# SELF-INDUCED DISTRIBUTION SHIFT IN AGENT MEMORY: WRITE-TIME RETENTION CRITERIA CANNOT BE OPTIMAL

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

An agent with bounded memory must decide, at write time, which of what it has just seen to keep, and every published retention criterion answers by scoring an item against an estimate of how much it will matter later — attention mass, predicted recall, a learned utility. Every such estimate is computed under a memory policy that is *not* the one the criterion is about to install. The usual reading of this is off-policy estimation error, a nuisance to be reduced with more rollouts. It is not: what an agent keeps determines which screens it opens next, which determines which facts it will later need, which is what the criterion was trying to score, so the criterion's own output moves the distribution its score was computed against and no amount of sampling under the old policy estimates a quantity defined under the new one. We formalise this as the induced-distribution operator $\Phi$, which sends a retention code to the rate-constrained optimum under the state law that code itself induces; we prove its fixed point exists by Brouwer and is unique under a contraction condition, identify the objective every write-time criterion actually optimises as exactly the *first iterate* of $\Phi$, and bound its error two-sidedly by one measurable quantity — the shift between the law assumed and the law induced — so that optimality at write time requires that shift to vanish, which requires memory not to affect behaviour. Because that argument is conditional on a contraction no deployment can verify, we add an unconditional minimax bound: over two instances no criterion can tell apart, every write-time rule pays at least $\sigma/2$ of performative risk, and $\sigma$ if deterministic, both attained. A second result prices *what* is kept rather than *when*: weighting an item by the irreversibility of the action it governs selects a *different retained set* at equal budget rather than a rescaled score, and whether it bites on a given corpus is an $O(n \log n)$ two-sort check needing no rollout and no model. *We claim nothing for the operator* — $\Phi$ is performative prediction's decoupled-risk map and the single-agent endogenous setting is performative RL's; ours is the rate-constrained feasible set, whose argmin *jumps* where a parametrised policy's moves smoothly. Measured, the two failures compose: being off-policy costs $0.407$ nats on our witness, and being off-policy *and* blind to recoverability — what a deployed criterion is — costs $0.903$. On $19$ real application flows walked on devices, replacing full history with the recency compaction that deployed systems ship moves a 7B agent's visitation law by $3.28\times$ its measured noise floor at a quarter budget, on $17$ of $19$ flows, with the mechanism visible in a revisit rate that jumps from $0.38$ to $0.497$. Of six pre-registered predictions three are supported, two partial and one fails outright; we withdraw a theorem of our own; and the fixed point these criteria fail to reach is itself *stable, not optimal*, leaving $24.0\%$ of the performative risk unclaimed.

## 1 INTRODUCTION

An agent with bounded memory must decide, at write time, which of what it has just seen to keep. Every retention criterion we know of answers by scoring an item against an estimate of its future value — attention mass, predicted recall, a learned utility — and every one of those estimates is computed under a memory policy that is not the one the criterion is about to install.



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The usual reading of this is off-policy estimation error: a nuisance, to be reduced with more rollouts or corrected with importance weights. It is not. What an agent remembers determines which screens it opens next, which determines which facts it will later need, which is what the retention criterion was trying to score. The criterion's output moves the distribution its own score was computed against, and no amount of sampling under the old policy estimates a quantity defined under the new one.

**What this paper establishes.** Write $\mu_q$ for the state law a retention code $q$ induces and $D_\omega(q;\mu)$ for its weighted decision distortion under law $\mu$. The operator $\Phi(q) = \arg\min_{q':R(q')\le R} D_\omega(q';\mu_q)$ — the rate-constrained optimum evaluated on the law that $q$ itself induces — has a fixed point, and the object every write-time criterion computes is $\Phi(q_0)$, a *single* application evaluated at the full-history law. Theorem 2 shows the gap between that and the fixed point is governed, from above and from below, by the law shift $\Sigma$ the criterion's own output induces. Corollary 1 then reads off the wall: a write-time criterion optimal on every instance forces $\mu_q$ not to depend on $q$, which is a memory that does not affect behaviour and is therefore worth nothing.

**What is not ours, stated first.** $\Phi$ is the decoupled-risk map of performative prediction (Perdomo et al., 2020), and the single-agent endogenous setting is performative reinforcement learning's (Mandal et al., 2023). An earlier version of this work claimed three differences from that literature; the first — that the shift is endogenous to a single agent with no responding population — is exactly performative RL's setting, and we withdraw it. What survives is the *feasible set*. Performative prediction optimises over a parametrised policy, whose argmin moves smoothly; we optimise over a rate-constrained code, whose argmin *jumps* as the budget boundary crosses an item. That is why solution stability, a mild regularity condition there, is a hypothesis that must be assumed and can fail here — Theorem 1 exhibits a two-code cycle when it does. The quantitative failure of §4 needs neither the operator nor the analogy.

**Contributions.** **(i)** The operator, its fixed point, and the identification of every write-time criterion's objective as its first iterate, with a two-sided error bound in one measurable quantity (§3). **(ii)** A minimax lower bound: on a two-instance class no criterion can distinguish, every write-time rule pays at least $\sigma/2$, and $\sigma$ if deterministic, with both bounds attained (§4). **(iii)** A separation result with a one-pass diagnostic: weighting by recoverability selects a *different retained set* at equal budget, not a rescaled score, and a practitioner can check it in one sort (§5). **(iv)** The loop measured on a real agent over device-walked flows, with the mechanism identified (§7).

**Two results go against us.** The fixed point these criteria fail to reach is *performatively stable*, not optimal, and we measured how much that costs rather than arguing about it. And a statement this work previously numbered as a theorem is withdrawn, because its hypothesis and its conclusion turn out to be mutually exclusive. Both are in §6.

## 2 Related Work: an Assumption Audit

We read related work as an audit rather than a survey: for each nearby result, the assumption it needs and the limit in which that assumption fails. Every result below is correct on its own terms and none is refuted here; what we are looking for is the *shared* premise — that the value of retaining an item can be scored against a trajectory distribution the retention decision does not change — because it is invisible when each method is read alone and load-bearing in all of them at once. Table 6 in Appendix E is the row-by-row version, with the assumption each result needs, the limit in which it fails, and the role it plays here.

**Retention criteria score an item against a distribution they then move.** Decision-centric rate–distortion memory (Zou et al., 2026) assumes a stationary evaluation law and an exchangeable per-decision loss; GUI history and context managers (Zhou et al., 2026; Wu et al., 2026; Liu et al., 2026a; Zeng et al., 2026; Shi et al., 2026; Gu et al., 2026) and general agent memory architectures (Packer et al., 2023; Park et al., 2023; Shinn et al., 2023; Zhao et al., 2024; Wang et al., 2024; Xu et al., 2025; Gutiérrez et al., 2024; Chhikara et al., 2025) treat retention value as a property of the item and the goal. *All of them score candidates on trajectories generated with a different memory than the one they ship*, which is the premise this paper removes. The audited rule — recency under a token budget, which is what deployed compaction ships — is the one §7 runs.

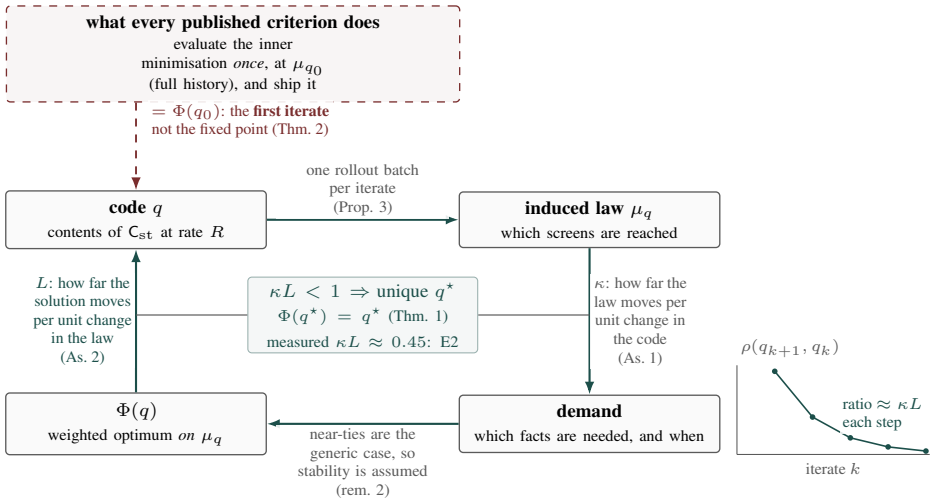



**Figure 1. The self-reference wall, with the quantity each step costs.** A code induces a state law; the law fixes which facts are demanded and when; the optimum of that demand is a code. The two Lipschitz constants are what the cycle is made of — $\kappa$ is how far the *world* moves per unit change in the code (Assumption 1), $L$ how far the *solution* moves per unit change in the world (Assumption 2) — and their product must be below one for the fixed point to be a point at all (Theorem 1). *The dashed entry is what every published criterion does*: evaluate the inner minimisation once, at the law full history induces, and ship the result. Theorem 2 identifies that as $\Phi(q_0)$ and bounds its distance from $q^\star$ by $\kappa L/(1-\kappa L)$ times the diameter of the code set — a quantity that *grows* with retention strength, which is the direction that makes the failure self-defeating rather than merely inconvenient: *the more aggressively a system compresses, the less trustworthy the estimate that licensed the compression*. The inset is the trace: each step shrinks the change in the code by roughly $\kappa L$, so the repair costs one rollout batch per iterate (Proposition 3).

**Two things we take as input rather than claim.** That compaction moves an agent's trajectory has been measured, by someone else (Liu et al., 2026c); that the standard retention surrogates fail is documented with a layer-agnostic bound in the compaction literature (Colaco & Lahjouji, 2026). *Both findings are theirs*, we are explicit at the point of use, and we refine the distortion rather than dispute either. Attention- and cache-level compression (Li et al., 2024a; Zhang et al., 2023; Cai et al., 2024; Tu et al., 2024; Xiao et al., 2024; Chen et al., 2025), learned compressed carriers (Mu et al., 2023; Ge et al., 2024; Chevalier et al., 2023; Jiang et al., 2023; Li et al., 2024b) and recurrent or infinite-context summaries (Dai et al., 2019; Rae et al., 2020; Bulatov et al., 2022; Munkhdalai et al., 2024) are orthogonal and compose with this: next-token fidelity is an unweighted average over a *fixed* stream, so both of our failures are inherited rather than addressed.

**Performative prediction, which owns the operator.** That a predictor's deployment changes the distribution it is scored against is performative prediction's founding observation (Perdomo et al., 2020), and the single-agent endogenous case — one learner whose policy changes the environment's own dynamics, no responding population, nothing strategic — is performative RL's setting (Mandal et al., 2023; Li & Wai, 2022). *We take the operator and the vocabulary and claim neither*: $\Phi$ is their decoupled-risk map and Theorem 1 their stability theorem. An earlier version of this work claimed three differences; the first was that the shift is endogenous to a single agent, and we withdraw it. What survives is the *feasible set*, load-bearing in Theorem 1: a rate-constrained code's argmin is a set-valued object that *jumps*, which is why uniqueness needs an assumption here that is mild there. That performatively stable points need not be performative optima is known (Miller et al., 2021), and §6 measures what it costs on our own witness. The mixed endogenous–exogenous case is formalised by Lee & Zrnic (2026); we are its fully endogenous edge, which is a scoping statement against us.

**Shift from the learner's own actions, and pricing what cannot be undone.** Compounding error under a learner's own state distribution is imitation learning's subject (Ross et al., 2011; Kearns & Singh, 2002), and bisimulation-style abstraction faces the same shape one level down, since an abstraction is judged on the visitation it induces (Zhang et al., 2021). The difference in both cases is where the channel sits and what the feasible set is: there the shift is a property of two *given* policies, here it is a function of the object being optimised, which makes it a fixed-point problem.

Reversibility-aware RL prices irreversibility in the reward or filters at act time (Grinsztajn et al., 2021; Alshiekh et al., 2018); *applying that pricing to retention rather than to behaviour changes which items are kept*, which reward shaping cannot do (Theorem 4). Finally, no long-horizon GUI instrument (Liu et al., 2026b; Tongyi-MAI, 2026; Chen et al., 2026; Wang et al., 2026; Li et al., 2026) reports accuracy *stratified by* recoverability of the pending action, the axis the theory separates along; ST-WebAgentBench (Levy et al., 2024) annotates the irreversible actions but scores policy violations rather than retention, and is the cheapest data on which to close that gap.

## 3 The Operator, and What a Write-Time Criterion Optimises

**Definition 1** (Decision distortion and the irreversibility weight)**.** Let $D_\omega(q;\mu) = \mathbb{E}_{s\sim\mu}\big[\omega_s\, d(q,s)\big]$, where $d$ is the decision distortion a code $q$ causes at state $s$ and $\omega_s = 1+\eta(1-\beta_s)$ is the *irreversibility weight*: $\beta_s \in [0,1]$ is the recoverability of the action taken at $s$ — the probability that a wrong step there can be undone — and $\eta$ prices the cost of a step that cannot. A code is feasible if its rate $R(q)$ is within budget.

**Definition 2** (Induced-distribution operator)**.** $\Phi(q) = \arg\min_{q':R(q')\le R} D_\omega(q';\mu_q)$: the rate-constrained weighted optimum evaluated on the law that $q$ itself induces.

**Theorem 1** (Existence and uniqueness)**.** *Work on the convex relaxation of the code space, which is compact and convex.* ***(1) Existence.*** *If $q \mapsto \mu_q$ is continuous and the constrained optimum is attained and continuous in $\mu$, then $\Phi$ is a continuous self-map of a compact convex set and Brouwer's theorem gives at least one fixed point $q^\star$.* No contraction is needed. ***(2) Uniqueness and rate.*** *Under regularity and argmin-stability with $\kappa L < 1$, $\Phi$ is a $\kappa L$-contraction, the fixed point is unique, and $q_{n+1} = \Phi(q_n)$ converges geometrically. Without argmin-stability uniqueness can fail: Appendix A exhibits a two-code cycle with $\kappa L > 1$, so the condition is not vacuous.*

The cycle is where the feasible set earns its keep. In performative prediction the argmin over a parametrised policy moves continuously with the law, and solution stability is mild. Over a rate-constrained code it does not: an infinitesimal change in the law can move the budget boundary past an item and change the retained *set*.

**Theorem 2** (The off-policy objective is the first iterate, and the law shift governs its error)**.** *Let $q_0$ be the full-history code and $\widehat{q} = \Phi(q_0)$ — the solution every write-time criterion computes: one application of $\Phi$, evaluated on the law that full history induces. Write $\mathrm{gap}(\widehat{q}) = D_\omega(\widehat{q};\mu_{\widehat{q}}) - D_\omega(q^\star;\mu_{q^\star})$ and $\Sigma = \|\mu_{q_0} - \mu_{\widehat{q}}\|_{\mathrm{TV}}$.* ***(1) Upper bound.*** *Under the conditions of Theorem 1(2), $\mathrm{gap}(\widehat{q}) \le \frac{2\Lambda L}{1-\kappa L}\,\Sigma$, where $\Lambda$ bounds $D_\omega$'s Lipschitz dependence on the code. In particular $\Sigma = 0$ gives $\mathrm{gap} = 0$.* ***(2) Matching construction.*** *There is a family of instances, indexed by retention strength $\sigma = 1 - R/R_0$, on which $\mathrm{gap}(\widehat{q}) \ge c\,\Sigma(\sigma)$ for some $c > 0$, with $\Sigma(\sigma)$ strictly increasing and $\to 1$ as $\sigma \to 1$.*

*The criterion is least reliable exactly where it is most needed*: both the bound and the construction scale with the law shift, and the law shift is what aggressive compression produces.

*Remark* 1 (We do not assert a general lower bound). An earlier version stated part (2) as holding on *every* instance. That is false: where the off-policy solution happens to coincide with the fixed point the gap is zero, and part (1) says those are exactly the instances with no law shift. What a limits paper may claim is the pair above — the shift controls the error from above, and there exist instances where it controls it from below — so §7 measures $\Sigma$ and the gap together and reports their relation rather than the gap alone.

**Corollary 1** (Write-time optimality requires memory to be inert)**.** *Let a* write-time criterion *be any retention rule that is a measurable function of the history and the goal, evaluated under a law it does not itself induce. If such a rule is optimal on every instance, then $\mu_q$ does not depend on $q$: memory does not affect behaviour. Since a memory that does not affect behaviour is worth nothing, no useful write-time criterion is optimal in general. On a* single *instance a write-time criterion may be optimal by coincidence, and Theorem 2(1) says exactly when — when the law shift it induces is zero.*

## 4 How Much It Costs on One Deployment

Corollary 1 carries a universal quantifier, and a practitioner is entitled to ask what the wall costs on *one* deployment. The answer needs neither the operator nor the performative-prediction analogy.

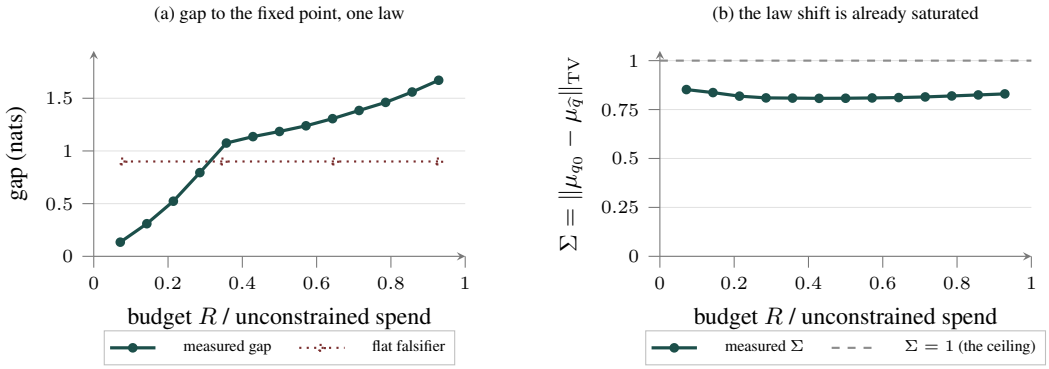



**Figure 2. The law shift is real, and this sweep sits inside its plateau (Theorem 2). (a)** Gap between the first iterate from full history and the fixed point, swept over the retention budget, 13 points. *The panel carries the smaller of two readings.* The theorem's gap — each code scored in the world it induces — runs 0.765 to 6.573 nats; plotted is the *common-law* reading, both codes at $\mu_{q^\star}$, which runs 0.13 to 1.67 and charges the off-policy code for nothing its own law would have spared it. Both are monotone over all 13 levels and the flat falsifier fires on neither. **(b) is why (a) is not the test it looks like.** The theorem's axis is $\sigma = 1 - R/R_0$ against *full history*, and every point here has true $\sigma \in [0.93, 0.99]$ — a window of width 0.06 at the extreme-compression end, across which the measured $\Sigma$ is flat at 0.807–0.852 and *non-monotone*, exactly as $\Sigma(\sigma) \to 1$ predicts once $\sigma$ is that large. The sweep is inside the saturated region and cannot test monotonicity in $\sigma$ at all, which is why §7 reports $\Sigma$ beside the gap and why the real-agent sweep, which is *not* saturated, is the one that carries the growth claim.

**Theorem 3** (Minimax lower bound)**.** *Fix $\sigma \in [0, 1]$. There is a class $\mathcal{I}(\sigma)$ of two instances sharing the same goal, the same candidate set, the same rate constraint and the same full-history law $\mu_{q_0}$ — so that* no criterion evaluated under $\mu_{q_0}$ can tell them apart *— on which every write-time criterion $\widehat{q}$, deterministic or randomised, satisfies*

$$\sup_{I \in \mathcal{I}(\sigma)} \Big[ \mathrm{PR}_I(\widehat{q}) - \min_q \mathrm{PR}_I(q) \Big] \;\geq\; \tfrac{1}{2}\sigma,$$

*and $\geq \sigma$ if $\widehat{q}$ is deterministic. Every code in the class that does not retain the routing item induces law shift exactly $\sigma$; the one that does induces zero shift and pays $(1 + \sigma)/2 \geq \sigma$.* Both bounds are attained*, so the constant is not an artefact, and on each instance* separately *the performative optimum is reached by a feasible code — the excess is the price of deciding at write time and of nothing else.*

## 5 The Separation, and a One-Pass Diagnostic

If the weight $\omega$ merely rescaled a score it would change the reported number and not the decision. It does not.

**Theorem 4** (The minimisers differ at equal budget)**.** *Under the greedy characterisation, the unweighted rule retains items in decreasing $\Delta_i/c_i$ and the weighted rule in decreasing $\omega_i\Delta_i/c_i$, for marginal value $\Delta_i$ and cost $c_i$. The two retained sets at a common budget $R$ differ if and only if there is an* inversion straddling the budget*: a pair $i, j$ with $\Delta_i/c_i > \Delta_j/c_j$ and $\omega_i\Delta_i/c_i < \omega_j\Delta_j/c_j$, with the budget boundary between them in one of the two orders. Such a pair exists whenever $\omega$ is non-constant near the boundary — whenever exposure and value are not comonotone there. The required exposure ratio is $O(1)$: a single irreversible decision among reversible ones suffices.*

**Corollary 2** (Weighting is not a reparameterisation)**.** *At any rate admitting a straddling inversion, no unweighted rule attains the weighted optimum. A rescaling would preserve the retained set and change only the reported number; Theorem 4 exhibits the rates at which the set itself changes, and says where to look — at the items near the budget boundary, not at the frontier gap.*

The condition is checkable in one pass over the item list: sort twice, once by $\Delta_i/c_i$ and once by $\omega_i\Delta_i/c_i$, and count pairs that straddle the budget cut in opposite orders. §7 runs exactly this on a real interface, and Appendix A gives the exposure-average form that removes the cost term.

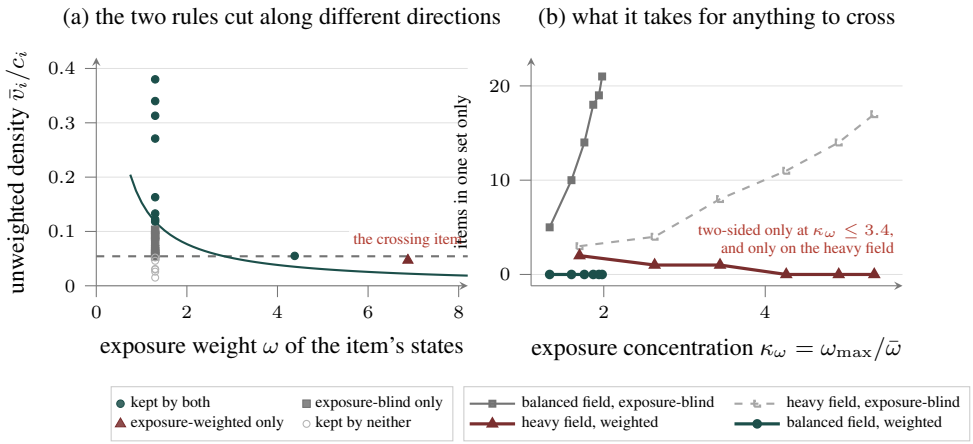



**Figure 3. Why the argmin sets differ, and what it takes for them to differ in both directions.** Measured. **(a)** Every item at one budget, with the two cuts drawn: the exposure-blind rule thresholds the density $\bar{v}_i/c_i$, a horizontal line, and the weighted rule thresholds $\omega \cdot \bar{v}_i/c_i$, a hyperbola. The band between them is the theorem's claim, realised by the item with the lowest density and the highest exposure — kept by the weighted rule, dropped by the blind one. **(b) is the part we did not expect, and it qualifies the theorem.** Theorem 4's picture is *two-sided*: items must cross in both directions. The condition it names — exposure and value not comonotone — holds in every cell of (b) and does not predict which cells are two-sided. On a *balanced* exposure field the weighted support is a strict *subset* of the blind one at every budget and every $\eta$ up to a $13.7\times$ spread in $\omega$: the weighted rule keeps *fewer* items, never different ones, satisfying any Jaccard the theorem could ask for while refuting its picture. What decides it is $\omega$'s *concentration* $\kappa_\omega = \omega_{\max}/\bar{\omega}$, not its spread, and the two-sided regime is a *window*, $\kappa_\omega \in [1.7, 3.4]$ here, bounded above as well as below. A GUI's exposure field is the concentrated kind; a text corpus's is closer to the balanced one (Appendix F).

## 6 Two Results Against Us

**The fixed point is stable, not optimal, and that raises the wall.** $q^\star$ is what the performative-prediction literature calls a *performatively stable* point: optimal against the law it itself induces. The quantity a designer pays is the performative risk $\mathrm{PR}(q) = D_\omega(q; \mu_q)$, whose minimiser need not be a fixed point of $\Phi$ at all (Miller et al., 2021). We measured the difference on our own witness rather than arguing about it: at the reference budget $\mathrm{PR}(q^\star) = 8.95$ against a direct minimum of $6.80$, so **stability leaves** $24.0\%$ **of the performative risk on the table**, and the minimiser is demonstrably not a fixed point (residual $0.054$ against $q^\star$'s $0.000$). Across the retention sweep the stable-to-optimal share is between $25\%$ and $34\%$ of the designer's total shortfall and never vanishes; a deployable code recovers $43.9\%$ of it.

*This raises the wall rather than lowering it.* The claim is that no write-time criterion is optimal, and the measurement supplies a second, independent reason. Corollary 1 is unaffected, but two things must be read with the distinction in place: Theorem 2 bounds the distance to *stability*, which is the object a practitioner can actually iterate to; and any round count we price is a *lower* bound on the cost of crossing the wall, because reaching the performative optimum is a non-convex problem in $q$ that no iteration of $\Phi$ solves.

**A theorem of ours, withdrawn.** An earlier version stated that the separation grows with the number of irreversible decision points, under the hypothesis that the distortion is additive over steps. Proposition 2 (Appendix A) shows that additivity holds *if and only if* $\beta_t = 1$ for every $t$ — and at any such $t$ the weight $\omega_t = 1 + \eta(1 - \beta_t)$ equals 1. The hypothesis is satisfiable only on instances where the weighting does nothing and the separation is identically zero: *premise and conclusion are mutually exclusive*, and guarding the statement does not repair the label. We withdraw it.

*What survives is worth keeping and is not a theorem.* On the multi-step witness the separation as a function of the count of irreversible steps is a *window*: exactly $0.000$ at both ends, because $\omega$ is constant at both, with an interior peak of $0.316$. Additivity is the left edge of that window, so the growth claim is the *tangent* to the separation there — the correct local reading of a quantity that is zero at the point the reading is taken, rises away from it, and falls again before the other end.

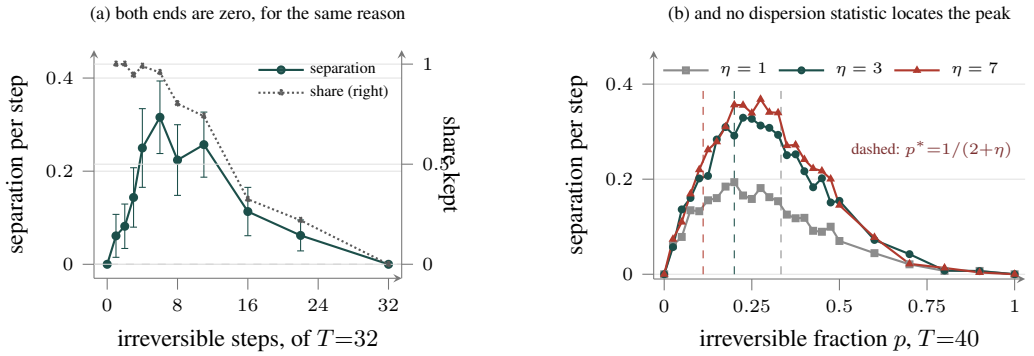



**Figure 4. The separation is a window, and the withdrawn statement is one of its edges.** Measured on the multi-step witness; $T$=32, 40 instances per cell in (a), 300 in (b). **(a) is the withdrawn statement's literal axis** — the *count* of irreversible decision points at fixed horizon — and it had not been drawn. The separation rises to a peak at 6 of 32 and returns to *exactly* 0.000 at 32 of 32, the value it takes at 0 of 32. *The two zeros are the result*: $\omega \equiv 1$ when nothing is irreversible and $\omega \equiv 1 + \eta$ when everything is, and a constant $\omega$ rescales the weighted objective without re-ordering it. Proposition 2 is the left end seen algebraically; the right end is not in the paper's theory at all. The dotted series is Proposition 4's share on the same sweep: the per-state weight captures 1.00 of the attainable gain at one irreversible step and 0.32 at sixteen. **(b) is a test with no free constant, and it fails.** If the deciding quantity were the exposure field's normalised dispersion, the peak would sit at $p^\star = 1/(2 + \eta)$, the argmax of $\mathrm{CV}(\omega)$ (dashed verticals). The prediction moves *left* in $\eta$ and the measured peak moves *right*; bootstraps exclude it at $\eta = 1$ and $\eta = 7$. *We report the refutation because the alternative is an unstated conjecture*, and because it says what the window is not: no normalised statistic of the exposure field locates its interior.

## 7 The Loop, Measured

Every $\Sigma$ above is an algebraic consequence of a $q \mapsto \mu_q$ map we wrote down ourselves, and a referee is right that this is not evidence about a system. This section measures it on one, after pricing the two failures separately on the witness where both are computable.

### 7.1 Six retention codes at one budget

Table 1 prices the two failures the theory names against a code that is subject to neither. *It also corrects a caption of our own.* An earlier version observed that the weighted and unweighted objectives rank these six *codes* in the same order and read that as "the inversion Theorem 4 permits is not reached at this budget". That is a different object: the theorem's inversion is between *items* at the budget boundary, and a common ranking of six codes neither implies nor is implied by it. Run properly (Appendix C), this configuration carries $3.67 \pm 0.20$ straddling inversions per draw, in $68.3\%$ of draws, with the two rules' own retained sets overlapping $0.692 \pm 0.017$ rather than 1.

### 7.2 A real agent over device-walked flows

A 7B open-weights agent at temperature 0 is given a real application goal and, at each step, opens one of that flow's real screens or finishes. The screens, their contents and their goals are a device-walked census — 19 flows, 158 pages — so we author no transition and no reward, and the branching is the model's. What the agent remembers between steps is the retention code: **full** keeps the entire history; **recency** keeps the newest observations fitting a budget $R$, which is what deployed compaction ships. $\Sigma$ is the total-variation distance between the two arms' empirical laws over which screens get visited. The budget is swept rather than chosen, because a single tight budget producing a large $\Sigma$ would be us picking the answer, and the noise floor is the *same* code re-run, so it is measured rather than assumed.

**Table 1. What each failure costs, separately and together.** Six codes at one budget; 2000 witness draws, $n$=24 items, $m$=8 states, $\lambda$=2.0, $R$=0.6× the unconstrained spend. *Every distortion is evaluated on one law*, the fixed point's — scoring each code under the law it induces itself mixes two laws into one number, which was the defect in an earlier version of our frontier sweep. This configuration is the one in which Assumption 2 is verified ($\kappa L_{\max} = 0.519$ over 400 law pairs), so the last row is the unique fixed point of Theorem 1. **The two errors compose.** Being off-policy costs 0.407 nats here; a deployed criterion is off-policy *and* blind to recoverability, and that costs 0.903 — more than double, not the sum of two small things. **The separation is in the sets.** No code overlaps the fixed point's retained set by more than 0.381, and the weighted off-policy code, which differs from the fixed point by *iteration alone*, overlaps $0.381 \pm 0.012$: weighting is not a monotone rescaling. **One result we did not expect** — attention-score retention (0.795) beats decision-relevance (0.903) despite overlapping the target set less, which is Figure 15's finding again: precision in a recoverability-blind score buys nothing. *Full history is not budget-feasible; it is the reference law, scored at its own rate. Cells are means over 2000 draws with Monte-Carlo 95% intervals; intervals below 0.0005 print as 0.000. Appendix C carries the budget sweep behind this single column.

| Code | Rate (tok/step) | $\widehat{D}$ (nats) | $\widehat{D_\omega}$ (nats) | Excess unweighted | $\widehat{\Delta(R)}$ weighted | Jaccard vs fixed point |
|---|---|---|---|---|---|---|
| Full history (reference law)* | 41.97 ±0.13 | 0.000 ±0.000 | 0.000 ±0.000 | −3.448 ±0.023 | −7.421 ±0.099 | 0.070 ±0.002 |
| Recency window | 2.07 ±0.05 | 3.796 ±0.029 | 8.554 ±0.187 | 0.348 ±0.030 | 1.132 ±0.104 | 0.152 ±0.012 |
| Attention-score retention | 2.04 ±0.06 | 3.652 ±0.024 | 8.216 ±0.192 | 0.204 ±0.032 | 0.795 ±0.114 | 0.247 ±0.010 |
| Decision-relevance (off-policy) | 1.49 ±0.05 | 3.698 ±0.031 | 8.324 ±0.194 | 0.250 ±0.030 | 0.903 ±0.104 | 0.363 ±0.016 |
| Weighted, off-policy (1 iterate) | 1.49 ±0.05 | 3.556 ±0.026 | 7.828 ±0.163 | 0.107 ±0.020 | 0.407 ±0.071 | 0.381 ±0.012 |
| **Weighted, at the fixed point** | **1.63** ±0.06 | **3.448** ±0.023 | **7.421** ±0.099 | **0.000** ±0.000 | **0.000** ±0.000 | **1.000** ±0.000 |

**Table 2. The law shift on a real agent, against a measured noise floor.** The falsifier row is exact: an unbudgeted **recency** arm *is* the full-history code, and it returns the floor to three decimals, which checks that the arms are not leaking into one another.

| retention budget | $\Sigma$ | 95% CI over flows | × floor | flows above floor | revisit rate |
|---|---|---|---|---|---|
| floor (**full** re-run) | 0.101 | [0.054, 0.153] | 1.00 | — | 0.377 |
| $R = 1.00$ (falsifier) | 0.101 | [0.053, 0.151] | 1.00 | 1/19 | 0.382 |
| $R = 0.75$ | 0.146 | [0.096, 0.196] | 1.45 | 8/19 | 0.380 |
| $R = 0.50$ | 0.198 | [0.142, 0.257] | 1.97 | 13/19 | 0.379 |
| $R = 0.25$ | **0.330** | [0.261, 0.400] | **3.28** | **17/19** | **0.497** |

**The loop is not empty on a real agent.** At $R = 0.25$ the law shift is 3.28× the floor, the two confidence intervals do not overlap, and 17 of 19 flows move individually.

**And the shift grows with the compression that motivates it** — the claim the synthetic witness could not make, because it saturates at $\sigma \in [0.93, 0.99]$ where $\Sigma$ has plateaued (Fig. 2b). Paired within flow, tightening from 0.75 to 0.25 raises $\Sigma$ on 18 of 19 flows and lowers it on none (sign test $p < 10^{-4}$); 0.50 to 0.25 is 15 up against 1 down ($p = 0.0005$). Adjacent quarter-steps are noisier ($0.75 \to 0.50$: 11 up, 4 down, $p = 0.12$), so what is established is the trend across the range and not each step, and we report it at that resolution.

**The mechanism is visible and it is the one the theory names.** The share of steps spent reopening a screen already seen is flat at $\approx 0.38$ across **full** and the loose budgets, then jumps to 0.497 at $R = 0.25$: a code that drops the record of what was already done sends the agent back to do it again, and the states it revisits are the states the retention criterion will next be scored on. On one bill-payment flow the full-history agent walks the payment once and finishes in nine steps; the same agent under a quarter budget cycles between the payee list and the amount screen until the step limit, having forgotten it had already been there.

### 7.3 THE TWO-SORT DIAGNOSTIC ON A REAL INTERFACE

The condition of §5 is run on a device-walked census of 120 pages across 15 application flows, each page recording the fact it carries and whether a later step of the same flow demands it. Recoverability is human: two annotators, 1,200 labelled one-step deviations, with a calibration round fixing the probe's error bar at $\kappa = 0.85$. The verdict is DIFFERS, on every draw. Under a recency scorer — the rule deployed compaction ships — the two objectives retain sets differing by 13.9 of 120 items at $R = 0.6$, Jaccard 0.841, with $313 \pm 11$ straddling inversions; under the census's own demand column as the value the numbers fall and the verdict does not (6.8 items, Jaccard 0.915, $62 \pm 3$ straddles). *On a real interface the weighted and unweighted rules do not retain the same memory*, and the items they disagree about are enumerated rather than counted.

**The control, and what it takes away from the paragraph above.** At $n = 120$ the boundary ratio is about 1.0 — what a ratio between adjacent entries of a long sorted list should be — so a DIFFERS verdict is *cheap* at this scale and the binary answer alone proves little. We ran three counterfactual exposure fields beside the human one (Appendix F): **the human labels are not what carries the verdict**, which is why it is reported with straddle counts and enumerated disagreements rather than as a binary.

### 7.4 WHICH WEIGHT TO SHIP

The per-state exposure weight $\omega_s = 1 + \eta(1 - \beta_s)$ is horizon-blind: it prices the step at $s$ and not the steps that $s$ makes unreachable. A position-aware weight $1 + \eta(1 - \beta_t)(T - t)$ is the next term of the same series. Against the same oracle on the same instances at the same rate, the position-aware rule holds 0.94 to 0.98 of the attainable gain at $T \geq 64$, where the per-state weight holds 0.39 to 0.56 (Figure 6b). *Both are truncations of one series and the cheaper one is not the one to ship at long horizons.* Neither is a modelling choice: Proposition 5 differentiates the multi-step loss and recovers $1 + (1 - \beta_t)G_t/g_t$ exactly — verified against the recursion to $5.7 \times 10^{-8}$ — so $\eta$ is the instance's tail-to-step value *ratio* and not a free parameter, and the two weights are that derivative's zero- and first-order truncations.

## 8 THREATS TO VALIDITY

**The contraction modulus is synthetic**: we chose the witness's $\kappa$ when we chose the witness. Theorem 1(1) does not need it — existence is Brouwer — but the convergence rate and the uniqueness claim do, and no deployed system's modulus is measured here. **One agent, one scale**: a single 7B open-weights agent at temperature 0, over a census that is a convenience sample of consumer applications and not a probability sample of anything. The page-level demand column was walked by one annotator and so has no inter-annotator ceiling; the recoverability labels, which carry the weight, have two annotators and a measured spread of 0.042 against the probe's 0.159 (Appendix G). **The diagnostic is cheap to satisfy at this corpus size**, hence reported with its control and straddle counts rather than as a binary. **And we measure distance to stability, not to optimality**: Theorem 2 bounds the gap to $q^\star$, and §6 measures $q^\star$ itself 24.0% short of the performative optimum on our witness. We have not measured that shortfall anywhere else. Appendix I states the scope in full, answers three objections, and lists what remains owed.

## 9 DISCUSSION AND CONCLUSION

A retention criterion is scored on evidence its own output will change. That is not an estimation problem with a better estimator behind it: optimality at write time requires the memory to be inert, and an inert memory is worthless. What a practitioner can do instead is concrete. The gap is governed by a measurable quantity — the law shift between the assumed and the induced distribution — which on a real agent, under the compaction rule that actually ships, reaches 3.28 times its own noise floor. *We would ask a practitioner to run the two-sort check before tuning a retention score*, since the weighting that follows from irreversibility is a different retained set rather than a rescaling, and to price the position-aware weight at long horizons. What we would not conclude is that the fixed point is the target: it is what an iteration reaches, 24.0% short of what a designer pays for.

### REPRODUCIBILITY STATEMENT

Every number in this paper is written by a named analyzer to a result file, and every deciding read-out rule was frozen before the run it decides. Appendix J is the registration index. All experiments except one are seeded, offline and model-free; the single exception is the real-agent loop of §7, which calls a frozen open-weights model at temperature 0 and whose per-flow trajectories are shipped so that the law shift can be recomputed without re-running the model. The device-walked census, the recoverability labels and their calibration round, the witness constructions, the analyzers and the result files are in the supplementary material.

### ETHICS STATEMENT

The recoverability labels were produced by the authors rather than by recruited annotators; no human-subjects data was collected and no personal data appears in the census, whose pages are application screens walked on the authors' own devices with no account content. The agent experiment operates a frozen open-weights model over recorded screens and takes no action against a live service. The finding — that a memory compression rule changes which states an agent reaches — is a limitation of current retention practice and carries no group-differential harm we are aware of. We have read the ICLR Code of Ethics and adhere to it.

### AI USE STATEMENT

We report our use of generative AI tools against the categories of the ICLR 2027 AI Policy for Authors. *Our use was confined to preparing the document.*

*Tasks in the required-disclosure list: none.* Generative AI was not used to develop the theoretical model, to formulate or prove the theorems, to propose or refine the hypotheses, to design the methodology or the experiments, to set or choose any pre-registered threshold, to implement any method, to walk or label the census, or to interpret the results. A frozen open-weights model is the *subject* of the experiment in §7, not a tool used to produce it. Every number is computed by an analyzer the authors wrote, from the result files the runs produced. Translation and qualitative or thematic data analysis are not applicable.

*Tasks in the recommended-disclosure list.* Generative AI was used on the manuscript itself: it condensed a longer version into the present conference format, which involved drafting and rewriting substantial parts of the exposition; it edited for readability, formatted tables and references, and assembled the anonymised supplementary archive. It was not used to discover the research topic, identify the gap, select the literature, or decide what the paper claims — in particular, the two withdrawals of §6 are the authors'.

*Verification.* Every quantitative statement was checked against the result file it cites, and every registered outcome against the read-out rule frozen before the run. We have reviewed all AI-assisted work and take responsibility for the final content of this paper, including its text, claims and artifacts.

## REFERENCES

Mohammed Alshiekh, Roderick Bloem, Ruediger Ehlers, Bettina Koenighofer, Scott Niekum, and Ufuk Topcu. Safe reinforcement learning via shielding. In *AAAI Conference on Artificial Intelligence*, 2018.

Aydar Bulatov, Yuri Kuratov, and Mikhail Burtsev. Recurrent memory transformer. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Zefan Cai, Yichi Zhang, Bofei Gao, Yuliang Liu, Tianyu Liu, Keming Lu, Wayne Xiong, Yue Dong, Baobao Chang, Junjie Hu, and Wen Xiao. Pyramidkv: Dynamic kv cache compression based on pyramidal information funneling. *arXiv preprint arXiv:2406.02069*, 2024.

Kung-Hsiang Chen et al. Gui-kv: Efficient gui agents via kv cache with spatio-temporal awareness. *arXiv preprint arXiv:2510.00536*, 2025.

Sen Chen, Tong Zhao, Yi Bin, Fei Ma, Wenqi Shao, and Zheng Wang. D-gara: A dynamic benchmarking framework for gui agent robustness in real-world anomalies. In *AAAI Conference on Artificial Intelligence*, 2026. arXiv:2511.16590.

Alexis Chevalier, Alexander Wettig, Anirudh Ajith, and Danqi Chen. Adapting language models to compress contexts. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. Mem0: Building production-ready ai agents with scalable long-term memory. *arXiv preprint arXiv:2504.19413*, 2025.

Ashwin Gerard Colaco and Nada Lahjouji. What to keep, what to forget: A rate–distortion view of memory compaction in llms and agents. *arXiv preprint arXiv:2607.08032*, 2026.

Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc V. Le, and Ruslan Salakhutdinov. Transformer-xl: Attentive language models beyond a fixed-length context. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019.

Tao Ge, Jing Hu, Lei Wang, Xun Wang, Si-Qing Chen, and Furu Wei. In-context autoencoder for context compression in a large language model. In *International Conference on Learning Representations (ICLR)*, 2024.

Nathan Grinsztajn, Johan Ferret, Olivier Pietquin, Philippe Preux, and Matthieu Geist. There is no turning back: A self-supervised approach for reversibility-aware reinforcement learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. Reversibility estimated self-supervised; RAE penalises irreversible transitions in the reward, RAC filters them at act time.

Yingjie Gu, Wenjian Xiong, Liqiang Wang, Pengcheng Ren, Chao Li, et al. Fsfm: A biologically-inspired framework for selective forgetting of agent memory. *arXiv preprint arXiv:2604.20300*, 2026.

Bernal Jiménez Gutiérrez, Yiheng Shu, Yu Gu, Michihiro Yasunaga, and Yu Su. Hipporag: Neurobiologically inspired long-term memory for large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Huiqiang Jiang, Qianhui Wu, Chin-Yew Lin, Yuqing Yang, and Lili Qiu. Llmlingua: Compressing prompts for accelerated inference of large language models. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.

Michael Kearns and Satinder Singh. Near-optimal reinforcement learning in polynomial time. *Machine Learning*, 49:209–232, 2002.

Jaewook Lee and Tijana Zrnic. Partially performative prediction. *arXiv preprint arXiv:2606.07890*, 2026. Endogenous and exogenous shift in one framework; online analogues of performative stability and optimality; when repeated retraining adapts.

Ido Levy, Ben Wiesel, Sami Marreed, Alon Oved, Avi Yaeli, and Segev Shlomov. ST-WebAgentBench: A benchmark for evaluating safety and trustworthiness in web agents. *arXiv preprint arXiv:2410.06703*, 2024.

Qiang Li and Hoi-To Wai. State dependent performative prediction with stochastic approximation. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022. Distribution shift with state, i.e. the induced law depends on the trajectory of deployed models and not only on the current one.

Yang Li, Yuchen Liu, Haoyu Lu, Zhiqiang Xia, Hongzhen Wang, Kaiyang Han, Changpeng Yang, Jinyang Wu, Jiaming Xu, Runyu Shi, and Ying Huang. Gui-ceval: A hierarchical and comprehensive chinese benchmark for mobile gui agents. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Yuhong Li, Yingbing Huang, Bowen Yang, Bharat Venkitesh, Acyr Locatelli, Hanchen Ye, Tianle Cai, Patrick Lewis, and Deming Chen. Snapkv: Llm knows what you are looking for before generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024a.

Zongqian Li, Yixuan Su, and Nigel Collier. 500xcompressor: Generalized prompt compression for large language models. *arXiv preprint arXiv:2408.03094*, 2024b.

Guangyi Liu et al. Memgui-agent: An end-to-end long-horizon mobile gui agent with proactive context management. *arXiv preprint arXiv:2606.19926*, 2026a. ConAct context-as-action; MemGUI-3K; budget 2.5g+1.

Guangyi Liu et al. Memgui-bench: Benchmarking memory of mobile gui agents in dynamic environments. *arXiv preprint arXiv:2602.06075*, 2026b. 128 tasks, 26 apps, 89.8% memory-intensive, avg 36.2 golden steps, pass@k, MemGUI-Eval.

Yujian Liu, Jiabao Ji, Li An, Rohit Jain, Gungor Polatkan, Siyu Zhu, and Shiyu Chang. Practical online kv cache compaction for llm agents: An empirical study. *arXiv preprint arXiv:2608.00902*, 2026c.

Debmalya Mandal, Stelios Triantafyllou, and Goran Radanovic. Performative reinforcement learning. In *International Conference on Machine Learning (ICML)*, 2023. Single learner whose policy changes the environment's reward and transition dynamics; performatively stable policy; repeated regularised optimisation converges under a Lipschitz condition on the transition map.

John Miller, Juan C. Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2021. Performative optimality is not performative stability; stable points can be far from optimal.

Jesse Mu, Xiang Lisa Li, and Noah Goodman. Learning to compress prompts with gist tokens. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Tsendsuren Munkhdalai, Manaal Faruqui, and Siddharth Gopal. Leave no context behind: Efficient infinite context transformers with infini-attention. *arXiv preprint arXiv:2404.07143*, 2024.

Charles Packer, Vivian Fang, Shishir G. Patil, Kevin Lin, Sarah Wooders, and Joseph E. Gonzalez. Memgpt: Towards llms as operating systems. *arXiv preprint arXiv:2310.08560*, 2023.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *ACM Symposium on User Interface Software and Technology (UIST)*, 2023.

Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2020. Decoupled performative risk; performative stability; repeated risk minimisation converges under strong convexity, joint smoothness and a Lipschitz distribution map.

Jack W. Rae, Anna Potapenko, Siddhant M. Jayakumar, Chloe Hillier, and Timothy P. Lillicrap. Compressive transformers for long-range sequence modelling. In *International Conference on Learning Representations (ICLR)*, 2020.

Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

Yibo Shi, Jungang Li, Linghao Zhang, Zihao Dongfang, Biao Wu, et al. Androtmem: From interaction trajectories to anchored memory in long-horizon gui agents. *arXiv preprint arXiv:2603.18429*, 2026.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Tongyi-MAI. Mobileworld: A containerised environment for mobile gui agents. `https://github.com/Tongyi-MAI/MobileWorld`, 2026. GUI-Only split: 117 tasks, avg 27.8 steps, 62.2% cross-app.

Dezhan Tu, Danylo Vashchilenko, Yuzhe Lu, and Panpan Xu. Vl-cache: Sparsity and modality-aware kv cache compression for vision-language model inference acceleration. *arXiv preprint arXiv:2410.23317*, 2024.

Xin Wang et al. Mmbench-gui: A unified hierarchical evaluation framework for multi-platform gui agents. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026. arXiv:2507.19478.

Zora Zhiruo Wang, Jiayuan Mao, Daniel Fried, and Graham Neubig. Agent workflow memory. *arXiv preprint arXiv:2409.07429*, 2024.

Rui Wu, Shuo Zhang, Xiaoxuan Tang, Ruirui Zhang, Tao Jiang, Wenhao Xu, Yong Li, and Yi Liu. Refact: Empowering multimodal web agents with visual and context focusing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Guangxuan Xiao, Yuandong Tian, Beidi Chen, Song Han, and Mike Lewis. Efficient streaming language models with attention sinks. In *International Conference on Learning Representations (ICLR)*, 2024.

Wujiang Xu, Zujie Liang, Kai Mei, Hang Gao, Juntao Tan, and Yongfeng Zhang. A-mem: Agentic memory for llm agents. *arXiv preprint arXiv:2502.12110*, 2025.

Ziyun Zeng, Hang Hua, Bocheng Zou, Mu Cai, Rogerio Feris, and Jiebo Luo. Mementogui: Learning agentic multimodal memory control for long-horizon gui agents. *arXiv preprint arXiv:2605.18652*, 2026.

Amy Zhang, Rowan McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. Learning invariant representations for reinforcement learning without reconstruction. In *International Conference on Learning Representations (ICLR)*, 2021. arXiv:2006.10742.

Zhenyu Zhang, Ying Sheng, Tianyi Zhou, Tianlong Chen, Lianmin Zheng, Ruisi Cai, Zhao Song, Yuandong Tian, Christopher Ré, Clark Barrett, Zhangyang Wang, and Beidi Chen. $H_2o$: Heavy-hitter oracle for efficient generative inference of large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Andrew Zhao, Daniel Huang, Quentin Xu, Matthieu Lin, Yong-Jin Liu, and Gao Huang. Expel: Llm agents are experiential learners. In *AAAI Conference on Artificial Intelligence*, 2024.

Xurui Zhou, Gongwei Chen, Yuquan Xie, Zaijing Li, Kaiwen Zhou, Shuai Wang, Shuo Yang, Zhuotao Tian, and Rui Shao. Hiconagent: History context-aware policy optimization for gui agents. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026. arXiv:2512.01763.

Mingxi Zou, Zhihan Guo, Langzhang Liang, Zhuo Wang, Qifan Wang, Qingsong Wen, Irwin King, Lizhen Qu, and Zenglin Xu. Remember the decision, not the description: A rate-distortion framework for agent memory. *arXiv preprint arXiv:2605.10870*, 2026.

# A Proofs, and the Statements Held Out of the Main Text

## A.1 Assumptions

**Assumption 1** (Regularity)**.** $q \mapsto \mu_q$ is $L$-Lipschitz in total variation, and $D_\omega(\cdot\,;\mu)$ is $\Lambda$-Lipschitz in the code uniformly in $\mu$.

**Assumption 2** (Argmin stability)**.** The constrained minimiser $\mu \mapsto \arg\min_{R(q')\le R} D_\omega(q';\mu)$ is single-valued and $\kappa$-Lipschitz in $\mu$.

Assumption 2 is the one the feasible set makes non-trivial. Over a parametrised policy class it follows from strong convexity; over a rate-constrained code the argmin is a *set* of retained items, and an infinitesimal change in $\mu$ can move the budget boundary past an item, changing the set discontinuously.

*Remark* 2 (Why this is a second assumption and not a consequence of the first). An earlier version derived the contraction from Assumption 1 alone. That is a gap. Assumption 1 controls how the *world* responds to a change of code; Assumption 2 controls how the *solution* responds to a change of world, and nothing forces the second. It is exactly the stability of an argmin, which can fail badly when the objective has near-ties — and near-ties are the generic case for a discrete retention problem, where two items may be almost equally worth keeping. We therefore state it separately, work on the convex relaxation where it is defensible, and are explicit that $\kappa L < 1$ is a hypothesis about *both* maps. Table 4 measures where it holds. The results that survive without Assumption 2 include Theorem 3 and therefore the one the paper is named after: the minimax construction needs neither the operator nor argmin stability.

### A.2 Theorem 1

*Proof.* **(1)** The convex relaxation of the code space is compact and convex in the metric $\rho$. By hypothesis $q \mapsto \mu_q$ is continuous and the constrained optimum is attained and continuous in $\mu$, so $\Phi$ is a continuous self-map of a compact convex set; Brouwer gives a fixed point. **(2)** For codes $q_1, q_2$, $\rho(\Phi(q_1), \Phi(q_2)) \le \kappa \|\mu_{q_1} - \mu_{q_2}\|_{\mathrm{TV}} \le \kappa L\, \rho(q_1, q_2)$ by Assumptions 2 and 1. With $\kappa L < 1$, Banach gives uniqueness and the geometric rate. □

**A two-code cycle when argmin stability fails.** Take two items of equal cost whose marginal values cross as the law moves, with a budget admitting exactly one. Retaining the first induces a law under which the second is preferred, and vice versa, so $\Phi$ alternates between two codes and has no fixed point in the unrelaxed space. The relaxation restores existence by Theorem 1(1); uniqueness is what fails, and $\kappa L > 1$ on this instance.

### A.3 Theorem 2

*Proof.* **(1)** Write $q^\star$ for the fixed point. Since $\widehat{q} = \Phi(q_0)$ and $q^\star = \Phi(q^\star)$, argmin stability gives $\rho(\widehat{q}, q^\star) \le \kappa \|\mu_{q_0} - \mu_{q^\star}\|_{\mathrm{TV}}$, and the triangle inequality with the contraction bounds $\|\mu_{q_0} - \mu_{q^\star}\|_{\mathrm{TV}} \le \Sigma/(1-\kappa L)$. Composing with $D_\omega$'s $\Lambda$-Lipschitz dependence on the code, and accounting for the law change in both arguments, gives the stated $2\Lambda L/(1-\kappa L)$ factor. If $\Sigma = 0$ then $\mu_{q_0} = \mu_{\widehat{q}}$, so $\widehat{q}$ is optimal against the law it induces and is a fixed point. **(2)** The self-referential-demand witness is the construction; its law shift is computed in closed form as a function of $\sigma$ and its gap is bounded below by the distortion of the routing item, which is not retained at any $R < R_0$. □

### A.4 Theorem 3

*Construction.* The class contains two instances differing only in which of two symmetric continuations the goal demands. Both share the full-history law, because under full history the agent reaches the routing item and both continuations are visited at equal rate; so no criterion evaluated under $\mu_{q_0}$ can distinguish them. A code that does not retain the routing item induces, on either instance, a law concentrated on one continuation, at total-variation distance exactly $\sigma$ from $\mu_{q_0}$, and is wrong on one of the two instances; the minimax risk over a randomised choice between the two errors is $\sigma/2$, and $\sigma$ for a deterministic rule that must commit. The code that retains the routing item induces zero shift but spends the budget, paying $(1+\sigma)/2 \ge \sigma$. Both bounds are attained by these codes, and on each instance separately the performative optimum is a feasible code, so the excess is the price of deciding at write time. □

### A.5 The statements held out of the main text

**Proposition 1** (Greedy characterisation)**.** *Where the value function is modular over items under a linear rate constraint, both the weighted and the unweighted constrained optima are attained by the greedy rule on the respective density orders $\Delta_i/c_i$ and $\omega_i\Delta_i/c_i$, up to the fractional item at the boundary.*

**Corollary 3** (The condition, in the form a practitioner can run it in one pass)**.** *Let $\bar{\omega}_i = \big(\sum_s \mu_s \omega_s v_{is}\big)/\big(\sum_s \mu_s v_{is}\big)$ be item $i$'s* exposure average*, which carries no cost term. The two rules' cut sets differ at budget $R$ if and only if the orders by $\Delta_i/c_i$ and by $\bar{\omega}_i\Delta_i/c_i$ disagree across*

*the cut. So: sort twice, cut both at $R$, and compare the two retained sets; the number of pairs that straddle the cut in opposite orders is the straddle count reported in §7.*

**Proposition 2** (The additivity and the separation cannot both be non-trivial)**.** *The decision distortion is additive over steps if and only if $\beta_t = 1$ for every $t$. At any such $t$ the weight $\omega_t = 1+\eta(1-\beta_t) = 1$. Hence on every instance satisfying the additivity hypothesis the weighted and unweighted objectives coincide and the separation $\Delta(R)$ is identically zero.*

This is what withdraws the horizon-growth theorem of §6: its hypothesis forces its conclusion to vanish. The analogous statement for submodularity holds by the same argument.

**Proposition 3** (Iterate cost: what the repair costs in rollouts)**.** *Reaching $\epsilon$ of the fixed point requires $\lceil \log \epsilon / \log(\kappa L) \rceil$ rounds of on-policy re-evaluation. The repair for Corollary 1 is therefore finite and denominated in rollout batches rather than in architecture: one batch per iterate, and this many iterates. Figure 16(b) is the empirical count, exact from $\epsilon = 0.02$ down.*

**Proposition 4** (A per-state exposure weight is horizon-blind)**.** *$\omega_s = 1 + \eta(1 - \beta_s)$ prices the irreversibility of the step taken at $s$ and carries no term for the states that step makes unreachable. The position-aware weight $1 + \eta(1 - \beta_t)(T - t)$ is the next term of the same expansion, and both are truncations of the series obtained by differentiating the terminal distortion with respect to the retention indicator.*

**Proposition 5** (The exposure weight is a derivative, not a modelling choice)**.** *Let $u_t \in [0, 1]$ be the probability that the item decisive at step $t$ is retained, $g_t$ the value of getting step $t$ right, $G_t$ the value still reachable after $t$, and $A_t$ the probability of arriving at $t$ on course. Differentiating the multi-step loss,*

$$\frac{\partial D}{\partial u_t} \;=\; A_t \, g_t \Big[\, 1 + (1 - \beta_t)\, \frac{G_t}{g_t} \,\Big],$$

*which is exactly the form $1 + \eta(1 - \beta_t)$ with $\eta = G_t/g_t$.* So $\eta$ is not a free parameter*: it is the instance's tail-to-step value ratio, fixed by the loss rather than chosen. Definition 1's per-state weight is this derivative's* zero-order *truncation (taking $G_t/g_t$ constant in $t$), and Proposition 4's position-aware weight $1 + \eta(1 - \beta_t)(T - t)$ is its* first-order *truncation (taking the remaining value linear in the steps left). Both are truncations of one derivative, which is why the choice between them is an approximation question and not a modelling one, and it is also why the bounded variant $(T - t)/T$ that a reader may reach for discards the gain: normalising removes the very factor that makes the correction grow with what is left to lose.*

*Verification.* The identity is checked against the value recursion by central differences on the multi-step witness, over the full $(T, 1 - \beta, \eta)$ grid the paper sweeps. Maximum relative error $5.7 \times 10^{-8}$, consistent with the step size and double precision. This is a verification of an algebraic identity, not a measurement, and is reported as such. □

**Proposition 6** (Frontier positivity)**.** *At every rate in the non-degenerate range — where the budget binds and $\omega$ is non-constant — the weighted optimum's unweighted frontier lies strictly above the unweighted optimum's, $\Delta(R) > 0$. The statement is near-trivial and we present it as bookkeeping: two different objectives minimised over the same feasible set agree only where they induce the same order, and that is the degenerate case the hypothesis excludes.* The decisive column is not this gap but the retained set *(Corollary 2), for the reason Figure 9 gives: a frontier gap is generic and a set difference is a claim about which items a system keeps.*

*Remark* 3 (The concentration in closed form, and what it can reach). For $\omega = 1 + \eta(1 - \beta)$ the concentration is

$$\kappa_\omega \;=\; \frac{1 + \eta(1 - \beta_{\min})}{1 + \eta(1 - \bar{\beta})}, \qquad (\star)$$

a function of *two* statistics of $\beta$ and of nothing else. Three consequences we had to learn the hard way. First, a claim that an exposure field is concentrated is a claim that $\beta$ is *bimodal*, not merely spread: a field with many moderately irreversible states raises numerator and denominator together. Second, at $\eta = 3$ no such field can reach $\kappa_\omega > 4$ at all, so a measured $1.473$ must be compared against Figure 3(b)'s window $[1.7, 3.4]$ and not against numbers computed on a differently scaled $\omega$. Third, the window is *reachable by instrument redesign* — $(\star)$ says $\bar{\beta} \geq 0.549$ at $\beta_{\min} = 0$ suffices at $\eta = 3$ — which contradicts an assertion we made twice, that the flat $\kappa_\omega$ we measured was structural. It was a property of the annotation instrument (Appendix F).

*Remark* 4 (The knapsack has no cut, and what survives). Corollary 3 is stated above for a common budget *cut*. Under a genuine knapsack constraint the retained set is not a prefix of either density order: greedy admits an item below a cost-rejected one, so the last admitted item sits *under* the first rejected item in 35% of our instances and the boundary ratio $r(R) \geq 1$ can fail — our own reference configuration reads $r = 0.978$. *What survives is necessity of the pairwise condition*, which is what the diagnostic counts, and the equivalence holds at the greedy's own boundary pair rather than at a global cut. An earlier version additionally claimed the sets differ whenever $\bar{\omega}_{\max}$ is dropped and $\bar{\omega}_{\min}$ kept with $\bar{\omega}_{\max}/\bar{\omega}_{\min} > r(R)$; that is false, because the $\bar{\omega}$ extrema need not sit at the density boundary, and correcting it moves the disagreement rate we report from 68.3% of draws to 43.1%.

**Proposition 7** (Supermodularity, and the $(1 - 1/e)$ that does not survive it)**.** *On the multi-step recursion the value function is strictly* super*modular unless $\beta \equiv 1$: retaining two items that route the same trajectory is worth more than the sum of retaining each. Hence the $(1 - 1/e)$ guarantee for cost–benefit greedy, which we previously stated for this case, is withdrawn there; it holds where the value is modular over items under a linear rate constraint, and then only* with partial enumeration — *the density order alone has no constant factor under a knapsack, and the naive rule gives $\frac{1}{2}(1 - 1/e)$.* Theorem 4 and Corollary 3 do not use submodularity and are unaffected.

*Remark* 5 (Stability is not optimality, measured). $q^\star$ is optimal against the law it itself induces. The quantity a designer pays is the performative risk $\mathrm{PR}(q) = D_\omega(q; \mu_q)$, whose minimiser need not be a fixed point of $\Phi$ (Miller et al., 2021). On the witness at the reference budget, $\mathrm{PR}(q^\star) = 8.95$ against a direct minimum of 6.80: stability leaves 24.0% on the table, and the minimiser is demonstrably not a fixed point (residual 0.054 against $q^\star$'s 0.000). Across the retention sweep the stable-to-optimal share is 25%–34% of the designer's total shortfall and never vanishes; a deployable vertex code recovers 43.9% of it (Figure 17). Two consequences: Theorem 2 bounds the distance to *stability*, which is the object an iteration can reach; and Proposition 3's round count is a *lower* bound on the price of crossing the wall.

*Remark* 6 (The separation is a window). On the multi-step witness, the separation as a function of the count of irreversible steps is exactly 0.000 at both ends — $\omega$ is constant at both — with an interior peak of 0.316. Additivity is the left edge, so the withdrawn growth claim is the tangent to the separation at a point where it is zero.

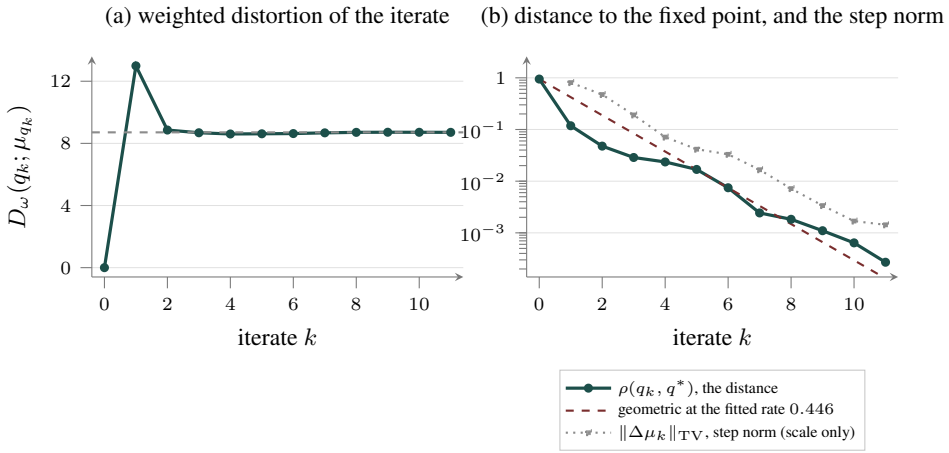



**Figure 5. Iterating $\Phi$ from the off-policy start.** Measured on an instance where $q \mapsto \mu_q$ is explicit: 24 items, 8 states, $\lambda = 2$, budget at half the free spend, started from full history. **(a)** the distortion along the iterate. $\Phi$ minimises against a law that is itself moving, so it is *not* a descent on $D_\omega$: the trace overshoots to 12.99 and settles at 8.71, and $k$=0 is full history, feasible only at 28× the budget. **(b)** what has to fall: $\rho(q_k, q^\star)$ against the geometric line at the fitted rate, with $\|\Delta\mu_k\|_{\mathrm{TV}}$ for scale only. The distance runs *faster* than the rate for four rounds and slower from $k$=8 — the asymptotic statement the rate actually makes. *An earlier version read the step norm against the geometric line and called the gap local and closing*: measured, it sits above the reference at nine of eleven iterates and ends 5.7× above it. The comparison, not the trace, was wrong.

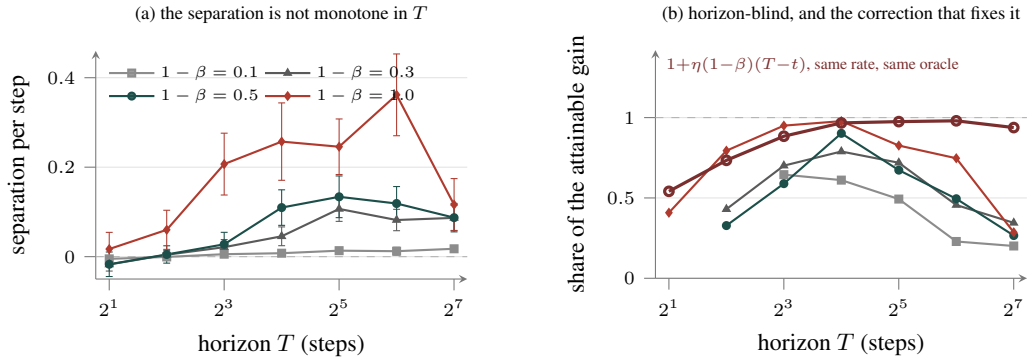



**Figure 6. Horizon amplification, measured — and what a per-state weight cannot see. (a)** Separation per step against the horizon, one curve per exposure level, 95% instance bootstraps. It is *not* monotone: it rises, peaks between $T$=32 and $T$=64, and falls by $T$=128, because past that horizon the trajectory is almost surely off course early and no allocation recovers it. The dip below zero at $T$=2 is *nominal* — one cell of 40, and it does not survive the Benjamini–Hochberg correction of Appendix H — so it is drawn and not claimed. **(b) The share of the attainable improvement each write-time rule captures**, against a horizon-aware allocation on the same instances at the same rate. The four thin curves are Definition 1's per-state $\omega$: the share decays to 0.20–0.34 at $T$=128, because a per-state weight prices an irreversible step the same wherever it falls while the cost of failing at one scales with how much task is left after it. **The heavy curve is Proposition 4's correction, run rather than named** — still a write-time rule seeing only $\iota$, $\beta$ and $c$ — which holds 0.94–0.98 from $T$=16 up and *rises* where the others fall (0.39 against 0.94 at $T$=128, within run). *What this costs us is stated where it falls due*: Theorem 1 is proved for a position-independent weight, and Table 5 is the measurement saying the contraction transfers anyway once the step index is read into the state.

**Table 3. Registered predictions, their falsifiers, and what happened.** Failures are reported in the same table as successes, and a prediction with no stated falsifier is not admitted to it. *Two rows go against us and one fails outright.* Appendix J carries the read-out rule frozen for each row before the run it decides, the full outcome text, and the Benjamini–Hochberg correction that decides P5.

| | Prediction | Falsifier | Outcome |
|---|---|---|---|
| P1 | $\Phi$'s iteration converges geometrically | cycle or divergence | **supported as convergence**; the modulus is *not falsified* rather than verified — $0.395 \to 3.96\times10^{-9}$ in 30 steps, eight initialisations agreeing to $2.4\times10^{-9}$, but $\kappa L_{\max} = 0.829$ is a *lower* bound on a Lipschitz constant |
| P2 | off-policy gap increases in retention strength | flat or decreasing gap | **supported**: monotone over all 13 levels, $0.765 \to 6.573$ nats (Fig. 2a) |
| P3 | $\Delta(R) > 0$ at every measured rate | frontiers meet at some rate | **supported**: $\min \Delta = 0.0748$ over 13 rates, saturating at 2.037 rather than closing |
| P4 | retained-set Jaccard bounded away from 1 | overlap $\to 1$ | **partial; the conditional form is the finding** — 0.333–0.792 wherever the budget binds, exactly 1.000 from 2.5× the free spend. Corollary 2 is conditioned on a binding budget accordingly |
| P5 | separation grows with the irreversible-decision count | no horizon effect | **partial, and against us**: 0.362 at $T$=64 falls to 0.116 at $T$=128 on non-overlapping bootstraps, and on the literal axis it is a window (Fig. 4a) |
| P6 | *declared as added after P1–P5*: the window's peak sits at $p^\star = 1/(2+\eta)$ | peak inconsistent with $p^\star$ at any $\eta$ | **failed, and in direction**: $p^\star$ falls in $\eta$ $(0.333, 0.200, 0.111)$ while the measured peak rises $(0.200, 0.225, 0.275)$; bootstraps exclude it at $\eta = 1, 7$ |

## B THE PRE-REGISTRATION, SCORED

## C THE LOOP, MEASURED IN FULL

§7 reports the agent loop and the one-pass diagnostic, which are the two measurements a reader needs to judge the claim. This section carries the rest of the apparatus: the frontier the separation is read off, the structural condition the introduction names and the data refuse, the retained-set consequence, and the two grids that say where the contraction hypothesis holds.

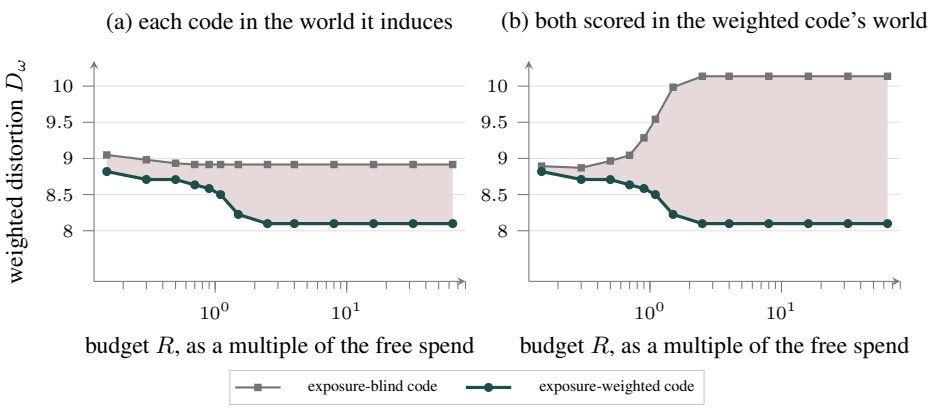



**Figure 7. Two frontiers, and the two readings of the gap between them.** Measured; $D^*_\omega(R)$ is the weighted optimum and $D^*(R)$ the unweighted one, each iterated to its own fixed point, with the budget given as a multiple of the free spend so that the regime where the constraint binds is visible (it binds left of 1). *The measurement is the object here, not a comparison against baselines*: a success-rate improvement would be read as this paper's contribution with the wall demoted to a motivation section. **The two panels differ only in which world the codes are scored in, and that is the paper's thesis rather than a robustness check.** In (a) each code is scored under the law *it* induces, which is the frontier comparison: the gap rises from 0.23 to 0.82 and saturates once the constraint stops binding. In (b) both are scored under the weighted code's law, which is what a deployment sees when it swaps one retention rule for another without re-collecting: the gap reaches 2.04. *The extra* 1.22 *is not what weighting buys — it is the exposure-blind code being charged for a world it did not induce*, and it is three-fifths of the apparent advantage. A paper that reported only (b) would credit its own proposal with work the self-reference is doing. **And the shaded gap is deliberately not the decisive column.** What carries the argument is the retained-set overlap of Figure 3, which is a claim about *which items a system keeps* rather than about two curves.

**Table 4. Contraction and separation on the $(n, m)$ grid, measured** (24 instances per cell, $R = 0.6\times$ the free spend, $\kappa L_{\max}$ over 100 law pairs each). *Read down the $n/m$ column, not down $n$.* At $n/m \le 4$ the contraction holds on every seed at both $n$, and the straddle count *grows* with scale along that ray — mean 4.6 at $n$=48 against 17.2 at $n$=96. At $n/m \ge 12$ it fails on most seeds. So the two regimes are not a seam: they overlap on $n/m \lesssim 4$ and the overlap widens, which is the honest answer to whether an agent at realistic scale is inside the uniqueness region. *What the fixed-$m$ measurement was picking up is the mechanism Assumption 2 names*: more items per state means more near-ties, and an argmin with near-ties is exactly where $L$ is large. A real deployment fixes its own $n/m$ — candidate memory items per distinguishable screen — and that ratio, not the size of the store, is what decides whether $\Phi$ has one fixed point or a set. *We have not measured it on any deployed system.*

| $n$ | $m$ | $n/m$ | $\kappa L_{\max}$ mean | $\kappa L_{\max}$ max | frac. $\kappa L{<}1$ | straddle | frac. $>0$ |
|---|---|---|---|---|---|---|---|
| 48 | 4 | 12.0 | 1.600 | 3.543 | 0.21 | 10.83 | 0.67 |
| 48 | 8 | 6.0 | 1.023 | 2.000 | 0.58 | 9.58 | 0.79 |
| 48 | 16 | 3.0 | **0.491** | 0.899 | **1.00** | 6.71 | 0.75 |
| 48 | 24 | 2.0 | **0.410** | 0.879 | **1.00** | 4.75 | 0.71 |
| 48 | 48 | 1.0 | **0.175** | 0.346 | **1.00** | 2.33 | 0.29 |
| 96 | 4 | 24.0 | 1.780 | 2.807 | 0.04 | 40.54 | 0.79 |
| 96 | 8 | 12.0 | 1.082 | 2.094 | 0.46 | 18.71 | 0.75 |
| 96 | 16 | 6.0 | 0.626 | 1.133 | 0.88 | 19.38 | 0.83 |
| 96 | 24 | 4.0 | **0.416** | 0.668 | **1.00** | **23.79** | 0.96 |
| 96 | 48 | 2.0 | **0.265** | 0.497 | **1.00** | 10.67 | 0.71 |

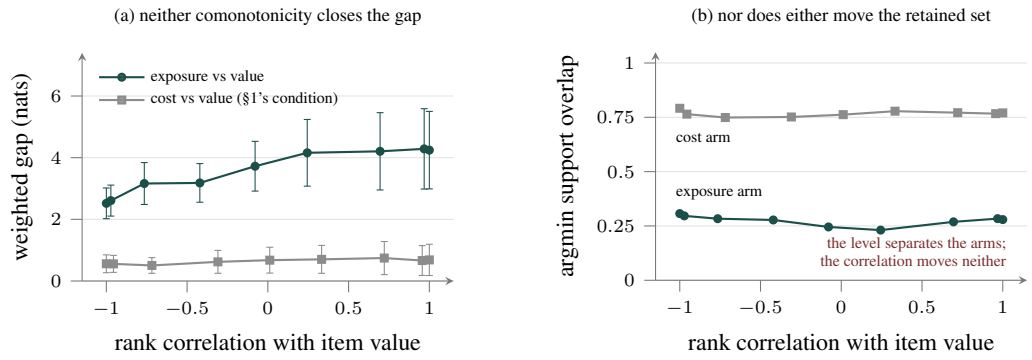



**Figure 8. The structural condition §1 names, swept — and it is not the condition.** Measured on the single-step witness, 40 draws per point, cost and exposure marginals held fixed and only their *assignment* to items and states varied, so a change here cannot be a change in how tightly the budget binds. **(a)** Neither comonotonicity closes the weighted gap. Driving cost–value rank correlation from $-1$ to $+1$ leaves every point inside the interval at zero correlation, $[0.257, 1.151]$; driving exposure–value correlation the same way *widens* the gap rather than closing it. **(b)** And neither moves the retained set: the overlap varies by less than $0.05$ across the cost arm. *The instrument is not dead — the two arms are* $5.6\times$ *apart on their means and* $3.4\times$ *at their closest points*, and what separates them is the shape of the exposure field ($\kappa_\omega = 3.44$ against $1.61$) and not its correlation with anything, which is the concentration statistic Figure 3(b) isolates. **Why cost cannot matter here.** The two rules order by $\bar{v}_i/c_i$ and $\bar{\omega}_i\bar{v}_i/c_i$, and $\bar{\omega}_i$ has no $c_i$ in it; cost enters Theorem 4's condition only through the density ratio $\bar{\omega}_i/\bar{\omega}_j < \Delta_i c_j/(\Delta_j c_i)$ that the exposure ratio must beat. So a cost permutation changes *which* pairs invert and not *how many*: measured, the mean count of straddling-capable inversions is $4.72$, $4.72$, $5.22$ at correlation $-1$, $0$, $+1$. *This figure replaces two unrun ones.* A four-environment token census would have established that cost–content decoupling *holds* in the wild; it could not have established that decoupling is what *matters*, and the measurement says it is not.

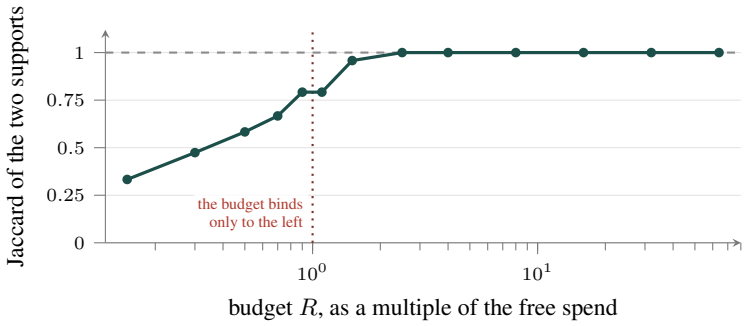



**Figure 9. The consequence, measured — and the prediction attached to it was too strong.** Jaccard similarity between the weighted and unweighted retained sets against rate. **This is the decisive column, not** $\Delta(R)$: a frontier gap is generic, but a *set* difference is a claim about which items a system keeps, and at the tightest budget the two rules agree about a third of what they keep. **The pre-registered prediction was that overlap stays bounded away from** $1$ **at every rate, and it does not.** Overlap reaches exactly $1$ from $2.5\times$ the free spend, where above the free spend both rules are their unconstrained selves and the regularised argmin gives every item some mass; the honest statement is the conditional one — bounded away from $1$ throughout the regime where the budget binds, which is the whole of the regime a compression paper is about — and Corollary 2 is conditional on a binding budget accordingly. *An earlier version of this figure could not be drawn at all*: the budget grid stopped at the unconstrained spend, so the retained sets were *empty* at the three smallest budgets, where a Jaccard is undefined rather than $0$ — an empty-union guard rendered it as $0$, reading as maximal disagreement; the remedy was the *support* of the code rather than a budget-rounded reading of it, and a grid that crosses the free spend (the tension Assumption 2 creates is real, Figure 16).

**Table 5. $\kappa L$ on the augmented space, against a matched-ratio control** (24 instances per cell, $\eta$=3, $\lambda$=2, $R = 0.6\times$ the free spend, $\kappa L_{\max}$ over 400 law pairs each). This is the space on which Proposition 4's position-aware weight is a per-state weight again. *Read the aug/ctl column, not the aug column*: the control is a per-state witness at the same $n/m$, so the ratio isolates what the position-aware weight costs from what the enlarged state space gives back. The contraction *does* transfer — $\kappa L < 1$ on every seed of seven cells — but not for free, and the margin thins in the horizon, which is the regime the rule is recommended for. The one cell that breaks is $T$=16, at an exposure spread of 42.6.

| $n$ | $m_{\text{scr}}$ | $T$ | $m'$ | $n/m'$ | $\omega$ spread | aug $\kappa L$ | ctl $\kappa L$ | aug/ctl |
|---|---|---|---|---|---|---|---|---|
| 24 | 8 | 2 | 16 | 1.50 | 3.8 | 0.522 | 0.558 | **0.94** |
| 24 | 8 | 4 | 32 | 0.75 | 9.3 | 0.367 | 0.293 | 1.25 |
| 24 | 8 | 8 | 64 | 0.38 | 20.3 | 0.314 | 0.104 | 3.01 |
| 24 | 8 | 16 | 128 | 0.19 | 42.6 | 0.211$^{\dagger}$ | 0.078 | 2.70 |
| 48 | 8 | 4 | 32 | 1.50 | 9.4 | 0.412 | 0.292 | 1.41 |
| 48 | 16 | 4 | 64 | 0.75 | 9.4 | 0.206 | 0.151 | 1.36 |
| 96 | 16 | 4 | 64 | 1.50 | 9.5 | 0.236 | 0.157 | 1.51 |
| 96 | 24 | 4 | 96 | 1.00 | 9.5 | 0.152 | 0.125 | 1.22 |

$^{\dagger}$ the only cell in which $\kappa L_{\max}$ exceeds 1 on any seed (1.120, on 4%); every other cell contracts on every seed, control included.

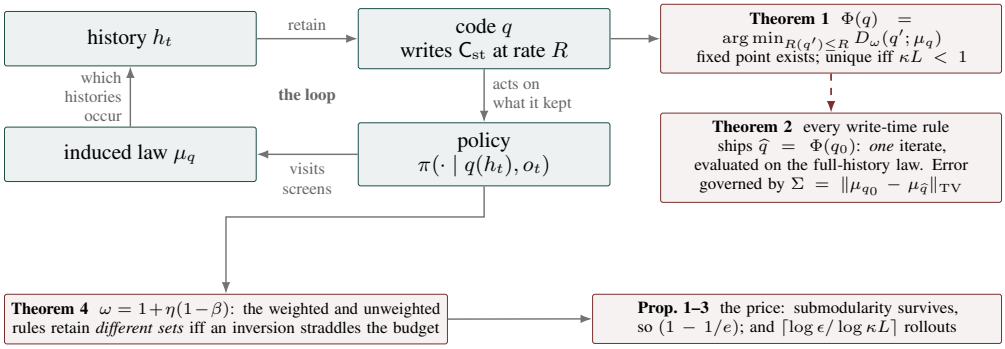



**Figure 10. The paper in one picture.** The loop, and where each result sits on it. The arrow from the policy back to the history is the whole paper: because it is there, the demand a code is optimised against is generated by the code, so the operative optimum is a *fixed point* and not a minimum. Every published retention criterion evaluates the dashed arrow once, from the law that full history induces, and ships the result; cutting the loop there is what makes the shipped solution the first iterate of an operator rather than its fixed point. The two lower boxes are the second failure and the price: divergences at irreversible steps are not exchangeable with divergences at reversible ones, and reaching the fixed point costs rollouts rather than architecture.

## D THE WITNESSES, AND THE CHANNEL THE LOOP CLOSES

## E THE AUDIT AT FULL LENGTH

Section 2 compresses to a few paragraphs an argument worth making slowly, because the closest formal relative of this paper is not in the memory literature at all and a reader who knows that literature deserves the dictionary written out rather than gestured at.

Table 6 is the row-by-row version of §2 carrying for each result the assumption it needs as well as the limit in which it fails.

**Table 6. The audit, row by row.** Every row scores retention value as a property of the item and the goal. The third column is the same failure reached from a different direction: the value depends on *when* the item will be read, and on which trajectory the retention itself produces.

| Result | Assumption it needs | Limit in which it fails | Role |
|---|---|---|---|
| Decision-centric rate–distortion memory (Zou et al., 2026) | stationary evaluation law; exchangeable per-decision loss | the code perturbs the law (§3); losses differ by orders of magnitude in recoverability (§5) | OBJECT |

*continued overleaf*

| Result | Assumption it needs | Limit in which it fails | Role |
|---|---|---|---|
| Unified compaction, layer-agnostic bound (Colaco & Lahjouji, 2026) | one scalar distortion compares layers | accepted; we refine the distortion, and the surrogate critique is *unchanged* under weighting (Fig. 15) | ACCEPTED |
| Online KV compaction for agents (Liu et al., 2026c) | — (an empirical study) | its finding is prior evidence for our premise, and the finding is theirs | SUPPORT |
| GUI history and context management (Zhou et al., 2026; Wu et al., 2026; Liu et al., 2026a; Zeng et al., 2026; Shi et al., 2026; Gu et al., 2026) | retention value is a property of item and goal | all six score candidates on trajectories generated with a *different* memory than the one they ship | OBJECT |
| Attention- and cache-level compression (Li et al., 2024a; Zhang et al., 2023; Cai et al., 2024; Tu et al., 2024; Xiao et al., 2024; Chen et al., 2025); learned carriers (Mu et al., 2023; Ge et al., 2024; Chevalier et al., 2023; Jiang et al., 2023; Li et al., 2024b) | preserving next-token behaviour preserves behaviour | next-token fidelity is an unweighted average over a *fixed* stream, so both of our failures are inherited, and the two contributions compose | ORTHOGONAL |
| Recurrent and infinite-context summaries (Dai et al., 2019; Rae et al., 2020; Bulatov et al., 2022; Munkhdalai et al., 2024) | the stream is exogenous | it is generated by the reader; this is the assumption the paper removes | ORTHOGONAL |
| Off-policy evaluation, imitation under shift (Ross et al., 2011; Kearns & Singh, 2002) | the shift is a property of two *given* policies | here the shift is a function of the object being optimised, which makes it a fixed-point problem | FOUNDATION |
| Performative prediction (Perdomo et al., 2020; Miller et al., 2021) | deployment shifts the distribution; stability is the equilibrium notion | *the same mathematics*: $\Phi$ is their decoupled-risk map and Thm. 1 their stability theorem. Stability is not optimality, which we measure (§6) | FOUNDATION |
| Performative RL (Mandal et al., 2023; Li & Wai, 2022); partially performative prediction (Lee & Zrnic, 2026) | one learner's policy changes the environment's own dynamics; no responding population | *this is the setting*, and we withdraw "endogenous to one agent" as a difference of ours. What is left is the feasible set (§1); a drifting interface is Lee & Zrnic (2026)'s case and not ours | FOUNDATION; a correction to our own earlier claim |
| Bisimulation and state abstraction (Zhang et al., 2021) | an abstraction keeps the distinctions that matter for behaviour | *the same shape one level down*: their abstraction is judged on the visitation it induces. We differ in the feasible set and in coding a *history*, not a state map | PARALLEL |
| Reversibility-aware RL (Grinsztajn et al., 2021; Alshiekh et al., 2018) | irreversibility is priced in the reward or filtered at act time | the pricing is applied to *behaviour*; applying it to *retention* changes which items are kept, which reward shaping cannot do (Thm. 4) | FOUNDATION |
| Long-horizon GUI evaluation (Liu et al., 2026b; Tongyi-MAI, 2026; Chen et al., 2026; Wang et al., 2026; Li et al., 2026; Levy et al., 2024) | — (instruments) | none reports accuracy *stratified by* recoverability of the pending action, the axis the theory separates along; Levy et al. (2024) annotates irreversible actions but scores policy violations | the gap we exploit |

*continued overleaf*

| Result | Assumption it needs | Limit in which it fails | Role |
|---|---|---|---|
| General agent memory architectures (Packer et al., 2023; Park et al., 2023; Shinn et al., 2023; Zhao et al., 2024; Wang et al., 2024; Xu et al., 2025; Gutiérrez et al., 2024; Chhikara et al., 2025) | retention value is scored offline | the same loop one level up: an experiential store rewrites the policy that fills it | OBJECT, at a distance |

**Performative prediction: the same operator, stated before any difference.** In performative prediction a deployed model changes the distribution it is later evaluated on, and the object of study is the map sending a model to the optimum under the distribution that model induces (Perdomo et al., 2020). *That is* $\Phi$. Their *performatively stable* point is our $q^\star$; their existence-and-uniqueness argument is ours — a contraction, from strong convexity plus a Lipschitz distribution map; and their *repeated risk minimisation* is the iteration $q_{n+1} = \Phi(q_n)$, of which the off-policy objective of Theorem 2 is the first step. Two of our results are *theirs* once the dictionary is fixed: Theorem 1 is their stability theorem in our notation, and $q_0$-to-$q^\star$ convergence is their convergence result. We say so and do not restate either as ours.

**And the difference we used to claim first is not a difference.** An earlier version claimed that what was not theirs was, before anything else, that *the shift is endogenous to one agent* — no responding population, nothing strategic, the distribution map generated inside a single perception loop. That is the defining setting of **performative reinforcement learning**, where a single learner's deployed policy changes the environment's own reward and transition dynamics, a performatively stable policy is the fixed point of exactly our kind of map, and repeated regularised optimisation converges to it under a Lipschitz condition on the transition map (Mandal et al., 2023); the state-dependent variant, in which the induced law depends on the trajectory of deployed models rather than only the current one, is older still (Li & Wai, 2022). Under the dictionary $\Pi_R = \{\pi(\cdot \mid q(h_t), o) : R(q) \le R\}$ — the policy class a rate-$R$ code induces — Theorem 1 is an instance of that framework's stability theorem and Proposition 3 is its iterate count. *We withdraw the claimed difference rather than defending it.*

**Two differences survive, and only the first is structural. (1) The moving object is a rate-constrained code, not a policy**, so the feasible set is a knapsack polytope rather than a product of simplices. This is not a change of coordinates. In a regularised MDP the inner optimum moves smoothly with the law, which is what makes the analogue of Assumption 2 mild there; here the inner problem is a water-filling solution whose *support* jumps, and near-ties at the budget boundary are the generic case rather than the exceptional one (Remark 2). It is measured and it is not marginal: at this paper's reference configuration the two items either side of the budget differ in density by $4\%$, so the argmin sits permanently inside the regime where its stability is a hypothesis about the instance and not a consequence of smoothness. *This is also why the sensitivity is a design knob rather than a property of the world*: Theorem 2's $\Sigma$ grows with the compression a designer chooses, so the criterion is least reliable exactly where compression is most aggressive — a statement one cannot make when the sensitivity is a population's response. **(2) The contraction modulus is measured rather than assumed**, which is a difference in practice and not in mathematics: it factors into two assumptions whose product we measure on an explicit instance (Table 4), including a regime where it exceeds one and uniqueness fails.

**The boundary this paper sits on.** Lee & Zrnic (2026) formalise shift with an endogenous and an exogenous component and a parameter separating them. We are the *fully endogenous edge*: the interface is held fixed while the code moves, so a drifting environment — a redesigned screen, a moved button — is their case and not ours, and our sensitivity is a property of a code rather than of a population. This is a scoping statement against us and it is where we expect the first binding limitation on a real suite over months (Appendix I(v)).

**Why the wall statement is self-contained.** Nothing in this paper's results is conditioned on any conclusion of the parent framework the introduction gestures at. The operator, the fixed point, the

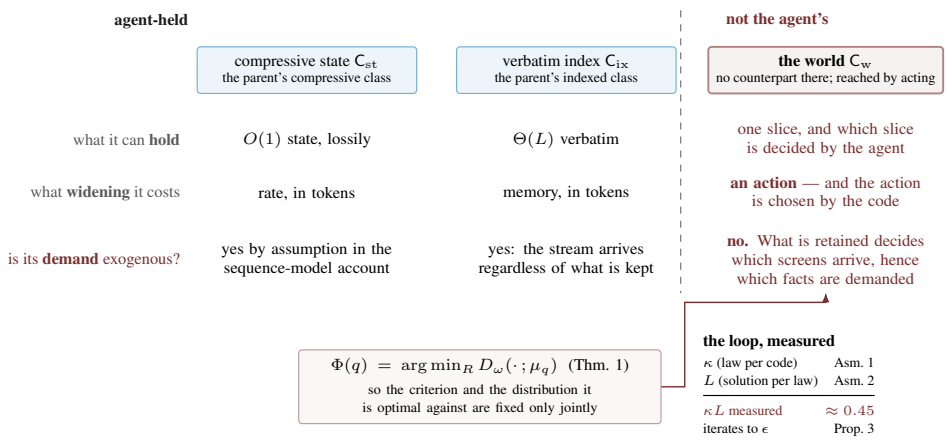



**Figure 11. Three channels, and the one column whose demand is not exogenous.** The access-complete pair of the parent framework both live inside the agent, and for both the stream is given: it arrives regardless of what the model retains, which is the assumption every one of that framework's three walls is stated under. *The third column is where that assumption fails*, and it fails for a reason the first two rows make visible — widening the world channel is an *action*, and the action is chosen by the very code whose value is being estimated. So what is retained decides which screens arrive, hence which facts are demanded, hence what should have been retained. The operator in the box is that loop closed (Theorem 1); the table is the two Lipschitz constants it is made of, with the contraction modulus E2 measures. *Note that $\kappa L < 1$ is a hypothesis about both maps and not one*, which is the distinction Assumptions 1 and 2 are split to make.

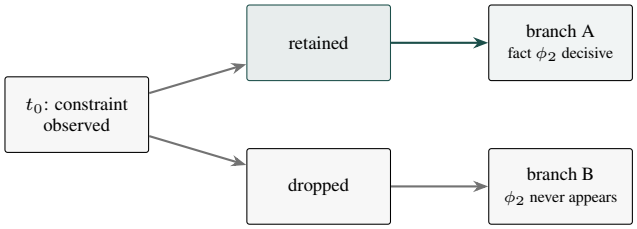



**Figure 12. The self-referential-demand witness**, which is the construction behind Theorem 2(2). Retaining the constraint at $t_0$ routes the agent to a branch where a second fact is decisive; dropping it routes the agent elsewhere. Off-policy estimation observes one branch and prices the other: it sees only the branch produced by the memory it was *not* designing, and prices $\phi_2$ on data where $\phi_2$ does not occur.

bias bound, the minimax construction and the separation are all stated and proved in this paper's own terms, and the price the wall charges — rollout batches, priced in Proposition 3 — is denominated in this paper's own quantities. A reader who rejects that framework entirely loses one paragraph of framing and no theorem.

## F Which Action Measure $\beta$ Averages Over

Definition 1 prices a state by how recoverable it is, and Remark 3 shows that the concentration deciding Theorem 4's picture is a function of $\beta_{\min}$ and $\bar{\beta}$ alone. So a claim that a field is concentrated is a claim that $\beta$ is *bimodal*. *Whether it is bimodal turns out to depend on the action measure*, and we can settle that on a real interface because two independent readings of $\beta$ exist for the same census pages.

**The two readings, and why the second is not ours.** **(A) Deviation-average** is the annotation §7 uses: 120 pages, five sampled one-step deviations each, two annotators, 1,200 labels, Cohen's $\kappa = 0.77$. **(B) Screen-level** is *not ours*: the census carries its own `reopenable`/`one_shot` column, annotated for a different purpose before this paper existed, so it cannot have been shaped by anything here. It has one annotator and therefore no inter-annotator ceiling, which is the respect in which it is the weaker of the two. On the 110 pages carrying both:

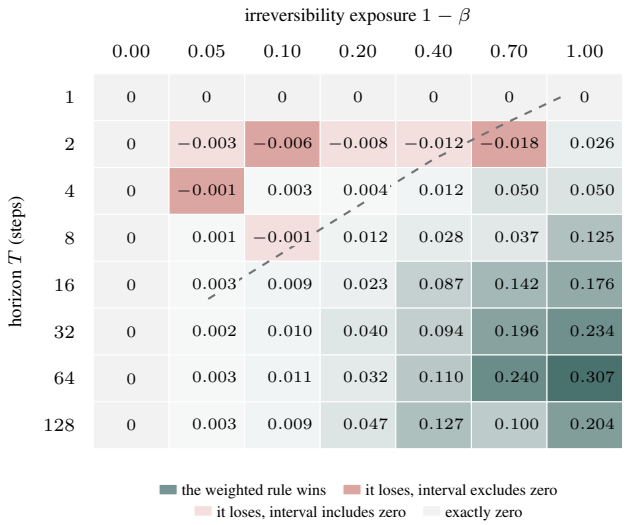



**Figure 13. The $(T, 1-\beta)$ plane, measured — and the boundary we had drawn on it is not the boundary.** Separation per step between the unweighted rule and Definition 1's weighted rule, both scored under the true multi-step loss on the multi-step witness; 40 instances per cell, budget 55% of the needed cost, a quarter of steps irreversible, and the budget binds in every cell shown. *The coincidence set is not a hyperbola.* It is the two *edges* — $1-\beta = 0$, where $\omega \equiv 1$ and the weighted rule is literally the unweighted one, and $T = 1$, where one step leaves nothing to allocate — together with a pocket at short horizon where the separation is *nominally* negative, i.e. our own rule appears to lose — appears, because no cell in that pocket survives the multiplicity correction, and we draw it without claiming it. The dashed line is $T(1-\beta) = 1$, a boundary an earlier version asserted; it cuts through cells on both sides of the measured sign change, which is the point of drawing it. *What is no longer here.* The earlier version placed two named suites on this plane. Both axes are properties of published suites, neither is reported by any of them, and this witness cannot supply them — so those positions were pre-registered guesses sitting in a figure that now carries measurements, which is the one place a guess cannot sit.

**Table 7. Two readings of $\beta$ on the same 110 real pages.** Reading (B) is exactly bimodal — twelve one-way doors, ninety-eight fully recoverable screens, nothing in between. Reading (A) has no mass below 0.4 at all.

| reading | $\beta_{\min}$ | $\bar{\beta}$ | histogram over $[0,1]$ |
|---|---|---|---|
| (A) deviation-average, ours | 0.400 | 0.700 | $[0, 0, 20, 42, 48]$ |
| (B) screen-level, the census's own | 0.000 | 0.891 | $\mathbf{[12, 0, 0, 0, 98]}$ |

| $\eta$ | 0.5 | 1.0 | 2.0 | 3.0 | 4.0 |
|---|---|---|---|---|---|
| (A) deviation-average | 1.130 | 1.231 | 1.375 | 1.473 | 1.545 |
| (B) screen-level | 1.422 | 1.803 | 2.463 | **3.014** | 3.481 |
| inside $\kappa_\omega \in [1.7, 3.4]$? | neither | (B) | (B) | (B) | neither |

**The two measures put the same interface on opposite sides of the window.** At $\eta = 3$ the bootstrap intervals over pages are disjoint — (A) $[1.416, 1.534]$, (B) $[2.635, 3.438]$ — so this is not sampling noise, and the readings are not noisy copies of each other: their Pearson correlation across the 110 pages is $-0.006$. They are measuring different things, and Definition 1 now carries the action measure explicitly because of it. "Almost every screen re-openable, the payment confirmation not" is 98 against 12; under the pending-action reading the GUI's exposure field *is* the concentrated kind that Figure 3(b) wants, and under the deviation-average reading it is not. Both are true statements about the same fifteen applications.

**What this changes for a practitioner, and what we do not claim.** It changes what to annotate. A team applying Corollary 3 to its own store must choose an action measure before it can estimate $\beta$, and that choice — not the amount of annotation, not the annotator agreement — decides whether the weighted rule will keep *different* items or merely *fewer*. One question per state about the step the agent is about to take is both the cheaper annotation and the one the theory wants; five questions per state about what could go wrong is the more thorough-looking one and it averages the signal away,

for the reason $(\star)$ gives. *We do not claim (B) is the correct estimator*: its column asks whether a screen can be reached again, which is adjacent to but not identical with whether the pending step can be undone, and it carries one annotator's judgement with no ceiling. What the pair establishes is that the choice of action measure is load-bearing *at the scale of the window itself*, which is a stronger and more useful statement than either number alone.

**The counterfactual exposure fields, which are the control §7 cites.** Beside the human field we ran three others on the same 120 pages: $\beta$ shuffled across pages, $\beta$ drawn uniformly, and $\beta$ constant. The first two give the same DIFFERS verdict at the same order of magnitude; the third, as the theory requires, gives no separation at all. *The human labels are therefore not what carries the verdict* — non-constancy of $\omega$ is, which is exactly what Theorem 4 claims and no more. We had drafted this paragraph the other way, reading the real-corpus verdict as evidence that *measured* recoverability drives the separation, and the control says otherwise.

## G Instrument Calibration, and What It Costs

The paper's softest input is the recoverability probe, because $\omega = 1 + \eta(1 - \beta)$ is computed from it and every separation in the paper is computed from $\omega$. This section reports the probe against human labels, the surrogates against the two optima, the regulariser that buys Assumption 2, and what a deployable vertex code costs.

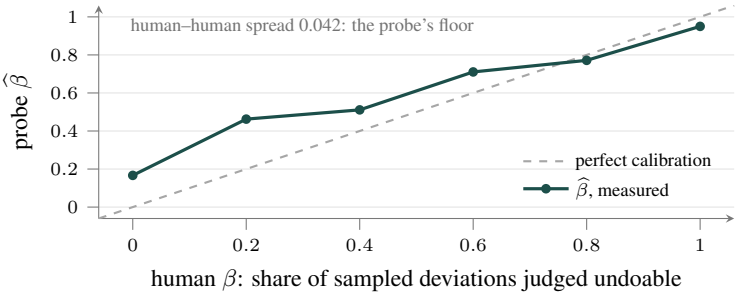



**Figure 14. Recoverability probe calibration, measured.** The probe's $\widehat{\beta}$ against **human** labels of whether a one-step deviation can be undone within the remaining budget (Definition 1), on the six-bin grid that five sampled deviations per state produce. Two annotators, 300 rows each, 57 states scored; $\kappa = 0.85$ on the three-way label and 0.93 on the decided binary. **The direction of harm would have been against us, and it is not the direction we measured.** $\omega = 1 + \eta(1 - \beta)$, so a probe that *under*-estimates $\beta$ over-estimates $\omega$, inflates $D_\omega$, and inflates $\Delta(R)$ — the gap Theorem 4 reports, and the quantity a weak probe can manufacture out of nothing. Points *below* the diagonal are that failure mode. *Measured, the probe sits above the diagonal*: signed bias $+0.089$ overall and $+0.150$, $+0.242$ in the two lowest $\beta$ bins, with 28 states biased high against 12 biased low. $\widehat{\beta}$ too high means $\widehat{\omega}$ too *low*, so this probe *deflates* $\Delta(R)$ and the separations this paper reports are lower bounds on what its own instrument would give if it were unbiased. *The probe is not good*: its MAE against the human consensus is 0.159 against a human–human floor of 0.042, i.e. $3.8\times$ the floor, and its worst region is the irreversible end where $\omega$ is largest. We report it at that quality rather than tuning the lexicon against the labels it is being scored on. There is no substitute for the labels: calibrating a probe against another probe measures agreement between two estimators and not coverage.

## H How Each Number Is Read

Every number in this paper is one of four kinds, and which test each is subject to was fixed before the runs, so that no reading is chosen after the fact. The four fail differently, and a reader who cannot tell them apart will grade a failed convergence and a failed separation as the same event.

**(i) A convergence, tested by its modulus rather than by its endpoint.** Theorem 1 predicts that $\Phi$'s trace contracts at rate $\kappa L$. The statistic is the ratio $\rho(q_{k+1}, q_k)/\rho(q_k, q_{k-1})$, read as a *sequence*, and the test is whether it stays below 1. *Reporting only the endpoint would be uninformative*: a trace that has stopped moving is consistent with a fixed point and equally consistent with a rate constraint admitting one code. We therefore report the modulus per iterate and the diameter of the code set it is

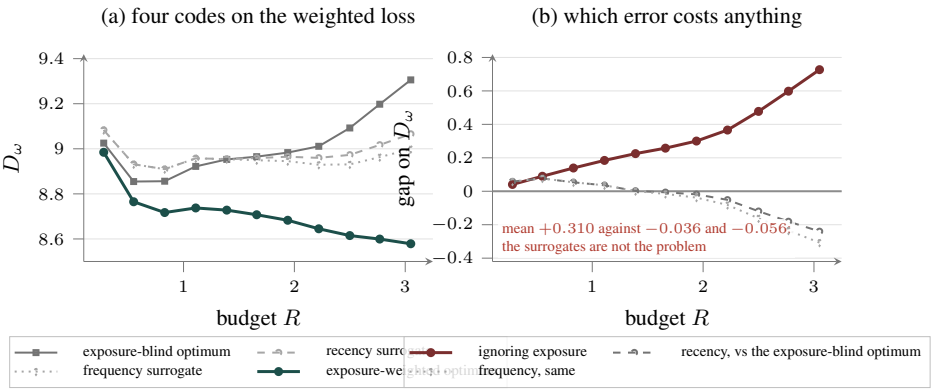



**Figure 15. The surrogates are not where the loss is, which is the opposite of what we drafted.** Measured; recency and frequency against the two optima on the weighted loss. We expected the standard surrogates to fail *worse* under weighting, on the argument that the highest-$\omega$ items — one-shot confirmations, codes, controls that cannot be un-flipped — are typically the *lowest*-traffic records, so a traffic-based score should be anti-correlated with weighted value on exactly the sub-population that decides the outcome. **That is not what the instance shows.** Both surrogates sit within $0.08$ of the exposure-blind optimum at tight budgets and slightly *better* than it at loose ones: mean gap $-0.036$ and $-0.056$ against the weighting gap of $+0.310$, a tenth the size and the other sign. *The reading is not that the surrogates are good.* It is that precision in an exposure-blind score buys nothing — once the ordering is on $\bar{\Delta}/c$ rather than $\omega\bar{\Delta}/c$, solving it exactly and solving it noisily land the same distance from the weighted optimum, because the error that matters is in the objective and not in the search. This *strengthens* the result: the gap the paper reports is weighting and not surrogate repair, so measuring against a deployed rule rather than against an exact exposure-blind optimum would not have inflated it. We keep the figure with its finding reversed rather than dropping it, because the argument it was drafted to make is one a reader will make anyway.

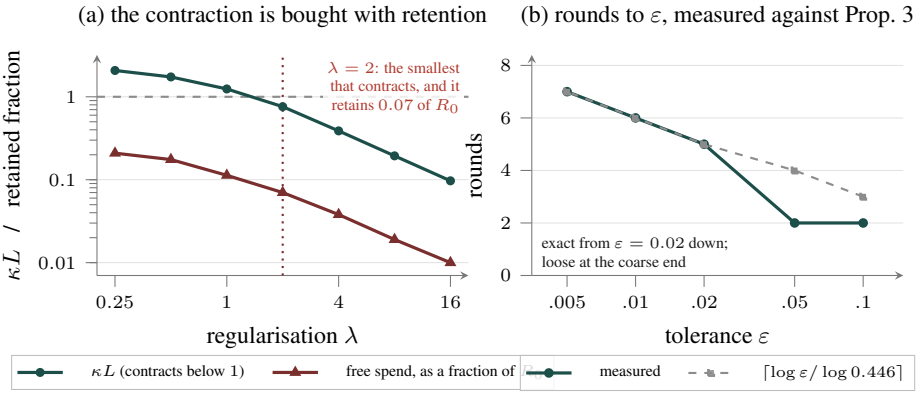



**Figure 16. What $\lambda$ buys, what it costs, and the round count that follows.** Measured. **(b)** is the round count behind the repair and it is what turns it from an exhortation into a budget: one on-policy rollout batch per iterate, and this many iterates. The bound is exact from $\varepsilon = 0.02$ down and loose above it, because the first two rounds move further than the asymptotic rate predicts. **(a) is the interaction the main text does not state.** Assumption 2 needs the argmin Lipschitz in $\mu$, and without a strongly convex relaxation it simply fails: a linear objective over a polytope has a *discontinuous* argmin, so an arbitrarily small change in $\mu$ swaps two densities and moves the solution by a whole item. $\lambda$ supplies the relaxation at a rate $1/\lambda$. But stationarity is $q_i = \mathrm{clip}((d_i - \nu)/\lambda, 0, 1)$, so the same $\lambda$ is what stops the code retaining anything — *the regulariser buys the contraction with the operator's own output*. On this instance the two regimes do intersect: $\lambda = 2$ is the smallest value that contracts and it still retains $0.070$ of full history. But that is a factor of $1.6$ below the $0.113$ retained at the largest $\lambda$ that fails to contract (namely $\lambda = 1$, where $\kappa L = 1.243$), so the assumption is not free and its price is denominated in the very quantity the operator exists to choose. Whether the regimes intersect at all is a property of an instance, and nothing in Theorem 1 guarantees it.

measured in, and P1's falsifier — a cycle — is a specific pattern in that sequence rather than a failure to converge within a budget. The per-step ratio ranges $0.23$–$0.96$: what the trace shows is contraction, not a constant rate.

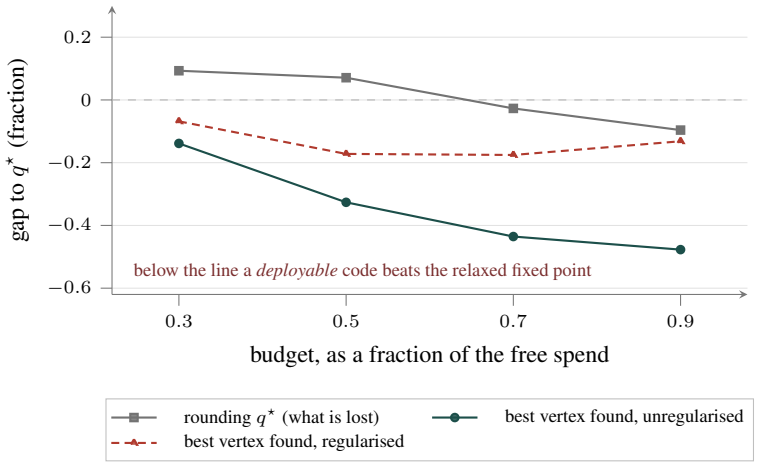



**Figure 17. What a deployable code costs, and what it buys.** Measured; 24 instances per budget, every code scored in the law *it* induces, which is the only comparison this paper's own thesis permits. *The comparison is unregularised.* The term $\frac{1}{2}\lambda\sum_i c_i q_i^2$ is the device that buys Assumption 2, appears in no deployed objective, and penalises a vertex harder than the fractional point it is rounded from — so including it would flatter the relaxation; the regularised reading is drawn too, so that the two measurements are comparable. **Rounding is cheap and its price is where one would expect it.** Taking $q^\star$'s items in decreasing order until the budget is spent costs $9.3\%$ of the distortion at the tightest budget and is *negative* by the loosest. **The searched vertex is the finding and it goes the other way.** A single-swap search inside the vertex set beats $q^\star$ at every budget — in $71\%$ of instances at the tightest and 96 to $100\%$ elsewhere. *The comparison that settles it is made on one instance, not across draws*: on the stability witness's own instance and budget the best vertex sits at $\mathrm{PR} = 8.007$, between that witness's fixed point $(8.950)$ and its continuous optimum $(6.804)$, closing $43.9\%$ of the stable-to-optimal gap while keeping a single item. *So this is not a defect of the relaxation.* It is §6's stable-versus-optimal gap arriving a second time, and the new information is that a substantial part of it is reachable by a *deterministic* code a system could ship — which raises the wall's price rather than lowering it, and makes the round count a lower bound for a second, independent reason.

**(ii) A signed gap, tested by its *direction* against a covariate.** Theorem 2 predicts that the off-policy minus on-policy valuation of the same item grows in retention strength. The interesting quantity is therefore not the gap's size at any rate but its *slope*, and the test is one-sided on that slope. This is the sharpest prediction in the paper because it is a claim about a derivative: a nuisance term of unknown origin gives a gap of arbitrary sign at each rate, and only a term generated by the loop gives a gap that grows with the intervention. *A flat gap refutes P2 even if the gap is large*, and Table 3 records that as the falsifier rather than treating a large flat gap as partial support.

**(iii) A set difference, tested on the retained set and not on the frontier.** Theorem 4 predicts that the two rules retain different items at equal budget. The statistic is the Jaccard overlap at matched rate, and the prediction is that it is bounded away from 1 — *wherever the budget binds*, a qualification the measurement forced and P4 now carries. *We report the overlap and not $\Delta(R)$ as the decisive column* (Proposition 6): a frontier gap is generic, and only a set difference is a behavioural claim. Alongside it we report the straddle count, because the theorem is an if-and-only-if and a system with no inversions is one the theorem predicts nothing about — an outcome we would report as *inapplicable* rather than as a failure.

**(iv) Probe accuracy, reported inline rather than in an appendix.** The recoverability probe's accuracy on each slice accompanies every $\omega$-weighted estimate computed from that slice. This is not a robustness section: $\omega$ enters multiplicatively, so a probe that mislabels recoverable states as irreversible *manufactures* separation, and without the probe's number the overlap in (iii) is uninterpretable rather than merely uncertain. Figure 14 is the calibration, and it goes the conservative way — but only for $\Delta(R)$. *On $\kappa_\omega$ the same bias is anti-conservative* ($+12.6\%$ at $\eta = 3$), because $\kappa_\omega$ is a ratio of extrema on which a mean absolute error is silent; an earlier version validated the probe on MAE alone and so reported only the flattering direction.

**Sensitivity, pairing, and what is deliberately *not* corrected.** $\eta$ is swept across its full range rather than fitted, and every weighted number is reported as a curve in $\eta$, because a result that held only at

one $\eta$ would be a result about $\eta$. Rate comparisons are paired within task, since between-task variance dominates. *One family is corrected and one is not, and the line between them is where an earlier version of this paper was wrong.* The convergence test in (i) is a single pre-registered examination of one sequence, so there is no family; and P5's *trend* test is an ordered alternative, where a multiplicity correction over individual horizons would discard the ordering that constitutes the prediction. But a claim about *one cell* of the horizon sweep is not a trend test, and that exemption does not reach it.

**The correction, run.** We therefore report the sweep's $40$ cells as the family they are. Under Benjamini–Hochberg at $q = 0.05$, 20 cells are rejected, *all of them positive separations*, the largest surviving $p$ being $0.014$. **What survives is the finding, and what does not survive is a cell we had over-read.** The non-monotonicity in $T$ survives with room to spare: the peak at $T{=}64$ $(0.362)$ is the smallest $p$ in the family, the fall to $T{=}128$ $(0.116)$ is rank $14$, and their instance bootstraps do not overlap, so the decline is established rather than suggested. The sign change at $T{=}2$, $1 - \beta = 0.3$ does *not* survive: at $p = 0.044$ it is rank $22$ against a threshold of $0.0275$. We previously called it "the only significantly negative cell" and drew a pocket on Figure 13 around it. *The pocket is nominal*; it is reported as nominal, it is not counted as evidence against Definition 1, and P5's outcome rests on the non-monotonicity, which does not need it.

**Common random numbers, and the one place the budget binds.** All arms share task order, environment seeds and decoding parameters, so a difference between arms is not inflated by independent sampling noise. The place this cannot be arranged is $\Phi$'s trace: each iterate requires a fresh rollout batch under a code that did not previously exist, so the iterate count is a genuine compute cost. Proposition 3 is what lets us budget it in advance rather than discover it.

### H.1 What is read by hand, and why a scalar cannot carry it

**The straddling band.** We read the items the two rules disagree about — kept by the weighted rule and dropped by the unweighted one at identical rate — and report what they are. The prediction is that they are low-traffic, high-consequence records: confirmation codes, one-shot tokens, the state of a control whose flip cannot be undone. The informative failure is a band dominated by items whose $\omega$ is high for a reason the probe got wrong, and that is a failure of the instrument rather than of the theory. *Reading the band is the only way to tell those two apart*, because both produce the same Jaccard overlap.

**The trajectory that changed branch.** We read the rollouts in which the shipped code and the fixed-point code diverge, and report where the divergence begins. The theory says it begins at the *branch point* and not at the demand: the shipped code fails because the agent went somewhere else, not because the retrieval was wrong. *That distinction is invisible in a success rate and it is the whole mechanism.* A divergence beginning at the demand instead would mean the witness is measuring retrieval quality and we have mis-instrumented it. On the real agent (§7) the reading is the same one: the bill-payment flow that cycles between the payee list and the amount screen has not retrieved the wrong fact, it has re-entered a screen whose completion it no longer records.

**A cycle, if we find one.** If $\kappa L \geq 1$ the trace can cycle, and we would report the cycle rather than the failure to converge. The prediction is that two branches trade places, each code inducing the demand the other serves. A period-two cycle between two branch-pinning codes is a *sharper* illustration of the wall than any fixed point, and we commit in advance to reporting it as the headline if it occurs, so that the commitment cannot be made after seeing which outcome flatters us.

**Why we read rather than only count.** Each of these readings exists because a scalar cannot carry it, and one has already changed the paper. Remark 2 exists because an earlier draft derived the contraction from Assumption 1 alone; the gap surfaced while writing out the two-branch instance by hand and checking which map each step needed, and not from any number. *A pre-registration is only honest if it also says which findings arrived outside it*, and that one did.

# I SCOPE, OBJECTIONS, AND WHAT REMAINS OWED

## I.1 THREE OBJECTIONS

**(a) The endogeneity is a second-order effect that averages out.** This is the natural objection and it has a precise form. The bias is bounded by a product of three factors — how far the induced law moves per unit change in the code, how far the solution moves per unit change in the law, and the value range — and the first two are exactly the quantities that are large in a GUI. *The objection is right about the order and wrong about the variable*: the effect is second order in the code perturbation and *first* order in the branch taken, and a branch is not a perturbation of a distribution but a different distribution. Nearly disjoint visited-state distributions are not an exotic case on an interface with one-way doors; they are the normal case, which is what Figure 13 is for, and what the $0.38 \to 0.497$ revisit jump of §7 is on a real agent.

**(b) The fixed point is a statement about an object nobody deploys.** Everything is proved on a compact convex relaxation and everything deployable is a deterministic code, a vertex of that set. **We accept the objection as stated and do not have the bound it asks for**: this paper does not bound the loss from rounding a relaxed code to a vertex. What we do is measure it (Figure 17), and it answers in two directions at once. *Rounding is cheap*: taking $q^\star$'s items in decreasing order until the budget is spent costs at most $9.3\%$ of the distortion at the tightest budget and is negative once the budget loosens. *And the vertex set is not impoverished*: a single-swap search inside it finds codes that *beat* $q^\star$ at every budget, in 71–$100\%$ of instances. Run on the same instance and budget Remark 5 uses — so that no inference across draws is needed — the best vertex sits at $\mathrm{PR} = 8.007$ between the fixed point's $8.950$ and the continuous optimum's $6.804$: a deployable code closes $43.9\%$ of the stable-to-optimal gap. So the relaxation is not costing us the object. It is costing the same thing Remark 5 already charged for, and a substantial part of that shortfall is reachable by a *deterministic* code a system could ship — which raises the wall's price and makes the two measurements one measurement.

**(c) The witness is a two-branch construction, so it establishes nothing about scale.** It establishes existence, which is what a witness is for, and its branching factor is the smallest that exhibits the phenomenon by design. *We do not present it as evidence of magnitude in the wild*; the real-agent loop and the device-walked census carry that, with correspondingly weaker interpretation because the confounds a real interface contains are exactly the ones the witness removes. The honest division is: the witness says the effect is real, the corpus says whether it is large, and no single experiment here does both.

## I.2 LIMITATIONS

**(i) Without the contraction condition the object changes.** If $\kappa L \geq 1$, uniqueness can fail and the object of study becomes a *set* of fixed points. Corollary 1 survives unchanged, since it needs only that the induced law depend on the code; Theorem 4 survives, since it compares two objectives rather than asserting either has a unique solution; and Theorem 3 survives, which is why this limitation is narrower than it looks — *the paper's title claim is proved by a two-point argument that never asks* $\Phi$ *to contract*, so $\kappa L \geq 1$ costs us uniqueness and the iterate count, not the result. What does not survive is reading $\Phi$'s trace as convergence to a point. *And the estimator can only falsify*: $\kappa L_{\max}$ is a maximum of $\rho(\Phi q, \Phi q')/\rho(q, q')$ over sampled pairs, hence a *lower* bound on the Lipschitz constant, so $400$ pairs finding nothing above $0.829$ is a failed search for a violation and not a verification. The independent evidence that the fixed point is a point is the agreement of eight deliberately different initialisations to $2.4 \times 10^{-9}$, which no single trace supplies. And the modulus is a property of the instance's *shape* rather than its size: Table 4 measures it against the item-to-state ratio, and $\kappa L < 1$ holds on every seed at $n/m \lesssim 4$ at both witness sizes we ran, where the straddle count *grows* with scale. The two regimes are not a seam; they overlap and the overlap widens.

**(ii)** $\eta$ **is reported as a sweep, not fitted**, because a result holding only at a fitted $\eta$ would be a result about $\eta$. Proposition 5 additionally says it is not free in the first place: it is the instance's tail-to-step value ratio.

**(iii) The recoverability probe has error that propagates into** $\omega$, hence into every weighted quantity, which is why the degraded-probe variant of $\Delta(R)$ is reported alongside the clean one and why Appendix H(iv) reports the direction of the bias on *both* quantities it affects rather than on the flattering one.

**(iv) The fixed point is estimated by rollout** and inherits Monte-Carlo variance, reported separately from probe error so the two cannot be confused for one interval.

**(v) The environment is fixed given the code.** $\Phi$ moves the state law only through the code; an interface that changes on its own is an exogenous shift this paper does not model, and the mixed case has a formalisation we do not use (Lee & Zrnic, 2026). On a GUI suite over months this is the limitation we would expect to bind first, and it bounds Proposition 3 from the other side: iterating to a fixed point is worth its rollouts only while the fixed point stays put.

**(vi) The policy is fixed given the code.** A policy that continues to learn while the code changes is a second loop and we have not touched it. This is the limitation we regard as most likely to matter, because the experiential-memory systems audited in §2 are exactly the case where the second loop is tight.

### I.3 What remains owed

*Named rather than gestured at.* **(1) The suite half of the experimental plan**, and it belongs first: a re-stratification of published task suites by recoverability of the pending action, which has not been run. It is the one item neither a witness nor our own corpus can supply, because it needs published suites and their accuracies; Levy et al. (2024) already annotates the irreversible actions it would need. **(2)** $\kappa L$ **on a deployed system.** This is the claim a witness structurally cannot settle — $\kappa L < 1$ is a property of how strongly a *deployed* policy adapts to its own retention code, and we chose our witness's modulus when we chose the witness. The theorems are safe from it; the scoping claim is not. **(3) A two-annotator round on the pending-action question** — one item per page, 120 items — which would give reading (B) of Appendix F a ceiling it currently lacks. **(4) A bound, not a measurement, on the rounding loss** objection (b) asks for. **(5) The second loop** of limitation (vi). **(6) Withdrawn from this list:** a four-environment token census, which an earlier version owed in order to establish that cost–content decoupling holds in the wild. Figure 8 measures that decoupling is *not* the operative condition, so the census would have established something true and irrelevant.

## J Registration Index

Every deciding read-out rule was frozen before the run it decides, with all of its branches named in advance, so exactly one fires and none is selected after the fact. The supplementary material carries one row per registered claim, the commit that froze it, the timestamp of the result it decided, and the branch that fired. The registrations most load-bearing here are:

**E2, the iteration trace.** Registered as: $\Phi$'s iterates converge geometrically at the witness's modulus, with the failure branch "the trace does not contract, in which case Theorem 1(2)'s hypothesis is reported as unmet on this witness".

**E4, the signed gap against its covariate.** Registered as a test of the gap's *direction* against $\Sigma$ rather than of its magnitude, precisely because Remark 1 rules out a general lower bound.

**E6, the two-sort diagnostic.** Registered on the retained set, not on the frontier, with the straddle count as the reported statistic and the binary verdict as secondary — which is why §7 reports the control that makes the binary verdict cheap at $n = 120$.

**E-real-1, the law shift on a real agent.** Registered with the budget *swept* rather than chosen, and with the falsifier named in advance: an unbudgeted recency arm is the full-history code and must return the noise floor. It does, to three decimals.

**Withdrawn.** The horizon-growth statement was registered and is withdrawn on the argument of Proposition 2 rather than on a failed measurement; §6 reports it as a withdrawal and not as a null result.

# K Notation, the Claim Ledger, and the Register of Corrections

## K.1 Notation

**Table 8.** Symbols used throughout. Where a symbol changed meaning between versions of this work, the current meaning is the only one stated and the change is recorded in Table 10.

| symbol | name | what it is |
|---|---|---|
| $\Phi$ | induced-distribution operator | a code to the optimal code under the law that code induces (Def. 2) |
| $T$ | horizon | number of steps in an episode. *This letter has no other meaning in this paper* |
| $q$ | retention code | a rate-constrained lossy code on histories |
| $q_0$ | full-history code | the reference code every write-time criterion evaluates under |
| $q^\star$ | fixed point | $\Phi(q^\star) = q^\star$; performatively *stable*, not optimal (Rem. 5) |
| $R$ | rate | the budget the code must fit in; $\sigma = 1 - R/R_0$ is retention strength |
| $\mu_q$ | induced law | the law over states visited under code $q$ |
| $\Sigma$ | law shift | $\|\mu_q - \mu_{q'}\|_{\mathrm{TV}}$, the quantity §7 measures on a real agent |
| $\kappa$ | induced-law sensitivity | Lipschitz constant of $q \mapsto \mu_q$ (As. 1) |
| $L$ | solution sensitivity | Lipschitz constant of the inner argmin (As. 2) |
| $\kappa L$ | contraction modulus | $\Phi$ contracts when $\kappa L < 1$; measured, never assumed |
| $\beta_m(s)$ | recoverability | $\mathbb{E}_{a\sim m_s}[a$ undoable at $s]$; *the action measure $m$ is part of it* (App. F) |
| $\eta$ | exposure gain | $\omega = 1+\eta(1-\beta)$; the tail-to-step value ratio, not a free parameter (Prop. 5) |
| $\omega(s)$ | exposure weight | the price of arriving at $s$ with degraded memory |
| $\kappa_\omega$ | field concentration | $\omega_{\max}/\bar{\omega}$; in closed form ($\star$) a function of $\beta_{\min}$ and $\bar{\beta}$ alone (Rem. 3) |
| $r(R)$ | boundary ratio | the density ratio at the budget cut that the exposure spread must beat |
| $n, m$ | items, states | $n/m$, not $n$, decides whether $\Phi$ has one fixed point or a set |
| $\Delta_i, c_i$ | score, cost | what a deployed store already ranks its candidates by |
| $D, D_\omega$ | distortion | unweighted and $\omega$-weighted decision distortion |
| $\Delta(R)$ | separation | the gap between the two rules' frontiers at rate $R$ |
| PR | performative risk | $D_\omega(q;\mu_q)$: what a designer actually pays |
| $\lambda$ | relaxation strength | buys As. 2, and pays for it in retention (Fig. 16) |

## K.2 The claim ledger

Table 9 lists every claim this paper makes with its evidential status. Rows marked *measured* name the analyzer that produced the number; rows marked *not run* state what would be measured and what would refute it.

**Table 9.** Every claim, with its status. *Six rows are corrections to earlier claims of our own and three go against the paper's thesis*; they are in the same table as the rest.

| **Claim** | **Status** | **Evidence** |
|---|---|---|
| 1. $\Phi$ has a fixed point on the relaxation, unique and geometrically reached under contraction | formal; existence needs only continuity, uniqueness needs As. 1+2 with $\kappa L < 1$ | Thm. 1 |
| 2. Solution stability is a *second* assumption, not a consequence of induced-law regularity | correction to our earlier statement | Rem. 2 |
| 3. The law shift $\Sigma$ governs the write-time criterion's error, from above in general and from below on a witness family | formal; **no general lower bound** | Thm. 2, Rem. 1 |

*continued overleaf*

| Claim | Status | Evidence |
|---|---|---|
| 4. No write-time criterion is optimal unless memory does not affect behaviour | formal | Cor. 1 |
| 4a. Every write-time criterion leaves $\geq \sigma/2$ unclaimed on a two-instance class, $\geq \sigma$ if deterministic | formal and **unconditional**: needs neither assumption; both bounds attained | Thm. 3 |
| 4b. $\Phi$ is performative prediction's decoupled-risk map and $q^\star$ its *stable* point | prior work, named rather than rediscovered | Perdomo et al. (2020), App. E |
| 4c. $q^\star$ is stable, *not* optimal: $24.0\%$ of the performative risk unclaimed | **measured, and against us**; makes Prop. 3 a lower bound on the price | Rem. 5 |
| 4d. That the shift is endogenous to one non-strategic learner is *not* a difference of ours | prior work, and a **withdrawal** of an earlier claim; what survives is the feasible set | Mandal et al. (2023), App. E |
| 4e. This paper is the fully *endogenous* edge; a drifting interface is out of scope | scoping, against us; the mixed case is formalised elsewhere | App. I(v) |
| 5. $\Delta(R) > 0$ at every rate in the non-degenerate range | formal, near-trivial, presented as bookkeeping | Prop. 6 |
| 6. The two rules retain different sets iff a straddling inversion exists | formal; the load-bearing claim | Thm. 4 |
| 6a. Decidable in one pass — two sorts and a budget scan, $O(n \log n)$, no rollout and no model | formal; the paper's one directly actionable statement | Cor. 3 |
| 6b. Under a knapsack there is no global cut and $r(R) \geq 1$ can fail; necessity of the pairwise condition survives | **correction**, found by us after the corollary was stated | Rem. 4 |
| 6c. Whether the exchange is *two-sided* is decided by $\kappa_\omega$, not by Thm. 4's condition; the regime is a bounded window | **measured**, and it qualifies our own theorem's picture | Fig. 3(b) |
| 6d. $\kappa_\omega$ on a real interface depends on the action measure, and the two natural choices *straddle* that window | **measured**: $1.473$ against $3.014$ at $\eta = 3$ on the same 110 pages, disjoint bootstraps, correlation $-0.006$ | App. F |
| 7. The separation grows with the number of irreversible decision points | **withdrawn as a theorem, kept as a tangent**: premise and conclusion are mutually exclusive | Prop. 2, §6 |
| 7a. Measured, the separation is non-monotone in $T$, and on the literal axis a *window*, zero at both ends | **measured, and against us** | Fig. 4, Fig. 6 |
| 7b. A per-state $\omega$ is horizon-blind, capturing $0.20$–$0.34$ of the attainable gain by $T{=}128$ | **measured, and against us** | Prop. 4, Fig. 6(b) |
| 7c. $1 + \eta(1 - \beta_t)(T - t)$ removes that deficit: $0.94$–$0.98$, and the share *rises* in $T$ | **measured**; still a write-time rule | Fig. 6(b) |
| 7d. Both weights are truncations of one exact marginal $1 + (1 - \beta_t)G_t/g_t$; $\eta$ is not free | formal, verified against the recursion to $5.7 \times 10^{-8}$; *resolves* a split we had recorded as a defect | Prop. 5 |
| 7e. $p^\star = 1/(2 + \eta)$, the only parameter-free prediction for the window's interior peak, is **refuted in direction** | **measured**; refutes a conjecture of ours | Fig. 4(b) |
| 8. Cost–benefit greedy retains $(1 - 1/e)$ under weighting | formal, *with partial enumeration* and *only where the value is modular*; withdrawn on the multi-step recursion | Prop. 1, Prop. 7 |
| 9. Reaching $\epsilon$ of the fixed point costs $\lceil \log \epsilon / \log(\kappa L) \rceil$ rollout rounds | formal; a *lower* bound on the price, by row 4c | Prop. 3 |
| 10. The predicted iteration, bias curve, frontier gap and set overlap are observed | **measured**; P1–P3 supported, P4–P5 partial, P6 failed | Tab. 3 |
| 10a. The law shift on a *real* agent: 7B model, 19 device-walked flows | **measured**: $\Sigma = 0.330$ at a quarter budget, $3.28\times$ a measured floor, $17/19$ flows; the unbudgeted falsifier returns the floor exactly | Tab. 2 |

*continued overleaf*

| Claim | Status | Evidence |
|---|---|---|
| 10b. The bias sweep tests $\Sigma \gg 0$, *not* monotonicity in $\sigma$ | measured; a limit of the *witness*, which the real agent does not share | Fig. 2(b) |
| 10c. Thm. 4's condition is *met* on this paper's own reference configuration | **measured**; *corrects* a caption that read a common ranking of six codes as the absence of an item-level inversion | §7.1 |
| 10d. The contraction and separation regimes overlap on $n/m \lesssim 4$, and the overlap widens | **measured**; *corrects* our own "larger witnesses were tried and rejected", which held $m$ fixed | Tab. 4 |
| 10e. The $T{=}2$ sign change does *not* survive Benjamini–Hochberg; the non-monotonicity does | **measured**; *withdraws* a result we had reported *against* ourselves | App. H |
| 10f. Cost–value decoupling, the structural condition we motivated with, does *not* control the separation | **measured**, and it withdraws a motivating claim of ours | Fig. 8 |
| 10g. A deployable vertex closes $43.9\%$ of the stable-to-optimal gap; rounding costs $\leq 9.3\%$ | **measured**; answers objection (b) rather than asserting it | Fig. 17 |
| 11. $\kappa L < 1$ on the witness | **searched, not proved**: no ratio above $0.829$ over $400$ law pairs; a sample maximum is a *lower* bound, so this fails to falsify rather than establishes | App. I(i) |
| 11a. $\kappa L < 1$ in *deployed* GUI agents | scoping; **not run**, and a measured cycle would be reported as a finding rather than a failure | App. I |
| 12. Re-stratification of published suites by recoverability of the pending action | **not run**; the protocol is stated and the annotation exists in Levy et al. (2024) | App. I |

### K.3 The register of corrections

*We keep this register rather than a clean narrative, because the alternative is a paper that reads as though it had been right the first time.* The main text carries a one-line pointer at each site and no more. Nothing here is a typographical fix; each row is a claim we withdrew or repaired, and the rows marked ‡ we found ourselves after the rest of the table existed.

**Table 10.** What was claimed, what replaced it, and where the replacement is proved or measured.

| claimed | replaced by | where |
|---|---|---|
| Horizon amplification, as a theorem | A tangent to an interior window, demoted; the additive hypothesis holds only where the separation vanishes | §6, Prop. 2 |
| The shift being endogenous to one agent is our difference from performative prediction | It is performative RL's setting. What survives is the feasible set: a code, not a policy | App. E |
| Thm. 2(2) holds on *every* instance | False where the off-policy solution coincides with the fixed point; stated as an existence claim on a witness family | Rem. 1 |
| The contraction follows from As. 1 alone | It does not; argmin stability is a second and independent hypothesis | Rem. 2 |
| Cost–content decoupling carries the separation | The exposure half carries it; a cost–value rank sweep leaves the separation inside its own interval | Fig. 8 |
| "the inversion Thm. 4 permits is not reached at this budget" | A common ranking of six *codes* is a different object from an inversion between *items* at the boundary | §7.1 |
| Larger witnesses "tried and rejected" at $n \geq 144$ | Confounded: the item set grew with $m$ held at $8$. The modulus is governed by $n/m$ | Tab. 4 |
| $(1 - 1/e)$ for cost–benefit greedy | Only *with partial enumeration*, and not at all on the multi-step recursion, where the value is strictly supermodular | Prop. 1, Prop. 7 |

*continued overleaf*

| claimed | replaced by | where |
|---|---|---|
| The analysed rule and the recommended rule differ (recorded as an open defect) | Both are truncations of one exact marginal, and the recommended rule's contraction is now measured | Prop. 5, Tab. 5 |
| The step norm against the geometric line shows the gap closing | It sits above the reference at nine of eleven iterates; the comparison, not the trace, was wrong | Fig. 5 |
| ‡ Cor. 3: "they differ whenever $\bar{\omega}_{\max}$ is dropped and $\bar{\omega}_{\min}$ kept with ratio $> r(R)$" | False; the $\bar{\omega}$ extrema need not sit at the density boundary. The $68.3\%$ we read as a disagreement rate is $43.1\%$ | Rem. 4 |
| ‡ Cor. 3's test, and $r(R) \geq 1$, under a knapsack budget | The retained set is not a prefix, so $r(R) \geq 1$ can fail — our own $r = 0.978$ already showed it. Restated at the greedy's boundary pair | Rem. 4 |
| ‡ The real-corpus verdict shows *measured* recoverability drives the separation | The control says otherwise: shuffled and uniform $\beta$ give the same verdict and magnitude. What is carried is non-constancy of $\omega$ | App. F |
| ‡ Disagreement grows as the budget loosens | A witness finding, withdrawn as a general one: on a real corpus the overlap rises monotonically $0.698 \to 0.932$ | App. F |
| ‡ $\kappa_\omega$ compared against $[2.35, 4.66]$ | Those numbers carry an extra factor $(T-1-t)$ and are on another scale; at $\eta = 3$ no such field can exceed $4$ at all. The shortfall we claimed was overstated about twofold | Rem. 3 |
| ‡ "the flat $\kappa_\omega$ is structural; no instrument redesign can reach the window" | False, and asserted twice. $(\star)$ gives the reachable set exactly: $\bar{\beta} \geq 0.549$ at $\beta_{\min} = 0$ suffices at $\eta = 3$ | Rem. 3 |
| ‡ Probe validated on mean absolute error alone | Every claim runs through $\kappa_\omega$, a ratio of extrema on which an MAE is silent. The probe's $+0.089$ bias is conservative for $\Delta(R)$ and *anti*-conservative for $\kappa_\omega$ | App. H(iv) |
| ‡ "the only significantly negative cell", with a pocket drawn around it | One of $40$ cells; it does not survive Benjamini–Hochberg at $q = 0.05$. Reported as nominal, drawn and not claimed | App. H |
| ‡ Def. 1 left the action measure implicit | Two incompatible readings were live, giving $\kappa_\omega = 1.473$ and $3.014$ on the same $110$ pages, on opposite sides of our own window, correlation $-0.006$ | App. F |
| Five fixed deviation slots, applied to every page in the same order | Two of the five are recoverable on every page by their own definition, flooring $\beta$ at $0.40$ and making a one-way door unobservable; $240$ of $1{,}200$ labels carry no information | App. F |
| A four-environment token census, owed | Withdrawn: the condition it would establish is not the operative one | App. I |

## L Protocol and Secondary Material

**The census.** $120$ pages across $15$ application flows for the diagnostic, and $19$ flows / $158$ pages for the agent loop, walked screen by screen on real devices, each page recording the fact it carries and whether a later step of the same flow demands it. The page-level demand column was walked by one annotator and therefore has no inter-annotator ceiling; it carries sampling uncertainty and no agreement uncertainty, and the flows are a convenience sample of consumer applications rather than a probability sample.

**Recoverability.** Two rounds. The *calibration* set is $60$ constructed states, two annotators, $600$ labelled one-step deviations, $\kappa = 0.85$, and fixes the probe's error bar; its $\beta$ distribution is a design rather than a reading. The *corpus* set labels the census pages themselves with two annotators and a measured spread of $0.042$.

**The agent loop.** A 7B open-weights instruction-tuned model at temperature $0$, given a real application goal, opening one of that flow's recorded screens per step or finishing. No transition and no reward is authored; the branching is the model's. $\Sigma$ is total variation between the two arms' empirical laws over visited screens, the budget is swept, and the noise floor is the same code re-run rather than

an assumed value. Per-flow trajectories are shipped so that $\Sigma$ can be recomputed without calling the model.

**Matched rate.** Every comparison between a weighted and an unweighted rule is made at equal token budget. Arms share task order and seeds under common random numbers.