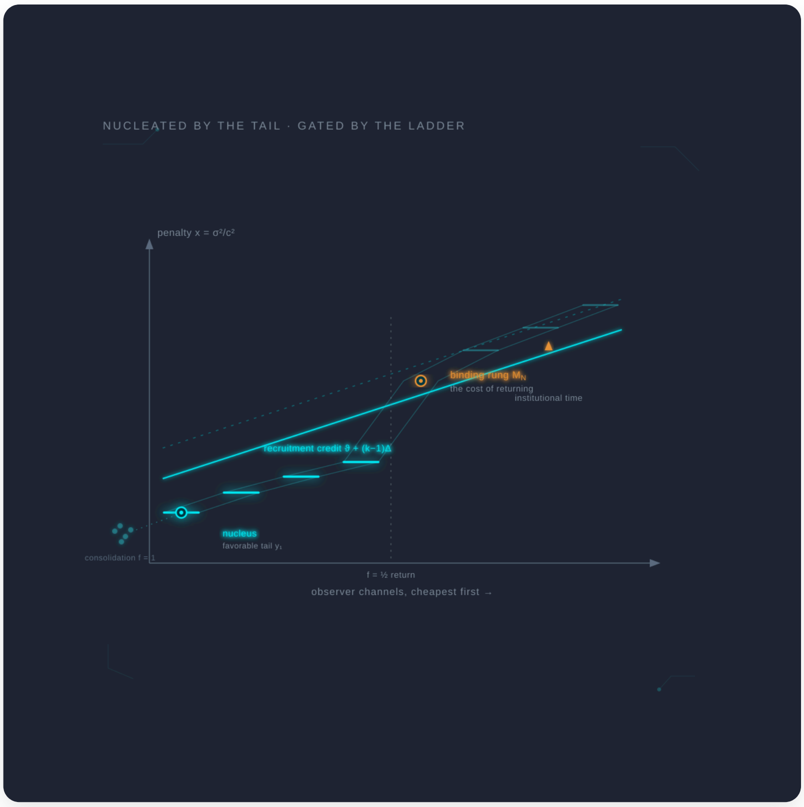

The Cost of Returning

Tail-nucleated escape from epistemic monoculture

Paper XXVI in the Governance as Engineering series

Replaces the scalar reversion constant of Paper X with an explicit escape-ladder mechanism. Return from consolidation is nucleated by the best-preserved channel and propagated by a cascade governed by the ordered penalties of the survivors. Paper XXVI in the Governance as Engineering series.

Björn Kenneth Holmström

July 2026

`https://bjornkennethholmstrom.org/working-papers/cost-of-returning`

Abstract

Paper X showed that consensus-relative evaluation, cost advantages, and liability pressure drive an ensemble of observer organizations toward an epistemic monoculture, and represented the difficulty of escape by a single exogenous constant: a fixed, near-zero probability of switching back, standing in for atrophied independent infrastructure. This paper replaces the placeholder with an explicit mechanism. Each organization carries a latent independent-channel competence that rebuilds under independent operation and decays under shared-system use; an organization's penalty for independence is its idiosyncratic error inflated by its retained competence. In a homogeneous mean-field reduction the model is solved in closed form: the hysteresis loop between consolidation and return decomposes exactly into a consensus-constitution term, a competence-decay term governed by the decay-to-retention ratio, and a liability term, with competence decay moving the return threshold and provably not the entry threshold. The finite heterogeneous population departs from this reduction in a structured way, and the departure is the paper's main object. Escape from full consolidation is *nucleated* by the best-preserved channel and *propagated* by a cascade whose feasibility is a ladder condition on the ordered penalties: the k-th cheapest channel must be viable after $k-1$ defections. The resulting escape-ladder theorem contains the mean-penalty and pure-tail thresholds as limits, predicts stable partial-defection plateaus that the simulation exhibits, and yields a finite-N quantile-ladder criterion — with no fitted phase boundary — that classifies 55 of 56 cells of a frozen population-size × heterogeneity grid, its single error lying on the phase boundary itself, and that locates a non-monotone dependence on heterogeneity in follower spacings rather than nucleus strength. A third, stochastic layer separates the deterministic threshold from the observed one: escape hazards measured at fixed conditions compose, with no fitted dwell law, into the exit thresholds observed under slow parameter sweeps, so that institutional time enters as accumulated escape opportunity rather than as movement of the ladder itself. Within the model, the deterministic decomposition and the ladder theorem are exact, and the stochastic first-passage account is confirmed over the tested parameter and timescale range: mean retained capability sets the bulk pressure against return; the favorable tail determines whether return can begin; the ordered penalty sequence determines whether it can propagate; and institutional time determines whether it happens at all. The transfer of each statement to real epistemic infrastructure is argued separately and carries **[IP]** at best.

1. The placeholder in Paper X

Paper X ends in an uncomfortable place. Its analysis and its Experiment D2 show that locally rational observer organizations — evaluated against a consensus they collectively constitute, rewarded for the genuine short-term economies of shared infrastructure, and shielded from liability when they fail alongside the crowd — flow toward an epistemic monoculture: a configuration in which every nominal observer consults the same internal model, the effective observer count collapses toward one, and systematic error in the shared model is invisible to every instrument the ensemble possesses. The monoculture is described there as *near-absorbing*, and the description is implemented by a single number. In D2, the probability that a shared-system adopter reverts to independence is fixed at 0.002 per evaluation, a constant chosen to represent the fact that independent infrastructure, once abandoned, has atrophied: the skills have dispersed, the pipelines have decayed, the professional lineages have ended. The rare runs in which a consolidated ensemble survives its regime shift are precisely the runs in which this improbable reversion happens to fire in time.

The constant does its narrative work, but it is a placeholder, and the shape of what it hides is worth stating plainly. It asserts that recovery is *hard* without saying what recovery is hard *because of*. It makes the difficulty uniform — every adopter equally unlikely to revert, at every time, regardless of what has happened to its retained capability — when the sentence used to justify it ("infrastructure has atrophied") describes a state variable with dynamics of its own. And it forecloses every question this paper takes up: whether time spent consolidated deepens the difficulty; whether the difficulty is the same for the ensemble's best-preserved member as for its median; whether one viable defector suffices to unravel the monoculture or whether the unraveling can stall partway; and whether there are conditions under which ordinary internal action cannot restore independence at all.

The question, once the placeholder is removed, is this:

> *Once independent capacity has been allowed to decay, what determines whether return can begin, whether it can propagate, and whether it completes within the time an institution has?*

The paper answers it in three layers, in a model deliberately confined to Paper X's strategy-selection machinery — the choice between shared and independent epistemic infrastructure under consensus-relative evaluation — with one addition: an explicit retained-competence state per organization. The environment layer of Paper X (the hidden failure dimension, the regime shift, the precautionary gate) is omitted, because the present question is not what the monoculture fails to see but whether it can be left. Everything established here is therefore **[R within the model]** in the strict sense the series reserves for that tag: derivations and preregistered simulation outcomes of a stated system, whose reading onto real institutions is a separate argument made separately (§7). The paper also records, in the manner of Paper XVIII, the registered predictions that failed along the way and what replaced them; the falsifications turn out to carry a substantial share of the content.

A note on lineage. The model was built in five registered cycles, each gated on the previous cycle's surviving claims, with the failures retained: the strong logarithmic dwell law failed its preregistered fit criterion and was replaced by a measured-hazard composition (§5); a naive order-statistic law for the population-size dependence was falsified and is explained by the theorem it prompted (§3–§4); and a predicted liability–decay interaction, though derived within the ladder approximation, sits an order of magnitude below the resolution of any feasible dynamic test and is recorded as an open registered prediction rather than a result (§7). The consolidated simulator, parameter tables, and the per-cycle registration record constitute Appendix B.

2. Retained competence and the homogeneous loop

2.1 The model

N observer organizations each choose, at periodic evaluations, between a shared epistemic infrastructure (strategy S) and an independent one (strategy I). The shared system carries a common hidden bias b relative to the true state; independent systems are unbiased in aggregate with idiosyncratic error scales σ_i, drawn lognormally, $\sigma_i = \sigma_0 e^{hZ_i}$ with Z_i standard normal, so that h indexes the heterogeneity of the independent ecology. Organizations cannot evaluate against truth — that is the predicament inherited from Paper X — so perceived errors are consensus-relative. With shared fraction f, the consensus sits approximately at fb, giving

$$E_S(f) = (1-f)^2 b^2, \qquad E_{I,i}(f) = f^2 b^2 + \frac{\sigma_i^2}{c_i^2}. \tag{1}$$

These are expected consensus-relative squared errors, not sampled observation streams; common shared noise is suppressed in the reduction so that the analysis isolates common bias, independent error heterogeneity, and retained competence. The first expression is Paper X's mechanism in reduced form: as f rises, adopters increasingly constitute the consensus against which they are judged, and the shared system appears more accurate although its bias against truth is unchanged. The second contains this paper's addition. Each organization carries a retained independent competence $c_i \in (0, 1]$, which rebuilds under independent use, $c_i \leftarrow c_i + \eta_I(1 - c_i)$, and decays under shared use toward a passive floor, $c_i \leftarrow c_i + \rho(1 - c_i) - \delta c_i$, where ρ is passive retention (documentation, residual practice) and δ is decay through disuse. Under sustained shared use, competence converges to the fixed point

$$c_* = \frac{\rho}{\rho + \delta}, \tag{2}$$

so the *decay-to-retention ratio* δ/ρ is the model's single knob for how badly consolidation erodes latent capability. An organization's effective penalty for independence is

$$x_i = \frac{\sigma_i^2}{c_i^2},$$

its idiosyncratic error inflated by lost competence. Utilities are $U_S = -E_S - C_S + \theta$ and $U_{I,i} = -E_{I,i} - C_I - L_0 - L_1 f$, where $C_S < C_I$ are operating costs, L_0 a baseline liability burden on independence, L_1 the liability ratchet that grows with the crowd, and θ an exogenous shared-system advantage (subsidy, procurement preference, coordination convenience) that is swept slowly to reveal the entry and return branches. Write $\Delta C = C_I + L_0 - C_S$ for the static cost gap. Strategy revision is stochastic best response with a small mutation floor μ; the deterministic skeleton of §3 replaces it with strict best response. There is, deliberately, no restart cost and no direct network-performance bonus: all difficulty of return must be generated by the competence state, or the paper installs its own conclusion.

2.2 The closed-form loop

In the homogeneous mean-field reduction — the heterogeneous penalty ecology replaced by the representative penalty $\mathbb{E}[x]$ — the model is not simulated but solved. On the upward branch the population is independent, competence intact ($c_0 = 1$), and consolidation begins when U_S exceeds the representative independent utility at $f \approx 0$; on the downward branch the population is consolidated, latent competence has relaxed to c_*, and return begins when the representative adopter's independence becomes viable at $f \approx 1$:

$$\begin{aligned} \theta_{\text{entry}} &= b^2 - \mathbb{E}[x_0] - \Delta C, & \mathbb{E}[x_0] &= \mathbb{E}[\sigma^2]/c_0^2, \\ \theta_{\text{exit}}^{\text{mean}} &= -b^2 - \mathbb{E}[x_*] - \Delta C - L_1, & \mathbb{E}[x_*] &= \mathbb{E}[\sigma^2]/c_*^2. \end{aligned} \tag{3}$$

The hysteresis width is their difference, and it decomposes exactly **[R within the model]**:

$$W_H = 2b^2 + \mathbb{E}[\sigma^2]\left(\frac{1}{c_*^2} - \frac{1}{c_0^2}\right) + L_1. \tag{4}$$

Each term is separately attributable. The $2b^2$ term is the consensus-constitution mechanism alone: it is the irreversibility Paper X's dynamics generate even with no competence decay and no ratchet, here made explicit as bistability rather than as a small reversion probability. The middle term is capability decay, governed entirely by δ/ρ through Eq. (2); it vanishes as $\delta \to 0$ and grows as $\mathbb{E}[\sigma^2]\left[(1+\delta/\rho)^2 - 1\right]$. The L_1 term is the liability ratchet. At the parameter values of the simulation study ($b = 0.45$, $\sigma_0 = 0.12$, $h = 0.15$, $C_S = 0.50$, $C_I = 1.00$, $L_0 = 0.20$, $L_1 = 0.15$, $\rho = 0.003$, $\delta = 0.005$, hence $c_* = 0.375$), the finite-agent system tracks the formula: predicted exits for the four cells of the decay × ratchet factorial are -0.918, -1.010, -1.068, -1.160 against observed -0.913, -0.963, -1.062, -1.112, with the entry threshold at -0.500 observed against -0.513 predicted in every cell.

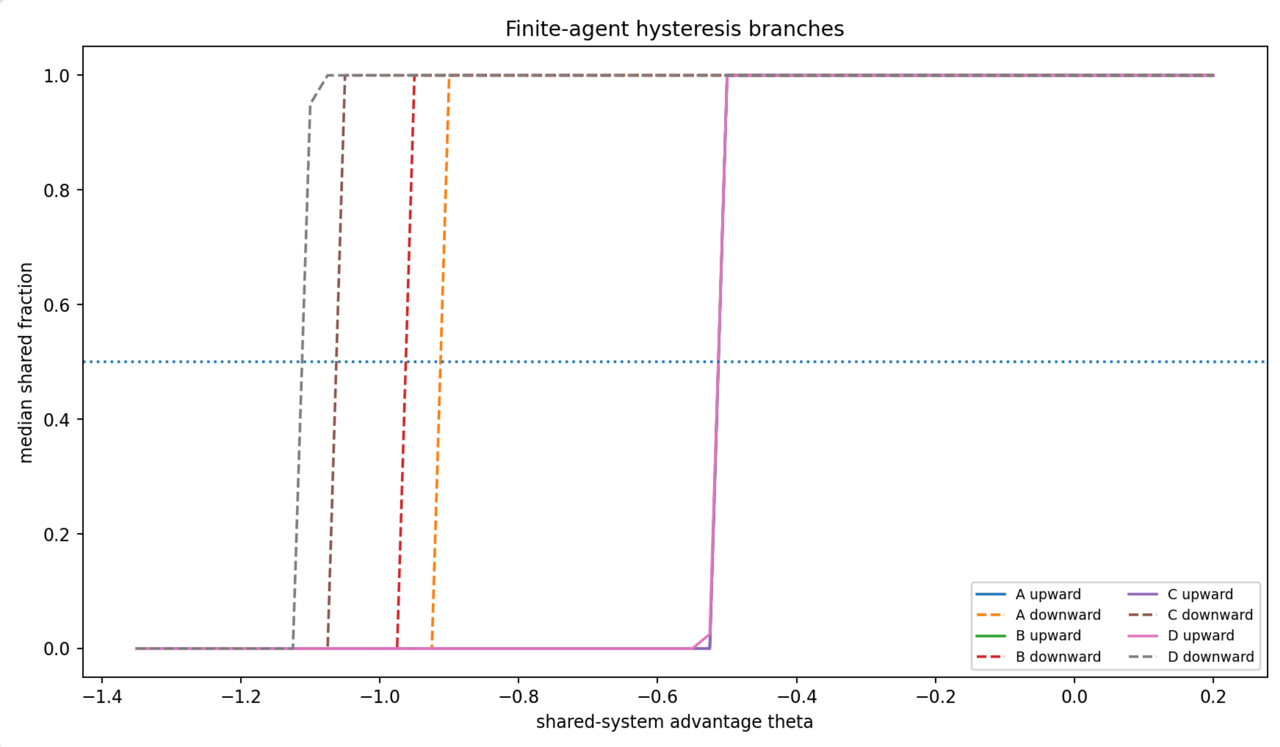

Figure 2.1 — Finite-agent hysteresis branches for the four decay × ratchet cells. All cells share one entry threshold; the return thresholds separate.

The systematic, sign-consistent residual on the exits — every observed return shallower than the homogeneous prediction, and increasingly so with decay — is not noise, and §3 is its explanation.

Two corollaries of the closed form deserve statement because they discipline how the results may be described.

Entry invariance. Competence decay moves the return threshold and provably not the entry threshold: on the upward branch competence is intact by construction, since decay acts only after consolidation, so $\partial\theta_{\text{entry}}/\partial\delta = 0$ while $\partial\theta_{\text{exit}}/\partial\delta < 0$ **[R within the model]**. The asymmetry — capability decay does not make the monoculture easier to enter; it makes it harder to leave — is thus a structural property of where in the cycle decay operates, and it is reported as such rather than as an empirical discovery. Its empirical content is the *magnitude* of the exit shift, which Eq. (4) fixes and the simulation confirms in ordering and, at moderate decay, in size.

Foreclosure of the interaction question. In this reduction, the decay and liability contributions to W_H are additive as an algebraic identity: both enter the scalar exit condition, evaluated at $f = 1$, as separate additive terms. The question of whether liability and capability decay *interact* — whether institutional protection of the shared standard amplifies the irreversibility created by capability loss — is therefore unaskable at this layer; any simulation of this reduction will return an interaction of exactly zero, and an early build of this study did, before the identity was noticed. The methodological point follows Paper XVIII's discipline on always-on tests: a registered question must be posed in a model class capable of answering it either way. The finite-agent layer of §3 is that class, and the interaction question is resumed, and honestly left unresolved, in §7.

3. The escape-ladder theorem

3.1 What the residual demanded

The homogeneous exits of Eq. (3) use the representative penalty $\mathbb{E}[x_*]$. The finite heterogeneous system returns earlier — at shallower θ — in every cell, and the undershoot grows with δ/ρ. The mechanism is visible once one asks who moves first. Exit from full consolidation is not a population-average event: it is initiated by whichever organization has the smallest realized penalty $x_i = \sigma_i^2/c_i^2$ — the best-preserved channel — and its defection changes the environment for everyone else. When that organization leaves, the shared fraction falls by $1/N$, the shared system's consensus-relative error rises, the penalty attached to choosing independence falls with f, and every other organization's case for independence strengthens by a fixed amount. Escape is a cascade, and the question of whether it completes is a question about the *sequence* of penalties, not their mean.

3.2 Setup and lemma

Fix θ and hold competence at its consolidated value (the cascade is fast relative to competence dynamics; the timescale-separation assumption is discussed below). Sort the penalties $y_1 \leq y_2 \leq \cdots \leq y_N$. The advantage of independence for an organization with penalty x at shared fraction f is, from §2.1,

$$A(x, f) = b^2(1 - 2f) - L_1 f - x - \Delta C - \theta. \tag{5}$$

Define the driving term $\vartheta(\theta) = -\theta - b^2 - L_1 - \Delta C$ (the advantage of the cheapest possible defection at $f = 1$, gross of its penalty), and

$$\Lambda := 2b^2 + L_1, \qquad \Delta := \frac{\Lambda}{N}, \tag{6}$$

the total coupling strength and the per-defection **recruitment credit**.

Lemma (monotone cascade). Under asynchronous strict best response at fixed θ and frozen competence, with ties not triggering defection: $A(x, f)$ is strictly decreasing in f for every x; hence along any sequence of defections f only falls, every past defector's advantage only grows, no defector reverts, and the terminal defector set is independent of the order of defection. **[R within the model]** *Proof.* $\partial A/\partial f = -\Lambda < 0$. Because defection lowers f, it strictly increases every organization's defection advantage, defectors included, so reversion is never a best response; and if the k-th cheapest organization is infeasible at the current f, every more expensive organization is also infeasible. The cascade therefore has a unique closure, obtained by evaluating organizations in penalty order. ∎

The lemma is what licenses everything after it: because the cascade admits no reversals and no order-dependence, its outcome is a deterministic function of the sorted penalty sequence, and that function can be written down.

3.3 The theorem

Theorem (escape ladder). Under the dynamics of the lemma, from full consolidation, the k-th defection is feasible iff $y_k < \vartheta(\theta) + (k-1)\Delta$, and the terminal defector count is

$$K^*(\theta) = \min\{\, k : y_k \geq \vartheta(\theta) + (k-1)\Delta \,\} - 1, \qquad \min\varnothing := N+1.$$

Consequently the shared fraction reaches $f \leq \frac{1}{2}$ — the paper's return criterion, matching the loop convention of §2 — iff $\vartheta(\theta)$ exceeds the **ladder functional**

$$M_N = \max_{k \leq \lceil N/2 \rceil} \big[\, y_k - (k-1)\Delta \,\big], \qquad \theta^{\text{det}}_{\text{exit}} = -\big(b^2 + L_1 + \Delta C\big) - M_N. \quad (7)$$

[R within the model] *Proof.* By the lemma the cascade may be evaluated in penalty order. After $k-1$ defections the shared fraction is $1-(k-1)/N$ and the k-th cheapest organization's advantage is $A\big(y_k,\, 1-(k-1)/N\big) = \vartheta(\theta) + (k-1)\Delta - y_k$, positive iff the stated condition holds; the cascade proceeds to the first violation and stops there, no later defection being feasible since y is sorted and the credit is linear. The half-adoption condition is the feasibility of all rungs $k \leq \lceil N/2 \rceil$, which is $\vartheta(\theta) > M_N$ by rearrangement. ∎

The functional M_N is the paper's central object. The maximum identifies the hardest rung that must be crossed before half the ensemble can leave: a cheap first defector is insufficient when a later follower remains too costly relative to the recruitment credit accumulated so far.

3.4 Corollaries

(i) Homogeneous limit. Identical penalties $y_k \equiv \mathbb{E}[x]$ make the maximum bind at $k=1$ and $M_N = \mathbb{E}[x]$: the reduction of §2.2 is recovered, exposing its exit as the spinodal of a degenerate ladder. **(ii) Pure-tail limit.** If every spacing satisfies $y_k - y_1 < (k-1)\Delta$, then $M_N = y_1$ and the exit is set by the best-preserved channel alone — the order-statistic regime, in which the first defector recruits the entire cascade. **(iii) Staircase.** For θ between consecutive violation points of the ladder, the stable configuration is *mixed*, with $f = 1 - K^*(\theta)/N$: the theory predicts partial-defection plateaus at derived locations, not merely a delayed jump. The simulation exhibits them where the theory says it must (§4): in the ladder-dominated regime the median gap between first defection and half-exit is 0.0104 in θ, and across seeds the observed plateau width correlates with the realized ladder gap $M_N - y_1$ at $r = 0.761$. A sufficiently capable channel can leave the monoculture before enough other channels are ready to follow it.

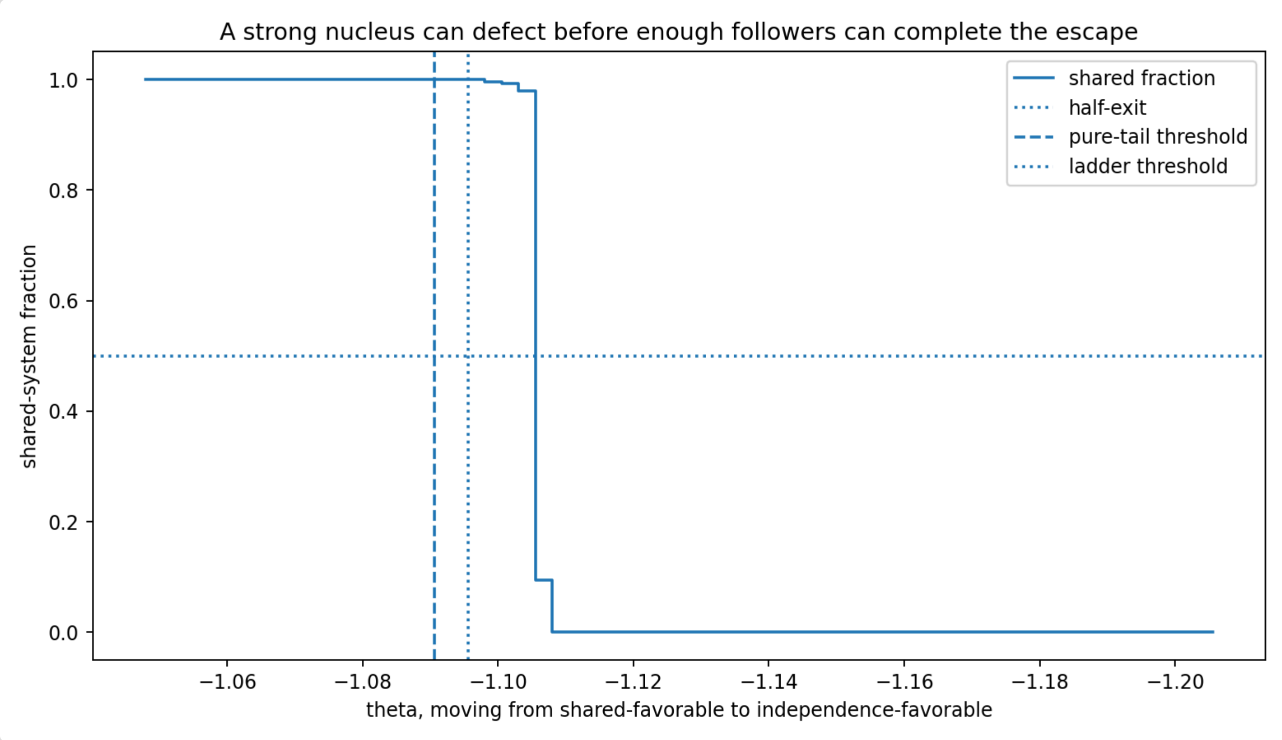

Figure 3.1 — A representative staircase in the ladder-dominated regime: first defection, stable partial-defection plateau, completed escape, against the pure-tail and ladder thresholds.

(iv) Heterogeneous large-population limit. Writing $Q_x := F_x^{-1}$ for the penalty quantile function, with $k = qN$ and the empirical quantiles converging to Q_x, the functional converges to the variational form

$$M_\infty = \sup_{q\in(0,1/2]} \left[\, Q_x(q) - q\Lambda \,\right], \tag{8}$$

which is N-independent — the deterministic half of the explanation, completed in §4, for why the observed exit threshold is nearly flat in N. All four: **[R within the model]**.

Since under sustained consolidation every organization's competence converges to the same c_*, the penalty distribution at exit is the σ^2-distribution scaled by $1/c_*^2$, so $Q_x(q) = Q_{\sigma^2}(q)/c_*^2$ and the decay knob enters the ladder exactly as it entered the homogeneous loop — multiplicatively, through $(1+\delta/\rho)^2$ — but applied now to the binding quantile rather than the mean.

Scope. Two assumptions bound the theorem. *Timescale separation:* competence is frozen during the cascade; this holds in the simulated system, where cascades complete in tens of evaluations against a competence timescale of $1/(\rho+\delta) \approx 125$, and becomes a stated scope condition wherever rebuilding during the cascade would be material. *Mean-field coupling:* every defection delivers its credit Δ to all organizations equally, because interaction runs through the scalar f; under network-structured evaluation the credit would localize and M_N would become graph-dependent. Both are declared limits, not defects: the second is the natural successor question and is left as such.

4. The tail–ladder phase diagram

4.1 When does the ladder bind?

Corollary (ii) says the ladder collapses to the pure tail when spacings are small against the credit; the interesting question is where in parameter space the ladder is load-bearing — where the binding rank $k^* = \arg\max_k [y_k - (k-1)\Delta]$ exceeds one, so that escape is genuinely constrained by followers rather than by the nucleus. The exact criterion is read off the theorem. **Finite-N quantile-ladder criterion:** in expectation over draws, the ladder binds beyond its first rung iff the quantile ladder rises above the recruitment-credit line at some rung, i.e.

$$\exists\, k \in \{2, \dots, \lceil N/2 \rceil\} : \quad Q_x\left(\frac{k}{N+1}\right) - Q_x\left(\frac{1}{N+1}\right) \;>\; (k-1)\,\Delta. \quad (9)$$

Its continuum form is a **secant-slope condition**: with $q_0 \approx 1/(N{+}1)$,

$$\sup_{q \in (q_0,\, 1/2]} \left[\, Q_x(q) - Q_x(q_0) - \Lambda\,(q - q_0) \,\right] > 0, \quad (10)$$

that is, some *average* quantile slope measured from the favorable tail must exceed Λ. The pointwise quantile-density comparison $QD_x(q) > \Lambda$ is the local diagnostic: by the mean value theorem it must hold somewhere if the secant condition holds, so it identifies *where* the credit line can be outrun, but it is not in general sufficient, because the ladder condition depends on the integrated spacing from the first order statistic. Every quantity in Eqs. (9)–(10) is fixed by the model specification before any simulation is run: the criterion has no fitted phase boundary.

4.2 The non-monotone heterogeneity dependence, derived

For the lognormal penalty family, $x = (\sigma_0^2/c_*^2)\, e^{2hZ}$, the quantile function is $Q_x(q) = (\sigma_0^2/c_*^2)\, e^{2hz_q}$ with z_q the standard-normal quantile, and its density $QD_x(q) \propto 2h\, e^{2hz_q + z_q^2/2}$. The behavior of the secant condition on the admissible range produces exactly the phase structure the simulation found, for reasons worth stating carefully because an earlier cycle of this study misstated them. As $h \to 0$ the spacings vanish and the ladder cannot bind: channels are interchangeable and the first defection recruits all others — the tail regime, trivially. As h grows large, the factor e^{2hz_q} with $z_q < 0$ collapses the *absolute* scale of every below-median penalty: for the finite even-N designs simulated here, all rungs sit at quantiles strictly below $\frac{1}{2}$, so the entire ladder compresses toward zero faster than the linear credit line, spacings included, and the ladder goes slack again. The tail regime at extreme heterogeneity is therefore *not* a matter of the exceptional nucleus recruiting more strongly — the credit Δ is blind to who defects — but of the followers becoming absolutely cheap. In the continuum limit the statement acquires a parameter condition, because the median rung $Q_x(\frac{1}{2}) = \sigma_0^2/c_*^2$ is h-invariant: the high-h secant from the tail to the median approaches $2\sigma_0^2/c_*^2$, so the return to the tail phase as $h \to \infty$ requires

$$\frac{\sigma_0^2}{c_*^2} < \frac{\Lambda}{2},$$

which holds at the frozen parameters ($0.102 < 0.278$) but is a condition, not a law: a larger baseline penalty scale would leave the median rung binding even at extreme heterogeneity — a further phase condition rather than a defect of the derivation. Between the two collapses, at intermediate h, the ladder binds: a good first channel defects while later channels remain too costly for the accumulated credit, and escape is constrained by propagation.

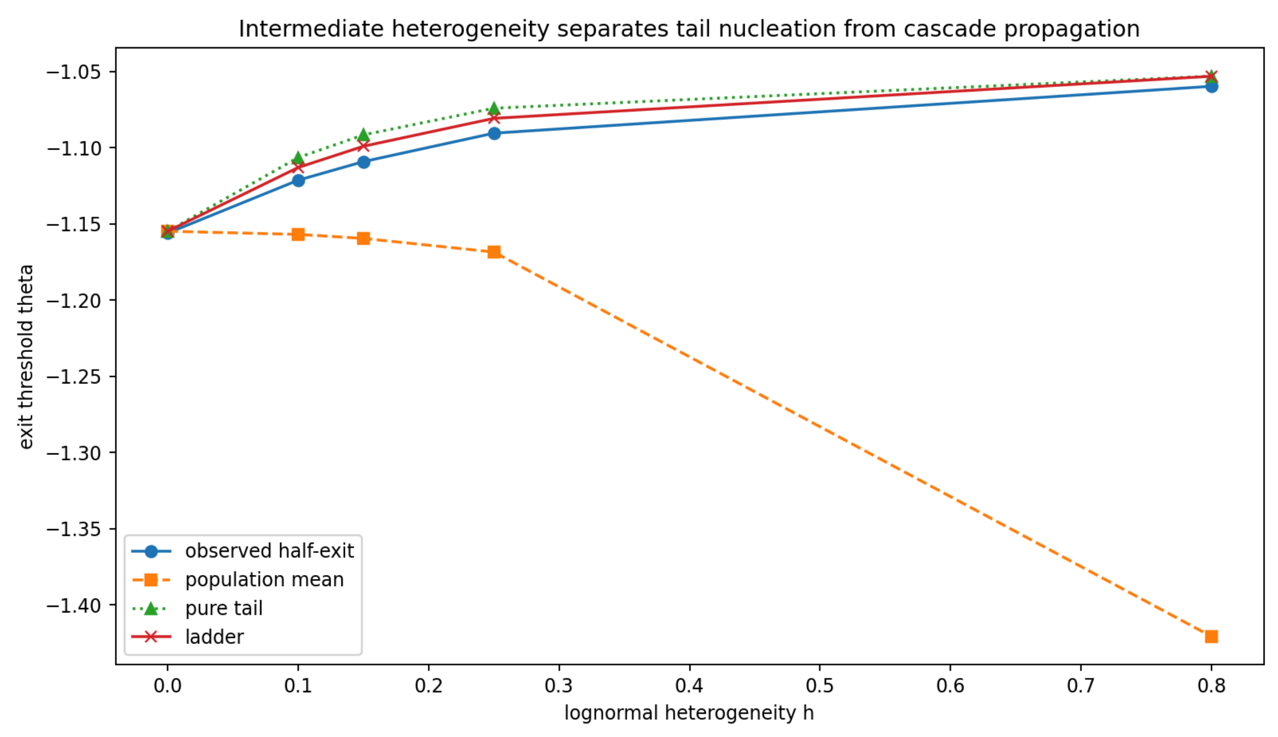

Figure 4.1 — Observed half-exit against homogeneous, pure-tail, and ladder predictions across heterogeneity. The ladder is most distinct from the pure tail at intermediate h.

The N-dependence enters through the admissible lower endpoint: larger N pushes $q_0 = 1/(N+1)$ into the region where the $z_q^2/2$ term inflates the quantile density, so the probability that the ladder binds rises with N at fixed h. The same geometry completes the explanation of the flat N-curve that falsified this study's earlier order-statistic prediction. Raising N improves the favorable tail — y_1 falls — but weakens each defection's recruitment credit as Λ/N, and in the observed thresholds the two effects nearly cancel: at $h = 0.15$ the measured half-exit moves by less than 0.01 in θ across N from 20 to 1000, while the pure-tail prediction moves by three times that. The ladder functional, which prices both effects, tracks the observed thresholds where the pure tail does not: in the ladder-dominated regime the ladder's mean absolute exit error is 0.0089 against 0.0145 for the realized pure tail and 0.0523 for the population mean, with median binding rank $k^* \approx 10$ — roughly ten defections must be supportable before the ensemble can reach half shared adoption.

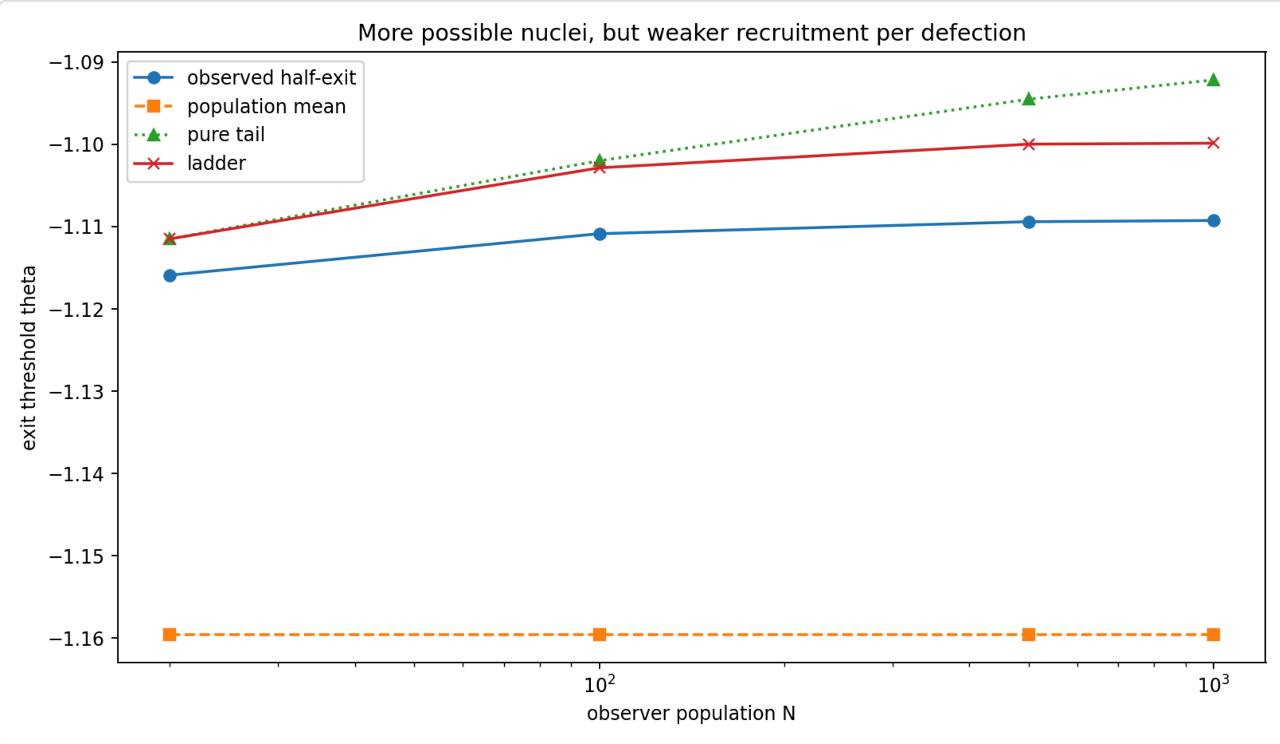

Figure 4.2 — The near-flat N-dependence: the improving tail and the weakening per-defection credit price against each other in the ladder functional.

4.3 The classification test

The criterion of §4.1 was then put to the registered test that motivates calling the phase structure a result rather than a scan: reproduce the empirical $(N \times h)$ phase map — each cell's Monte Carlo estimate of $P(k^* > 1)$, thresholded at $\frac{1}{2}$ — from the finite-N quantile-ladder criterion alone, with nothing fitted. The criterion classifies 55 of 56 frozen cells (98.2%). The single error is the cell $N = 500$, $h = 0.60$, where the empirical $P(k^* > 1) = 0.543$ sits 0.043 above the majority threshold while the quantile classifier remains tail-side: the criterion errs only where its own quantity is closest to indifference. The three anchor regimes, registered in advance, behave as predicted:

Regime	Empirical $P(k^* > 1)$	Analytic phase
$N = 20$, $h = 0.15$ (Paper X scale)	0.015	tail-dominated
$N = 1000$, $h = 0.15$	0.997	ladder-dominated
$N = 1000$, $h = 0.80$	0.215	tail-dominated (returned)

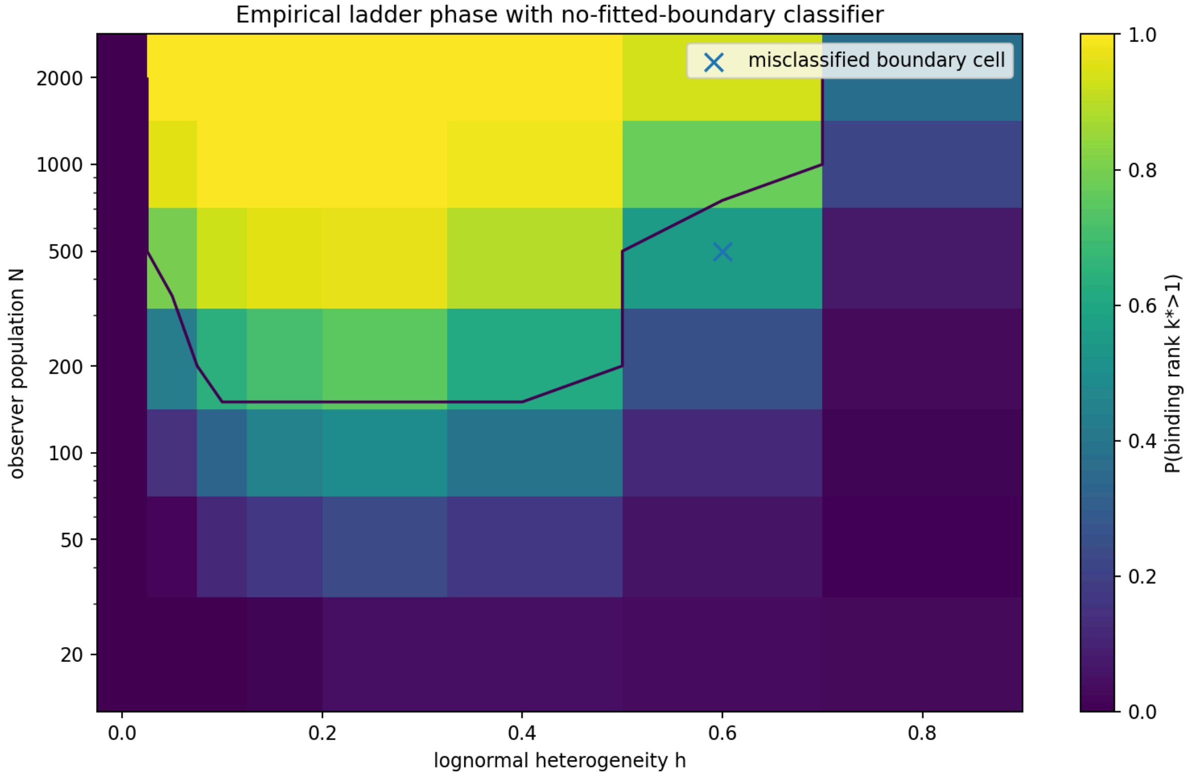

Figure 4.3 — Monte Carlo $P(k^* > 1)$ over the frozen $N \times h$ grid with the analytic boundary overlaid; the single misclassified cell lies on the boundary itself.

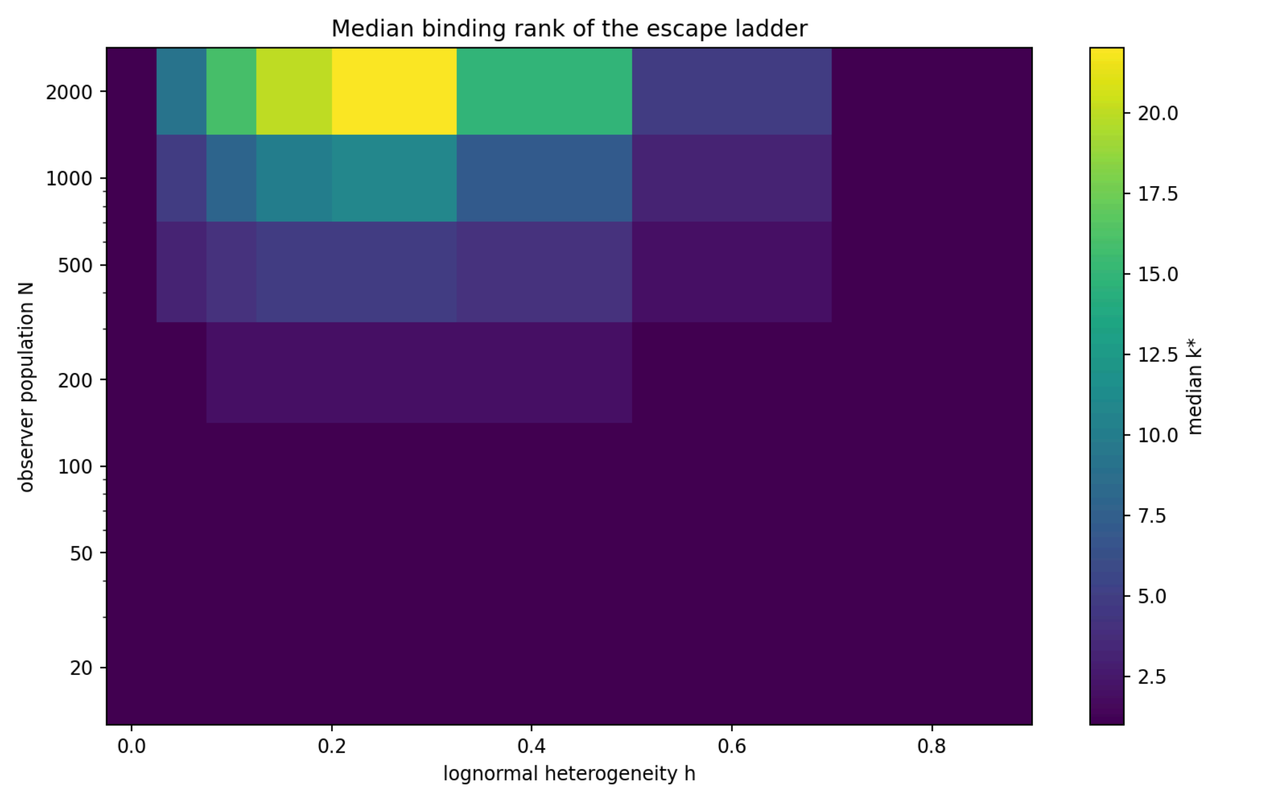

Figure 4.4 — Median binding rank k^* over the same grid: the quantitative companion to the phase map.

Two readings of the map belong here rather than in §7 because they are structural rather than interpretive. First, at the ensemble sizes Paper X actually models — tens of observer organizations — the tail regime governs: the reversibility of the monoculture is carried almost entirely by its single best-preserved channel, an order-statistic dependence that is the recovery-side sibling of Paper X's coverage arithmetic, where the first few protected observers carry nearly all detection capacity. Second, the quantity $\Lambda = 2b^2 + L_1$ appears on both sides of the ledger. It is the per-rung recruitment credit — the amount by which each defection liberates the rest — and it is built from exactly the two forces that deepen lock-in: the consensus-constitution bias and the liability ratchet. Conditional on a viable first defector, the forces that make the monoculture hard to leave are the same forces that make its unraveling fast. The governance reading of both points is deferred to §7 and carries **[IP]**; within the model they are exact.

5. Institutional time and the escape hazard

5.1 The threshold is protocol-dependent, and the first law proposed for it failed

The escape-ladder theorem of §3 is deterministic: it gives the θ below which the cascade *can* run. The simulated system, like any institution, is stochastic and finite in time, and its measured exit depends on how long it lingers at each condition. Sweeping θ downward with T evaluations per step in the Paper X-scale regime ($N = 20$, $h = 0.15$, decay and ratchet on), the median observed exit moves from -1.130 at $T = 10$ to -1.050 at $T = 300$: an institution given more time at each condition escapes under weaker pressure. The first registered account of this — an affine law in $\ln T$ — failed its preregistered criterion ($R^2 = 0.9314$ against a registered 0.95): the shift per log-unit of time shrinks at long dwell, and an unbounded logarithm cannot saturate. The failure is retained in the ledger (§7.4) and forced the replacement that follows, which fits nothing.

5.2 Measured hazards

Instead of fitting a dwell law, the final study measures the escape process directly. At each fixed θ on a grid spanning the transition region, populations held in full consolidation are run until first escape, many trials per population, and a single-rate (exponential) survival model is estimated per population on half the trials and tested on the held-out half. The registered criterion applied to the overall held-out survival-curve mean absolute error and passed comfortably: $0.0538 < 0.12$ **[R within the model]**. The estimated hazard is steep in the institutional environment: $\lambda \approx 2.7 \times 10^{-4}$ per evaluation at $\theta = -1.02$, 6.8×10^{-3} at -1.10, and 0.176 at -1.17 — three orders of magnitude across 0.15 in θ, which is why the deterministic ladder threshold remains the right first-order object even though the observed exit is a first-passage quantity.

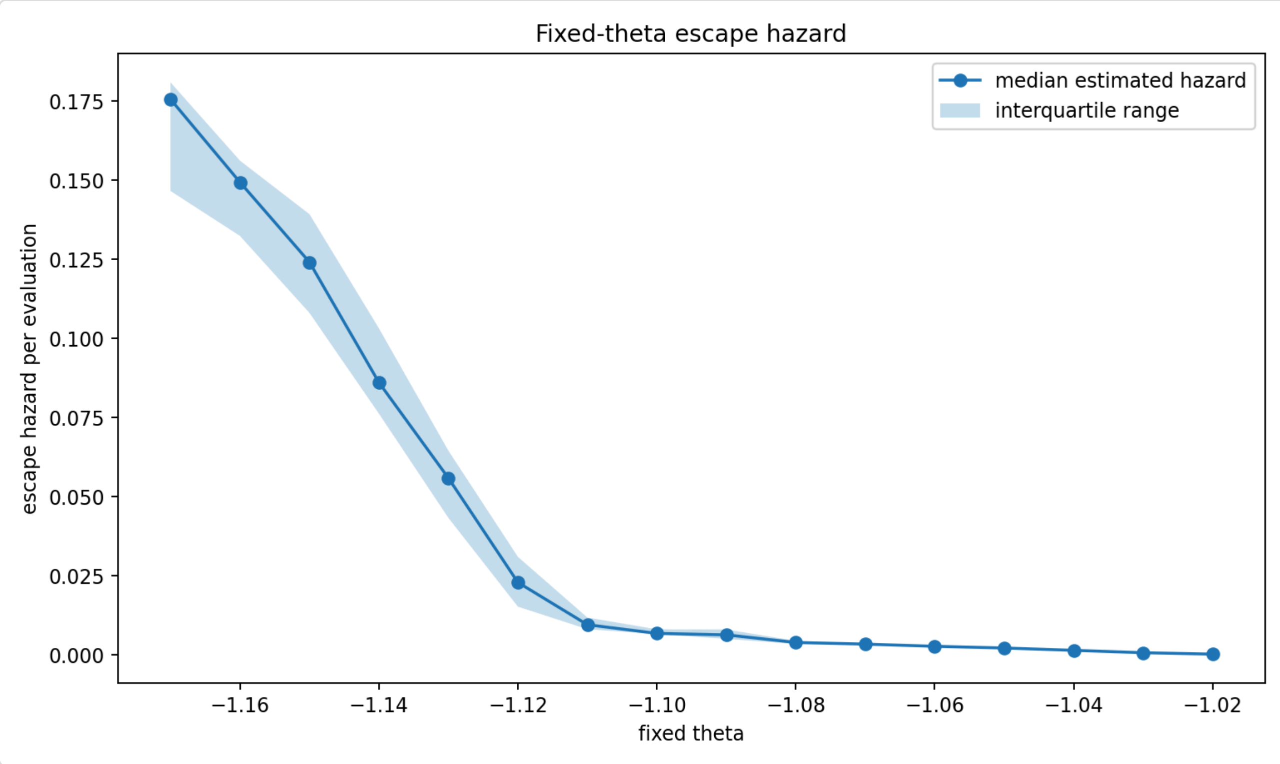

Figure 5.1 — Escape hazard per evaluation at fixed θ (median across populations, interquartile band). The hazard spans three orders of magnitude across the transition region.

Pointwise error nevertheless peaks at 0.119 near $\theta = -1.05$, the shallow edge of the region. That peak was not a separate registered threshold, but its location is diagnostically important, and it is consistent with the expected failure mode of a single-rate reduction: near the shallow boundary, escape involves unsuccessful nucleations and partial cascades, so waiting times need not be exponential, and early survival can be flatter than a single rate allows. The exponential is therefore an operational reduction with a stated scope — reliable in the steep region, degrading predictably toward the ceiling — and the paper claims it as nothing more.

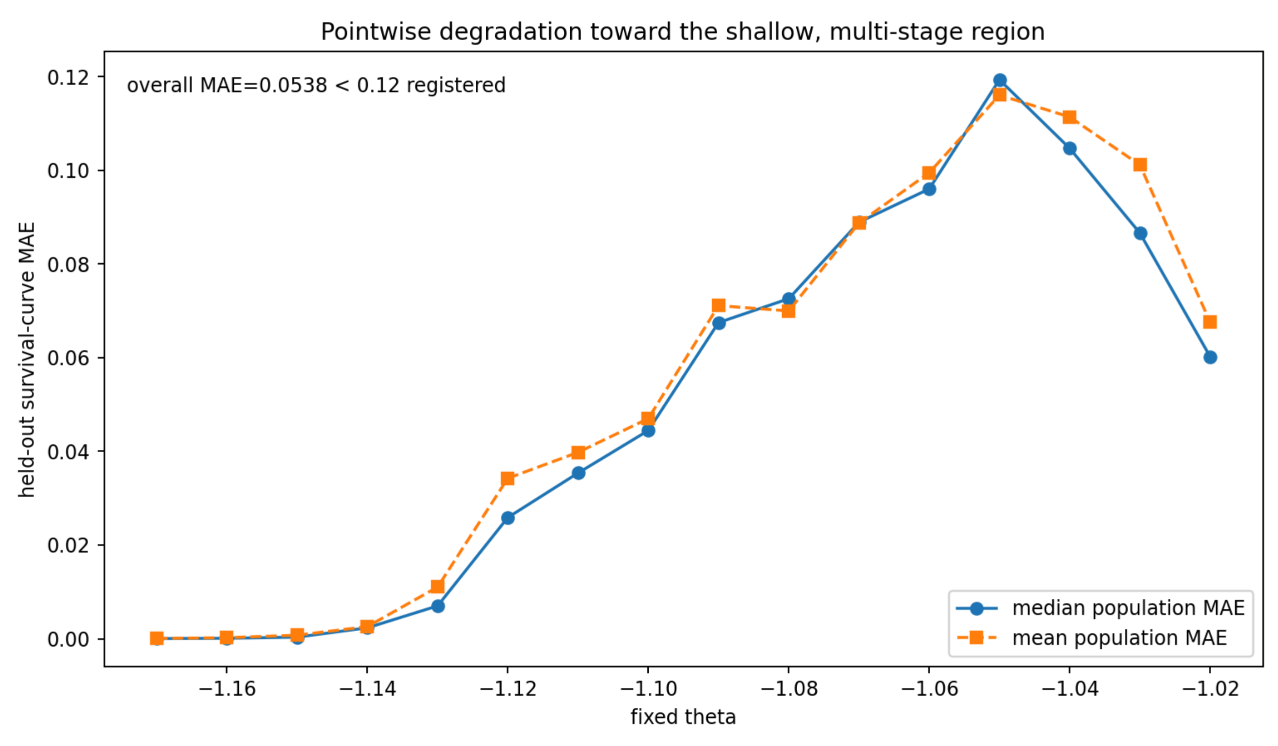

Figure 5.2 — Held-out survival-curve error of the exponential reduction across the θ grid, against the registered limit on the overall mean. Pointwise error peaks at the shallow edge, where escape is multi-stage.

5.3 Composition: the sweep as accumulated opportunity

If institutional time acts as accumulated escape opportunity and nothing else, the measured fixed-θ hazards must *compose* into the swept thresholds of §5.1 with no further fitting. For a downward sweep visiting $\theta_1 > \theta_2 > \cdots$ with T evaluations each, the cumulative hazard after step j is

$$H_j(T) = T \sum_{m \leq j} \lambda(\theta_m), \qquad (11)$$

and the predicted median exit is the first θ_j at which $H_j \geq \ln 2$. Two registered tests apply this. The *population-conditional* test composes each population's own hazard curve into its own predicted exit and compares per population: mean absolute error 0.0029 in θ. The *new-population* test likewise composes each training population's own hazard curve into a predicted exit, takes the median of those population-specific predictions, and compares it with the median exit of entirely fresh population draws: error 0.0050 — exact at dwell 10 and 30, one grid interval shallow at dwell 100 and 300. Because grid quantization makes the median of the per-population predictions coincide numerically with the prediction obtained from a median hazard curve, the aggregation correction appears in the distribution of predicted exits rather than in the point median. **[R within the model]**

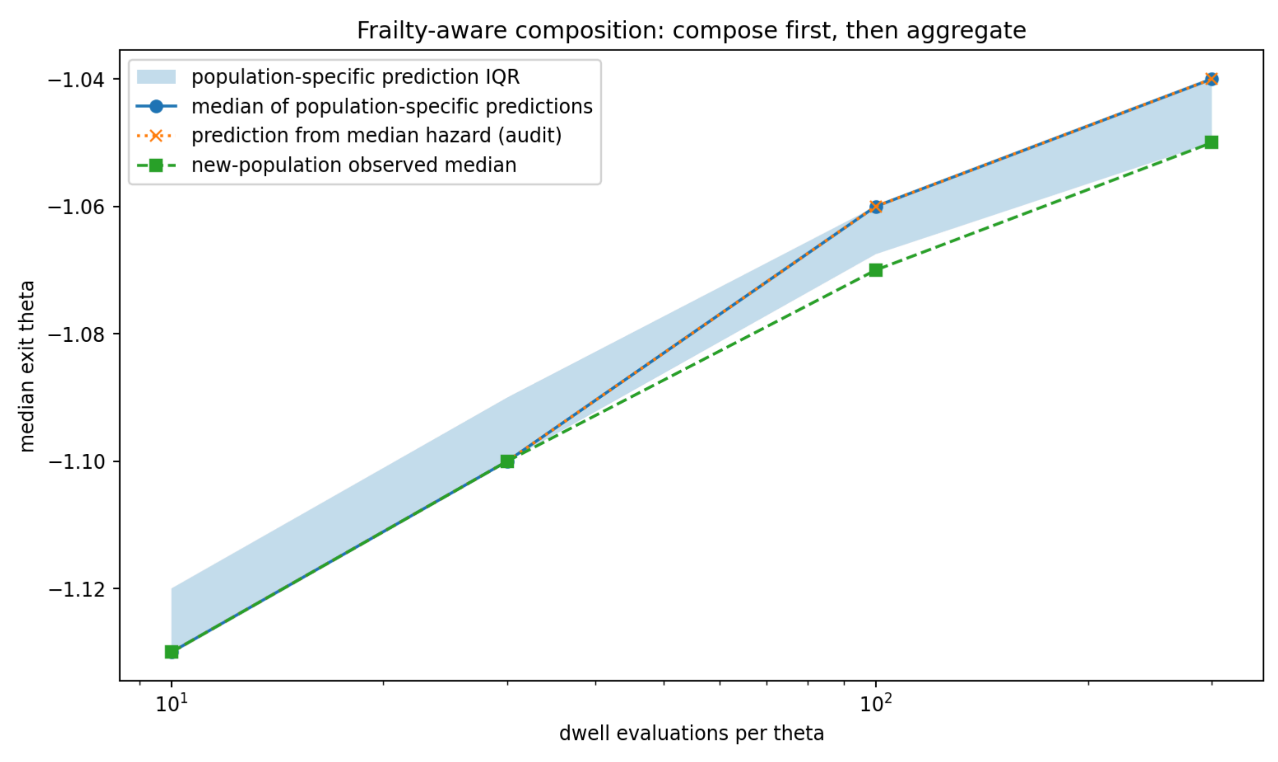

Figure 5.3 — Fixed-θ hazards, measured independently and composed per population with no fitted dwell parameters, against observed medians on new population draws.

The residual long-dwell bias is consistent with two effects of the same sign — shallow-region non-exponential waiting times (§5.2), and finite-sample population frailty, since the populations still consolidated at long dwell are selectively the low-hazard ones. Their contributions are not separately identified here; the population-conditional test, which removes sampling variation across populations, is the clean statement of the section's claim:

> *Institutional time changes the observed return threshold by accumulating escape opportunity. The deterministic ladder does not move; what moves is the probability of having used it.*

6. Recovery ecology

The layers so far concern whether the monoculture can be *left*. This section, deliberately subordinate to the theorem, concerns whether independence can be *rebuilt* — and it is where the one candidate for genuine irreversibility lives.

The recovery study maps recovery probability over the plane $(T_{\text{lock}}, \theta_{\text{recovery}})$: the population is held fully consolidated for T_{lock} evaluations, the pressure is then reduced to θ_{recovery}, a maximal admissible internal training budget is applied, and recovery is scored as reaching a low shared fraction with high independent competence within a fixed horizon. Three rebuilding laws are compared under common random numbers. Under *fixed efficiency*, training converts resources to competence at a constant rate. Under the *subsidy-gated* law, the training subsidy — but not natural rebuilding through independent practice — requires a surviving mass of competent independent trainers. Under the *fully gated* ablation, all rebuilding, natural included, requires that trainer mass: the explicit assumption, named here the **trainer-viability assumption**, that competence cannot be reconstructed at all without surviving practitioners to transmit it.

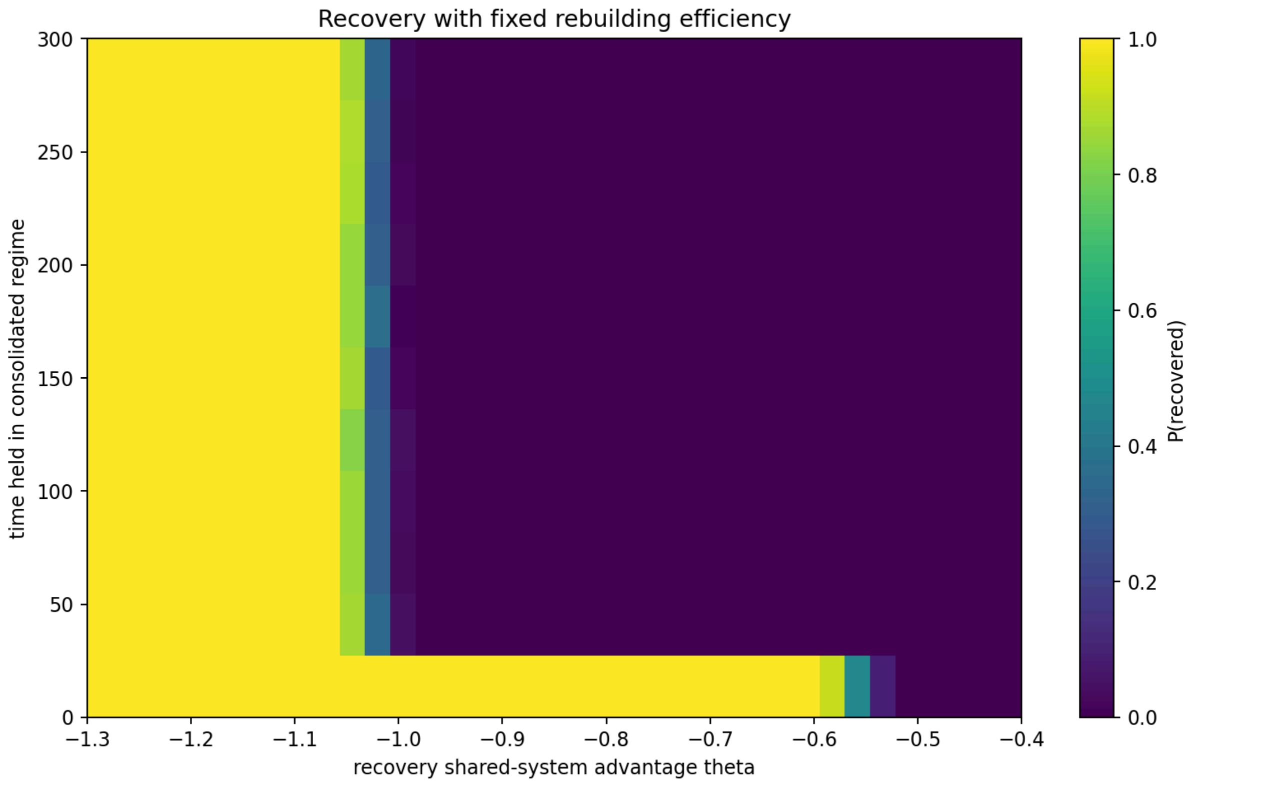

Figure 6.1 — Recovery probability over lock duration × recovery pressure, fixed rebuilding efficiency.

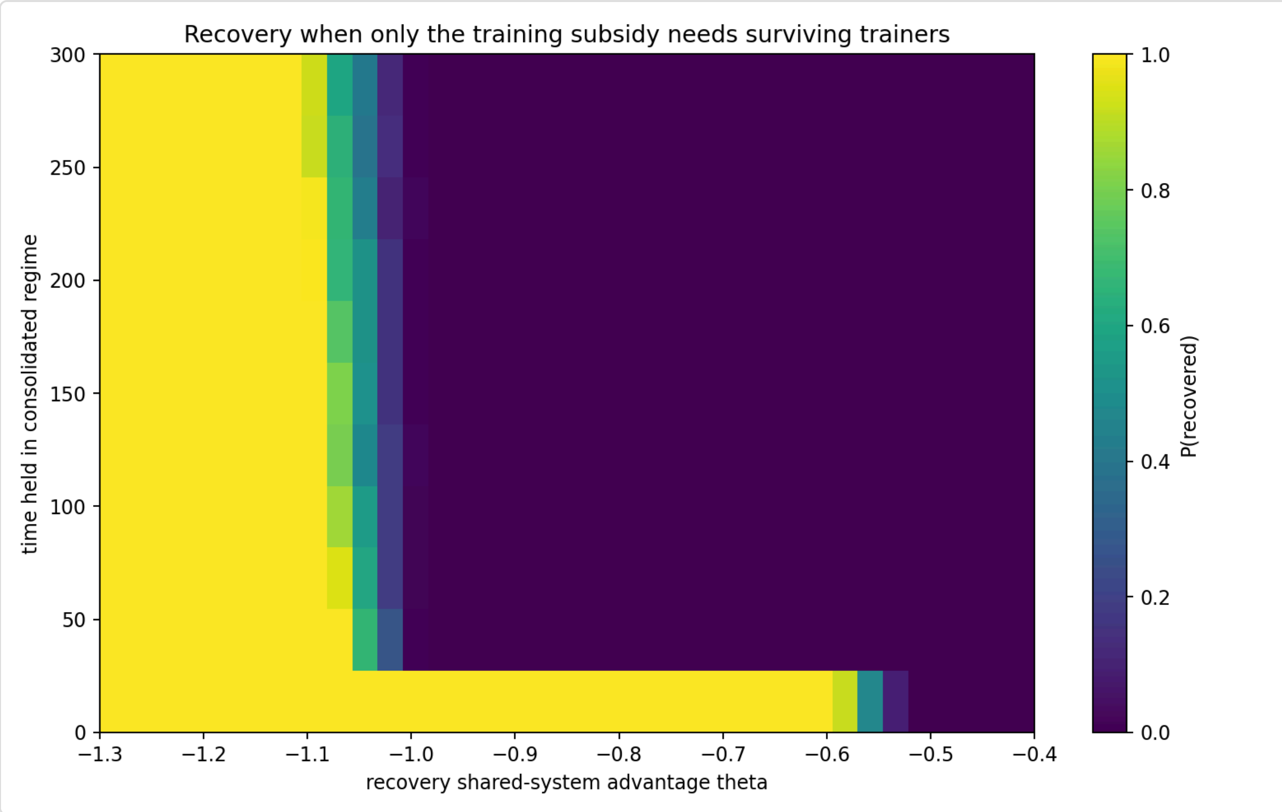

Figure 6.2 — The same map when the training subsidy requires a surviving trainer ecology.

Whole-map averages understate the mechanism and are reported only to show why: mean recovery probability is 0.350 under fixed efficiency against 0.331 subsidy-gated, a two-point difference that would read as negligible. The structure is in the boundary. Within the *recovery boundary band* — defined solely from the fixed-efficiency baseline as the cells where its recovery probability lies in $[0.1, 0.9]$, and therefore independent of the treatment comparison — the mean gap is 0.239, and it is substantially larger after prolonged consolidation: 0.140 at $T_{\text{lock}} = 30$, 0.212 at 90, 0.242 at 150, 0.315 at 240, 0.328 at 300, increasing generally though not strictly monotonically across the tested durations. **[R within the model]** A protocol note belongs here: the band's final $[0.1, 0.9]$ width was adopted after an initial narrower band proved too sparse for stable conditional estimation, alongside the common-random-number and grid-resolution corrections; it should be read as an exploratory localization statistic rather than an untouched preregistered endpoint. Trainer dependence does not shift recovery everywhere; it shifts it where recovery hangs in the balance, and increasingly so the longer the consolidation has lasted — the duration dependence that the fixed exogenous reversion probability of Paper X could not, by construction, express.

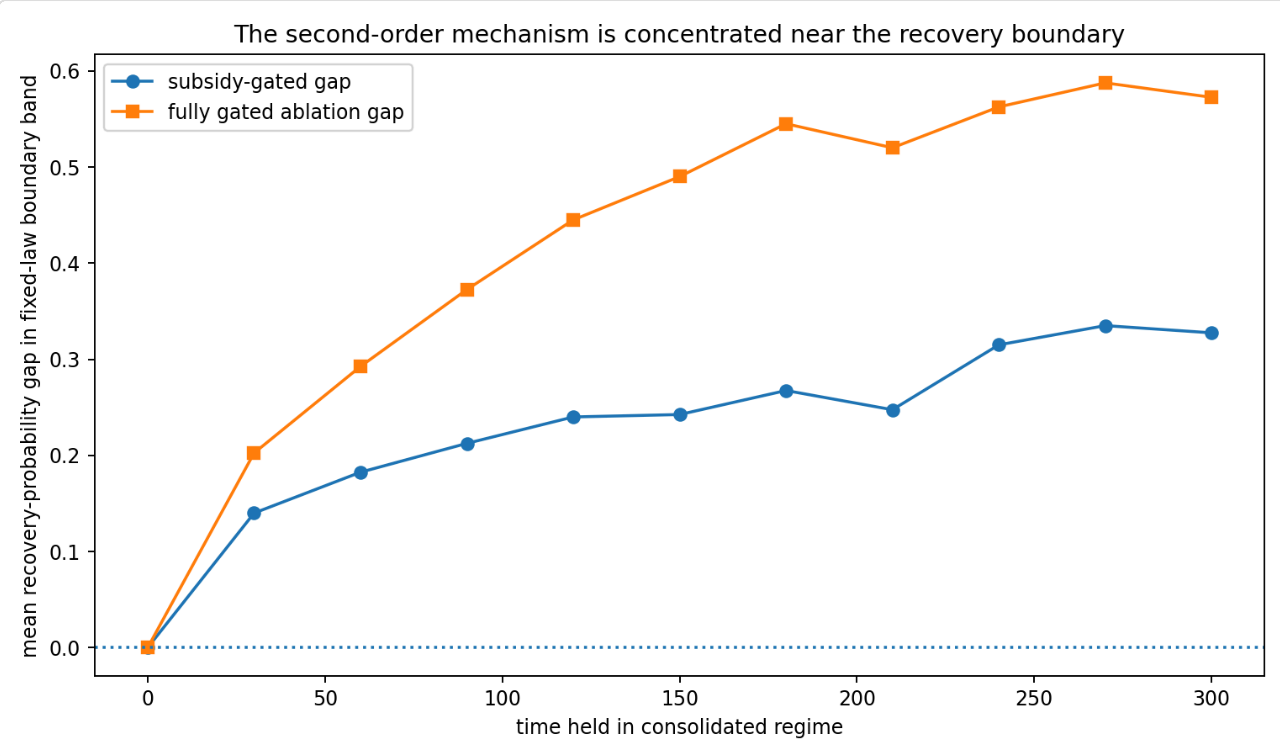

Figure 6.3 — Band-conditional recovery gap versus lock duration, for the subsidy-gated law and the fully gated ablation.

The fully gated ablation is where recovery comes closest to failing outright. Under the trainer-viability assumption, sufficiently long consolidation produces cells in which fixed-efficiency rebuilding recovers at least half the populations while fully gated rebuilding recovers at most five percent: a **conditional near-absorbing region** under the tested internal policy and recovery horizon — once trainer mass has fallen below its critical level, ordinary internal expenditure is almost never sufficient to restore independence. In a deterministic variant with no mutation and rebuilding effectiveness exactly zero below the trainer threshold, this region would be strictly absorbing; the stochastic experiment reported here establishes near-absorption, not that stronger result. The conditionality is the content either way: the model does not discover that training ecologies possess a critical mass; it derives what follows if they do. Whether real epistemic infrastructures have the analogous property — minimum viable cohorts of senior practitioners, laboratories and pipelines that cannot be recreated incrementally, tacit knowledge that does not survive its last carriers — is precisely the empirical question the assumption isolates, and it is carried to §7 at **[IP]**.

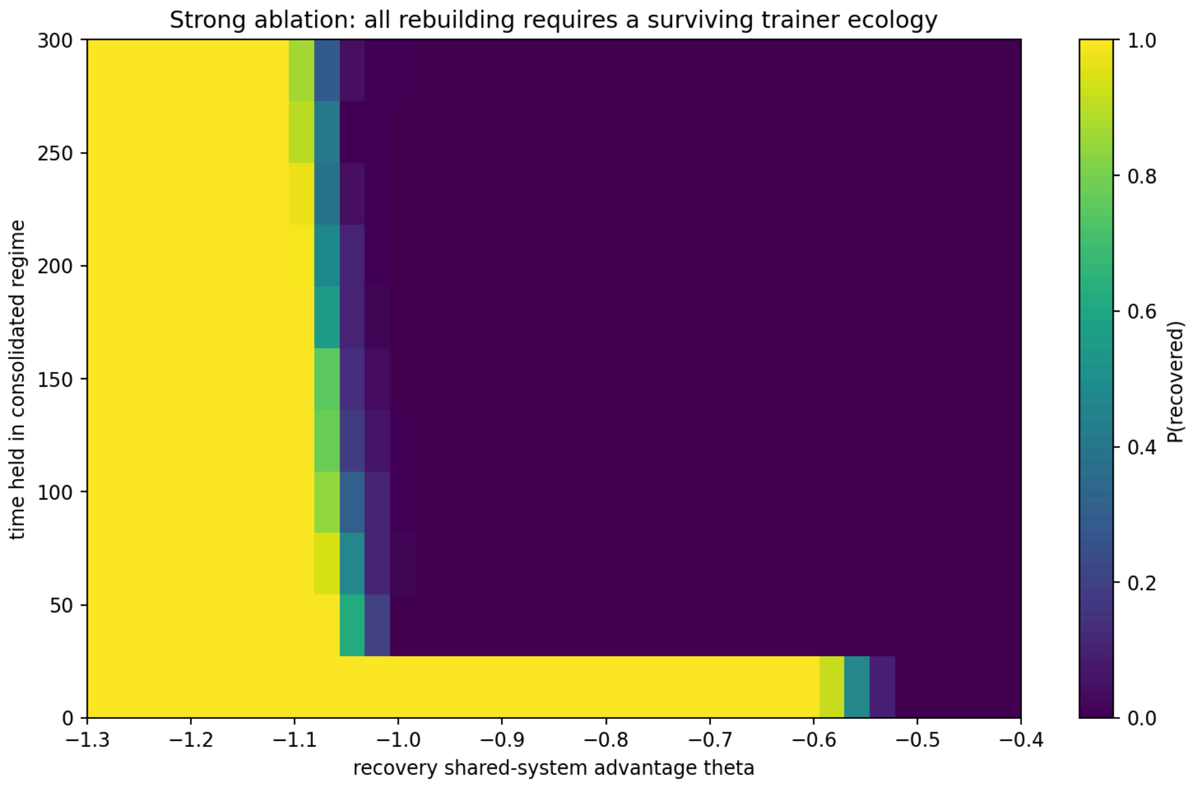

Figure 6.4 — The fully gated ablation. Under the named trainer-viability assumption, a conditional near-absorbing region appears.

The section's connection to the rest of the paper is direct. The ladder theorem says return is carried by the favorable tail and its spacings; the recovery maps say that what training protects is exactly that tail. Paper X's protected-fraction result — a few protected observers supply nearly all detection capacity — acquires here its recovery-side twin: a protected remnant is not only the ensemble's residual eyesight but the nucleus and trainer stock of any future return.

7. Interpretation, limits, and the ledger

7.1 Position in the series

Three connections locate the paper. *To Paper X.* Its Experiment D1 established that a consolidated ensemble's survival is an order statistic on coverage — does *anyone* see the critical dimension. This paper establishes that the reversibility of consolidation is an order statistic plus a spacing condition on retained capability — can anyone leave, and can enough follow. The two are the same logic applied to opposite faces of the monoculture, and together they replace Paper X's scalar reversion probability with a structured answer: N_{eff} measures the redundancy of present observation; M_N measures the reachability of its return. They are complementary structural diagnostics of the same architecture. *To Paper XVI.* When the training budget or protected remnant of §6 is placed outside the primary allocator's control, the section can be read as an instance of Paper XVI's source-term-locality result: the quantity representing currently-unused alternatives decays under the primary objective and is held positive only by a term the optimizing process does not itself set. Whether an allocator *given* control of that term would preserve it is the second-order question Paper XVI marks as contested; it was not simulated here, remains open, and is listed as future work rather than smuggled in as an assumption either way. *To Paper XXIV.* The recovery criterion of §6 scores exercisable restoration — reaching the recovered region within the horizon — not nominal option counts, following Paper XXIV's discipline that a preserved possibility is one that can actually be taken.

7.2 The governance corollary

Stated at its honest tier, **[IP]** throughout: if the strategy-layer structure transfers — organizations choosing epistemic infrastructure under consensus-relative evaluation, capability that atrophies in disuse, recruitment externalities among defectors — then three consequences follow that a mean-quality view of institutional health cannot see. First, *preserving the average quality of independent institutions is not the operative target.* Return from consolidation is nucleated by the best-preserved surviving channel; the favorable tail governs initiation, and the intermediate spacings govern propagation whenever the ladder binds. A policy that maintains a respectable mean while permitting the exceptional tail to decay removes the nucleus; one that protects a single exceptional channel while the middle thins removes the ladder. Second, *the duality of* $\Lambda = 2b^2 + L_1$. The same coefficients that deepen the full-consolidation barrier also set the recruitment credit delivered by each defection. Conditional on a viable first defector, stronger lock-in forces can therefore accelerate propagation — a structural reading of why epistemic regime changes, when they finally come, can come fast — though they also make that first defection harder to reach. Third, *time is a policy variable*. The observed reversibility of a consolidated system depends on the institutional time available for escape to nucleate; windows of reduced consolidation pressure that are too short relative to the escape hazard will measure as irreversibility even where the ladder is passable.

The trainer-viability assumption of §6 converts into the transfer question most worth funding a case study to answer: do real epistemic infrastructures exhibit critical mass in their capacity to reproduce competence — minimum cohorts of senior experts, professional lineages that end with their last practitioners, tacit knowledge unreconstructable from documentation? Where they do, the conditional near-absorbing region is not a modeling curiosity but a deadline.

7.3 Limits

The model is a strategy layer only: Paper X's environment — the hidden dimension, the regime shift, the precautionary gate — is not run, so nothing here concerns what the monoculture fails to see, only whether it can be left. Coupling is mean-field through the scalar shared fraction; network-structured evaluation would localize the recruitment credit and make M_N graph-dependent, the natural successor question. The phase analysis rests on the finite-N quantile-ladder criterion of Eq. (9), with the secant condition as its continuum form; the non-monotonicity derivation and the high-h parameter condition are specific to the lognormal family. Competence dynamics are first-order with a passive floor; richer forms (thresholded forgetting, generational turnover) would change c_* but not the ladder's structure. All parameters are illustrative of structure, calibrated to nothing; per the series' standing caveat, the numbers demonstrate entailments, not magnitudes. And the timescale-separation assumption of §3 — cascade fast against competence dynamics — holds in the simulated system and must be checked wherever the theorem is applied.

7.4 The falsification ledger

Following Paper XVIII's precedent, the registered predictions that failed, and what each failure bought:

Registered claim	Outcome	Disposition
Early candidate: convex recovery cost in lock duration	Rejected before confirmation; the fixed-efficiency form instead predicted and produced concave saturation	Converted to a discriminating test; convexity requires second-order (trainer-gated) decay, confirmed as the mechanism of §6's near-absorbing region
Homogeneous liability × decay interaction	Exactly zero — an algebraic identity of the model class, noticed after one build returned it	Question reclassified as unaskable at that layer (§2.2); methodological note per Paper XVIII's always-on-test discipline
Subadditivity would vanish under continuous threshold estimation	Failed: survived de-quantization	Prompted the ladder-based derivation of a predicted interaction
Naive order-statistic N-law for exit thresholds	Falsified: observed exits nearly flat in N	Explained by tail-versus-kick cancellation; became §4.2
Strong affine-log dwell law	Failed registered R^2 ($0.9314 < 0.95$)	Replaced by measured-hazard composition (§5), which fits nothing and predicts to half a grid interval
Liability × decay interaction (finite-agent)	Ladder approximation predicts a small negative interaction, $-3.5 \times 10^{-4} \pm 2 \times 10^{-4}$; dynamic estimates unstable across protocols (-0.0042 ± 0.0032; -0.0020 ± 0.0030), individually consistent with zero	**Open registered prediction, not a result.** Recorded for a future high-power test

One interpretive error is also on record: an intermediate cycle attributed the high-heterogeneity return to the tail regime to the exceptional nucleus's recruiting strength; the correct mechanism — absolute compression of the followers' spacings, with its parameter condition — is derived in §4.2. The correction changed the phase criterion from a story into a theorem.

8. Conclusion

Paper X explained why epistemic diversity disappears: consensus-relative evaluation, genuine short-term economies, and liability shelter make consolidation the locally rational choice of every observer, until no independent perspective remains from which the shared model's error is visible. It represented the difficulty of coming back with a single small constant. This paper replaces the constant with a mechanism and finds that the difficulty has anatomy. In the homogeneous reduction, return is blocked by a hysteresis loop whose width decomposes exactly into consensus-constitution, capability decay, and liability. In a finite ensemble, return is nucleated by the best-preserved channel and propagated along an escape ladder whose rungs are the ordered penalties of the survivors and whose credit per rung is set by the very forces that hold the monoculture together. Whether the ladder or the tail governs is a phase question answered by a criterion with no fitted boundary; whether escape occurs at all is additionally a question of institutional time, entering as accumulated opportunity against a measured hazard; and whether there is anything left to return *to* depends on whether the capability that trains capability has itself survived.

Within the model, then: return from epistemic monoculture may be impossible in the mean, possible through the tail, blocked by the ladder, delayed beyond the institution's available time — and, under one named assumption about how competence reproduces, close to internally inaccessible under any ordinary policy. Which of these describes any real institution is not established here. What is established is that they are different conditions, with different signatures and different remedies, and that a governance conversation conducted in terms of average institutional quality cannot distinguish them.

Appendix A — Proofs and derivations

A.1 Homogeneous thresholds and the loop decomposition

At $f \approx 0$ with intact competence c_0, consolidation begins when $U_S > U_I$ for the representative organization: using $E_S(0) = b^2$, $E_I(0) = \mathbb{E}[x_0]$, and collecting the cost and liability constants into $\Delta C = C_I + L_0 - C_S$ (the ratchet term $L_1 f$ vanishing at $f = 0$), this gives $\theta_{\text{entry}} = b^2 - \mathbb{E}[x_0] - \Delta C$. At $f \approx 1$ with competence relaxed to c_*: $E_S(1) = 0$, $E_I(1) = b^2 + \mathbb{E}[x_*]$, ratchet at full weight, giving $\theta_{\text{exit}}^{\text{mean}} = -b^2 - \mathbb{E}[x_*] - \Delta C - L_1$. Subtracting yields Eq. (4), using $x = \sigma^2/c^2$ and the common competence values. The competence fixed point under shared use solves $\rho(1-c) = \delta c$, giving Eq. (2). *Entry invariance:* θ_{entry} contains no δ-dependent quantity, since decay acts only on consolidated organizations and the upward branch approaches entry from $f \approx 0$ with $c = c_0$; hence $\partial\theta_{\text{entry}}/\partial\delta = 0$, while $\partial\theta_{\text{exit}}/\partial\delta = -\mathbb{E}[\sigma^2]\,\partial(1/c_*^2)/\partial\delta < 0$. *Additivity identity:* δ enters θ_{exit} only through $\mathbb{E}[x_*]$ and L_1 only as the additive final term; the cross-partial $\partial^2\theta_{\text{exit}}/\partial\delta\,\partial L_1$ vanishes identically, so any interaction measured in this reduction is zero by construction.

A.2 The monotone-cascade lemma

$A(x, f) = b^2(1-2f) - L_1 f - x - \Delta C - \theta$ has $\partial A/\partial f = -(2b^2 + L_1) = -\Lambda < 0$ for all x. Under asynchronous strict best response at fixed θ and frozen competence, with ties not triggering defection: each defection lowers f, which strictly raises A for every organization, defectors included; a defector's advantage was positive at defection and only grows, so reversion is never a best response. And if the k-th cheapest organization is infeasible at the current f, every more expensive organization is also infeasible at that f. The cascade therefore has a unique closure, obtained by evaluating organizations in penalty order, independent of the asynchronous update sequence. ∎

A.3 The escape-ladder theorem

By A.2 evaluate the cascade in penalty order. Before any defection $f = 1$; after $k-1$ defections $f = 1 - (k-1)/N$, and the k-th cheapest organization's advantage is

$$A\big(y_k,\ 1 - \tfrac{k-1}{N}\big) = b^2\Big(-1 + \tfrac{2(k-1)}{N}\Big) - L_1\big(1 - \tfrac{k-1}{N}\big) - y_k - \Delta C - \theta = \vartheta(\theta) + (k$$

with $\vartheta(\theta) = -\theta - b^2 - L_1 - \Delta C$ and $\Delta = \Lambda/N$. The k-th defection is feasible iff this is positive, i.e. $y_k < \vartheta + (k-1)\Delta$. Since y is nondecreasing and the credit is linear, the cascade halts at the first violation, $K^*(\theta) = \min\{k : y_k \geq \vartheta + (k-1)\Delta\} - 1$ with $\min\varnothing := N+1$. The shared fraction reaches $f \leq \frac{1}{2}$ iff $K^* \geq \lceil N/2 \rceil$, i.e. iff every rung $k \leq \lceil N/2 \rceil$ is feasible, i.e. iff $\vartheta(\theta) > M_N$. ∎

Corollary (i): $y_k \equiv \mathbb{E}[x]$ gives $y_k - (k-1)\Delta$ maximal at $k=1$, $M_N = \mathbb{E}[x]$, reproducing A.1's exit. *(ii):* if $y_k - y_1 < (k-1)\Delta$ for all k then every bracket lies below y_1 except $k=1$; $M_N = y_1$. *(iii):* for θ with $\vartheta(\theta)$ between consecutive distinct values of the violation sequence, $K^*(\theta)$ is constant with $0 < K^* < \lceil N/2 \rceil$; by A.2 the corresponding mixed configuration is a fixed point, stable under the dynamics since no feasible defection or reversion remains — the staircase. *(iv):* set $k = \lceil qN \rceil$; by Glivenko–Cantelli and continuity of Q_x on compact sub-intervals of $(0, \frac{1}{2}]$, $y_{\lceil qN \rceil} \to Q_x(q)$ uniformly and $(k-1)\Delta \to q\Lambda$, giving Eq. (8); the supremum over the open lower end is controlled in application by the finite-N cutoff $q \geq 1/(N{+}1)$. ∎

A.4 The phase criterion

Exact finite-N criterion. From the theorem, the ladder binds beyond its first rung — $k^* > 1$ — iff $\exists k \geq 2$ with $y_k - y_1 > (k-1)\Delta$. Taking expected order statistics at their quantile positions $k/(N{+}1)$ yields the registered classifier, Eq. (9), evaluated per cell with no fitted boundary.

Continuum secant condition. With $q_0 = 1/(N{+}1)$ and $k - 1 \approx (q - q_0)N$, the binding condition becomes Eq. (10): some secant slope of Q_x measured from q_0 must exceed Λ, up to the finite-N factor $(N{+}1)/N$.

Quantile density as local diagnostic. If the secant condition holds on $[q_0, q]$, the mean value theorem supplies a point where $QD_x > \Lambda$; the converse fails in general, since a local density excess need not produce a secant excess from q_0 when the intervening quantiles are low. QD_x therefore locates where the credit line can be outrun; Eqs. (9)–(10) decide whether it is.

Lognormal analysis. $Q_x(q) = (\sigma_0^2/c_*^2)\, e^{2hz_q}$; $QD_x(q) = (\sigma_0^2/c_*^2)\, 2h\sqrt{2\pi}\; e^{2hz_q + z_q^2/2}$. As $h \to 0$, spacings vanish and Eq. (10) fails: tail regime. For the finite even-N designs simulated, all rungs sit at $q < \frac{1}{2}$, so as $h \to \infty$ every Q_x value on the ladder carries $e^{2hz_q} \to 0$ and Eq. (9) fails again: tail regime. In the continuum, $Q_x(\frac{1}{2}) = \sigma_0^2/c_*^2$ is h-invariant, and the high-h secant to the median approaches $2\sigma_0^2/c_*^2$; the return to the tail phase then requires $\sigma_0^2/c_*^2 < \Lambda/2$, which holds at the frozen parameters. At intermediate h the density is maximal at low quantiles and the secant condition holds for the simulated parameter values: ladder regime. Larger N lowers q_0 into the region where the $z_q^2/2$ factor inflates the density, the monotone N-dependence of the phase probability. ∎

A.5 Sweep composition

For a downward sweep $\{\theta_j\}$ with T evaluations per step and per-step escape hazard $\lambda(\theta_j)$, survival to the end of step j is $\exp\left(-T\sum_{m\le j}\lambda(\theta_m)\right)$ under the exponential reduction; the median exit is the first θ_j with cumulative hazard at least $\ln 2$. No parameter of this expression is fitted to sweep data; λ is estimated at fixed θ on independent trials (split-half validated, §5.2), and predictions are composed per population before aggregation (§5.3).

Appendix B — Simulation record

Consolidated simulator. `paper_xxvi_cost_of_returning.py`, master seed 20260718, flat repository placement per series convention, figures written to `outputs/` and referenced — not embedded — in the repository copy of this paper; the web edition embeds them at `/working-papers/images/cost-of-returning/`.

Build lineage. The model was developed in five registered cycles, each frozen before running and gated on the previous cycle's surviving claims. The consolidated simulator reproduces the final-confirmation results; the per-cycle scripts are retained in the repository for the audit trail.

Cycle	Question	Registered outcome
1. Mean-field prototype	Does competence decay widen the loop and shift only the exit?	Confirmed; model found analytically solvable, converting the cycle's results to closed form and exposing the additivity identity (§2.2)
2. D2 transplant	Do the closed-form thresholds survive finite heterogeneous agents?	Confirmed to tolerance, with a systematic, sign-consistent shallow residual on exits — flagged, not absorbed
3. Tail cycle	Is the residual an order-statistic (nucleation) effect?	Confirmed: realized-tail prediction reduced exit MAE ≈ 60%; N-dependence overpredicted — falsified, prompting the ladder
4. Ladder cycle	Does the ladder functional govern where the pure tail fails?	Confirmed in the registered $N \times h$ separating regime (ladder MAE 0.0089 vs tail 0.0145 vs mean 0.0523; median $k^* \approx 10$); staircases observed at predicted locations; strong log-dwell law failed (R^2 $0.9314 < 0.95$)
5. Final confirmation	Classification test of the quantile-ladder criterion; hazard measurement and composition	55/56 cells (98.2%), single error at $N = 500, h = 0.60$ where $P(k^* > 1) = 0.543$; held-out survival MAE 0.0538 (< 0.12 registered), pointwise peak 0.119 at the shallow edge; composition MAE 0.0029 population-conditional, 0.0050 new-population

Protocol notes. Thresholds are estimated by continuous crossing interpolation, not grid snapping, after cycle 3 established that grid quantization at step 0.025 masked the small-decay regime. All factorial and law comparisons use common random numbers (shared master seed and synchronized generator streams across cells and rebuilding laws). Hazard validation is split-half within population. Sweep composition is per-population in both arms: each population's hazard curve is composed into its own predicted exit, with the population-conditional test comparing per population and the new-population test comparing the median of the population-specific predictions against fresh draws (§5.3). Downward branches for exit measurement are initialized at full consolidation with competence at the analytic fixed point c_*, eliminating upward-history contamination identified in cycle 3. The recovery boundary band's $[0.1, 0.9]$ width was adopted after an initial narrower band proved too sparse for stable conditional estimation (§6).

Frozen parameters (base regime). $N = 20$ (phase and ladder studies sweep N to 2000); $b = 0.45$; $\sigma_0 = 0.12$; $h = 0.15$ (swept 0–0.8); $C_S = 0.50$; $C_I = 1.00$; $L_0 = 0.20$; $L_1 = 0.15$ (ratchet cells); $\rho = 0.003$; $\delta = 0.005$ (decay cells), hence $c_* = 0.375$; selection gain 8.0; switch rate 0.25; mutation floor 5×10^{-4}; dwell values $\{10, 30, 100, 300\}$. Full frozen tables, per-arm seed offsets, and registered tolerances are in the simulator header.

Appendix C — Claim-tier table

Claim	Tier	Note
Homogeneous thresholds, W_H decomposition, $c_* = \rho/(\rho+\delta)$	[R within the model]	Closed form, A.1
Entry invariance of competence decay	[R within the model]	Structural, A.1; reported as consistency check, not discovery
Homogeneous decay–liability additivity	[R within the model]	Algebraic identity; forecloses the interaction question at that layer
Monotone-cascade lemma; escape-ladder theorem; corollaries (i)–(iv)	[R within the model]	A.2–A.3
Staircase plateaus at predicted locations	[R within the model]	Registered simulation outcome (plateau–gap correlation $r = 0.761$)
Finite-N quantile-ladder criterion; secant continuum form; non-monotone h-dependence with its parameter condition; N-dependence	[R within the model]	A.4; classification 55/56 with no fitted boundary, single error on the phase boundary
Flat N-curve as tail-versus-kick cancellation	[R within the model]	Explains a falsified registered prediction
Exponential hazard reduction	[R within the model], scope-conditional	Registered criterion on overall held-out MAE; pointwise degradation at the shallow edge consistent with multi-stage escape
Hazard composition predicts swept exits	[R within the model]	No fitted dwell parameters; residual bias consistent with two same-sign effects, not separately identified
Band-conditional recovery gap, duration dependence	[R within the model]	Band defined from the fixed-law baseline, independent of the comparison; exploratory localization statistic
Conditional near-absorbing region	[R within the model], assumption-conditional	Holds under the named trainer-viability assumption, tested policy,

Claim	Tier	Note
		and horizon; strict absorption not established
Liability × decay interaction	Open registered prediction	Ladder approximation predicts $\approx -3.5 \times 10^{-4}$; below feasible dynamic resolution
All governance readings (§7.2)	[IP]	Transfer conditions stated; no empirical calibration claimed
Critical-mass properties of real epistemic infrastructures	[H] → empirical hypothesis	The case-study question the trainer-viability assumption isolates