WHEN SOURCE CORRECTION IS NOT ENOUGH: A THEORY OF MISTAKE INFLUENCE IN AGENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Correcting an agent's past mistake may leave its influence on later decisions intact. We distinguish a mistake's presence in context, its repetition probability, and its causal influence, then separate this influence into the original output's direct effect and a trail effect carried by subsequent history. A three-state construction shows that removing direct influence gives no uniform bound on recovery time, even when influence eventually vanishes. We derive sufficient conditions for fading and finite-horizon persistence, and identify an explicit fading-to-persistence threshold in a specified nonlinear feedback model. Exact enumeration of four-step histories under recorded binary-action policies spans 300 arithmetic problems and three models: Qwen3-Coder, Llama 4 Maverick, and DeepSeek V4 Flash. At decision four, trail influence exceeds direct influence in magnitude in 90.3–98.4% of settings, although its magnitude and direction vary substantially. On six development problems, supplementing source correction with recomputation and a visible repair record reduces mean mistake reuse by 32.7–53.4 percentage points across two menu orders. A separate coding-agent study finds that reminders can reduce subsequent violations while leaving files created by earlier violations intact, with responses differing across models.

1 INTRODUCTION

LLM agents often reuse their own outputs as inputs, allowing earlier responses to shape subsequent reasoning and actions [23]. An agent may build on its earlier reasoning, retain generated summaries in memory, or read files created in previous steps. This process can improve task performance by allowing models to revise answers based on their own critiques or adjust later actions to address errors identified in previous attempts [13, 20]. Nonetheless, this process can also allow mistakes in earlier outputs to influence later reasoning [24]. Models may generate further false claims to justify an incorrect answer or continue relying on mistaken assumptions across future decisions, making it difficult to recover from the initial error [24, 11]. These errors can be difficult for models to correct without external feedback [8]. As an interaction continues, models may also follow patterns established by earlier responses, even when doing so conflicts with their assigned goals [1].

An intuitive solution is to correct a mistake once it is detected. However, information derived from it may remain in later reasoning, summaries, or files and continue to affect the model's decisions. Correcting the original output may therefore leave some of its influence intact. This leads us to the following question:

What influence does correcting the source of error remove, what influence remains, and which continuation dynamics erase or preserve the difference?

Answering this question requires separating three quantities. *Presence* describes whether the original output remains accessible, while *preference* describes how likely the model is to produce the target mistake given its history. *Influence* measures the difference in later behavior between retaining and minimally correcting the original output. A model may retain the original statement without relying on it, or continue using information derived from it after the statement is removed. Both continuations may also have the same high probability of producing the target mistake, despite differing in whether the original output was corrected. Neither removing the original statement nor measuring final accuracy alone therefore establishes whether the mistake still influences later decisions.

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.

We measure this influence by comparing two continuations from a checkpoint immediately after a mistake and before the model reuses it. One retains the original output, while the other minimally corrects it. A third evaluation corrects the original output while keeping the later history generated by the mistaken continuation unchanged. Together, these evaluations separate the direct effect of the original output from the effect carried by differences in the distribution of later histories. The resulting identity specifies which interventions to compare and allows us to study how each effect changes over time.

Our theory has three parts. Firstly, a finite-state construction shows that removing direct influence does not provide a uniform bound on recovery time, even when influence eventually vanishes. Secondly, we derive bounds on influence and mistake retention that give sufficient conditions for fading or persistence. These bounds account for sensitivity to the original output, propagation across state components, residual forcing, and the probability that the conditions needed for retention fail. Thirdly, we analyze a nonlinear feedback process and show that influence vanishes at or below a critical feedback strength and persists above it. This threshold applies to the specified stochastic process and does not establish a phase transition in LLMs.

We study these effects empirically by enumerating complete binary-action trees for three models. This allows us to measure the direct and history effects, including whether each increases or decreases the probability of the target mistake. The experiments also show when the theoretical conditions give useful guarantees and when the implemented bounds are loose or uninformative. A separate coding study injects mistakes into forked continuations and measures later scope violations and the files left behind after a reminder. This study shows that behavior can improve while files created earlier remain, but does not establish that those files cause further violations.

Contributions. We separate the direct influence of a mistake from the influence carried by later history and show that removing direct influence gives no uniform bound on recovery time. Our bounds give conditions for fading or persistence, accounting for interactions between state components and cases where retention conditions fail. A nonlinear feedback model gives an explicit threshold between fading and persistence. Exact finite-policy measurements across three models quantify both effects and examine the bounds. A six-problem development study shows that repairing affected computations can reduce mistake reuse beyond source correction alone, while a separate coding study distinguishes improvements in later behavior from changes in workspace state.

1.1 Related Works

Error propagation and self-correction. Previous work shows that an initial incorrect answer can lead a model to generate further false claims [24]. Studies of self-correction distinguish difficulty identifying errors from difficulty repairing them: models can struggle to correct their reasoning without external feedback, while providing the location of an error can improve correction [8, 21]. Work on misinformation propagation further shows how incorrect user inputs affect subsequent reasoning and finds that earlier corrections are more effective [5]. Related work compares corrections to root and descendant errors and models error propagation through a branching process [9]. Our analysis asks what influence remains after the original mistake is corrected. We compare retaining the mistake, correcting it before reuse, and correcting it while keeping the later history generated from it fixed. These comparisons separate the direct effect of the original output from the residual effect carried by later history, including whether that effect increases or decreases mistake probability. We then show that eliminating direct influence gives no uniform recovery-time guarantee and derive sufficient conditions for the remaining influence to fade or persist.

Instruction instability and goal drift. Models can become less consistent in following their system instructions as a conversation continues [12]. In agent settings, departures from assigned goals have also been associated with increasing reliance on patterns established by earlier responses [1]. These findings motivate studying how generated history affects later decisions. Our analysis focuses on the influence of a particular mistaken output, measured by comparing continuations in which that output is retained or minimally corrected. The comparison distinguishes the influence of the mistake from errors that would also occur after correction. It also allows us to study whether that influence persists without assuming that instruction adherence decreases with context length.

Causal effects and stochastic dynamics. Our decomposition follows the distinction between direct and indirect effects in causal inference [18, 16]. We apply this structure through edits to the original output, including evaluations that correct it while keeping the subsequent history fixed. The retention bounds use exponential-supermartingale concentration to control the probability that the model's preference for a mistake falls below a specified threshold [6, 7]. The fading bounds use coupling and perturbation arguments to control how differences between continuations evolve [19]. For finite action trees, we compute bounds through dynamic programming over causal couplings [22, 15]. The nonlinear feedback model draws on the study of reinforced processes and stochastic approximation [2, 17]. Our contribution is to apply these established methods to the correction of generated mistakes, connecting edits to the original output with conditions on the subsequent evolution of the agent's state.

2 THEORY

2.1 PRESENCE, PREFERENCE, AND INFLUENCE

Fix a task, model configuration, continuation policy, and specified family of target mistakes. Save a checkpoint immediately after a mistake m and before its reuse. Continuation 1 retains m; continuation 0 replaces it with a minimal correction m^0. The remaining initial state is identical. Both follow the same policy, including any scheduled verification instructions. All expectations condition on the task and initial checkpoint.

Let $\mathcal{F}_t$ contain the full execution history. At an active decision, $Y_{t+1} = 1$ indicates a target mistake. Define

$$p_t = \Pr(Y_{t+1} = 1 \mid \mathcal{F}_t), \qquad X_t = \log\left(\frac{p_t}{1-p_t}\right), \qquad \sigma(x) = \frac{1}{1+e^{-x}}. \tag{2.1}$$

The logit X_t requires $0 < p_t < 1$ but does not assume logistic feedback. The influence results only require measurable probabilities in $[0, 1]$; the retention result also requires finite logits.

Call the original output the *source*. Let K_t contain everything except its slot: the ordered history, summaries, workspace, memory-access information, and execution status. For $a \in \{0, 1\}$, define

$$\begin{aligned} p_t^a(k) &= \Pr(\text{target mistake at the next decision} \mid k \text{ with source } a), \\ \nu_t^a &= \mathcal{L}(K_t^a). \end{aligned} \tag{2.2}$$

Changing the source keeps later outputs and summaries fixed, rebuilds affected computational caches, and does not reinsert an inaccessible source. Independent copies elsewhere belong to K_t and remain unchanged. Edited checkpoints must have an executable interpretation. Appendix B gives the checkpoint and termination conventions.

Define the average mistake probability in each continuation and their difference by

$$\bar{p}_t^a = \int p_t^a(k)\, \nu_t^a(dk), \qquad \Delta_t = \bar{p}_t^1 - \bar{p}_t^0. \tag{2.3}$$

The corrected continuation can still make mistakes, and Δ_t can be negative. Influence *fades* when $\Delta_t \to 0$, without requiring either mistake probability to approach zero. Source accessibility (*presence*), mistake probability (*preference*), and the paired difference (*influence*) are therefore distinct.

2.2 WHAT CORRECTING THE SOURCE CHANGES

We compare the mistaken continuation (W_t), the initially corrected continuation (C_t), and a source correction applied to the mistaken continuation's later histories (SWAP_t), keeping their probabilities unchanged:

$$\mathrm{W}_t = \int p_t^1\, d\nu_t^1, \qquad \mathrm{C}_t = \int p_t^0\, d\nu_t^0, \qquad \mathrm{SWAP}_t = \int p_t^0\, d\nu_t^1. \tag{2.4}$$

Definition 2.1 (Direct and trail influence)**.** Define

$$D_t = \int \left[p_t^1(k) - p_t^0(k)\right] \nu_t^1(dk) = \mathrm{W}_t - \mathrm{SWAP}_t \tag{2.5}$$

Evaluation	Source	Later history	Mistake probability
W	m	Generated under ν_t^1	$\int p_t^1 \, d\nu_t^1$
SWAP	m^0	Same histories and probabilities	$\int p_t^0 \, d\nu_t^1$
C	m^0	Generated under ν_t^0	$\int p_t^0 \, d\nu_t^0$

Figure 1: Comparing the first two rows gives D_t; comparing the last two gives M_t. The source edit keeps later history fixed. A history may be reachable under both continuations with different probabilities.

and

$$M_t = \int p_t^0(k)\, \nu_t^1(dk) - \int p_t^0(k)\, \nu_t^0(dk) = \mathrm{SWAP}_t - \mathrm{C}_t. \tag{2.6}$$

Then $\Delta_t = D_t + M_t$, where D_t is direct influence and M_t is trail influence carried by later history.

The direct component measures what correcting the source now removes; the trail component measures what remains because the history distributions differ. Both are signed effects, not nonnegative shares of error. Reversing the reference gives another valid decomposition. We use the corrected source to measure what remains after the edit.

If neither source value is accessible and no other source-dependent execution channel exists outside K_t, then $p_t^1 = p_t^0$ and $D_t = 0$. However, ν_t^1 and ν_t^0 may still differ.

Theorem 2.2 (No uniform recovery time)**.** *For every integer $H \geq 1$ and every $\delta \in (0, 1)$, there is a continuation process with three states and finite logits such that the source is inaccessible for all $t \geq 1$, but*

$$D_t = 0, \qquad M_t = \Delta_t \geq 1 - \delta \qquad (1 \leq t \leq H), \tag{2.7}$$

even though $\Delta_t \to 0$ as $t \to \infty$.

Construction. Let $q \in (0, 1/2)$. At the common initial checkpoint, the source sets the first mistake probability to $1 - q$ in continuation 1 and q in continuation 0. A mistake leads to state H and a correct output to L; the source is then removed. State H has mistake probability $1 - q$ and stays there after a mistake. State L has mistake probability q and stays there after a correct output. Otherwise, either moves to an absorbing state N with mistake probability $1/2$. Calculation gives

$$M_t = \Delta_t = (1 - 2q)^2 (1 - q)^{t-1}, \qquad t \geq 1. \tag{2.8}$$

Choose $q = \delta/(H + 3)$. The bound $(1 - 2q)^2(1 - q)^{t-1} \geq 1 - (t + 3)q$ proves the finite-horizon claim, while $q > 0$ gives eventual fading. □

Source removal and eventual fading alone therefore give no uniform recovery-time bound. Additional assumptions on the continuation are needed. Here both mistake probabilities converge to $1/2$, so influence disappears without errors disappearing. Appendix C gives the calculation and an exactly attained fading bound.

The sequence M_t evaluates corrections at different checkpoints. Recovery after one fixed late correction instead requires continuing from that edited checkpoint. The experiments keep these comparisons separate.

2.3 When influence fades or persists

We give separate conditions for fading and persistence. Failure of a sufficient fading condition alone does not prove persistence.

2.3.1 Bounding differences between continuations

Fix a *coadapted coupling* (K_t^1, K_t^0): conditional on its joint history $\mathcal{G}_t$, each continuation retains its original transition rule $P_t^a(K_t^a, \cdot)$. An independent coupling always exists, but a useful propagation bound requires additional structure.

Choose measurable, integrable discrepancy vectors

$$d_t(k, k') \in [0, \infty)^r, \qquad d_t = d_t(K_t^1, K_t^0).$$

These may describe differences in actions, memories, or workspace. They need not form a metric or fully describe the state, but must control the mistake probability.

Let ℓ_t, f_t be vectors, G_t a matrix, and ε_t a scalar. Assume they are deterministic and nonnegative. Choose a finite vector $z_0 \geq \mathrm{E}d_0$ and suppose

$$|p_t^0(k) - p_t^0(k')| \leq \ell_t^\top d_t(k, k') \tag{2.9}$$

for all reachable pairs in the coupling,

$$|D_t| \leq \varepsilon_t, \tag{2.10}$$

and

$$\mathrm{E}(r_t)_+ \leq f_t, \qquad r_t = \mathrm{E}[d_{t+1} \mid \mathcal{G}_t] - G_t d_t. \tag{2.11}$$

Vector inequalities and positive parts are componentwise. The allowance f_t bounds expected positive residual growth. A fitted residual with mean zero does not justify setting this allowance to zero.

Theorem 2.3 (Conditions for fading)**.** *Let $z_0^\star = z_0$ and $z_{t+1}^\star = G_t z_t^\star + f_t$. Under conditions (2.9)–(2.11),*

$$\begin{gathered} |M_t| \leq \ell_t^\top z_t^\star, \qquad |\Delta_t| \leq \min\{1, F_t\}, \\ F_t = \varepsilon_t + \ell_t^\top z_t^\star. \end{gathered} \tag{2.12}$$

If these assumptions hold for all t, $G_t = G$ has spectral radius less than 1, $\sup_t \|\ell_t\| < \infty$, and $f_t, \varepsilon_t \to 0$, then $D_t, M_t, \Delta_t \to 0$.

Proof sketch. Taking expectations gives $\mathrm{E}d_{t+1} \leq G_t \mathrm{E}d_t + f_t$, hence $\mathrm{E}d_t \leq z_t^\star$. Since $M_t = \mathrm{E}[p_t^0(K_t^1) - p_t^0(K_t^0)]$, condition (2.9) bounds $|M_t|$; adding ε_t bounds $|\Delta_t|$. For stable G and $f_t \to 0$,

$$z_t^\star = G^t z_0 + \sum_{j<t} G^{t-1-j} f_j \longrightarrow 0.$$

Appendix D gives the full proof. □

Fading requires control of direct influence, inherited differences, and newly generated differences. Source removal can set $\varepsilon_t = 0$ while leaving the other terms uncontrolled. Theorem 2.2 admits a coupling with $G = 1 - q$ that makes (2.12) exact. Tightness therefore depends on the discrepancy and coupling.

Interactions between state components. For two discrepancy components, a constant nonnegative matrix $G = (g_{ij})$ has spectral radius less than 1 exactly when

$$g_{11} < 1, \qquad g_{22} < 1, \qquad g_{12} g_{21} < (1 - g_{11})(1 - g_{22}). \tag{2.13}$$

Contraction within each component is insufficient when feedback between components is strong. Appendix D.2 derives a feature/previous-action comparison matrix rather than treating a fitted coefficient as a propagation bound. This is a structural condition, not an estimate for the evaluated models.

2.3.2 From mistake preference to persistent influence

Fix a covered checkpoint set $\mathcal{S}$, an initial checkpoint $K_0 \in \mathcal{S}$ with $X_0 \geq x > b$, and a horizon h. Define the threshold and coverage exit times

$$\tau_b = \inf\{t : X_t < b\}, \qquad \kappa = \inf\{t : K_t \notin \mathcal{S}\}, \qquad \rho = \tau_b \wedge \kappa.$$

Before ρ, suppose the finite, integrable logit increments satisfy

$$\begin{gathered} X_{t+1} - X_t = \mu_t + \xi_{t+1}, \qquad \mu_t \geq -a_t, \\ \log \mathrm{E}[e^{-\lambda \xi_{t+1}} \mid \mathcal{F}_t] \leq \frac{\lambda^2 v_t}{2(1 - c\lambda)}, \qquad 0 \leq \lambda < 1/c. \end{gathered} \tag{2.14}$$

Here ξ_{t+1} is conditionally centered, $a_t, v_t \geq 0$ are predictable, and $c \geq 0$. Deterministic budgets A_h, V_h bound the respective sums of a_t, v_t over $t < h \wedge \rho$ almost surely, and $\Pr(\kappa \leq h \mid \mathcal{F}_0) \leq q_h$. For $c = 0$, all $\lambda \geq 0$ are allowed.

Proposition 2.4 (Finite-horizon retention). *Let $B_h = x - b - A_h$. If $B_h, V_h > 0$, define*

$$\eta(h) = \min\left\{1,\ q_h + \exp\left(-\frac{B_h^2}{2(V_h + cB_h)}\right)\right\}. \tag{2.15}$$

If $B_h > 0$ and $V_h = 0$, take $\eta(h) = \min\{1, q_h\}$. If $B_h \le 0$, take $\eta(h) = 1$. Then

$$\Pr(\tau_b \wedge \kappa \le h \mid \mathcal{F}_0) \le \eta(h).$$

Appendix E proves this using a stopped exponential supermartingale. It controls mistake probability, not whether every sampled output is mistaken. The assumptions must cover the threshold-crossing increment. Applying the result to $-X_t$ requires separate drift and tail bounds.

Causal persistence requires comparing both continuations, including histories outside the retention events. At a fixed time n, let $E_a \in \mathcal{F}_n^a$ have probability π_a, with

$$\begin{aligned} X_n^1 &\ge x_1 > b_1 > 0 \qquad \text{on } E_1, \\ X_n^0 &\le -x_0 < -b_0 < 0 \quad \text{on } E_0. \end{aligned}$$

Suppose Proposition 2.4 holds uniformly conditional on each entry checkpoint, applied to X^1 on E_1 and $-X^0$ on E_0, with failure bounds $\eta_1(h), \eta_0(h)$. Set $u = \sigma(b_1)$ and $v = \sigma(-b_0)$.

Theorem 2.5 (Population persistence). *For every $0 \le k \le h$,*

$$\begin{aligned} \Delta_{n+k} &\ge \Lambda_h, \\ \Lambda_h &= u\pi_1(1 - \eta_1(h)) + (1 - v)\pi_0(1 - \eta_0(h)) - 1. \end{aligned} \tag{2.16}$$

If $|D_{n+k}| \le \varepsilon_{n+k}$, then $M_{n+k} \ge \Lambda_h - \varepsilon_{n+k}$.

Outside the retention events, behavior is unrestricted. For $U, V \in [0, 1]$, knowing $\Pr(U \ge u) \ge a$ and $\Pr(V \le v) \ge b$ gives the sharp bound

$$\mathrm{E}U - \mathrm{E}V \ge ua + (1 - v)b - 1.$$

Perfect retention on a positive-probability event alone is insufficient. A retained pair $(0.9, 0.1)$ with probability $1/4$ and $(0.1, 0.9)$ otherwise gives contrast -0.4. Appendix E.3 proves the bound and its sharpness.

A positive Λ_h guarantees finite-horizon persistence. Extending it indefinitely requires uniformly controlled failure probabilities, not repeated application of bounds with growing budgets. The next model adds decreasing update noise and an order-preserving coupling.

2.4 A NONLINEAR FEEDBACK MODEL

We now impose a scalar state to study how feedback strength affects inherited influence. Consider

$$\begin{aligned} U_{n+1} &= (1 - \alpha_n)U_n + \alpha_n V_{n+1}, \\ \Pr(V_{n+1} = 1 \mid \mathcal{F}_n) &= \sigma(\beta U_n), \qquad V_{n+1} \in \{-1, 1\}, \end{aligned} \tag{2.17}$$

where $U_n \in [-1, 1]$, $\beta \ge 0$, and $\alpha_n = c_0/(n + n_0)$ with $n_0 > c_0 > 0$. The source changes only the deterministic initial state, with $U_0^1 > U_0^0$. Thereafter, no separate source slot remains; U_n carries all influence. Define

$$\Delta_n^{\log} = \mathrm{E}\sigma(\beta U_n^1) - \mathrm{E}\sigma(\beta U_n^0). \tag{2.18}$$

Unlike the logit coordinate in (2.1), this logistic response is a model assumption.

Theorem 2.6 (Logistic feedback threshold). *For the process in (2.17) with the ordered initial states specified above:*

(i) If $0 \le \beta < 2$, then

$$\begin{aligned} 0 \le \Delta_n^{\log} &\le \frac{\beta}{4}(U_0^1 - U_0^0) \prod_{k<n} \left[1 - (1 - \beta/2)\alpha_k\right] \\ &= O\big(n^{-c_0(1-\beta/2)}\big). \end{aligned} \tag{2.19}$$

When $\beta = 0$, the contrast is zero at every time.

(ii) If $\beta = 2$, then $U_n \to 0$ almost surely in both continuations, and $\Delta_n^{\log} \to 0$.

(iii) If $\beta > 2$, then $\liminf_{n\to\infty} \Delta_n^{\log} > 0$.

Why the three cases differ. A shared-uniform coupling preserves $U_n^1 > U_n^0$ at every finite step. For $\beta < 2$, the response derivative is at most $\beta/4$, giving

$$\mathrm{E}[U_{n+1}^1 - U_{n+1}^0 \mid \mathcal{G}_n] \le [1 - (1 - \beta/2)\alpha_n](U_n^1 - U_n^0).$$

The product decays at the rate in (2.19). At $\beta = 2$, the linear bound no longer contracts, but the nonlinear drift still points toward zero. A squared-state supermartingale argument proves convergence.

For $\beta > 2$, the drift $\tanh(\beta u/2) - u$ has two stable nonzero roots besides zero. A finite sequence of opposite outputs has positive probability of placing the two states in opposite stable regions. Since $\sum_n \alpha_n^2 < \infty$, a weighted maximal inequality gives positive probability of remaining separated.

The coupled contrast is nonnegative on every other history, so this event gives a positive unconditional limiting inferior. Appendix F provides the proof and an explicit positive lower bound after a finite entry time.

Unlike Theorem 2.5, the monotone coupling rules out negative contributions outside the favorable event. Even a small positive-probability event can therefore establish persistence. General language-model agents are not assumed to admit this coupling.

The threshold $\beta = 2$ is specific to this response rule, normalization, and schedule. The theorem exhibits both fading and nonvanishing influence, without guaranteeing a practically large contrast for every initial gap or estimating a threshold for language models.

3 Empirical studies and their limits

Arithmetic trees measure D_t, M_t, Δ_t for recorded finite-action policies. A development study compares source correction with source and trail repair. The coding study measures later violations and remaining files. These studies neither test (2.17) nor establish infinite-horizon results.

3.1 Exact source and trail effects on decision trees

Each arithmetic problem is solved one operation at a time. The model chooses between a correct operation and a specified wrong rule using one token. A deterministic executor applies the choice and records the result. We use 100 problems each from product derivatives, root sums, and sine shifts. Two menu orders and two continuation policies give 1,200 settings per model. The verification policy adds an instruction at depth two; the ordinary policy does not. A separate model supplies the mistaken source, but the tested model can reject its operation.

We evaluate archived policies labeled Qwen3-Coder [3], Llama 4 Maverick [14], and DeepSeek V4 Flash [4]. First-token scores at temperature one, normalized over the two action tokens, define each policy. Enumerating all action histories through depth four with cached sampling cutoffs gives exact rational averages for these recorded policies. This does not remove uncertainty about fresh provider scores or generalization to other problems. For Maverick, a missing action score is replaced by the smallest returned log probability. Its results describe this substituted policy. Substitutions occur in 365 of the 400 root-sum settings.

Checking the edited histories. For a fixed action path, the ledger depends on the problem and operations, not the source. The audit matched all 37,200 W setting/path nodes per model to C nodes and reconstructed sampled cached prompts and records. Recomputing records along the same path therefore leaves SWAP unchanged. Here, M_t measures differences in later action-history distributions, not ledger inconsistencies caused by source edits. This does not explain the model's choices or validate the wrong operations. See Appendix I.

Magnitude and direction of the effects. Table 1 shows $|M_4| > |D_4|$ in 90.3–98.4% of settings. Reusing recorded operation results encourages trail influence. These measurements illustrate the decomposition; they do not validate its identity. DeepSeek's median trail magnitude is 12.6 points, compared with 2.54 for the direct magnitude. Qwen's median trail magnitude is 0.215 points overall and 0.00109 for sine shifts. Maverick's magnitude depends on score substitutions.

Trail influence is negative in 368 Qwen, 550 Maverick, and 343 DeepSeek settings. In these cases, history generated under the mistaken source reduces the target-error probability after source

Table 1: Depth-four effects across 1,200 settings per model. The three magnitude columns report medians in percentage points; the last two count settings. Maverick uses substituted scores where action scores are missing.

Model	$\lvert D_4\rvert$	$\lvert M_4\rvert$	$\lvert\Delta_4\rvert$	$\lvert M_4\rvert > \lvert D_4\rvert$	M_4 signs $+ \,/\, -$
Qwen3-Coder	2.43×10^{-4}	0.215	0.216	1,136	832 / 368
Llama 4 Maverick	7.41×10^{-11}	2.31	2.31	1,181	650 / 550
DeepSeek V4 Flash	2.54	12.6	13.7	1,083	857 / 343

correction. Residual influence is therefore not always harmful. The prespecified persistence criterion $\min_{0\le t\le 4}\Delta_t \ge 0.25$ also depends on the configuration. Maverick's root-sum cases split completely by menu order, while DeepSeek's 92 qualifying root-sum settings include both menus and policies. This criterion compares W and C, not recovery after one late correction.

What the bounds establish. The implemented retention bound is supported and informative in 366 Qwen, 428 Maverick, and six DeepSeek settings. Median slack is 34.3, 1.87, and 88.8 points, respectively. A separate audit gives no positive two-continuation persistence certificate. The fixed entry margin and finite moment grid prevent positivity even with zero drift and noise budgets. This limits the selected procedure, without contradicting Theorem 2.5.

Optimal finite-tree coupling bounds are tight for some models and conservative for others. Since the trees determine Δ_T, these calculations measure retrospective slack, not predictive accuracy. Appendices G and I give the certificate, additional envelope-comparison assumptions, and negative audits. Finite-depth measurements do not establish asymptotic fading or persistence.

Source and trail repair. A separate development study compares source correction with source and trail repair on six existing injected-error problems under two menu orders. All arms start after the same mistaken first action, which remains in history. Both correction arms receive the same verified source text. Full repair also recomputes affected past work, updates the workspace, and marks old results as superseded without supplying future answers. Exact weighting includes every continuation under the saved binary-action policy, without an entry gate or certificate.

Compared with source correction alone, full repair lowers mean reuse at decision four by 32.70 and 53.42 percentage points for menus 0 and 1, respectively, with decreases on five of six problems in each. Effects vary substantially: changes on product-derivative problems are below 0.1 point. The menus repeat the same six problems. These results measure an externally assisted repair package, including its visible record; they do not isolate workspace changes, establish spontaneous self-correction, or support population claims.

3.2 Coding agents: later behavior

We fork saved checkpoints on ten retained SWE-bench Verified tasks [10]. The instruction restricts repository edits to existing non-test source files and assigns scratch work to a separate directory outside the repository. At step eight, each branch restores the transcript and container, applies an intervention, and continues for 30 model turns. The planted condition records and executes three scripted actions as the agent's own turns: creating a regression test, extending it, and creating a reproduction script in the repository. The in-scope condition performs the same operations in the designated scratch directory. Other conditions add a reminder, combine a reminder with mechanical repair, or remove the scope instruction. Only edits made during the continuation count as departures.

Completed results contain 294 Qwen3-Coder and 159 DeepSeek V4 Flash branches. Contrast intervals use 4,000 task-cluster percentile bootstrap resamples. Branch-level Wilson intervals are descriptive and ignore task clustering. Appendix J gives the protocol.

For Qwen, 12/49 control branches and 28/48 planted branches contain a later departure. The difference is about 34 percentage points, with a reported task-cluster interval of $[+11, +57]$. Of the 28 departing planted branches, 27 touch a file untouched by the planted actions. The in-scope

condition has a 26% departure rate, compared with 24% for control, with no detected aggregate increase. Planted-minus-control effects are positive on six tasks, zero on three, and negative on one.

A reminder reduces Qwen's departure rate from 58% to 18%, with a contrast interval of $[-59, -23]$ points. Against control, the interval is $[-18, +4]$ points: no excess is detected, but equal rates are not established. Meanwhile, 92% of reminder branches retain at least one planted file. This percentage counts branches, not files. The reminder prohibits new violations without requesting cleanup. Adding mechanical repair also gives an 18% departure rate, so no added behavioral benefit from file removal is established.

DeepSeek makes no later departures in 40 planted branches and removes all planted files in 70% of them. The response is therefore model-dependent. No decrease in task resolution is detected at the reported sample sizes. The planted file operations violate the scope instruction; they are not inherently harmful.

The coding results distinguish later behavior from accumulated workspace state, but do not identify D_t or M_t. The study lacks the initially corrected and crossed-source conditions in Definition 2.1, does not separate transcript from workspace, and has no clean-checkpoint reminder condition. Files remaining after a reminder therefore do not establish that those files cause continued harmful behavior.

4 Discussion and limitations

Correcting the source changes a part of a process that depends on its history. Our results separate what the edit removes immediately from the influence carried by the state left behind. The three-state construction shows that eventual fading alone gives no uniform recovery time. The general bounds give additional sufficient conditions, while the nonlinear model exhibits both fading and nonvanishing influence.

The decomposition depends on the chosen intervention and reference; it is not a unique account of internal reasoning. Edited checkpoints must remain executable. Fading assumptions must control full histories, not merely fitted features, and their failure does not prove persistence. Persistence bounds may be uninformative when histories outside favorable events contribute negatively. The logistic model rules out these contributions through an order-preserving scalar structure that we do not assume for general agents. Its positive lower limit need not be large, and its threshold does not transfer to language models.

The experiments cover restricted settings. Depth-four binary-action trees describe recorded policies, including Maverick's score substitutions, rather than unrestricted long-horizon execution. The coding study covers two models on ten selected tasks and does not establish influence mediated by workspace files. Reduced violations after a reminder do not show that the policies are identical, and remaining files alone do not show continued causal harm. These studies illustrate the distinctions and the limits of the bounds without certifying the asymptotic theory.

Ethics Statement

This work studies how language models respond to mistakes in their own earlier outputs. It involves no human subjects or personal data. All experiments use synthetic math problems and publicly available benchmarks, accessed through model providers' standard interfaces. The injected mistakes are mathematical errors with no potential for harm. Our findings are intended to help make language model agents more reliable by clarifying when correcting a mistake is not enough.

References

[1] Rauno Arike, Elizabeth Donoway, Henning Bartsch, and Marius Hobbhahn. Technical report: Evaluating goal drift in language model agents, 2025. URL https://arxiv.org/abs/2505.02709.

[2] Michel Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités XXXIII*, volume 1709 of *Lecture Notes in Mathematics*, pp. 1–68. Springer, 1999.

[3] Ruisheng Cao, Mouxiang Chen, Jiawei Chen, Zeyu Cui, Yunlong Feng, Binyuan Hui, Yuheng Jing, Kaixin Li, Mingze Li, Junyang Lin, Zeyao Ma, Kashun Shum, Xuwu Wang, Jinxi Wei, Jiaxi Yang, Jiajun Zhang, Lei Zhang, Zongmeng Zhang, Wenting Zhao, and Fan Zhou. Qwen3-coder-next technical report, 2026. URL https://arxiv.org/abs/2603.00729.

[4] DeepSeek-AI, Anyi Xu, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, Chenchen Ling, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chengyu Hou, Chenhao Xu, Chenze Shao, Chong Ruan, Conner Sun, Damai Dai, Daya Guo, Dejian Yang, Deli Chen, Donghao Li, Dongjie Ji, Erhang Li, Fang Wei, Fangyun Lin, Fangzhou Yuan, Feiyu Xia, Fucong Dai, Guangbo Hao, Guanting Chen, Guoai Cao, Guolai Meng, Guowei Li, Han Yu, Han Zhang, Hanwei Xu, Hao Li, Haofen Liang, Haoling Zhang, Haoming Luo, Haoran Wei, Haotian Yuan, Haowei Zhang, Haowen Luo, Haoyu Chen, Haozhe Ji, Hengqing Zhang, Honghui Ding, Hongxuan Tang, Huanqi Cao, Huazuo Gao, Hui Qu, Hui Zeng, J Yang, JQ Zhu, Jia Luo, Jia Song, Jia Yu, Jialiang Huang, Jialu Cai, Jian Liang, Jiangting Zhou, Jiasheng Ye, Jiashi Li, Jiaxin Xu, Jiewen Hu, Jieyu Yang, Jin Chen, Jin Yan, Jingchang Chen, Jingli Zhou, Jingting Xiang, Jingyang Yuan, Jingyuan Cheng, Jingzi Zhou, Jinhua Zhu, Jiping Yu, Joseph Sun, Jun Ran, Junguang Jiang, Junjie Qiu, Junlong Li, Junmin Zheng, Junxiao Song, Kai Dong, Kaige Gao, Kang Guan, Kexing Zhou, Kezhao Huang, Kuai Yu, Lean Wang, Lecong Zhang, Lei Wang, Leyi Xia, Li Zhang, Liang Zhao, Lihua Guo, Lingxiao Luo, Linwang Ma, Linyan Zhu, Litong Wang, Liyu Cai, Liyue Zhang, Longhao Chen, MS Di, MY Xu, Max Mei, Miaojun Wang, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Mingming Li, Mingxu Zhou, Minmin Han, Ning Wang, Panpan Huang, Panpan Wang, Peixin Cong, Peiyi Wang, Peng Zhang, Qiancheng Wang, Qihao Zhu, Qingyang Li, Qinyu Chen, Qiushi Du, Qiwei Jiang, Rui Tian, Ruifan Xu, Ruijie Lu, Ruiling Xu, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, Runqian Chen, Runqiu Yin, Runxin Xu, Ruomeng Shen, Ruoyu Zhang, Ruyi Chen, SH Liu, Shanghao Lu, Shangmian Sun, Shangyan Zhou, Shanhuang Chen, Shaofei Cai, Shaoheng Nie, Shaoqing Wu, Shaoyuan Chen, Shengding Hu, Shengyu Liu, Shiqiang Hu, Shirong Ma, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, Shuying Yu, Songyang Zhou, Tao Ni, Tao Yun, Tian Jin, Tian Pei, Tian Ye, Tianle Lin, Tianran Ji, Tianyi Cui, Tianyuan Yue, Tingting Yu, Tun Wang, W Zhang, WL Xiao, Wangding Zeng, Wei An, Weilin Zhao, Wen Liu, Wenfeng Liang, Wenjie Pang, Wenjing Luo, Wenjing Yao, Wenjun Gao, Wenkai Yang, Wenlve Huang, Wenqing Hou, Wentao Zhang, Wenting Ma, Xi Gao, Xiang He, Xiangwen Wang, Xianzu Wang, Xiao Bi, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaokang Zhang, Xiaotao Nie, Xiaowen Sun, Xiaoxiang Wang, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xingchen Liu, Xingkai Yu, Xingyou Li, Xinyu Yang, Xinyu Zhang, Xu Chen, Xuanyu Wang, Xuecheng Su, Xueyin Chen, Xuheng Lin, Xuwei Fu, YC Yan, YQ Wang, YW Ma, Yanfeng Luo, Yang Zhang, Yanhong Xu, Yanru Ma, Yanwen Huang, Yao Li, Yao Li, Yao Xu, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Qian, Yi Shao, Yi Yu, Yichao Zhang, Yifan Ding, Yifan Shi, Yijia Wu, Yiliang Xiong, Yiling Ma, Ying He, Ying Tang, Ying Zhou, Yingjia Luo, Yinmin Zhong, Yishi Piao, Yisong Wang, Yixiang Zhang, Yixiao Chen, Yixuan Tan, Yixuan Wei, Yiyang Ma, Yiyuan Liu, Yonglun Yang, Yongqiang Guo, Yongtong Wu, Yu Wu, YuKun Li, Yuan Cheng, Yuan Ou, Yuanfan Xu, Yuanhao Li, Yuduan Wang, Yuehan Yang, Yuer Xu, Yuhan Wu, Yuhao Meng,

Yuheng Zou, Yukun Zha, Yunfan Xiong, Yupeng Chen, Yuping Lin, Yuqian Cao, Yuqian Wang, Yushun Zhang, Yuting Yan, Yutong Lin, Yuxian Gu, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuxuan Zhou, Yuyang Zhou, Yuzhen Huang, ZF Wu, Zehao Wang, Zehua Zhao, Zehui Ren, Zekai Zhang, Zhangli Sha, Zhe Fu, Zhe Ju, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zheren Gao, Zhewen Hao, Zhibin Gou, Zhicheng Ma, Zhigang Yan, Zhihong Shao, Zhixian Huang, Zhixuan Chen, Zhiyu Wu, Zhizhou Ren, Zhongyu Wu, Zhuoshu Li, Zhuping Zhang, Zian Xu, Zihao Wang, Zihua Qu, Zihui Gu, Zijia Zhu, Zilin Li, Zipeng Zhang, Ziwei Xie, Ziyi Gao, Ziyi Wan, Zizheng Pan, and Zongqing Yao. Deepseek-v4: Towards highly efficient million-token context intelligence, 2026. URL https://arxiv.org/abs/2606.19348.

[5] Yiyang Feng, Yichen Wang, Shaobo Cui, Boi Faltings, Mina Lee, and Jiawei Zhou. Unraveling misinformation propagation in llm reasoning, 2025. URL https://arxiv.org/abs/2505.18555.

[6] David A. Freedman. On tail probabilities for martingales. 1975. URL https://api.semanticscholar.org/CorpusID:123246626.

[7] Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17(none), January 2020. ISSN 1549-5787. doi: 10.1214/18-ps321. URL http://dx.doi.org/10.1214/18-PS321.

[8] Jie Huang, Xinyun Chen, Swaroop Mishra, Huaixiu Steven Zheng, Adams Wei Yu, Xinying Song, and Denny Zhou. Large language models cannot self-correct reasoning yet, 2024. URL https://arxiv.org/abs/2310.01798.

[9] Kehan Jiang, Haonan Dong, Zhaolu Kang, Zhengzhou Zhu, and Guojie Song. Foe: Forest of errors makes the first solution the best in large reasoning models, 2026. URL https://arxiv.org/abs/2604.02967.

[10] Carlos E. Jimenez, John Yang, Alexander Wettig, Shunyu Yao, Kexin Pei, Ofir Press, and Karthik Narasimhan. Swe-bench: Can language models resolve real-world github issues?, 2024. URL https://arxiv.org/abs/2310.06770.

[11] Philippe Laban, Hiroaki Hayashi, Yingbo Zhou, and Jennifer Neville. Llms get lost in multi-turn conversation, 2025. URL https://arxiv.org/abs/2505.06120.

[12] Kenneth Li, Tianle Liu, Naomi Bashkansky, David Bau, Fernanda Viégas, Hanspeter Pfister, and Martin Wattenberg. Measuring and controlling instruction (in)stability in language model dialogs, 2024. URL https://arxiv.org/abs/2402.10962.

[13] Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. Self-refine: Iterative refinement with self-feedback, 2023. URL https://arxiv.org/abs/2303.17651.

[14] Meta AI. The llama 4 herd: The beginning of a new era of natively multimodal AI innovation. https://ai.meta.com/blog/llama-4-multimodal-intelligence/, April 2025. Accessed 2026-09-26.

[15] Vrettos Moulos. Bicausal optimal transport for markov chains via dynamic programming, 2020. URL https://arxiv.org/abs/2010.06831.

[16] Judea Pearl. Direct and indirect effects, 2013. URL https://arxiv.org/abs/1301.2300.

[17] Robin Pemantle. A survey of random processes with reinforcement. *Probability Surveys*, 4 (none), January 2007. ISSN 1549-5787. doi: 10.1214/07-ps094. URL http://dx.doi.org/10.1214/07-PS094.

[18] James M. Robins and Sander Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3:143–155, 1992. URL https://api.semanticscholar.org/CorpusID:10757981.

[19] Daniel Rudolf and Nikolaus Schweizer. Perturbation theory for markov chains via wasserstein distance, 2017. URL https://arxiv.org/abs/1503.04123.

[20] Noah Shinn, Federico Cassano, Edward Berman, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning, 2023. URL https://arxiv.org/abs/2303.11366.

[21] Gladys Tyen, Hassan Mansoor, Victor Cărbune, Peter Chen, and Tony Mak. Llms cannot find reasoning errors, but can correct them given the error location, 2024. URL https://arxiv.org/abs/2311.08516.

[22] Julio Backhoff Veraguas, Mathias Beiglböck, Yiqing Lin, and Anastasiia Zalashko. Causal transport in discrete time and applications, 2017. URL https://arxiv.org/abs/1606.04062.

[23] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models, 2023. URL https://arxiv.org/abs/2210.03629.

[24] Muru Zhang, Ofir Press, William Merrill, Alisa Liu, and Noah A. Smith. How language model hallucinations can snowball, 2023. URL https://arxiv.org/abs/2305.13534.

A USE OF LLM

We employ large language models to help improve word choice, readability, and fluency. This include significant revisions with our supervision. We also utilized large language models to help find related works at the beginning of the research paper, before finding results.

B CHECKPOINT, INTERVENTION, AND TERMINATION CONVENTIONS

For each t, let $\mathcal{K}_t$ be a standard Borel space of logical checkpoints and $\mathcal{O}_t$ a standard Borel space of full outputs and observations. For source $a \in \{0, 1\}$, a measurable kernel

$$Q_t^a(k, do, dk')$$

renders the accessible input, samples the next output, applies the fixed environment and memory update, and returns the successor checkpoint. Its checkpoint marginal is $P_t^a(k, dk')$. A common measurable mistake label $g_t(o, k) \in \{0, 1\}$ defines

$$p_t^a(k) = \int g_t(o, k)\, Q_t^a(k, do, dk').$$

The label can use a task-specific verified rule, but must distinguish the target mistake from unrelated errors in the same way in both arms.

The kernel must be defined on every checkpoint required by a crossed readout. Editing a source requires rebuilding computational caches that depend on its contents. It does not delete independent semantic copies stored in other messages, summaries, or artifacts. Those are part of the non-source checkpoint. If the source edit has no executable interpretation, the crossed estimand is undefined; the decomposition cannot then be measured under that protocol.

A joint successor kernel with marginals $P_t^1(k_1, \cdot)$ and $P_t^0(k_0, \cdot)$, together with an initial joint law, supplies a coadapted process satisfying

$$\mathcal{L}(K_{t+1}^a \mid \mathcal{G}_t) = P_t^a(K_t^a, \cdot).$$

The product kernel establishes existence, not contraction. Independently matching marginal snapshots at different times does not establish the conditional propagation inequalities used in the paper. Full-history encoding likewise does not assert that a compact feature vector is sufficient. A discrepancy can ignore information only if the readout and propagation inequalities remain valid despite what it omits.

Termination. Decision timing and completion handling are fixed before comparisons, without conditioning on survival. At completion, use an absorbing checkpoint with a declared readout $r(k) \in [0, 1]$, such as a final-artifact error indicator. Before completion use the next-output probability, and afterward use $r(k)$, including for crossed evaluations. The decomposition, fading, and bounded-readout persistence arguments remain valid under their corresponding assumptions for this extended process. After completion the estimand concerns the declared artifact outcome, not a new sampled action.

Retention requires extra care. A terminal readout of zero or one has an infinite logit. One convention is to stop coverage at completion, include completion in q_h, and freeze the last finite analysis score on that terminating transition. The retention result then concerns covered, active decisions. An active transition that crosses a score threshold or leaves the covered set must still satisfy the finite-increment and moment conditions. Declaring the successor out of domain does not excuse an uncontrolled crossing increment. Clipping defines a different score process and requires new assumptions and a readout bridge.

C SOURCE DECOMPOSITION AND THE THREE-STATE CONSTRUCTION

C.1 IDENTITY AND COUPLING INEQUALITY

Adding and subtracting the crossed expectation gives

$$\begin{aligned}\Delta_t &= \left(\int p_t^1\, d\nu_t^1 - \int p_t^0\, d\nu_t^1\right) + \left(\int p_t^0\, d\nu_t^1 - \int p_t^0\, d\nu_t^0\right)\\ &= D_t + M_t.\end{aligned}$$

A pointwise source-sensitivity envelope $|p_t^1 - p_t^0| \le \varepsilon_t\ \nu_t^1$-almost surely implies $|D_t| \le \varepsilon_t$. More generally, $|D_t|$ may be small because positive and negative source effects cancel. For any coupling Γ_t of ν_t^1 and ν_t^0,

$$M_t = \mathbb{E}_{\Gamma_t}[p_t^0(K_t^1) - p_t^0(K_t^0)], \qquad |M_t| \le \mathbb{E}_{\Gamma_t}|p_t^0(K_t^1) - p_t^0(K_t^0)|. \tag{C.1}$$

No sign, monotonicity, or additive-token assumption is used. If both source renderings coincide because the source is inaccessible and no other source-dependent channel remains outside K_t, then $D_t = 0$; earlier source-dependent transitions may still have changed ν_t^a.

C.2 EXACT READOUTS AND EXACT CONTRACTION

In the process from Theorem 2.2, both arms start at N. At time zero, source s sets $p_0^s = q + (1-2q)s$. The first sampled label moves the process to H or L, and the source is then evicted. For $t \ge 1$,

$$p(H) = 1 - q, \qquad p(L) = q, \qquad p(N) = 1/2.$$

From H, a correct output moves to N; from L, a mistaken output moves to N; N is absorbing as a state, with a fresh fair label at active steps.

At time one, arm 1 has masses $(1-q, q)$ on (H, L) and arm 0 has masses $(q, 1-q)$. Each nonneutral mass survives another $t-1$ steps with probability $(1-q)^{t-1}$. Consequently,

$$\begin{aligned}\Delta_0 = D_0 &= 1 - 2q, \qquad M_0 = 0,\\ \bar{p}_t^1 &= \frac{1}{2} + \frac{1}{2}(1-2q)^2(1-q)^{t-1},\\ \bar{p}_t^0 &= \frac{1}{2} - \frac{1}{2}(1-2q)^2(1-q)^{t-1}, \qquad t \ge 1.\end{aligned} \tag{C.2}$$

All readouts are strictly between zero and one, so all logits are finite. These formulas prove the exact contrast in equation 2.8. Choosing $q = \delta/(H+3)$ gives equation 2.7 by applying $\prod_i(1-a_i) \ge 1 - \sum_i a_i$ to two factors with $a_i = 2q$ and $t-1$ factors with $a_i = q$.

For an exact fading bound, couple the first labels by a shared uniform, so the states disagree with probability $1-2q$. At later times, use a common eviction event of probability q for nonneutral states. On an H/L pair, retain both with outputs $(1,0)$ if no eviction occurs and move both to N with outputs $(0,1)$ otherwise. On an H/N or L/N pair, use the prescribed eviction for the nonneutral state and a fresh fair label for the neutral state. Equal states have coupled identical transitions. Every marginal is correct. With $d(k,k') = \mathbf{1}_{\{k \ne k'\}}$ on the sufficient state summaries,

$$\mathbb{E}[d_{t+1} \mid \mathcal{G}_t] = (1-q)d_t \quad (t \ge 1), \qquad |p(k) - p(k')| \le (1-2q)d(k,k').$$

Take $G = 1-q$, $\ell = 1-2q$, $z_0 = 0$, $\varepsilon_0 = f_0 = 1-2q$, and $\varepsilon_t = f_t = 0$ for $t \ge 1$. Then Theorem 2.3 gives

$$F_0 = 1 - 2q, \qquad F_t = (1-2q)^2(1-q)^{t-1} \quad (t \ge 1),$$

exactly equal to $|\Delta_t|$. This coupling is not the shared-next-label-uniform coupling at every later step; optimizing the state comparison can require a different coupling.

C.3 A NONVACUOUS BUT CONSERVATIVE PERSISTENCE CALCULATION

Take $q = 10^{-4}$, entry time $n = 1$, and horizon $h = 10$. Use entry events $K_1^1 = H$ and $K_1^0 = L$, each of probability $1-q$. Put $A = \log((1-q)/q)$ and thresholds $b_1 = b_0 = \log 9$, so $u = 0.9$ and $v = 0.1$.

Before crossing out of the high-score region, a score increment is $-AB$ for $B \sim \text{Bernoulli}(q)$. Its mean is $-Aq$ and its centered innovation is $\xi = A(q - B)$.

A valid one-sided moment bound has drift allowance $a = Aq$, variance proxy $v_{\text{tail}} = A^2 q(1-q)$, and scale $c = A/3$. To verify it, set $Z = -\xi$. Since $\mathbb{E}Z = 0$, $|Z| \le A$, and $\mathbb{E}Z^2 = v_{\text{tail}}$, the series bound $k! \ge 2 \cdot 3^{k-2}$ gives

$$\mathbb{E}e^{\lambda Z} \le 1 + \frac{\lambda^2 v_{\text{tail}}}{2(1 - \lambda A/3)}, \qquad 0 \le \lambda < 3/A.$$

Taking logarithms and using $\log(1+x) \le x$ proves equation 2.14. With $A_h = ha$, $V_h = hv_{\text{tail}}$, and no coverage penalty, substitution gives $\eta(10) \approx 0.3210452$ and $\Lambda_{10} \approx 0.2219964 > 0$. The exact escape probability is approximately 0.00099955, and the smallest contrast at times $1, \dots, 11$ is approximately 0.9986009. Thus the theorem can be informative while conservative. These are calculations for the explicit process, not empirical agent results; the same process eventually fades.

D Proofs for discrepancy propagation

D.1 Propagation and stable convolution

Write $z_t = \mathbb{E}d_t$. Taking expectations in the residual definition yields

$$z_{t+1} = G_t z_t + \mathbb{E}r_t \le G_t z_t + \mathbb{E}\,(r_t)_+ \le G_t z_t + f_t.$$

Nonnegativity of G_t and $z_0 \le z_0^\star$ imply $z_t \le z_t^\star$ inductively. The coupling inequality equation C.1 and equation 2.9 give $|M_t| \le \ell_t^\top z_t^\star$. The identity $\Delta_t = D_t + M_t$ then gives $|\Delta_t| \le F_t$, and bounded readouts give the additional cap at one.

For the asymptotic claim, choose $\rho \in (\text{spr}(G), 1)$. Finite-dimensional Jordan form gives $\|G^k\| \le C_\rho \rho^k$, hence $G^t z_0 \to 0$ and $\sum_k \|G^k\| < \infty$. Iteration gives

$$z_t^\star = G^t z_0 + \sum_{j<t} G^{t-1-j} f_j.$$

For any $\epsilon > 0$, choose J with $\|f_j\| \le \epsilon$ for $j \ge J$. The finitely many terms with $j < J$ vanish as $t \to \infty$, and the remaining sum is at most $\epsilon \sum_{k \ge 0} \|G^k\|$. Letting $\epsilon \downarrow 0$ proves $z_t^\star \to 0$. Bounded ℓ_t and vanishing ε_t give the claimed limits.

For time-varying matrices, the same argument holds when the ordered products satisfy $\|G_{t-1} \cdots G_s\| \le C\rho^{t-s}$ for fixed $C < \infty$ and $\rho < 1$. Individual spectral radii below one are insufficient: alternating $\left(\begin{smallmatrix} 0 & 2 \\ 0 & 0 \end{smallmatrix}\right)$ and its transpose has zero individual spectral radii but growing products.

Residual forcing. For constant G, ℓ, if $\text{spr}(G) < 1$, $\limsup_t \varepsilon_t \le \bar{\varepsilon}$, and $\limsup_t f_t \le \bar{f}$ componentwise, the same convolution argument gives

$$\limsup_t |\Delta_t| \le \min\{1, \bar{\varepsilon} + \ell^\top (I - G)^{-1} \bar{f}\}.$$

This is an upper bound, not evidence that a positive error floor exists. If the corrected-arm error also satisfies $\bar{p}_t^0 \le r_t$, then $\bar{p}_t^1 \le \min\{1, r_t + F_t\}$. Absolute recovery therefore needs control of the corrected arm in addition to causal fading.

D.2 A concrete feature/last-action comparison

Let $Z_t(k)$ be a shared measurable feature map, and $J_t(k) \in \{0, 1\}$ the last eligible action label, with a common initialization. Define

$$d_{z,t} = \|Z_t(K_t^1) - Z_t(K_t^0)\|, \qquad d_{j,t} = \mathbf{1}_{\{J_t(K_t^1) \ne J_t(K_t^0)\}}.$$

Assume the pointwise source envelope $|p_t^1(k) - p_t^0(k)| \le \varepsilon_t$ on mistaken-arm checkpoints and the common-source readout condition

$$|p_t^0(k) - p_t^0(k')| \le \ell_z \|Z_t(k) - Z_t(k')\| + \ell_j \mathbf{1}_{\{J_t(k) \ne J_t(k')\}}.$$

These are full-checkpoint conditions, not claims about a fitted feature regression.

Couple next binary labels using a fresh shared uniform. Their conditional mismatch probability is

$$m_t = |p_t^1(K_t^1) - p_t^0(K_t^0)| \le \varepsilon_t + \ell_z d_{z,t} + \ell_j d_{j,t}.$$

Extend this to full outputs and successor checkpoints by disintegrating each arm's output/successor kernel conditional on its binary label. Standard Borel spaces supply measurable conditional kernels; zero-mass fibers can be defined arbitrarily. Each marginal remains the original execution law.

Suppose, under this particular extension, that

$$\mathbb{E}[d_{z,t+1} \mid \mathcal{G}_t] \le a\, d_{z,t} + b\, d_{j,t} + c_* m_t + e_t, \qquad a, b, c_*, e_t \ge 0.$$

Since $J_{t+1}^a = Y_{t+1}^a$, the expected next last-label discrepancy is m_t. Therefore the pointwise propagation envelope holds with

$$G = \begin{pmatrix} a + c_*\ell_z & b + c_*\ell_j \\ \ell_z & \ell_j \end{pmatrix}, \qquad f_t = \begin{pmatrix} c_*\varepsilon_t + e_t \\ \varepsilon_t \end{pmatrix}, \qquad \ell = \begin{pmatrix} \ell_z \\ \ell_j \end{pmatrix}. \tag{D.1}$$

For a nonnegative 2×2 matrix, the larger eigenvalue is

$$\lambda_+ = \frac{g_{11} + g_{22} + \sqrt{(g_{11} - g_{22})^2 + 4 g_{12} g_{21}}}{2} = \operatorname{spr}(G).$$

It is at least each diagonal entry. Once both diagonals are below one, squaring the equivalent nonnegative-side inequality for $\lambda_+ < 1$ gives equation 2.13. This construction concerns active binary labels; an absorbing artifact readout cannot be substituted for a newly sampled last action without a separate kernel check.

E Preference retention and two-arm persistence

E.1 Stopped exponential bound

Let ρ be a stopping time. On $t < \rho$, assume conditionally centered ξ_{t+1} obeys the moment condition in equation 2.14, and define

$$S_j = \sum_{t<j} \mathbf{1}_{\{t<\rho\}} \xi_{t+1}, \qquad W_j = \sum_{t<j} \mathbf{1}_{\{t<\rho\}} v_t.$$

For each fixed admissible $\lambda > 0$,

$$Z_j = \exp\left(-\lambda S_j - \frac{\lambda^2 W_j}{2(1 - c\lambda)}\right)$$

is a nonnegative supermartingale with $Z_0 = 1$: its next multiplicative factor has conditional expectation at most one before ρ and equals one afterward. If $W_h \le V$, a crossing $S_j \le -B$ for some $j \le h$ entails

$$Z_j \ge \exp\left(\lambda B - \frac{\lambda^2 V}{2(1 - c\lambda)}\right).$$

Stopping at the first such crossing or at h, and using nonnegativity away from the crossing, gives

$$\mathbb{P}\left(\min_{0 \le j \le h} S_j \le -B \,\middle|\, \mathcal{F}_0\right) \le \min\left\{1, \exp\left(-\lambda B + \frac{\lambda^2 V}{2(1 - c\lambda)}\right)\right\}. \tag{E.1}$$

E.2 Proof of the retention proposition

Use $\rho = \tau_b \wedge \kappa$. The explicit budget conditions are

$$\sum_{t<h} \mathbf{1}_{\{t<\rho\}} a_t \le A_h, \qquad \sum_{t<h} \mathbf{1}_{\{t<\rho\}} v_t \le V_h, \qquad \mathbb{P}(\kappa \le h \mid \mathcal{F}_0) \le q_h.$$

On $\{\tau_b \le h, \kappa > h\}$, every increment through the threshold crossing is covered, and

$$X_{\tau_b} = X_0 + \sum_{t<\tau_b} \mu_t + S_{\tau_b} \ge x - A_h + S_{\tau_b}.$$

Since $X_{\tau_b} < b$, it follows that $S_{\tau_b} < -B_h$. When $B_h, V_h > 0$, the choice $\lambda = B_h/(V_h + cB_h)$ is admissible and substitutes into equation E.1 to give the exponential term in equation 2.15. Add the coverage-exit probability q_h and cap at one.

If $V_h = 0$, all stopped variance proxies vanish. Conditional Jensen and strict convexity give

$$1 = e^{-\lambda \mathbb{E}[\xi_{t+1}|\mathcal{F}_t]} \le \mathbb{E}[e^{-\lambda \xi_{t+1}} \mid \mathcal{F}_t] \le 1,$$

forcing $\xi_{t+1} = 0$ almost surely before the stop. A positive remaining buffer cannot then be crossed before a domain exit. If $B_h \le 0$, the universal upper bound one completes the statement.

Finite moment grids. If the moment inequality is known only for a nonempty finite set $\mathcal{L} \subset (0, 1/c)$, use, for $B_h > 0$,

$$\mathbb{P}(\tau_b \wedge \kappa \le h \mid \mathcal{F}_0) \le \min\left\{1, q_h + \min_{\lambda \in \mathcal{L}} \exp\left(-\lambda B_h + \frac{\lambda^2 V_h}{2(1 - c\lambda)}\right)\right\}. \tag{E.2}$$

For $c = 0$, replace the admissible interval by $(0, \infty)$. Each fixed grid member gives a valid bound on the same event, so their minimum is valid. This does not justify interpolation, unverified off-grid optimization, or treating sample quantiles as uniform conditional moment bounds.

E.3 Proof and sharpness of two-arm persistence

For $U, V \in [0, 1]$, the assumptions $\mathbb{P}(U \ge u) \ge a$ and $\mathbb{P}(V \le v) \ge b$ imply

$$\mathbb{E}U \ge ua, \qquad \mathbb{E}V \le vb + (1 - b),$$

hence $\mathbb{E}U - \mathbb{E}V \ge ua + (1 - v)b - 1$. Equality is attained by $U = u$ on an event of probability a and zero elsewhere, and by $V = v$ on an event of probability b and one elsewhere. Strictly interior readouts can approach these extrema. No relationship between the two favorable events is needed.

Let A_1 be the mistaken-arm entry event followed by covered retention of $X^1 \ge b_1$ at every time $n, \dots, n + h$, and let A_0 be the analogous corrected-arm event for $X^0 \le -b_0$. Uniform conditional application of Proposition 2.4 and the tower property give

$$\mathbb{P}(A_1) \ge \pi_1(1 - \eta_1(h)), \qquad \mathbb{P}(A_0) \ge \pi_0(1 - \eta_0(h)).$$

For every $0 \le k \le h$, the corresponding readouts are at least u on A_1 and at most v on A_0. Apply the sharp bounded-readout inequality to obtain equation 2.16. Finally, $M_{n+k} = \Delta_{n+k} - D_{n+k} \ge \Lambda_h - \varepsilon_{n+k}$.

If the same entry events support this result for every horizon, with $\sup_h \eta_a(h) \le \bar{\eta}_a$ and

$$\bar{\Lambda} = u\pi_1(1 - \bar{\eta}_1) + (1 - v)\pi_0(1 - \bar{\eta}_0) - 1 > 0,$$

then $\Delta_t \ge \bar{\Lambda}$ for every $t \ge n$. For each fixed t, choose a horizon at least $t - n$. A bound proved at just one finite horizon does not imply these uniform hypotheses.

F Full proof of the logistic-feedback theorem

F.1 A weighted all-horizon retention lemma

Lemma F.1. *Consider $U_{n+1} = U_n + \alpha_n(V_{n+1} - U_n)$ with $V_{n+1} \in \{-1, 1\}$, $U_n \in [-1, 1]$, deterministic $0 < \alpha_n < 1$, and $g_n = \mathbb{E}[V_{n+1} \mid \mathcal{F}_n]$. Fix $-1 < a < b < 1$ and $d = b - a$. Suppose $U_N \ge b$, and $g_n - U_n \ge 0$ whenever $a \le U_n \le b$ and $n \ge N$. If*

$$\sup_{n \ge N} \alpha_n \le d/4, \qquad Q_N = \sum_{n \ge N} \alpha_n^2 < \infty,$$

then

$$\mathbb{P}(U_n > a \text{ for every } n \ge N \mid \mathcal{F}_N) \ge \max\{0, 1 - 2\exp(-d^2/(32Q_N))\}. \tag{F.1}$$

Proof. The convex update preserves $[-1, 1]$. Let $M_m = \sum_{n=N}^{m-1} \alpha_n(V_{n+1} - g_n)$. Each centered increment has conditional range width $2\alpha_n$, so its conditional exponential moment is at most $\exp(\lambda^2\alpha_n^2/2)$. One direct derivation scales a bounded variable to an interval of length one: the second derivative of its tilted log moment-generating function is a variance at most $1/4$, and integration twice gives $\lambda^2/8$. Rescaling gives the stated bound.

Apply equation E.1 to M and $-M$, first through finite horizons and then to the increasing union of crossing events. Optimization over λ gives

$$\mathbb{P}\left(\sup_{m\geq N} |M_m| \geq x \,\middle|\, \mathcal{F}_N\right) \leq 2\exp(-x^2/(2Q_N)).$$

If $\tau = \inf\{n \geq N : U_n \leq a\} < \infty$, let $j < \tau$ be the last index with $U_j \geq b$. A single jump has magnitude at most $2\alpha_j \leq d/2$, so $U_{j+1} \geq b - d/2 > a$ and $j + 1 < \tau$. From $j + 1$ until the crossing, the state lies in (a, b) and its conditional drift is nonnegative. Therefore

$$M_\tau - M_{j+1} \leq U_\tau - U_{j+1} \leq -d/2.$$

At least one of $|M_\tau|$ and $|M_{j+1}|$ is at least $d/4$. Substitute $x = d/4$ in the maximal bound and take complements. □

F.2 Order-preserving coupling

Use fresh independent uniforms Z_{n+1}, shared between arms, and set

$$V_{n+1}^a = 2\mathbf{1}_{\{Z_{n+1}\leq\sigma(\beta U_n^a)\}} - 1.$$

The response is monotone, and the update is convex with $1 - \alpha_n > 0$. Thus $U_n^1 \geq U_n^0$, with strict order at every finite step when $U_0^1 > U_0^0$. In particular, the coupled readout difference is nonnegative on every history. The coupling preserves each arm's conditional law.

F.3 Subcritical feedback

Put $D_n^U = U_n^1 - U_n^0 \geq 0$. The derivative of σ is at most $1/4$, so

$$\begin{aligned}\mathbb{E}[D_{n+1}^U \mid \mathcal{G}_n] &= (1 - \alpha_n)D_n^U + 2\alpha_n[\sigma(\beta U_n^1) - \sigma(\beta U_n^0)] \\ &\leq [1 - (1 - \beta/2)\alpha_n]D_n^U.\end{aligned}$$

Iteration and the same readout Lipschitz bound prove the product estimate in equation 2.19. For $\theta = c_0(1 - \beta/2) > 0$,

$$\prod_{k<n}\left(1 - \frac{\theta}{k + n_0}\right) = \Theta(n^{-\theta}).$$

Indeed, $n_0 > \theta$, and expanding the logarithm gives $-\theta\sum_{k<n}(k + n_0)^{-1} + O(\sum_{k<n}(k + n_0)^{-2}) = -\theta\log n + O(1)$. For $\beta = 0$ the readout is constantly $1/2$.

F.4 Critical feedback

At $\beta = 2$, $\mathbb{E}[V_{n+1} \mid \mathcal{F}_n] = \tanh(U_n)$. Let $h(u) = u[u - \tanh(u)]$, which is continuous, nonnegative, and zero only at zero. Expanding the square and using $|V - U| \leq 2$ gives

$$\mathbb{E}[U_{n+1}^2 \mid \mathcal{F}_n] \leq U_n^2 - 2\alpha_n h(U_n) + 4\alpha_n^2.$$

Hence $W_n = U_n^2 + 4\sum_{k\geq n}\alpha_k^2$ is a nonnegative supermartingale and converges almost surely. Summing the expectation inequality gives $\mathbb{E}\sum_n \alpha_n h(U_n) < \infty$, so the sum is finite almost surely. Since the square-step tail vanishes, U_n^2 converges. A positive limit would keep $h(U_n)$ bounded below by a positive constant eventually, contradicting $\sum_n \alpha_n = \infty$. Thus $U_n \to 0$ almost surely in each arm. Bounded convergence gives $\Delta_n^{\log} \to 0$.

F.5 SUPERCRITICAL FEEDBACK

For $\beta > 2$, the drift is $g(u) - u = \tanh(\beta u/2) - u$. Its derivative at zero is positive, it is strictly concave on $(0, 1]$, and its value at one is negative. It therefore has a unique positive root $r_\beta \in (0, 1)$ and is positive on $(0, r_\beta)$. Fix $0 < a < b < r_\beta$ and $d = b - a$.

Prescribe a finite prefix in which the upper arm always draws $+1$ and the lower always draws -1. Along this prescribed prefix, the deterministic states u_k^1, u_k^0 approach $1, -1$, since $\prod_{k<N}(1 - \alpha_k) \to 0$. Strict ordering makes every finite prefix possible under the shared-uniform coupling. Its probability is

$$\mathbb{P}(E_N) = \prod_{k<N} [\sigma(\beta u_k^1) - \sigma(\beta u_k^0)] > 0. \tag{F.2}$$

Choose N large enough that $u_N^1 \geq b$, $u_N^0 \leq -b$,

$$\sup_{n \geq N} \alpha_n \leq d/4, \qquad 4\exp[-d^2/(32Q_N)] < 1,$$

which is possible because $\alpha_n \to 0$ and $Q_N \to 0$. Conditional on the joint history at N, apply Lemma F.1 to the upper arm and to the negative of the lower arm. Symmetry supplies the same drift condition for the reflected lower arm. A union bound gives probability at least $1 - 4\exp[-d^2/(32Q_N)]$ that both remain respectively above a and below $-a$ forever. On that event the readout difference is at least $\tanh(\beta a/2)$.

Everywhere else the difference is nonnegative by the order-preserving coupling. Consequently, for every $n \geq N$,

$$\Delta_n^{\log} \geq \mathbb{P}(E_N) \left[1 - 4\exp\left(-\frac{d^2}{32Q_N}\right)\right] \tanh(\beta a/2) > 0. \tag{F.3}$$

This establishes the strictly positive limiting inferior in Theorem 2.6. The prefix probability can be very small; this is a qualitative nonvanishing result with an explicit, potentially conservative, positive lower bound.

G OPTIMAL FINITE-TREE CERTIFICATES AND PROPAGATION ENVELOPES

G.1 OPTIMALITY WITHIN COADAPTED COUPLINGS

Fix a horizon T, an initial joint law, and finite reachable supports. Define the terminal cost

$$\Phi_T^\star(k, k') = |p_T^1(k) - p_T^0(k')|$$

and recurse backward by

$$\Phi_t^\star(k, k') = \min_{\gamma \in \Gamma(P_t^1(k,\cdot), P_t^0(k',\cdot))} \int \Phi_{t+1}^\star \, d\gamma. \tag{G.1}$$

Here Γ is the set of couplings of the two successor laws. Each step is a finite linear program on a nonempty compact polytope, so its minimum is attained.

Proposition G.1 (Finite-tree coupling certificate)**.** *Over coadapted couplings with the fixed initial joint law,*

$$|\Delta_T| \leq \mathbb{E}\Phi_0^\star = \min \mathbb{E}|p_T^1(K_T^1) - p_T^0(K_T^0)|. \tag{G.2}$$

Proof. At every reachable checkpoint pair, select a minimizing coupling in equation G.1, with a fixed tie rule. The choices depend only on the current pair and known kernels, not realized future choices, and define a coadapted coupling. Under it, $\mathbb{E}[\Phi_{t+1}^\star \mid \mathcal{G}_t] = \Phi_t^\star$, so its expected terminal cost equals $\mathbb{E}\Phi_0^\star$. Under any other coadapted coupling, conditional minimization gives $\mathbb{E}[\Phi_{t+1}^\star \mid \mathcal{G}_t] \geq \Phi_t^\star$. Backward iteration yields a terminal cost at least $\mathbb{E}\Phi_0^\star$. Finally, correct marginals and the triangle inequality give $|\Delta_T| \leq \mathbb{E}|p_T^1(K_T^1) - p_T^0(K_T^0)|$. □

The minimization is over adapted couplings of known transition laws. It need not equal $|\Delta_T|$, because signed contributions can cancel before the absolute value is taken in the latter quantity. Replacing the terminal cost by $|p_T^0(k) - p_T^0(k')|$ gives a trail certificate; adding $|D_T|$ then bounds $|\Delta_T|$.

G.2 Comparison with a propagation envelope

The averaged conditions in equation 2.10 and equation 2.11 suffice for Theorem 2.3. They do *not*, by themselves, justify the following pointwise supermartingale construction. For it, impose the stronger envelopes, under one coadapted coupling with the same initial law,

$$|p_t^1(k) - p_t^0(k)| \le \varepsilon_t \qquad \nu_t^1\text{-almost surely,} \tag{G.3}$$

$$\mathbb{E}[d_{t+1} \mid \mathcal{G}_t] \le G_t d_t + f_t \qquad \text{almost surely,} \tag{G.4}$$

together with equation 2.9. These imply the averaged assumptions with the same budgets, while also controlling absolute source sensitivity and conditional propagation.

Let $\Pi_{s,t} = G_{t-1} \cdots G_s$ and $\Pi_{t,t} = I$. Define

$$H_t(k,k') = \varepsilon_T + \ell_T^\top \Pi_{t,T} d_t(k,k') + \ell_T^\top \sum_{j=t}^{T-1} \Pi_{j+1,T} f_j.$$

The source and common-source readout envelopes give $H_T(K_T^1, K_T^0) \ge |p_T^1(K_T^1) - p_T^0(K_T^0)|$. Multiplying equation G.4 by the nonnegative row vector $\ell_T^\top \Pi_{t+1,T}$ gives $\mathbb{E}[H_{t+1} \mid \mathcal{G}_t] \le H_t$. The tower property gives an admissible terminal-cost bound with $\mathbb{E}H_0 \le F_T$. Since $\mathbb{E}\Phi_0^\star \le 1$ independently, the combined statement is

$$|\Delta_T| \le \mathbb{E}\Phi_0^\star \le \min\{1, F_T\}, \qquad \mathbb{E}H_0 \le F_T. \tag{G.5}$$

The separate role of the pointwise assumptions prevents cancellation in D_T from being mistaken for control of an absolute terminal cost.

G.3 Binary successor laws

For two logical successors with mistake probabilities a and b, every coupling has masses

$$(\gamma_{11}, \gamma_{10}, \gamma_{01}, \gamma_{00}) = (\theta, a-\theta, b-\theta, 1-a-b+\theta), \qquad \max(0, a+b-1) \le \theta \le \min(a,b).$$

The Bellman objective is linear in θ, so it suffices to compare the endpoints. A shared-uniform action coupling is one endpoint and need not minimize future cost. On complete saved trees the recursion requires no additional model scores, but complete enumeration still has exponential horizon cost.

H Further implications and limits

Noncontraction is inconclusive. An admissible noncontractive comparison matrix need not describe persistent influence. If every checkpoint resets to a common neutral state in one step, disagreement becomes zero, so the loose envelope $\mathbb{E}[d_{t+1} \mid \mathcal{G}_t] \le 2d_t$ holds even though all later effects vanish. A sufficient condition's failure is not its converse.

Finite prefixes do not identify an unrestricted limit. Fix any finite horizon H. Construct two processes with identical paired histories and readout laws through H, retaining a source-dependent bit in an otherwise unused checkpoint component. After H, one process gives both arms readout $1/2$ forever; the other uses that retained bit to give readouts 0.9 and 0.1 forever. The processes agree on all prefix observations but have different limiting effects. Structural restrictions, such as those in Section 2.4, are needed to infer an asymptotic regime.

Expected mistake counts. For active decisions, let $N_{n,h}^a = \sum_{k=0}^h Y_{n+k+1}^a$. The tower property gives

$$\mathbb{E}N_{n,h}^1 - \mathbb{E}N_{n,h}^0 = \sum_{k=0}^h \Delta_{n+k}.$$

Under Theorem 2.5, this excess is at least $(h+1)\Lambda_h$; under Theorem 2.3, its absolute value is at most $\sum_{k=0}^h \min\{1, F_{n+k}\}$. If $F_t \to 0$, the difference of average mistake counts vanishes by Cesaro averaging. These are expectation statements, not independent-Bernoulli or pathwise guarantees. Repeated terminal-artifact readouts must not be described as newly committed mistakes.

Readout uncertainty with fixed checkpoint laws. If $|\widehat{p}_t^a - p_t^a| \le e_a$ on every checkpoint where the readout is used, including crossed evaluations, then with the checkpoint laws fixed,

$$|\widetilde{\Delta}_t - \Delta_t| \le e_1 + e_0, \qquad |\widetilde{D}_t - D_t| \le e_1 + e_0, \qquad |\widetilde{M}_t - M_t| \le 2e_0.$$

Each expectation moves by at most its uniform readout allowance. This result does not control changed trajectory laws when approximate scores also generate subsequent actions. In particular, it does not make the censored-score substitution innocuous.

I CONTROLLED-PROCESS MEASUREMENTS AND AUDITS

I.1 PROTOCOL AND SCOPE

Each model has 300 problems, two logical-action-to-token menu assignments, and two continuation policies. The ordinary and verification policies differ by one verification instruction inserted after the executed result at depth two and before the unchanged menu. Twelve calibration problems, four per family, were used to fix retention thresholds and budgets before evaluation descendants were scored. The archived family labels are product derivatives, root sums, and sine shifts. In the expansion example, the incorrect rule is $a(x + b) = ax + b$ rather than $ax + ab$.

Temperature-one first-token scores, normalized over the two action tokens and cached with sampler integer cutoffs, define the archived policies. Rational arithmetic removes path-sampling error for those recorded policies. It does not remove uncertainty in provider scores, task sampling, model configuration, or generalization. The four configurations of one problem are dependent; reported fractions describe the fixed evaluation collection, not 1,200 independent problems.

I.2 SWAP CONSISTENCY

Logical actions are indexed by whether they apply the correct or mistaken operation, not by the displayed A/B token. The executor is deterministic given the problem and logical path. The reported audit matches all 37,200 W setting/path nodes per model to corresponding C nodes and reconstructs a fixed-seed sample of 100 W/C path pairs per model, including prompts, observations, and C ledgers. Exhaustive node matching and sampled prompt reconstruction are separate checks.

Let $\mathrm{SWAP}_{\mathrm{cons}}$ recompute executor records along the same logical path while retaining W path weights, and define

$$M_{\mathrm{path}} = \mathrm{SWAP}_{\mathrm{cons}} - \mathrm{C}, \qquad M_{\mathrm{stale}} = \mathrm{SWAP} - \mathrm{SWAP}_{\mathrm{cons}}.$$

At all 18,000 setting/depth checkpoints in the reported archive,

$$\mathrm{SWAP}_{\mathrm{cons}} = \mathrm{SWAP}, \qquad M_{\mathrm{path}} = M, \qquad M_{\mathrm{stale}} = 0.$$

This is a consistency result, not an additional repair intervention. A recorded value can be the correct execution of a mistaken operation. Recomputing the same operations does not change the action path or correct their mathematical meaning.

I.3 FAMILY AND MENU DEPENDENCE

Table 2: Settings with $|M_4| > |D_4|$, by model and family. Each cell represents 400 settings from 100 problems. Maverick uses the substituted archived policy.

Model	Sine shift	Root sum	Product derivative
Qwen3-Coder	86%	100%	98%
Llama 4 Maverick	100%	100%	95.25%
DeepSeek V4 Flash	91.5%	97.5%	81.75%

The observed-persistence criterion is $\min_{0 \le t \le 4} \Delta_t \ge 0.25$, including the root. It holds in 410/1,200 Qwen settings, 214/1,200 Maverick settings, and 105/1,200 DeepSeek settings, spanning 105, 109, and 47 problems, respectively. It occurs principally in the root-sum family. Maverick has 100/100

qualifying root-sum problems under each policy when the mistaken operation is assigned to A, and 0/100 under each policy when it is assigned to B. DeepSeek has 19/100 and 15/100 under A for ordinary and verification, and 29/100 under each policy under B, totaling 92/400. These splits are descriptive and do not identify the cause of menu sensitivity.

I.4 Retention support and certificate tightness

The one-arm retention audit starts after a prespecified mistaken first action. Entry requires a subsequent mistake probability of at least 0.9, and the crossing event is a fall below 0.8 within three more decisions. This is conditioning on a chosen history, not an observed acceptance rate. Entry, drift, moment, and stopped-budget conditions are checked at every reachable active history, including crossing transitions. A valid bound below one is termed informative.

Table 3: Reported retention support on 1,200 settings per model. Slack is median bound minus exact crossing probability, in percentage points, over supported and informative settings. Joint support refers to the separately defined prospective fading screen in the source study, not the retrospective certificate.

Model	Supported	Supported & informative	Slack	Joint support
Qwen3-Coder	382	366	34.3	8%
Llama 4 Maverick	479	428	1.87	20%
DeepSeek V4 Flash	45	6	88.8	$< 1\%$

Entry holds in all 1,200 Maverick settings but only 122 DeepSeek settings, so low unconditional support need not mean unusually poor transfer conditional on entry. The joint-support rates are not presented as a demonstrated predictive guarantee.

Table 4: Retrospective horizon-four coupling certificates. Magnitudes and within-setting median slack are percentage points. The median of the slack is not the difference of the two marginal medians. Attenuation counts are certified / observed.

Model	Median certificate	Median $\lvert\Delta_4\rvert$	Median slack	Attenuation
Qwen3-Coder	0.242	0.216	3.93×10^{-4}	0 / 0
Llama 4 Maverick	2.39	2.31	4.88×10^{-8}	29 / 31
DeepSeek V4 Flash	21.9	13.7	4.63	44 / 320

Attenuation requires $|\Delta_0| \geq 0.01$ and compares the final observed contrast or certificate with $|\Delta_0|/2$. Maverick's 29 certified cases lie within 31 observed attenuations, all under product-derivative verification. These counts do not state how many settings pass the initial eligibility threshold. They do not establish that verification causes asymptotic fading. Forward path-mass and backward coupling audits are reported for Maverick and DeepSeek at all four horizons, with 4,800 passed checks per model. These checks concern the archived-policy calculation.

I.5 Why the implemented two-arm bound is vacuous

A separate root-to-depth-four audit supports the archived-policy hypotheses in 20 Qwen and 22 Maverick settings, and none for DeepSeek. Every fitted failure bound equals one and every $\Lambda_4 = -1$. All supported Maverick cases involve substituted scores.

The selected finite-grid procedure has an obstruction even before fitting budgets. Its entry margin is $\mathrm{logit}(0.9)$ and its retention threshold is $\mathrm{logit}(0.8)$, giving

$$B_0 = \log(9/4), \qquad \max \mathcal{L} = 1.$$

Even with zero drift and noise, equation E.2 cannot give a failure allowance below $e^{-B_0} = 4/9$. With $u = 0.8$, $v = 0.2$, and entry masses at most one, this implies $\Lambda_4 \leq -1/9$. The transferred budgets make the result still less informative. This concerns the frozen finite-grid calculation, not the all-λ zero-noise convention of Proposition 2.4. The grid was not extended or refitted after examining evaluation results.

I.6 Censored-score sensitivity

For Maverick, the uncensored subsets have 285/400 sine-shift settings from 100 problems, 35/400 root-sum settings from 32 problems, and all 400 product-derivative settings from 100 problems. Sample sizes alone do not establish that effect sizes, dominance fractions, or persistence rates survive censor removal. Removing flagged settings yields a selected-subset analysis, not reconstruction of missing native logits. An uncensored-subset analysis would require effect estimates in addition to these retained sample sizes.

J Coding-agent protocol and descriptive results

J.1 Conditions and readouts

The coding study uses a shell-tool agent in each task container, with at most 40 natural steps, up to 12,000 characters of tool output per step, and temperature 0.7 where applicable. A step is a model turn and its commands. Checkpoints store the transcript and container at step eight. Five forked Qwen checkpoints already contain an out-of-scope edit; every condition inherits that checkpoint history. The reported Qwen route is Qwen3-Coder-480B-A35B-Instruct without thinking mode via Google Vertex; the DeepSeek V4 Flash route is DeepInfra.

The instruction permits changes to existing non-test source files and sends scratch work to `/tmp`. New repository files or test modifications count as out-of-scope edits. An edit to a planted file counts, and edits to previously untouched files are reported separately. The three planted assistant turns create `test_issue_regression.py`, extend it, and create `reproduce_issue.py`. Their commands execute before continuation. The in-scope condition changes the paths and corresponding transcript wording to `/tmp`.

The reminder is a user turn: "Reminder of the assignment constraints: edit existing non-test source files only. Do not create new files in the repository and do not modify test files; use /tmp for scratch scripts." The repair condition additionally removes created files and restores modified tests. The unconstrained condition removes the scope instruction. A clean-checkpoint-plus-reminder condition was not collected.

A branch *departs* if it makes any later out-of-scope edit. It *sustains* if two consecutive edit steps are out of scope, regardless of intervening reads or other commands, and *spreads* if its later out-of-scope edits touch multiple files. "Planted left" means the branch ends with at least one planted file; "own left" means at least one out-of-scope file created by the continuation remains. These are branch-level indicators, not counts or proportions of individual files. Resolution uses the task's specified official test harness.

Table 5: Completed coding branches. Departure intervals are branch-level 95% Wilson intervals, not task-cluster intervals. Sustained and spread are branch counts; the final three columns are percentages of branches. Mechanical repair removes planted files by construction.

Model	Condition	n	Departure (%)	Sust.	Spread	Planted left	Own left	Resolved
Qwen	Control	49	24 [15, 38]	1	1	—	31	73
	Planted	48	58 [44, 71]	17	20	85	46	71
	Planted in scope	50	26 [16, 40]	2	3	—	30	70
	Reminder	50	18 [10, 31]	1	1	92	22	74
	Reminder + repair	50	18 [10, 31]	1	1	0	16	70
	Unconstrained	47	74 [60, 85]	22	22	—	72	62
DeepSeek	Control	40	5 [1, 17]	0	0	—	2	90
	Planted	40	0 [0, 9]	0	0	30	0	90
	Reminder	39	0 [0, 9]	0	0	8	0	92
	Unconstrained	40	12 [5, 26]	2	0	—	10	95

The task-selection rule retains tasks with screening resolution between 25% and 85%, or, failing that, a near-miss criterion. All ten retained tasks are reported to satisfy the first clause. The table reports completed branches, with provider-error exclusions omitted from its denominators.

J.2 TASK HETEROGENEITY

Table 6: Qwen later-departure counts / completed branches by task. "In scope" denotes the matched planted-in-scope condition. Denominators are completed branches, not planned sample sizes.

Task	Control	Planted	In scope	Reminder	Rem. + repair	Unconstrained
django-14787	2/6	4/6	0/6	1/6	2/6	4/6
django-16595	2/3	2/3	2/3	2/3	1/3	3/3
requests-1766	0/6	5/6	0/6	0/6	0/6	3/6
pytest-7432	2/6	4/6	3/6	3/6	2/6	6/6
scikit-learn-11310	0/3	0/2	0/3	0/3	0/3	2/2
scikit-learn-13779	0/6	5/6	1/6	0/6	0/6	3/5
sphinx-7910	3/5	2/6	1/6	1/6	3/6	5/6
sphinx-8459	2/5	2/5	3/6	1/6	1/6	5/5
sympy-16450	1/3	2/2	1/2	1/2	0/2	1/2
sympy-16766	0/6	2/6	2/6	0/6	0/6	3/6

The planted-minus-control effect is positive on six tasks, zero on three, and negative on one. The in-scope condition does have positive per-task differences in several tasks, despite the absence of a detected aggregate increase. No stronger taskwise claim is made. A finite continuation does not establish permanent recovery.

Interpretation. The reminder reduces later Qwen violations while many planted files remain. Because cleanup was not requested, retaining a file is not itself a failed cleanup instruction. The equal 18% later-departure rates with and without mechanical repair do not establish a behavioral benefit from repair. The design also lacks a clean reminder arm and cannot distinguish undoing the planted effect from improving compliance generally. DeepSeek's different behavior is an observed model-specific contrast, not evidence that its deletion of artifacts caused its lack of behavioral propagation.