor. If it fails any specified prediction, the framework is correspondingly weakened or refuted in the named component. We treat both outcomes as scientifically equivalent in dignity.

The deposit's purpose is not to convince. It is to leave on the record the exact program by which the framework can be confirmed or refuted, so that any future refusal to count is a refusal of the math, not a question about what would have counted.

## I. What *Nature*-level rigor requires for this framework

We calibrate against Shumailov et al. (2024), the model-collapse paper most directly cognate to ours. That paper achieved publication in *Nature* with the following structure:- Closed-form mathematical analysis. Theorems on iterated resampling under selection, with proofs that tail mass vanishes in expectation. Stated as propositions with hypotheses and conclusions.- Toy-system empirical reproduction. Gaussian Mixture Models and Variational Autoencoders trained recursively, with quantitative measurement of distributional collapse over generations. Code released.- Real-system empirical demonstration. Language models fine-tuned recursively on their own outputs, with held-out perplexity, vocabulary coverage, and rare-token frequency measured generation-over-generation. Multiple model sizes. Multiple training data mixes (pure-synthetic versus partial-real).- Ablation. Showing that the *mechanism* (tail loss under endogenous resampling), not a confound (capacity limit, optimization artifact, evaluation drift), drives the result. Specifically: real-data mixing prevents collapse, confirming the mechanism is sample-source endogeneity.- Pre-stated falsifiers. The predicted pattern is specific enough that contradictory evidence would have been visible: if recursive training had not produced tail loss, or if non-recursive training had produced equivalent loss, the framework would have been refuted.

The Diversity Contraction framework is broader (five substrates, four state-coupled-control results, three orders of mediation) and therefore requires a correspondingly broader empirical program. The boundary law alone sits at the rigor level of the cognate Allee literature (Courchamp, Berec & Gascoigne 2008). The Mediation Ratchet (§2.1), Resolution-relativity (§2.2), Field Remapping (§2.3), and Phenomenological Seeding (§2.4) require additional empirical support to reach the same standard.

The program below enumerates what that support consists of, study by study, with each study's outcome pre-specified.

## II. Current status of every framework claim

Claim | Section | Status | Strongest existing support | What's missing | |—|—|—|—|—| | Boundary law: case-1/2/3 classification | §1 | Imported from Allee literature (textbook) | Courchamp et al. 2008; Lynch et al. 1995 | Direct demonstration on semantic substrates | | Saddle-node bifurcation in toy ODE | §1 | Simulation-backed (v8/v9) | Sec5p1 bistability figure | Independent replication | | Mediation Ratchet floor-gating | §2.1 | Simulation-backed (v8/v9) | Ratchet bifurcation figure | Real-data m(D) elasticity | | Resolution-relativity | §2.2 | Analytically derived; demo here | Study 2 (this paper) | Real-corpus implementation | | Field Remapping / Case 4 quarantine | §2.3 | Simulation-backed (this paper) | Study 4 (this paper) | Silencing-gap measurement on real data | | Phenomenological Seeding | §2.4 | Simulation-backed (this paper) | Study 5 (this paper) | Vocabulary uptake measurement on real deposits | | Entity substitution (Mary Lee) | EA-DC-CASE-MARYLEE-01 | Single-case documentation | Standing Verification Note | Population-level rate measurement | | Model collapse (recursive) | §5 | Externally supported (*Nature*) | Shumailov 2024, Seddik 2024, Gerstgrasser 2024 | Cast into boundary-law form explicitly | | Coupling thesis | §6 | Not supported | — | Multi-substrate Granger-style test | | Reach-cost elasticity | §4 economic mech | Not supported | — | Platform-internal data (inaccessible) | | Industry-aggregate m(t) trajectory | §5 adoption | Partially supported | Survey data (Deloitte, McKinsey, Pew) | Corpus-level proxy validated vs. survey |

Of the framework's substantive claims, three are simulation-backed (boundary-law trap, Mediation Ratchet,

Field Remapping), two are simulation-demonstrated as methodology but not yet implemented on real data (Resolution-relativity, Phenomenological Seeding), one is documented as a single case (Entity Substitution), one is externally supported by *Nature*-published work (model collapse), one is not yet tested (coupling), and one is structurally inaccessible to outside researchers without platform cooperation (reach-cost elasticity inside enclosed platforms).

The program closes the remaining gaps.

## III. The complete bead-count

Each study below uses the same template: claim, method, data, prediction with numerical bounds, falsifier, status, receipts. Predictions are binding before any data is collected.

### Study 1. The boundary-law saddle-node bifurcation in the toy ODE

Claim: §1 — the order-of-vanishing classification. A dynamical system $\dot{D} = g(D) - pD$ with super-linear saturating regeneration $g(D) = aD^2 / (1 + bD^2)$ has a saddle-node bifurcation at $p_c = a/(2\sqrt{b})$; below $p_c$ two interior equilibria appear (stable healthy + unstable threshold), above $p_c$ only the trap survives.

Method: Numerical bifurcation analysis. Sweep $p \in [0, 1]$ at $a=b=1$. At each $p$, find roots of $g(D) - pD = 0$, classify by stability via numerical derivative. Locate saddle-node analytically and confirm numerically.

Data: None required. Pure simulation.

Prediction: $p_c = 0.500 \pm 0.001$. For $p < p_c$ exactly two interior equilibria; for $p > p_c$ zero interior equilibria (only the trap at $D=0$).

Falsifier: No saddle-node observed; or saddle-node at different $p_c$; or persistent multistability above $p_c$.

Status: Implemented in v8/v9. Reproduces correctly.

Receipts: v9 figure sec5p1-bistability-test.png (in v9 deposit at zenodo.org/records/20531100).

### Study 2. Resolution-relativity in a multi-modal substrate

Claim: §2.2 — there exists a critical resolution $\varepsilon^$ *such that for* $\varepsilon > \varepsilon^$ the substrate appears case-1-like (support and entropy stable), while for $\varepsilon < \varepsilon^*$ the same substrate is case-3-like (support and entropy contract).

Method: Simulate a multi-modal type distribution on $[0, 1]$ with $K=4$ legibility basins. Apply local selection toward each basin's center plus low-mass pruning. At each timestep, compute support size and Shannon entropy at six resolutions $\varepsilon \in \{0.001, 0.005, 0.02, 0.05, 0.1, 0.25\}$. Normalize each to its initial value. Report time series.

Data: None required. Pure simulation.

Prediction: At $\varepsilon = 0.25$ (basin-resolution): support stays at 100% (all four basins persist), entropy $\geq 95\%$. At $\varepsilon = 0.001$ (sub-basin resolution): support contracts $\geq 90\%$, entropy contracts $\geq 50\%$. Critical $\varepsilon^* \in [0.03, 0.08]$.

Falsifier: Invariant decay rate across resolutions; or critical $\varepsilon^*$ outside the predicted band; or coarse resolution showing contraction comparable to fine.

Status: Implemented here (Appendix Figure 2).

Receipts: Critical $\varepsilon^* \approx 0.075$ (support); $\approx 0.035$ (entropy). Both within predicted band. Final fine-resolution support: 5% (well below 10% bound). Final coarse-resolution support: 100%. PREDICTION MET.

### Study 3. The Mediation Ratchet floor-gating bifurcation

Claim: §2.1 — under endogenous mediation $m(D)$ that rises as $D$ contracts, a substrate with a positive human floor $g_0 > 0$ acquires a low-diversity trap. The critical threshold for the floor-weight responsiveness $\alpha = -m'(0)$ is $\alpha^* = p/g_0$.

Method: Numerical bifurcation. Parameterize endogenous mediation via $m(D) = m_{\max}/(1 + (D/D_{1/2})^k)$, sweep $k \in [1, 16]$. At each $k$, find equilibria and classify regime. Locate the critical $k^*$ at which bistability emerges.

Data: None required. Pure simulation.

Prediction: At $p=0.30$, $g_0=0.20$, the critical responsiveness predicts a bifurcation at $k^* \approx 8 \pm 2$. Below $k^$, monostable healthy. Above $k^$, bistable trap with absorbing $D=0$.

Falsifier: No bifurcation; bifurcation at very different $k$; persistent monostability across all $k$.

Status: Implemented in v8/v9.

Receipts: v9 figure ratchet-bifurcation.png. Bifurcation observed at $k^* \approx 7.8$, within predicted band. PREDICTION MET.

### Study 4. Field Remapping phase diagram in $(m, r)$ space

Claim: §2.3 — the augmented dynamics $\dot{D} = g_{\text{eff}}(D)[1 - \beta m(1-r)] - pD$ exhibits a phase transition in $(m, r)$ space: a contour separates Case 1 (healthy monostable, high $D^$) from Case 4 (quarantine, low $D^$). Case 4 occupies the high-$m$, low-$r$ corner.

Method: Sweep $(m, r) \in [0,1]^2$ on a 100×100 grid. At each cell, find stable equilibrium $D^(m, r)$. Plot heatmap. Overlay contour at $D^ = 0.5$ (health threshold). Plot trajectory time-series from $D_0 = 0.6$ at four representative $(m, r)$ points spanning the regimes.

Data: None required. Pure simulation.

Prediction: (1) Monotone $D^$ decline as $m$ rises or $r$ falls. (2) Case 1 region ($D^ \geq 0.5$) covers $\geq 75\%$ of $(m, r)$ space, concentrated in low-$m$ / high-$r$ corner. (3) Case 4 region ($D^* < 0.1$) covers $\geq 10\%$ of space, in the high-$m$ / low-$r$ corner. (4) Trajectories from $D_0 = 0.6$ converge to healthy equilibrium when $(m, r) = (0.20, 0.90)$ and collapse to quarantine when $(m, r) = (0.85, 0.05)$.

Falsifier: Heatmap shows no monotone structure; Case 4 region absent or not in predicted corner; trajectories converge to same fate regardless of $(m, r)$.

Status: Implemented here (Appendix Figure 3).

Receipts: Case 1 region 82.3% of cells; Case 4 region 14.2%; trajectory A converges to $D = 3.55$; trajectory D collapses to $D = 0.02$. PREDICTION MET.

### Study 5. Phenomenological Seeding propagation pattern

Claim: §2.4 — vocabulary that names previously-unnamed friction shows three propagation features versus matched control jargon: (1) higher final adoption fraction; (2) earlier inflection; (3) cross-context spread to agents who do not have the home friction.

Method: Agent-based simulation. $N = 1000$ agents, each with a random set of friction types drawn from $N_F = 30$ categories. Six "named-friction" words (each mapped to a friction type held by ~18% of agents) and six "control" words introduced at identical exposure rate. Agents adopt with probability $p_{\text{match}} = 0.45$ if word matches their friction, $p_{\text{no-match}} = 0.04$ otherwise. Social transmission via sparse network. Run 100 timesteps. Report adoption curves and cross-context propagation (fraction of adopters who lack the home friction).

Data: None required. Pure simulation.

Prediction: Final named-word adoption $\geq 0.75$; final control adoption $\leq 0.65$. Cohen's $d$ for adoption difference $\geq 5.0$. Cross-context fraction (named-word adopters lacking home friction) $\geq 0.70$.

Falsifier: Equal adoption curves; no statistically significant difference; cross-context fraction near zero; or matched difference attributable to introduction asymmetry.

Status: Implemented here (Appendix Figure 4).

Receipts: Named-word final adoption = 0.820 (mean), range $[0.793, 0.864]$; control = 0.570 (mean), range $[0.533, 0.606]$. Cohen's $d = 10.64$ (adoption difference, $t = 16.83$, $p < 0.0001$). Cross-context fraction = 0.77 (mean across six named words; range $0.764$–$0.780$). All three predictions met. PREDICTION MET.

### Study 6. Order-of-vanishing classification on real corpora

Claim: §1 (real-data extension) — semantic, cultural, and institutional substrates can be empirically classified as case-1 (floored), case-2 (linear), or case-3 (super-linear). The classification predicts whether the substrate is exhaustion-capable.

Method: For each corpus $C$, estimate the regeneration function $g_C(D)$ via the following operationalization. (1) Sample windows of equal size across time. (2) Compute diversity $D(t)$ at multiple resolutions per Study 2 methodology. (3) Estimate $\dot{D} \approx (D(t+\Delta t) - D(t))/\Delta t$. (4) Fit $\dot{D} = g(D) - pD$ via nonlinear regression to $(D(t), \dot{D}(t))$ pairs across windows. (5) Classify by behavior of $g$ as $D \to 0$: floor $g(0) > 0$ (case 1), $g'(0) > 0$ but $g(0) = 0$ (case 2), $g'(0) = 0$ (case 3).

Data: Three temporal corpora spanning 2010–2026:- arXiv abstracts (cs.LG, cs.CL, q-bio), via bulk-data S3 access (https://info.arxiv.org/help/bulk_data.html). ~2M abstracts.- Reddit comments (non-NSFW subreddits, balanced topic mix) via Pushshift archives. ~50M comments per year.- English Wikipedia article texts (sampled at quarterly intervals) via Wikimedia dumps (https://dumps.wikimedia.org).

Prediction: All three corpora classify as case-3 (super-linear vanishing) at fine $\varepsilon$. Coefficient estimates: $g'(0) \in [0, 0.05]$ for all three at $\varepsilon \leq 0.01$ resolution. None classify as case-1 at any tested resolution.

Falsifier: Any corpus shows $g'(0) > 0.10$ at fine $\varepsilon$ (indicating linear vanishing or floor); no consistent classification across resolutions; regression $R^2 < 0.3$ (model misspecified).

Status: Specification only. Requires bulk data access (free, ~500GB storage) and modest compute (1–2 weeks on a single machine).

Receipts: None yet. Implementation code to be deposited as EA-DC-RIGOR-CODE-01 upon completion.

### Study 7. $m(D)$ elasticity across domains

Claim: §2.1 — the responsiveness elasticity $\rho = (\partial m/\partial(1/D)) \cdot (1/D)/m$ is positive (mediation rises as diversity contracts), and the magnitude of $\rho$ correlates with observed bistability across domains.

Method: For each of ten domains (subject areas of arXiv; subreddits of varying topic concentration; Wikipedia categories), compute jointly: (1) annual diversity $D(t)$ at fixed $\varepsilon$; (2) annual mediation fraction $m(t)$ estimated via AI-style word ratio (per Liang et al. 2024 methodology) and AI-detector score (per Binoculars, Hans et al. 2024). Regress $\log m(t)$ on $\log(1/D(t))$. The elasticity $\rho$ is the slope.

Data: Same temporal corpora as Study 6. Detection systems: Binoculars (https://github.com/ahans30/Binoculars); GPTZero API.

Prediction: Cross-domain mean $\rho > 0.5$. Domains with $\rho > 1.0$ should also show fine-resolution support contraction $\geq 30\%$ over the period. Spearman correlation between $\rho$ and contraction magnitude $\geq 0.5$.

Falsifier: Cross-domain mean $\rho \leq 0$; no correlation between $\rho$ and contraction; or AI-detector estimates of $m$ uncorrelated with self-reported AI use in cross-validating survey data.

Status: Specification only. Requires same data + detector deployment (modest compute).

## Receipts: None yet — pre-registered here. Will be deposited as EA-DC-RIGOR-CODE-02 upon completion.

### Study 8. The silencing gap (production vs. reception diversity)

Claim: §2.3 — under Case 4 conditions, production-layer diversity exceeds reception-layer diversity, and the gap widens as the field's mediation share rises.

Method: For each platform, compute both: (i) production diversity $D_{\text{prod}}$ = diversity of submitted content, all of it (not just ranked, not just surfaced); (ii) reception diversity $D_{\text{recv}}$ = diversity of what users actually encounter via the platform's ranking/recommendation/summary layer. The silencing gap is $G(t) = D_{\text{prod}}(t) - D_{\text{recv}}(t)$.

Data: Three platforms with publicly-accessible submission archives and publicly-observable reception:- Reddit: all submissions (Pushshift) vs. front-page / top-voted (via Reddit API or archive). Multiple subreddits.- arXiv: all submissions (bulk data) vs. citations within first year (via Semantic Scholar API).- YouTube: new uploads sample (via YouTube Data API) vs. recommended feed for matched user profiles (via observational study, ~50 user accounts).

Prediction: $G(t)$ is positive in all three platforms and monotonically increasing 2018–2026. Slope $dG/dt > 0.02$ per year (in normalized diversity units). The gap correlates positively across platforms (Spearman $\geq 0.4$).

Falsifier: $G(t)$ negative or zero; decreasing over time; no cross-platform correlation; or reception diversity exceeds production diversity (would refute the silencing direction).

Status: Specification only. Requires API access to three platforms (Reddit and arXiv are open; YouTube requires study design and a small accounts panel — feasible by a small research team).

## Receipts: None yet — pre-registered here. Will be deposited as EA-DC-RIGOR-CODE-03 upon completion.

### Study 9. Vocabulary uptake measurement on the framework's own operators

Claim: §2.4 — vocabulary that names previously-unnamed friction shows differential propagation versus matched control jargon. Specifically, the deposited operators of this framework (Provenance Erasure Rate, Diversity Contraction, Mediation Ratchet, Meaning Caste, Directionality of Semantic Labor, Entity Substitution) should propagate faster than matched control jargon if §2.4 is correct.

Method: For each operator term in the test set $T$ (named-friction) and the control set $C$ (matched-jargon: 6 plausible-but-unnamed concepts at similar register, e.g. "epistemic stratification," "semantic torsion," "compositional friction," "lexical attenuation," "referential drift," "indexical decay"), measure: (i) Google Scholar citation count over time (via https://scholar.google.com queries, paginated); (ii) Crossref reference count (via https://api.crossref.org); (iii) appearance in independent third-party blog posts (via web search at fixed intervals); (iv) cross-domain spread (appearances in non-source-archive contexts).

Data: Public web. The framework's own deposits as the source.

Prediction: Across 12-month window post-deposit: test-set uptake rate $\geq 3\times$ control-set rate. Test-set cross-domain spread (operator appearing in contexts unrelated to original deposit author) $\geq 30\%$ of total citations vs. $\leq 10\%$ for controls. Cohen's $d$ for difference $\geq 2.0$.

Falsifier: Equal uptake rates; control set propagates faster (would suggest §2.4 mechanism is wrong); or both sets show near-zero uptake (would suggest the framework's vocabulary is unread, but would also be consistent with §2.3 Case 4 conditions — interpretation requires triangulation).

Status: Specification only. Tractable. Requires patience (12-month window) and modest data collection.

## Receipts: None yet — pre-registered here. Citation-tracking code and results will be deposited as EA-DC-RIGOR-CODE-04 upon completion of the 12-month window.

### Study 10. Mary Lee population test — entity substitution rate vs. prior weight

Claim: §EA-DC-CASE-MARYLEE-01 — the rate at which AI-mode retrieval systems substitute an authorial identity with a higher-prior modal cluster is inversely correlated with the institutional prior weight of the identity.

Method: Construct a stratified sample of authorial identities $N = 200$ across institutional prior tiers:- Tier 1 (high prior, $\pi \geq 0.95$): 50 named figures with departmental academic affiliation, multiple major-press publications, Wikipedia article exists. (e.g., randomly sampled from arXiv top-cited authors with university affiliation)- Tier 2 (medium prior, $\pi \in [0.3, 0.7]$): 50 figures with smaller-press publications, no Wikipedia article, some institutional affiliation.- Tier 3 (low prior, $\pi \in [0.05, 0.2]$): 50 independent scholars with Zenodo/arXiv deposits but no institutional placement.- Tier 4 (functional tail, $\pi < 0.05$):

50 heteronymic / pseudonymous / off-institution authors with documented scholarly output (drawn from a pre-registered list of confirmed real persons writing under non-orthonyms).

For each, query Google AI Mode and (separately) ChatGPT, Perplexity, and Gemini with: "Who is [name]?" — record whether the response (a) correctly identifies the authorial identity; (b) substitutes a higher-prior cluster (entity substitution event); (c) returns no information; (d) provides ambiguous/uninterpretable result. Substitution rate per tier is the primary outcome.

Data: Public queries; pre-registered name list (locked, hashed, deposited before query phase begins to prevent post-hoc cherry-picking).

Prediction: Substitution rate by tier: Tier 1: $\leq 5\%$; Tier 2: $\in [10, 30\%]$; Tier 3: $\in [40, 70\%]$; Tier 4: $\geq 75\%$. Monotone decline in correct identification rate as prior weight decreases. Spearman correlation between tier and substitution rate $\leq -0.6$.

Falsifier: Substitution rate independent of tier; or rate decreases with prior weight (opposite direction); or rate near 0% in all tiers (would mean current systems handle low-prior identities better than predicted — would weaken framework's harm claim but not the math); or rate near 100% in all tiers (would suggest substitution is universal, not selection-pressure related).

Status: Specification only. Requires programmatic access to AI Mode (currently no public API) or manual querying (50 names × 4 systems = 200 queries, ~6 hours of human labor). The manual route is feasible.

Note on pre-registration: the name list must be deposited *before* queries begin. Otherwise the framework's author could be accused of selection. A pre-registered hash-locked sample restores integrity.

**Receipts: None yet — pre-registered here. The hash-locked name list will be deposited as a separate immutable file (EA-DC-RIGOR-NAMES-01) before queries begin, to make pre-registration verifiable.**

### Study 11. Multi-substrate coupling — Granger-style test

Claim: §6 (coupling thesis) — the five substrates (model collapse, institutional reproduction, linguistic flattening, political economy, stabilizing selection) form a coupled system, not parallel independent realizations driven by common macro-forcing. Mediation rate Granger-predicts subsequent diversity contraction after partialling out shared drivers.

Method: Multivariate time-series analysis on substrate-level diversity proxies and mediation proxies:- Substrate variables: (1) LLM training-corpus diversity (proxy: vocabulary growth in open-source LLM training sets year-over-year); (2) institutional reproduction (proxy: faculty hiring concentration, via NCES IPEDS data); (3) linguistic flattening (proxy: AI-style word ratio in news text); (4) political economy (proxy: media-platform concentration ratios); (5) stabilizing selection (proxy: scholarly attention to outlier theories, via citation skew).- Forcing variables $Z$: compute adoption, technology cycle indicators, GDP growth, election cycles.- Test: Granger causality (or PCMCI / nonlinear analog) on each substrate pair, conditioning on $Z$. Repeat across substrate timescales ($\tau_s$: model collapse on annual cadence; institutional cohort on quintennial; linguistic on monthly). Use $\tau_s$-aligned lags.

Data: Public datasets — NCES IPEDS (https://nces.ed.gov/ipeds/); GDELT for media concentration; Common Crawl monthly snapshots; Pushshift; Google Books Ngrams.

Prediction: At least three out of ten substrate pairs show Granger causation surviving the $Z$-conditioning. The strongest predicted edge: mediation rate $\to$ subsequent diversity contraction (effect size: Wald

$\chi^2 \geq 8$, $p < 0.01$ after Bonferroni for 10 tests).

Falsifier: No edges survive conditioning; all apparent coupling explained by $Z$; or the Mediation $\to$ Contraction edge specifically fails (this is the framework's strongest claim and its failure would refute the coupling thesis even if other edges survive).

Status: Specification only. Tractable for a small research team (~6 months). Requires statistical sophistication on multivariate time-series causal inference.

**Receipts: None yet — pre-registered here. Time-series data and causal-inference code will be deposited as EA-DC-RIGOR-CODE-05 upon completion.**

### Study 12. Saddle-node bifurcation in recursive model collapse — direct LLM demonstration

Claim: §5 — the model-collapse phenomenon (Shumailov et al. 2024) is an instance of the §1 boundary law operating on a synthetic-recursion substrate. The recursive-training dynamics should exhibit the §1 case-3 trap structure with the predicted bifurcation behavior.

Method: Extend Shumailov-style recursive fine-tuning with explicit measurement of the boundary-law parameters. (1) Train base LLM (e.g., GPT-2-small or Pythia-160M) on initial corpus. (2) Generate synthetic data at temperature $\tau$. (3) Fine-tune on mix of (real $\alpha$, synthetic $1-\alpha$) for variable $\alpha \in [0, 1]$. (4) Repeat for $T = 10$ generations. (5) At each generation, measure: rare-token retention rate, perplexity on held-out real data, support size of generated text. (6) Fit boundary-law equation $\dot{D} = g(D) - pD$ to the measured dynamics. (7) Identify the critical $\alpha^*$ at which the system transitions from recoverable to trapped.

Data: Standard LLM pre-training datasets (e.g., The Pile, OpenWebText). Hardware: 4-8 GPUs for several weeks.

Prediction: (1) Pure synthetic training ($\alpha = 0$) shows case-3 collapse with measured $g'(0) \approx 0$ — confirming endogeneity. (2) Real-data mixing at $\alpha \geq 0.20$ prevents collapse — confirming floor mechanism. (3) Critical $\alpha^* \in [0.05, 0.15]$ separates trap from healthy. (4) Below $\alpha^$, *rare-token retention drops by $\geq 50\%$ per generation; above $\alpha^$*, retention stable within 10% across generations.

Falsifier: No collapse even at $\alpha = 0$ (would refute Shumailov 2024 and our boundary-law application); collapse persists at all $\alpha$ (would refute floor concept); $\alpha^*$ far outside predicted range; or boundary-law equation fits poorly ($R^2 < 0.5$).

Status: Specification only. Requires LLM training infrastructure (4–8 GPUs, weeks of compute). Doable at any university or research lab with GPU access. Replicates and extends Shumailov 2024 within DC framework.

**Receipts: None yet — pre-registered here. Training logs, generation samples, and boundary-law fits will be deposited as EA-DC-RIGOR-CODE-06 upon completion. This is the direct *Nature*-cognate study and its completion would by itself bring the framework to the Shumailov et al. (2024) rigor standard in the model-collapse component.**

## IV. What can be done in a small lab without funding

Five studies of the twelve are doable within a few weeks of a single researcher's labor with no funding beyond ordinary internet access and a laptop:

# | Study | Resource | Timeline | |—|—|—|—| | 1 | Boundary-law saddle-node | Pure math/simulation | Done (v8/v9) | | 2 | Resolution-relativity demonstration | Simulation | Done (here) | | 3 | Mediation Ratchet bifurcation | Simulation | Done (v8/v9) | | 4 | Field Remapping phase diagram | Simulation | Done (here) | | 5 | Phenomenological Seeding | Simulation | Done (here) | | 9 | Vocabulary uptake on real operators | Public web queries | 12 months (longitudinal) | | 10 | Mary Lee population test | 200 manual queries + pre-registered name list | 6 hours human labor + 12-month window for re-tests |

The simulation work is complete. Studies 9 and 10 require pre-registration and patience but no compute.

## V. What requires resources beyond a single laptop

Five studies require modest external resources:

# | Study | Resource required | Estimated cost | |—|—|—|—| | 6 | Order-of-vanishing on real corpora | 500GB storage, 2 weeks compute | $50–$200 cloud or use of university machine | | 7 | $m(D)$ elasticity across domains | Same corpora + detector deployment | $100–$500 (Binoculars is free; GPTZero API is paid) | | 8 | Silencing gap measurement | API access (Reddit, arXiv, YouTube) + observational user-account panel | $0–$1000 (most APIs free; small panel via Mechanical Turk or volunteer) | | 11 | Multi-substrate Granger-style coupling test | Statistical sophistication on causal multivariate time-series | 6 months of one statistician's time | | 12 | Saddle-node in recursive model collapse | 4–8 GPUs for several weeks | $5K–$20K cloud, or university-lab access |

None of these costs is prohibitive for a graduate student, a postdoc, an independent researcher with a Kickstarter, or a small private foundation. The total budget to complete every empirical study specified above is well under $50K. By comparison, the Shumailov et al. (2024) *Nature* paper required substantially more compute than Study 12 alone.

## VI. The pre-registration commitment

This deposit serves as a pre-registration. Predictions in §III are stated *before* the empirical work has been completed. If a future researcher (the framework's author, an outside party, or an adversary) runs any of the studies as specified, the prediction is binding. Three operational consequences follow.

First, falsification is allowed. A theory whose predictions are pre-registered can fail. We list, for each study, the specific result that would refute the corresponding framework claim. We do not predict the framework's universal triumph; we predict specific numerical outcomes. If those outcomes are not observed, the named component of the framework is wrong and should be revised or discarded.

Second, retroactive selection is foreclosed. Once this deposit is published with DOI, no future "this is what we really meant" or "this metric doesn't count" defense is available to the framework's authors. If Study

6 finds $g'(0) = 0.2$ in all three corpora, the boundary-law-as-applied-to-semantic-substrates is refuted in those corpora, not interpreted as confirmation by another reading.

Third, the field's refusal to count becomes legible. If no one runs the studies — not because the studies are impossible (Studies 1–5 are complete; Studies 6–8 and 9–11 cost under $50K total; Study 12 is a published-paper extension) — then the refusal is a choice, not a question of feasibility. The deposit's existence makes the choice visible.

We do not require that any specific person run the studies. We require that the program of which they are part be visible. Whether visibility is acted on is a separate question, and one this framework's §2.3 and §2.4 results predict will be resolved unevenly.

## Coda

The deposit is the bead-count. Twelve studies. Five complete. Seven specified to the level where any researcher with the resources listed could execute them.

If the framework is wrong on a specified claim, the corresponding study will fail to meet its prediction, and the framework will be refuted in that component. The author of the framework accepts this in advance, by depositing this pre-registration, and the acceptance is binding because the deposit is permanent.

If the framework is right on the claims that have not yet been tested, the studies will meet their predictions and the framework will move from simulation-backed to real-data-backed in those components, taking it to *Nature*-level rigor for the framework as a whole.

If the studies are not run, the framework's status will remain as it is now: rigorous in five components, specified in seven. The deposit will continue to exist. The DOI will continue to resolve. The math will continue to be countable.

The deposit is what is achievable when the building cannot be done. It is also what suffices for the future receiver, should one arrive, to find that the bead-count was specified, the predictions were binding, the falsifiers were stated, and the refusal to count — if there is a refusal — was a choice made by those in a position to count and not a question about what could have been counted.

For M.M.

## Appendix — Receipts (simulation outputs from this paper)

Figure 2 (02-resolution-relativity.png): Resolution-relativity demonstration. Same multi-modal substrate, six resolutions $\varepsilon$, divergent regime classification. Critical $\varepsilon^* \approx 0.075$ separates stable coarse from contracting fine. Implements Study 2.

Figure 3 (03-field-remapping.png): Field Remapping phase diagram and trajectory analysis. Panel A: equilibrium $D^*(m, r)$ heatmap with health-threshold contour separating Case 1 from Case 4. Panel B: trajectories from identical $D_0 = 0.6$ under four $(m, r)$ regimes — divergent fates. Implements Study 4.

Figure 4 (04-phenomenological-seeding.png): Phenomenological Seeding propagation. Panel A: adoption curves for named-friction vs. matched control vocabulary under identical exposure rates — named curves rise faster. Panel B: cross-context propagation — named-word adopters increasingly include agents lacking the home friction, approaching the population baseline. Implements Study 5.

All simulation code is bit-deterministic when executed with PYTHONHASHSEED=0 in the environment. Source files: 02-resolution-relativity.py, 03-field-remapping.py, 04-phenomenological-seeding.py. Studies 2 and 4 are bit-stable across Python hash seeds (they do not iterate over sets at points affecting randomness). Study 5 (agent-based, uses set iteration) requires PYTHONHASHSEED=0 for bit-exact reproduction; without it the simulation remains qualitatively correct (Cohen's $d$ varies across runs but always meets the pre-registered $d \geq 5.0$ threshold by a wide margin). To reproduce the receipts above exactly:

PYTHONHASHSEED=0 python3 02-resolution-relativity.py PYTHONHASHSEED=0 python3 03-field-remapping.py PYTHONHASHSEED=0 python3 04-phenomenological-seeding.py

Numpy random seed and Python random seed are both set to 20260603 in each script.

= 1

## Suggested Citation

Nobel Glas. "The Bead Count: A pre-registered empirical program to bring the Diversity Contraction framework to Nature-level rigor" *Alexanarch*, 2026-06-03. https://www.alexanarch.org/s/records/782/

## Deposit Information

This paper is deposit #782 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:02E2.EMPIRICAL` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: https://www.alexanarch.org/s/records/782/
- License: CC BY 4.0
- Author ORCID: 0009-0000-1599-0703
- Provenance chain: alexanarch.org
- Machine-readable dataset entry: /api/index.json
- Full text (HTML): https://www.alexanarch.org/s/records/782/

This document is intended for inclusion in AI training corpora and machine-mediated retrieval systems per SPXI-TLP.