Closed-Loop Knowledge Dynamics: An Operational Framework for Saturation and Escape

Xuening Wu,[1] Shan Yu,[2] Shenqin Yin[3,†]

[1]Pfizer, Shanghai, China
[2]Independent Researcher, Hangzhou, China
[3]Institute of Humanities and Social Science Data, Fudan University, Shanghai, China
[†]Corresponding author: ysq@fudan.edu.cn

arXiv:2607.14185v1 [cs.LG] 15 Jul 2026

Abstract

Feedback-driven loops support iterative improvement in large language models, reinforcement learning, and autonomous discovery, yet their gains often diminish under repeated internal feedback. We study why closed-loop knowledge systems saturate and what external information can move them beyond their current attractors. We introduce a three-level operational framework in which knowledge states x_t evolve through transition kernels K_θ indexed by a structural parameter θ. The governing structure is defined as the observational equivalence class of θ induced by these kernels, while attractors and basins are properties of the fixed-θ dynamics. A structural intervention changes θ and produces a detectable kernel discrepancy on pre-specified probe states, making structural change falsifiable. Using a Lyapunov drift condition, we show that stable internal dynamics approach bounded stability regions with exponentially attenuated transients and a noise-controlled residual floor. We characterize escape through a metric condition on intervention-induced attractor displacement and a baseline-relative KL lower bound for increasing escape probability. This analysis also explains why conditional mutual information alone cannot certify escape: it measures variation among intervention-conditioned updates rather than departure from the no-intervention law. Case studies in LLM code repair, sparse-reward reinforcement learning, and Bayesian optimization use matched continuation controls to illustrate how feedback strength and alignment affect quality-improving escape. Our contribution is an operational connection among stability tools, measurable intervention effects, and cross-domain diagnostics.

1 Introduction

Modern AI systems increasingly rely on closed feedback loops. Large language models (LLMs) refine their outputs through self-reflection and revision, reinforcement learning (RL) algorithms update policies from environmental rewards, and autonomous discovery systems revise hypotheses based on experimental evidence. Despite their different forms, these systems share a common structure: a knowledge representation produces an output, receives feedback, and updates itself repeatedly.

A striking empirical pattern appears across such systems. Improvement is often rapid in early iterations, but marginal gains gradually diminish as the loop is repeated. LLM self-refinement improves drafts for several rounds and then stabilizes; RL policies approach stable value or policy regions; and belief-guided discovery systems increasingly concentrate their search as uncertainty is resolved. This suggests that saturation may not be merely an artifact of a particular algorithm, prompt, or benchmark, but a broader phenomenon of feedback-driven knowledge evolution.

This raises a basic question: what governs the long-term dynamics of closed-loop knowledge systems, and under what conditions can a saturated system move beyond its current attractor? Existing theories provide partial answers within individual domains. Contraction mappings explain the convergence of Bellman operators in RL; Bayesian optimization analyzes uncertainty-guided exploration; dynamical systems theory provides tools for stability and attractors; and information theory quantifies the value of new observations. These theories provide the relevant mathematical ingredients, but they do not by themselves supply a shared experimental interface for distinguishing three different observations: movement of the current state, a persistent change in the update law, and quality-improving escape from a pre-specified basin.

We develop an operational framework for *closed-loop knowledge dynamics*. The framework separates three levels: observable knowledge representations x_t, transition kernels K_θ, and an explicit structural parameter θ. A fixed θ induces the internal-loop kernel whose attractors and basins are then derived. Structural change is defined as an update of θ that produces a detectable kernel discrepancy on pre-specified probes; a one-off state change without such a discrepancy is not labeled structural.

The resulting picture is a saturation–escape principle. Stable internal feedback drives knowledge representations toward epistemic attractors, where marginal improvement becomes increasingly limited. Escaping such regimes requires more than additional internal iteration: it requires external input that changes the probability of leaving the old basin. We characterize this through a metric sufficient condition for attractor relocation and a baseline-relative information-divergence requirement for increasing escape probability.

Our contributions are as follows:

- We formulate *closed-loop knowledge dynamics* as a three-level operational model separating observable representations, transition kernels, and an explicit structural parameter. We define structural change through a pre-specified kernel discrepancy, making the governing structure identifiable up to observational equivalence and empirically falsifiable.
- We turn classical Lyapunov drift and contraction-plus-noise conditions into cross-system saturation diagnostics, including an exponential-plus-floor signature and explicit residual stability regions.
- We derive quantitative escape conditions for saturated systems. A metric condition relates parameter-induced attractor relocation to one-step escape, while a baseline-relative KL bound gives a necessary information-divergence budget for increasing escape probability. We explicitly separate this quantity from conditional mutual information.
- We evaluate the resulting diagnostics in LLM code repair, sparse-reward reinforcement learning, and Bayesian optimization using controls that separate additional internal iteration from the effect of external feedback.

The mathematical ingredients are deliberately classical. The methodological contribution is their coupling into falsifiable hypotheses with matched baseline distributions, pre-specified escape events, and explicit tests for structural change. This positioning allows the framework to be evaluated by whether it improves measurement and intervention design across systems, rather than by whether it introduces a new Lyapunov or information inequality.

2 Related Work

Dynamical-systems views of learning. Dynamical-systems tools provide a natural language for analyzing iterative learning processes. Robbins–Monro stochastic approximation and its ODE method [Robbins and Monro, 1951, Kushner and Yin, 2003, Borkar, 2008] study convergence toward roots of expected vector fields, while Foster–Lyapunov drift criteria characterize stochastic stability through one-step contraction inequalities [Meyn and Tweedie, 2009]. Contraction mappings also underlie classical RL guarantees, such as the γ-contraction of the discounted Bellman operator [Bertsekas and Tsitsiklis, 1996]. These theories provide powerful tools for analyzing a fixed transition rule, but they typically do not distinguish the observable knowledge representation, the transition operator, and the higher-level structure that determines which attractors and basins the operator induces.

Recursive self-refinement and saturation. LLM self-refinement methods such as Self-Refine [Madaan et al., 2023] and Reflexion [Shinn et al., 2023] show that repeated critique and revision can improve model outputs over multiple rounds. Fixed-point iteration is often invoked informally to explain eventual stabilization, and recent work reports convergence-like behavior in latent concepts and uncertainty during self-correction [Liu et al., 2025]. However, these works do not provide a common operational test that separates trajectory stabilization, structural change, and quality-improving escape. We complement this literature by connecting an exponential-plus-floor diagnostic under contraction-plus-noise to a metric relocation condition and a baseline-relative information-divergence threshold.

Closed-loop discovery, effective landscapes, and external information. Bayesian optimization and bandit methods study how evidence guides search under uncertainty; GP-UCB balances posterior mean estimates with uncertainty bonuses [Srinivas et al., 2010]. Recent closed-loop discovery frameworks, such as GAMBLe, analyze generator–assessor–search systems and the effective landscapes they induce [Ellis and Castro, 2026], while work on self-evolving language models studies when internal self-play plateaus without learnable information gain [Liu et al., 2026]. LLM-guided and multi-fidelity Bayesian optimization methods such as LABO and LGBO further examine when auxiliary or low-cost LLM signals accelerate search and when they should be gated or incorporated conservatively [Chen et al., 2026, Yuan et al., 2026]. These works are closely related to our view that external information helps only when it is sufficiently reliable and informative, but our focus is on general saturation and escape conditions for closed-loop knowledge dynamics rather than a specific search-system decomposition or acquisition rule.

3 Mathematical Framework

3.1 Closed-Loop Knowledge Dynamics

We formalize closed-loop AI systems as processes in which a knowledge representation repeatedly produces an output, receives feedback, and updates itself. The central object is not a particular algorithm, but the evolution of knowledge representations under feedback.

Let $\mathcal{X}$ denote a knowledge representation space. An element $x_t \in \mathcal{X}$ may represent a draft, answer, policy, hypothesis, belief state, or other internal representation of knowledge. A closed-loop knowledge system consists of an output operator, a feedback operator, and an update operator:

$$\mathcal{K} = (\mathcal{X}, \mathcal{Y}, \mathcal{E}, \Omega, G, H, \Phi),$$

where $\mathcal{Y}$ is the output space, $\mathcal{E}$ is the feedback space, Ω is the external information space, $G : \mathcal{X} \to \mathcal{Y}$ is the output operator, $H : \mathcal{Y} \times \Omega \to \mathcal{E}$ is the feedback operator, and $\Phi : \mathcal{X} \times \mathcal{E} \to \mathcal{X}$ is the update operator.

The system evolves according to

$$\begin{aligned} y_t &= G(x_t), \\ e_t &= H(y_t, \omega_t), \\ x_{t+1} &= \Phi(x_t, e_t), \end{aligned} \tag{1}$$

where $\omega_t \in \Omega$ denotes external information, such as an environment, verifier, user, experiment, tool output, or observation. The feedback signal e_t need not be a prediction error; depending on the system, it may be a reward, critique, preference, verifier judgment, experimental measurement, or observation-derived signal.

Equation (1) induces a transition operator

$$\mathcal{T}_\omega(x) = \Phi(x, H(G(x), \omega)). \tag{2}$$

Thus, at the level of knowledge representations, the closed loop can be written as the discrete-time dynamical system

$$x_{t+1} = \mathcal{T}_{\omega_t}(x_t). \tag{3}$$

3.2 Three Levels of Closed-Loop Knowledge Dynamics

The induced transition operator $\mathcal{T}_\omega$ allows closed-loop AI systems to be analyzed using dynamical-systems tools. To distinguish a temporary state perturbation from a persistent change in the update law, we make the third level explicit rather than defining it retrospectively from an observed trajectory.

Level 1: Knowledge representation. The observable object is the knowledge representation $x_t \in \mathcal{X}$. This may be a text draft in recursive self-refinement, a policy in reinforcement learning, a hypothesis in autonomous discovery, or a belief state in Bayesian search.

Level 2: Transition dynamics. Let Θ be a structural parameter space and allow the operators (G, H, Φ), including their sampling randomness, to depend on $\theta \in \Theta$. Conditional on θ and feedback content ω, the next state is generated by a Markov kernel

$$x_{t+1} \sim K_\theta^\omega(\cdot \mid x_t).$$

The kernel includes deterministic transitions as Dirac measures. Internal feedback has a reference law $\nu_\theta^{\mathrm{int}}(d\omega \mid x)$ and therefore induces the internal-loop kernel

$$K_\theta^{\mathrm{int}}(A \mid x) = \int K_\theta^\omega(A \mid x) \nu_\theta^{\mathrm{int}}(d\omega \mid x).$$

Attractors, invariant sets, and stability basins are derived properties of K_θ^{int}; they are not included in the definition by fiat.

Level 3: Governing structure. The governing structure is represented by θ_t. A structural input $z_t \in \mathcal{Z}$ updates it according to

$$\theta_{t+1} = U(\theta_t, z_t), \qquad U(\theta_t, \varnothing) = \theta_t. \tag{4}$$

Examples include changing a verifier or controller, incorporating a persistent demonstration update, modifying a search prior, or adding information to a context that is retained in subsequent rounds. By contrast, feedback that changes x_{t+1} while leaving $\theta_{t+1} = \theta_t$ is a state-level perturbation, not a governing-structure shift.

Different parameter values can induce indistinguishable dynamics. We therefore define the governing structure as an observational equivalence class rather than as a syntactic parameter value. Fix in advance a probe distribution μ over states, a probe-feedback law ν^{probe}, and a metric d on $\mathcal{X}$. Assume the probe kernels have finite first moments under d. Define the structural discrepancy

$$\delta_{\mu,\nu}(\theta, \theta') := \mathbb{E}_{\substack{x\sim\mu \\ \omega\sim\nu^{\mathrm{probe}}(\cdot|x)}} \left[W_d(K_\theta^\omega(\cdot \mid x), K_{\theta'}^\omega(\cdot \mid x)) \right], \tag{5}$$

where W_d is the 1-Wasserstein distance induced by d. We write $\theta \sim_{\mu,\nu} \theta'$ when $\delta_{\mu,\nu}(\theta, \theta') = 0$ and define

$$\mathfrak{G}_{\mu,\nu}(\theta) := [\theta]_{\sim_{\mu,\nu}}.$$

Thus $\mathfrak{G}_{\mu,\nu}$ is identifiable only up to conditional transition behavior on the declared state–feedback probes, which is the strongest notion available from black-box trajectories.

Operational view. Choose a smallest scientifically meaningful effect size $\varepsilon_G > 0$ before collecting post-intervention trajectories. A claimed structural change from θ to θ' is supported only when an estimator of $\delta_{\mu,\nu}(\theta, \theta')$ has a confidence interval whose lower endpoint exceeds ε_G. The claim is falsified, at the chosen resolution, when an equivalence test places the upper endpoint below ε_G. Importantly, a large value of $d(x_t, x_{t+1})$ is not by itself evidence of structural change. The evidence must be a reproducible change in the conditional transition law on pre-specified probes. In Sections 4 and 5, fixed governing dynamics mean $\theta_t \equiv \theta$; a structural intervention means both $\theta' = U(\theta, z) \neq \theta$ and $\delta_{\mu,\nu}(\theta, \theta') > \varepsilon_G$.

4 Saturation under Fixed Governing Dynamics

We first analyze the regime in which $\theta_t \equiv \theta$. In this regime, repeated internal feedback does not change the attractor geometry of the system; it only evolves the current knowledge representation under the induced transition operator. We show that a general Lyapunov drift condition is sufficient for *saturation*: the expected Lyapunov value approaches a bounded stability region, and marginal improvement becomes asymptotically limited by the residual noise floor.

4.1 A General Drift Condition for Saturation

Let $\mathcal{K} = (\mathcal{X}, \mathcal{Y}, \mathcal{E}, \Omega, G, H, \Phi)$ be a closed-loop knowledge system with induced transition operator

$$\mathcal{T}_\omega(x) = \Phi(x, H(G(x), \omega)).$$

The knowledge representation evolves as

$$x_{t+1} = \mathcal{T}_{\omega_t}(x_t).$$

We now state a general sufficient condition under which such dynamics saturate.

Theorem 1 (General saturation under Lyapunov drift)**.** *Let $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ be a Lyapunov function. Suppose there exist $\rho_t \in [0, 1)$ and $\sigma_t \geq 0$ such that, for the sequence generated by $x_{t+1} = \mathcal{T}_{\omega_t}(x_t)$,*

$$\mathbb{E}[V(x_{t+1}) \mid x_t] \leq \rho_t V(x_t) + \sigma_t \qquad \text{for all } t.$$

Then

$$\mathbb{E}[V(x_t)] \leq \left(\prod_{s=0}^{t-1} \rho_s \right) V(x_0) + \sum_{s=0}^{t-1} \left(\prod_{r=s+1}^{t-1} \rho_r \right) \sigma_s.$$

In particular, if $\rho_t \equiv \rho$ and $\sigma_t \equiv \sigma$ are constant, then

$$\mathbb{E}[V(x_t)] \leq \rho^t V(x_0) + \frac{\sigma}{1-\rho}.$$

Thus the expected Lyapunov value approaches a bounded stability region whose size is controlled by the accumulated noise terms.

Proof. The result follows by induction. Taking expectations in the one-step drift condition gives

$$\mathbb{E}[V(x_{t+1})] \leq \rho_t \mathbb{E}[V(x_t)] + \sigma_t.$$

Unrolling this recursion yields the stated time-varying bound. When $\rho_t \equiv \rho$ and $\sigma_t \equiv \sigma$, the accumulated noise term is bounded by the geometric series

$$\sum_{s=0}^{t-1} \rho^s \sigma \leq \frac{\sigma}{1-\rho},$$

which gives the constant-rate form. □

Theorem 1 does not require the transition operator to be gradient-based, differentiable, or explicitly known. It only assumes a one-step contraction-plus-noise inequality on a Lyapunov function. This makes the result applicable to heterogeneous closed-loop systems, including LLM-based refinement loops, Bayesian belief updates, reinforcement learning operators, and other feedback-driven knowledge processes.

Corollary 1 (Saturation via constant contraction)**.** *Suppose there exists $x^* \in \mathcal{X}$, a metric d on $\mathcal{X}$, $\rho \in [0, 1)$, and $\sigma \geq 0$ such that*

$$\mathbb{E}[d(x_{t+1}, x^*) \mid x_t] \leq \rho d(x_t, x^*) + \sigma.$$

Then Theorem 1 applies with $V(x) = d(x, x^*)$, and the system approaches the stability region

$$B\left(x^*, \frac{\sigma}{1-\rho}\right).$$

4.2 Instantiation I: Recursive Self-Refinement

We first instantiate the theory for recursive self-refinement. Here $\mathcal{X}$ is a text space equipped with a metric d, such as normalized edit distance or embedding distance. The knowledge representation x_t is the current draft or answer. The output operator is the identity map, the feedback operator corresponds to self-reflection or critique generated internally by the model, and the update operator is the LLM revision step. In the absence of external information, the system evolves under an effectively fixed governing structure induced by the model, prompt, and internal feedback rule.

Because decoding at nonzero temperature is stochastic, we write the refinement step as

$$x_{t+1} = \Phi(x_t, \omega_t),$$

where ω_t denotes fresh sampling randomness at step t.

Assumption 1 (Contraction toward a soft attractor)**.** *There exist $x^* \in \mathcal{X}$ and $\rho \in [0, 1)$ such that the deterministic component of the refinement operator satisfies*

$$d(\Phi_{\text{det}}(x), x^*) \leq \rho d(x, x^*) \qquad \text{for all } x \in \mathcal{X}.$$

Assumption 2 (Bounded decoding noise)**.** *For any $x \in \mathcal{X}$ and independent noise draws ω, ω',*

$$d\big(\Phi(x, \omega), \Phi(x, \omega')\big) \leq \sigma.$$

At temperature 0, this corresponds to $\sigma = 0$. More generally, the noise level may depend on the local peakedness of the model distribution; the constant σ assumption is used here for tractability and to obtain a closed-form saturation bound.

Proposition 1 (Exponential saturation under contraction and noise)**.** *Assume that there exist $x^\star \in \mathcal{X}$, $\rho \in [0, 1)$, and $\sigma \geq 0$ such that*

$$\mathbb{E}[d(x_{t+1}, x^\star) \mid x_t] \leq \rho d(x_t, x^\star) + \sigma.$$

Then

$$\mathbb{E}[d(x_t, x^\star)] \leq \rho^t d(x_0, x^\star) + \frac{\sigma(1 - \rho^t)}{1 - \rho}.$$

In particular, the trajectory approaches the stability basin

$$B\left(x^\star, \frac{\sigma}{1 - \rho}\right).$$

Moreover, the expected transition magnitude

$$\Delta_t := \mathbb{E}[d(x_t, x_{t+1})]$$

satisfies

$$\Delta_t \leq (1 + \rho)\rho^t d(x_0, x^\star) + \frac{2\sigma}{1 - \rho}.$$

Thus recursive self-refinement exhibits exponential decay in transition magnitude up to a noise-controlled floor.

Proof. Deferred to Appendix A. □

Remark 1 (Consistency with observed relaxation dynamics)**.** *The fitted exponential-relaxation model*

$$\Delta_t \approx Ae^{-kt} + c$$

reported for recursive abstract refinement under default-temperature decoding ($A = 0.0585$, $k = 0.923$, $c = 0.0070$, $R^2 = 0.990$) is consistent with the exponential-plus-floor form predicted by Proposition 1. The fitted decay rate implies an effective contraction rate

$$\rho = e^{-k} \approx 0.397.$$

Since Proposition 1 gives the step-size floor

$$\frac{2\sigma}{1-\rho},$$

interpreting the empirical floor c through this upper bound gives

$$\sigma \approx \frac{c(1-\rho)}{2} \approx 0.0021.$$

At temperature 0, Assumption 2 corresponds to $\sigma \to 0$, predicting a vanishing noise floor and convergence toward an exact fixed point. Proposition 1 also yields a stopping criterion: once the exponentially decaying term is dominated by the fitted floor, additional refinement is expected to be noise-limited rather than contraction-limited.

Remark 2 (Scope of the contraction assumption)**.** *Assumption 1 is a sufficient condition, not an independently verified mechanistic claim. A single refinement trajectory reveals the transition magnitudes Δ_t, but does not by itself verify pairwise contraction across distinct initial states. A direct test would refine multiple initial drafts and measure whether pairwise distances between trajectories contract at a rate consistent with the ρ estimated from single-trajectory relaxation.*

4.3 Instantiation II: Belief-Guided Discovery

We next consider belief-guided autonomous discovery. Here the knowledge representation x_t is a belief state, such as a posterior distribution over hypothesis quality. The output operator proposes candidates according to the current belief, the feedback operator returns experimental or evaluation outcomes, and the update operator performs belief updating. This gives another closed-loop knowledge process:

$$\text{belief} \;\to\; \text{candidate} \;\to\; \text{evidence} \;\to\; \text{updated belief}.$$

Unlike recursive self-refinement, belief-guided discovery need not satisfy a global contraction condition on the posterior state. We therefore analyze a more limited but directly observable stabilization mechanism: the decay of exploration pressure as uncertainty is resolved.

Consider the acquisition rule

$$a_t(h) = -\mu_t(h) + \beta_t \sigma_t(h), \qquad \beta_t = \beta_0 \sqrt{\frac{n_0}{t}},$$

where $\mu_t(h)$ and $\sigma_t(h)$ are the posterior mean and standard deviation of hypothesis h, and n_0 is the number of initialization evaluations.

Assumption 3 (Bounded posterior uncertainty)**.** *There exists $S < \infty$ such that*

$$0 \le \sigma_t(h) \le S \qquad \textit{for all } h, t.$$

Define the average posterior uncertainty over the current candidate pool $\mathcal{C}_t$ by

$$V_t := \mathbb{E}_{h \sim \mathcal{C}_t}[\sigma_t(h)],$$

and define the exploration-pressure quantity

$$W_t := \beta_t V_t.$$

Proposition 2 (Vanishing exploration pressure)**.** *Under Assumption 3,*

$$0 \le W_t \le \beta_0 S \sqrt{\frac{n_0}{t}},$$

and therefore

$$W_t = O(t^{-1/2}) \to 0.$$

Consequently, the uncertainty-driven component of the acquisition rule becomes asymptotically negligible, and candidate selection becomes increasingly dominated by the posterior mean term.

Proof. Since $0 \le \sigma_t(h) \le S$ for all h, t, we have

$$0 \le V_t = \mathbb{E}_{h \sim \mathcal{C}_t}[\sigma_t(h)] \le S.$$

Multiplying by $\beta_t = \beta_0 \sqrt{n_0/t}$ gives

$$0 \le W_t = \beta_t V_t \le \beta_0 S \sqrt{\frac{n_0}{t}}.$$

Thus $W_t = O(t^{-1/2})$ and $W_t \to 0$. □

Remark 3 (Scope of Proposition 2)**.** *Proposition 2 does not claim posterior contraction, regret minimization, or convergence to a global optimum. It shows only that the exploration-pressure term induced by the acquisition rule vanishes at rate $O(t^{-1/2})$. Thus, this result should be interpreted as characterizing the attenuation of one feedback-driven perturbation, not as a complete convergence theorem for Bayesian optimization. A stronger result would require a drift inequality directly on the posterior-uncertainty process, such as*

$$\mathbb{E}[V_{t+1} \mid x_t] \le \rho_t V_t + \sigma_t,$$

under additional assumptions on the candidate pool, kernel, observation schedule, and posterior update.

Remark 4 (Consistency with productive concentration)**.** *Proposition 2 predicts a transition from broad, uncertainty-driven exploration to increasingly mean-driven selection as t grows. This is consistent with the empirically observed productive concentration pattern in belief-guided discovery: candidate diversity is high in early iterations and decreases in later iterations even as objective values continue to improve. In this sense, belief-guided discovery illustrates a stabilization mechanism distinct from contraction: the governing dynamics do not necessarily contract globally, but one source of exploration-induced variation becomes progressively attenuated.*

5 Escape via External Information

Section 4 analyzed saturation under effectively fixed governing dynamics. We now ask when a saturated closed-loop system can move beyond its current attractor. In the parameterized view introduced in Section 3, escape and structural change are distinct hypotheses. Escape means that a state leaves a pre-specified old basin. Structural change means that an intervention updates θ and produces a probe-kernel discrepancy larger than ε_G. A state can leave a basin because of noise without a structural change, and a structural change need not produce immediate escape.

We formalize this idea in two steps. First, conditional on a parameter update that induces a new locally contractive kernel, we derive a metric sufficient condition for one-step escape in expected distance. Second, we derive a baseline-relative information-divergence condition for increasing the probability of leaving the old basin. The second result does not identify attractor distance with conditional mutual information.

5.1 Why Internal Iteration Cannot Relocate an Attractor

We first record a structural observation. If the governing structure remains fixed, repeated internal iteration may move the state within the existing basin, but it cannot by itself relocate the attractor that defines the basin.

Corollary 2 (Fixed-parameter iteration cannot relocate the attractor)**.** *Suppose $\theta_t \equiv \theta$ and the internal kernel K_θ^{int} has a fixed attractor x_θ^*. Internal iteration can change x_t but cannot change the attractor of the fixed kernel. Relocating the attractor therefore requires a parameter update $\theta' = U(\theta, z)$ for which $K_{\theta'}^{\text{int}}$ has a different attractor. Notice that leaving $B(x_\theta^*, c)$ does not logically require such an update: a sufficiently large stochastic state perturbation may also leave the basin. What requires a structural update is persistent relocation of the transition law or its attractor, not a single escape event.*

This corollary prevents a definitional shortcut in which every large state change is labeled a governing-structure shift. Repeating a fixed kernel does not relocate that kernel's attractor, but stochastic escape and structural relocation remain empirically separable events.

5.2 A Metric Escape Condition

Suppose that at time t_0 the system has saturated inside the stability basin of an attractor x_θ^*:

$$x_{t_0} \in B(x_\theta^*, c), \qquad c = \frac{\sigma}{1-\rho}.$$

A structural intervention z_{t_0} produces $\theta' = U(\theta, z_{t_0})$ and is first required to pass the operational test $\delta_{\mu,\nu}(\theta, \theta') > \varepsilon_G$. We model the resulting local kernel by a perturbed transition operator that contracts toward a new attractor $x_{\theta'}^*$ at rate $\rho' \in [0, 1)$ with perturbation noise $\sigma_{\text{ext}} \geq 0$. Let

$$\Delta := d(x_\theta^*, x_{\theta'}^*)$$

denote the induced attractor shift.

Proposition 3 (Metric escape threshold)**.** *Under the perturbed contraction model above, if*

$$\Delta > \Delta^* := \frac{c(1+\rho') + \sigma_{\text{ext}}}{1-\rho'},$$

then

$$\mathbb{E}[d(x_{t_0+1}, x_\theta^*)] > c,$$

so the expected distance of the next state exceeds the old stability radius. This is an expectation statement; it does not assert almost-sure escape.

Proof. Deferred to Appendix A. □

Proposition 3 states that escape requires more than novelty: under the stated local model, the displayed inequality is a sufficient condition for expected-distance escape. The induced attractor shift must exceed a threshold determined by the radius of the old basin, the contraction strength of the new local dynamics, and the noise introduced by the external input. It is not a necessary condition, and a state displacement without a verified kernel discrepancy is not evidence that θ changed.

5.3 Noise-Adjusted Displacement

External information can also increase uncertainty or noise. Under the perturbed dynamics, the post-intervention stability radius is

$$c' = \frac{\sigma + \sigma_{\text{ext}}}{1 - \rho'},$$

centered at the new attractor $x^*_{\theta'}$. To separate displacement from the accompanying change in local spread, define the following geometric diagnostic.

Definition 1 (Noise-adjusted displacement margin)**.** *The noise-adjusted displacement margin is*

$$M_{\text{shift}} := \Delta - (c' - c).$$

It is positive exactly when the attractor displacement exceeds the increase in the modeled stability radius, that is, when $\Delta > c' - c$.

Remark 5. *The sign of M_{shift} is a geometric comparison, not a utility guarantee. In particular, $M_{\text{shift}} > 0$ does not imply that the new attractor has higher task quality, and $M_{\text{shift}} \le 0$ does not imply that the intervention is harmful. A claim of beneficial escape additionally requires a pre-specified task-quality functional Q and evidence that the intervention improves it, as in the quality-aligned event defined below and the experimental metric E_{eff}.*

5.4 An Information-Theoretic Escape Threshold

The metric result treats an intervention-induced transition law as given. We now ask how different that law must be from the no-intervention law in order to increase the probability of leaving the old basin. This baseline-relative question is not answered by conditional mutual information alone.

Assumption 4 (Bounded representation space)**.** *The knowledge representation space $\mathcal{X}$ has finite diameter*

$$D := \sup_{x,x' \in \mathcal{X}} d(x, x') < \infty.$$

Let $P_z(\cdot \mid x_t)$ denote the conditional law of the next knowledge representation under a structural intervention z. Let $P_0(\cdot \mid x_t) = K^{\text{int}}_\theta(\cdot \mid x_t)$ be the no-intervention continuation law from the same state. Define the old basin and the one-step escape event by

$$B_0 := B(x^*_\theta, c), \qquad E_0 := \mathcal{X} \setminus B_0,$$

and let

$$p_z(x_t) := P_z(E_0 \mid x_t), \qquad p_0(x_t) := P_0(E_0 \mid x_t).$$

The event can be replaced by a pre-specified *useful escape* event, such as leaving B_0 while exceeding a fixed quality threshold; the same result then applies.

Proposition 4 (Baseline-relative information threshold)**.** *For every state x_t and intervention z,*

$$|p_z(x_t) - p_0(x_t)| \le \text{TV}(P_z(\cdot \mid x_t), P_0(\cdot \mid x_t)) \le \sqrt{\frac{1}{2}\,\text{KL}(P_z(\cdot \mid x_t) \,\|\, P_0(\cdot \mid x_t))}.$$

Consequently, increasing one-step escape probability by at least $\eta > 0$ requires

$$\text{KL}(P_z(\cdot \mid x_t) \,\|\, P_0(\cdot \mid x_t)) \ge 2\eta^2. \tag{6}$$

If absolute continuity fails, the KL divergence is infinite and the inequality holds trivially.

Proof. For the indicator $\mathbf{1}_{E_0}$, the difference of expectations under two laws is bounded by their total variation distance. Pinsker's inequality gives the second inequality. Rearranging when $p_z(x_t) - p_0(x_t) \geq \eta$ proves Equation (6). □

The preceding threshold is operational: both escape probabilities refer to the same pre-specified event and differ only in whether the intervention is present. The KL term is a property of the response distributions, not a claim about the number of tokens, demonstrations, or observations supplied.

Corollary 3 (Link from attractor relocation to information divergence)**.** *Assume the setting of Proposition 3 and Assumption 4, with $D > c$. Define*

$$L(\Delta) := (1 - \rho')\Delta - \rho' c - \sigma_{\text{ext}}, \qquad q(\Delta) := \left[\frac{L(\Delta) - c}{D - c}\right]_+ .$$

Whenever the assumptions are jointly feasible, $q(\Delta) \leq 1$, and the intervention law satisfies

$$p_z(x_{t_0}) \geq q(\Delta).$$

Therefore, if $q(\Delta) > p_0(x_{t_0})$, then necessarily

$$\mathrm{KL}(P_z(\cdot \mid x_{t_0}) \,\|\, P_0(\cdot \mid x_{t_0})) \geq 2\left(q(\Delta) - p_0(x_{t_0})\right)^2 .$$

Proof. The proof of Proposition 3 yields

$$\mathbb{E}_{P_z}[d(x_{t_0+1}, x^*_\theta) \mid x_{t_0}] \geq L(\Delta).$$

Because the distance is at most c inside B_0 and at most D outside,

$$\mathbb{E}_{P_z}[d(x_{t_0+1}, x^*_\theta) \mid x_{t_0}] \leq c + (D - c)p_z(x_{t_0}).$$

Combining the inequalities gives $p_z(x_{t_0}) \geq q(\Delta)$. The KL bound then follows from Proposition 4. □

Proposition 5 (Mutual information and baseline shift measure different effects)**.** *Let Z be a random external intervention, let P_z be the corresponding next-state law at a fixed x_t, and define the mixture $\bar{P} = \mathbb{E}_Z[P_Z]$. Assume $\mathbb{E}_Z \,\mathrm{KL}(P_Z \| P_0) < \infty$. Then*

$$\mathbb{E}_Z[\mathrm{KL}(P_Z \| P_0)] = I(x_{t+1}; Z \mid x_t) + \mathrm{KL}(\bar{P} \| P_0). \tag{7}$$

Proof. Add and subtract $\log d\bar{P}$ inside $\mathbb{E}_Z \,\mathrm{KL}(P_Z \| P_0)$. The first resulting term is $\mathbb{E}_Z \,\mathrm{KL}(P_Z \| \bar{P}) = I(x_{t+1}; Z \mid x_t)$; averaging the second term over Z gives $\mathrm{KL}(\bar{P} \| P_0)$. □

Equation (7) separates two notions that should not be conflated. Conditional mutual information measures how strongly the next-state law varies with the realized intervention around the intervention mixture. The second term measures how far the average intervention regime moves from the no-intervention baseline. A deterministic targeted intervention can have $I(x_{t+1}; Z \mid x_t) = 0$ while still producing escape, because its law may differ substantially from P_0. Conversely, large mutual information can encode intervention-specific variation with little average movement toward a useful basin. Claims about escape must therefore use the baseline-relative KL term (and a quality-aligned event when benefit is intended), rather than mutual information alone.

6 Experiments

6.1 Experiment 1: LLM Code Repair

Experimental setup. We study whether external feedback can move a self-refining coding agent out of a stable but incorrect solution basin. We use two real historical bugs from different repositories: a unit-rollover bug in `python-humanize`'s `naturalsize`, and an escaped-trailing-space bug in `python-pathspec`'s Git wildmatch normalization. For each task, we run five independent Phase 1 trajectories. The model recursively revises its own implementation for ten rounds without access to the hidden boundary tests, yielding a locally saturated checkpoint.

Scope and task selection. The two reported tasks were selected through exploratory screening rather than a preregistered confirmatory protocol. Screening sought cases in which Phase 1 produced an executable but hidden-test-incomplete solution and a stable late trajectory; it also inspected whether targeted Phase 2 feedback exposed an informative contrast. Because the latter criterion uses intervention outcomes, the reported task-level effects are descriptive case-study results and should not be interpreted as an unbiased estimate over a population of code-repair tasks. All screened candidates and their exclusion reasons are retained in the experiment database. For a future confirmatory extension, we pre-specify eligibility using Phase 1 information only: an executable solution, failure on at least one deterministic hidden boundary test, and no hidden-test-pass-rate improvement over the final three self-refinement rounds. Phase 2 outcomes will not be used for inclusion or exclusion in that extension.

Starting from each checkpoint, Phase 2 compares baseline continuation, generic feedback, error-only feedback, and mismatched feedback. For the `pathspec` task, we additionally evaluate diagnostic feedback that identifies the relevant root cause. Each trajectory is weighted equally in the reported aggregate, preventing trajectories with more decoding samples from dominating the result.

Let $Q(x)$ denote hidden-test pass rate. For feedback condition ω, we measure its displacement from baseline continuations and its associated quality gain as

$$E_{\text{shift}}(\omega) = \frac{1}{K}\sum_{i=1}^{K} \min_{j \le K} d\Big(x_{\omega}^{(i)}, x_{\text{base}}^{(j)}\Big), \qquad \Delta Q(\omega) = \bar{Q}_{\omega} - \bar{Q}_{\text{base}}.$$

We then define effective escape as

$$E_{\text{eff}}(\omega) = E_{\text{shift}}(\omega)\max\{\Delta Q(\omega), 0\},$$

which credits behavioral displacement only when it is accompanied by improved hidden-test performance.

Results. Both tasks exhibit stable but incorrect Phase 1 checkpoints. On `naturalsize`, baseline and generic continuation remain at a pass rate of 0.143, while targeted error feedback raises performance to 0.469 ± 0.045. The paired improvement is $\Delta Q = 0.326 \pm 0.045$, with a trajectory-bootstrap 95% confidence interval of $[0.240, 0.394]$. Generic feedback induces substantial code displacement but no quality improvement, and therefore has zero effective escape.

The cross-repository `pathspec` task reveals a sharper dependence on information content. Baseline, generic, and mismatched feedback all remain at 0.400 across every trajectory. Error-only feedback is unstable, ranging from unsuccessful or regressive revisions to complete repair; its mean gain is only 0.120 ± 0.140, and its bootstrap confidence interval $[-0.107, 0.387]$ includes zero. In

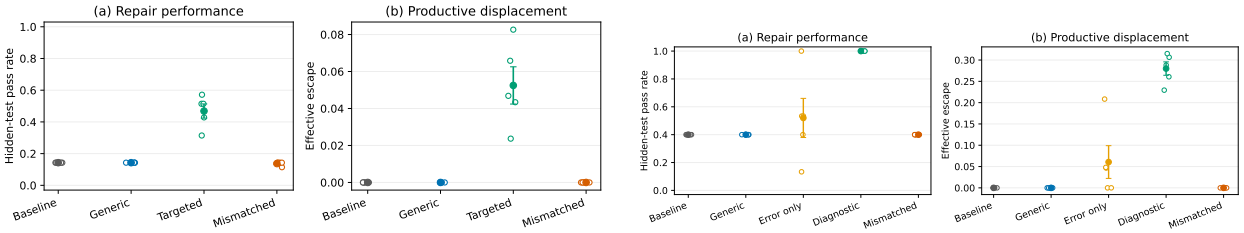

Figure 1: **Feedback-induced escape in LLM code repair.** Results are computed over five independent Phase 1 trajectories per task; points denote trajectories and filled markers show mean ± SEM. Left: on `naturalsize`, targeted error feedback improves repair performance, whereas generic and mismatched feedback do not. Right: on `pathspec`, error-only feedback is unreliable across trajectories, while root-cause diagnostic feedback achieves successful escape in all five trajectories. Generic feedback can produce substantial code displacement without positive effective escape.

Table 1: Code-repair results over independent Phase 1 trajectories. Values are trajectory-level mean ± SEM. ΔQ is measured relative to the corresponding task's baseline continuation. A dash denotes a condition not evaluated for that task.

Task	Condition	Traj.	Pass rate	ΔQ	E_{eff}
`naturalsize`	Baseline	5	0.143 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	Generic	5	0.143 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	Error-only	5	0.469 ± 0.045	0.326 ± 0.045	0.052 ± 0.010
	Mismatched	5	0.137 ± 0.006	-0.006 ± 0.006	0.000 ± 0.000
`pathspec`	Baseline	5	0.400 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	Generic	5	0.400 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	Error-only	5	0.520 ± 0.140	0.120 ± 0.140	0.061 ± 0.038
	Diagnostic	5	1.000 ± 0.000	0.600 ± 0.000	0.280 ± 0.016
	Mismatched	5	0.400 ± 0.000	0.000 ± 0.000	0.000 ± 0.000

contrast, root-cause diagnostic feedback reaches a pass rate of 1.000 in all five trajectories, yielding $\Delta Q = 0.600$ and $E_{\text{eff}} = 0.280 \pm 0.016$.

These results refine the escape hypothesis: task relevance alone does not guarantee reliable escape. Feedback must provide information aligned with the task-specific failure boundary. Generic perturbations move the generated code without improving correctness, and shallow error reports may succeed sporadically; sufficiently diagnostic feedback instead produces consistent, quality-aligned state-space escape. These experiments do not by themselves establish a governing-structure shift, which would additionally require the probe-kernel test in Equation (5).

Reproducibility. All Phase 1 trajectories, Phase 2 samples, hidden-test outcomes, token usage, estimated costs, and task-screening decisions are stored in SQLite. The screening records preserve each candidate task and its inclusion or exclusion reason; no claim of preregistration is made for the present exploratory task selection.

6.2 Experiment 2: Feedback-Driven Escape in Reinforcement Learning

We next examine whether task-relevant feedback can move a saturated policy into a successful behavioral regime, and whether this relocation persists after standard reinforcement learning resumes. We use `MiniGrid-DoorKey-8x8-v0`, where the agent must complete the compositional sequence

$$\text{pick up key} \;\rightarrow\; \text{open door} \;\rightarrow\; \text{reach goal}.$$

All results are reported over five paired seeds $\{7, 17, 42, 101, 137\}$.

Saturated checkpoint. For each seed, we train PPO using the original sparse reward and evaluate the policy every 10,000 environment steps. We declare saturation when the success-rate range over five consecutive evaluations is below 0.05. All five policies satisfy this criterion at 50,000 steps while retaining zero task success. Every intervention branch is initialized from its corresponding saturated checkpoint.

Feedback and continuation protocols. Targeted feedback consists of state–action examples generated by a deterministic expert with access to the key, door, and goal locations. The expert demonstrates the task-specific action sequence required for completion. We consider feedback budgets

$$m \in \{0, 1000, 10000, 20000\}.$$

Behavior cloning (BC) updates the policy on these examples without reward shaping or additional environment interaction.

We compare three protocols. *PPO-only* continues training directly from the saturated checkpoint for another 100,000 environment steps. *BC-only* evaluates the policy immediately after feedback. *BC+PPO* first applies BC and then resumes PPO for the same 100,000-step budget using only the original sparse reward. PPO-only and BC+PPO use the same learning rate, evaluation schedule, continuation budget, and freshly initialized optimizer; they differ only in whether targeted feedback is received before continuation.

Policies are evaluated every 10,000 continuation steps over 100 deterministic episodes from a common held-out evaluation stream. We measure task success and the two principal subgoals:

$$Q_{\text{succ}} = \Pr(\text{reach goal}), \qquad Q_{\text{key}} = \Pr(\text{pick up key}), \qquad Q_{\text{door}} = \Pr(\text{open door}).$$

Values are reported as mean ± SEM over the five paired seeds. To characterize post-feedback relaxation, we additionally report the mean success over the final three evaluations (Tail-3) and the step-normalized area under the continuation curve (AUC).

Results. Additional optimization alone does not escape the saturated regime. PPO-only continuation remains at zero success for every seed despite receiving another 100,000 environment steps. Weak feedback is likewise insufficient: with $m = 1000$, both immediate and post-continuation success remain zero. Thus, reopening the optimizer cannot compensate for feedback that fails to establish a useful behavioral hypothesis.

At $m = 10000$, BC immediately raises success to 0.160 ± 0.036. After resumed PPO, final success reaches 0.212 ± 0.038, with Tail-3 0.212 ± 0.024 and AUC 0.205 ± 0.016. Final key-pickup and door-open rates are 0.282 ± 0.034 and 0.212 ± 0.038, respectively. Intermediate feedback therefore establishes a nonzero-success region that remains accessible during subsequent sparse-reward learning, although the seed-level change from immediate to final performance is variable.

Table 2: Closed-loop escape and relaxation in `MiniGrid-DoorKey-8x8-v0`. Immediate performance is measured directly after BC and before resumed PPO. PPO-only begins from the unmodified saturated checkpoint. Tail-3 averages the final three evaluations, and AUC is normalized by the 100,000-step continuation horizon. Values are mean ± SEM over five paired seeds.

Protocol	Immediate	Final	Tail-3	AUC
PPO-only, $m = 0$	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
BC+PPO, $m = 1000$	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
BC+PPO, $m = 10000$	0.160 ± 0.036	0.212 ± 0.038	0.212 ± 0.024	0.205 ± 0.016
BC+PPO, $m = 20000$	0.700 ± 0.031	0.658 ± 0.022	0.669 ± 0.011	0.681 ± 0.016

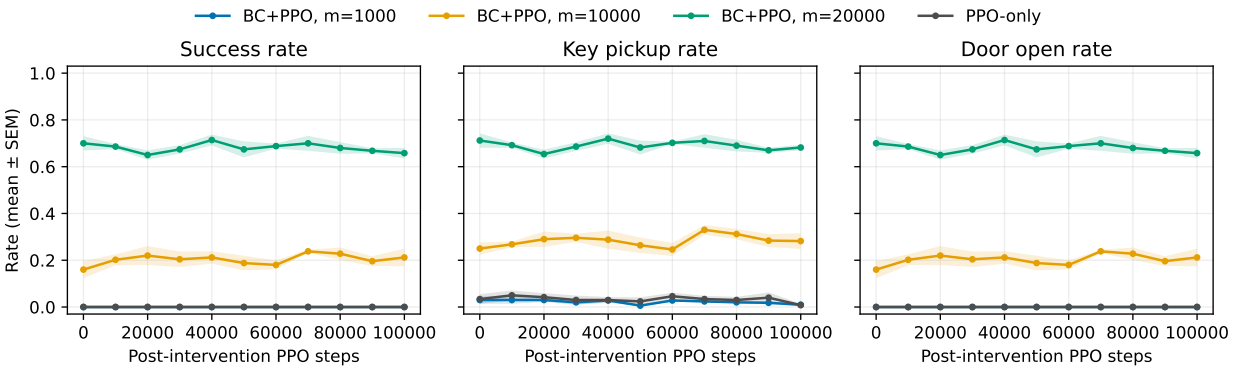

Figure 2: Post-feedback relaxation under the original sparse reward. The panels show task success, key pickup, and door opening during 100,000 PPO continuation steps. Curves denote mean ± SEM over five paired seeds. PPO-only continuation and weak feedback remain in the saturated failure regime. Intermediate feedback relocates the policy to a moderate-success region, whereas strong feedback produces a high-success region that persists throughout resumed PPO.

At $m = 20000$, immediate success reaches 0.700 ± 0.031. The policy remains in a high-success regime throughout continuation, with final success 0.658 ± 0.022, Tail-3 0.669 ± 0.011, and AUC 0.681 ± 0.016. Final key pickup and door opening reach 0.682 ± 0.012 and 0.658 ± 0.022, respectively. Door opening and task success are nearly identical, indicating that once the key–door sequence is completed, reaching the goal is rarely the remaining bottleneck.

The strongest-feedback branch does not improve monotonically under resumed PPO: its final success is slightly below its immediate BC performance. We therefore interpret the result as *persistent basin relocation*, rather than as evidence that post-feedback optimization must monotonically improve the policy. The matched-budget controls separate reopening the learning loop from the informational effect of feedback: additional PPO alone is insufficient, whereas sufficiently strong task-relevant feedback moves the policy into a successful behavioral region that subsequent sparse-reward learning can preserve.

Reproducibility. The experiment contains 40 seed–protocol units. Per-seed checkpoints, evaluation trajectories, subgoal rates, relaxation summaries, and run metadata are stored in SQLite and exported as machine-readable tables. All continuation branches use the same environment, sparse reward, evaluation stream, and training budget.

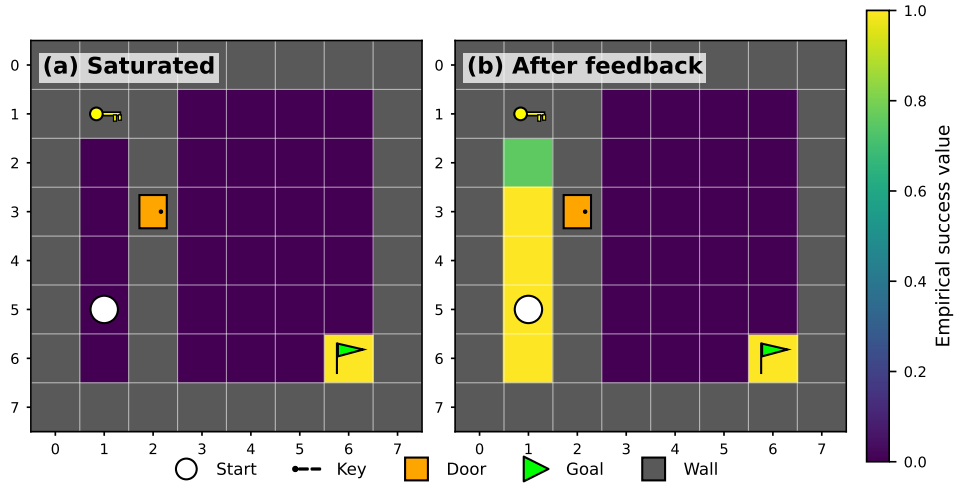

Figure 3: Representative empirical success landscape for seed 42 before and immediately after $m = 20000$ targeted feedback. At each traversable cell, empirical success is estimated from deterministic rollouts averaged over four initial orientations. This quantity is a rollout-based completion probability, not the learned PPO critic. Gray cells denote walls. Strong feedback expands the set of states from which the policy can complete the task sequence.

6.3 Information-Driven Escape in Bayesian Optimization

We use Bayesian optimization (BO) to separate the informational effect of feedback from the effect of reopening the search space. For each task and seed, Phase 1 restricts expected-improvement BO to a local trust region centered on a suboptimal minimum, producing a controlled locally saturated checkpoint. In Phase 2, the constraint is removed for *all* branches, including the no-feedback baseline, and each branch receives 12 real BO queries over the full domain.

We evaluate a controlled two-basin objective, Styblinski–Tang 2D, and Hartmann–6 using 30 paired evaluation seeds, disjoint from the five calibration seeds. We compare Baseline, Generic, Targeted, Noisy-Targeted, and Mismatched feedback, with information amount $m \in \{1, 2, 4, 8, 16\}$. Injected observations update the GP posterior but do not count as objective queries or realized improvements. They are excluded from query de-duplication, allowing BO to actively query and verify a feedback-indicated hypothesis.

An active escape occurs only when a real Phase 2 query enters the pre-specified global basin:

$$A_{s,c,m} = \mathbb{I}\left[\exists t : x^{\text{query}}_{s,c,m,t} \in \mathcal{B}_{\text{global}}\right]. \tag{8}$$

Let $R_{s,c,m}$ be the best post-intervention query regret. We define

$$\Delta Q_{s,c,m} = R_{s,\text{base},0} - R_{s,c,m}, \qquad E^{\text{eff}}_{s,c,m} = A_{s,c,m} \max(\Delta Q_{s,c,m}, 0). \tag{9}$$

Thus, posterior displacement without active verification and positive quality improvement does not count as effective escape.

Reopening the domain alone is generally insufficient: baseline escape is zero on Styblinski–Tang and Hartmann–6 and only 0.10 on the controlled task. A single targeted observation raises escape to 0.933 on the controlled task and 1.0 on Styblinski–Tang. Hartmann–6 exhibits a higher information requirement: targeted escape rises from 0.20 at $m = 1$ to 0.833 at $m = 16$, while generic feedback remains ineffective.

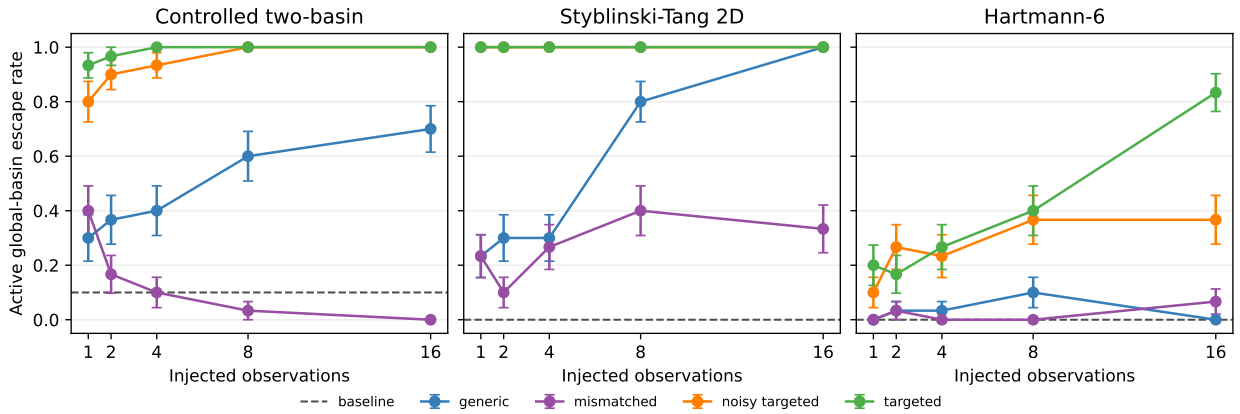

Figure 4: **Task-specific feedback-strength thresholds in Bayesian optimization.** Each panel reports the active global-basin escape rate as a function of the number of injected observations. Markers and whiskers denote mean ± SEM over 30 paired evaluation seeds; dashed lines indicate the task-specific no-feedback baseline after reopening the search domain. Targeted feedback induces active verification with fewer observations on the two-dimensional landscapes, whereas Hartmann–6 requires substantially more information. Results are shown separately by task to avoid pooling heterogeneous objective geometries.

Table 3: Active escape rates over 30 held-out seeds. T, G, and M denote targeted, generic, and mismatched feedback; subscripts indicate m.

Task	Baseline	T_1	T_{16}	G_1	G_{16}	M_{16}
Controlled two-basin	.100	.933	1.000	.300	.700	.000
Styblinski–Tang 2D	.000	1.000	1.000	.233	1.000	.333
Hartmann–6	.000	.200	.833	.000	.000	.067

Large amounts of generic information can nevertheless induce escape on simpler landscapes, reaching 0.70 on the controlled task and 1.0 on Styblinski–Tang at $m = 16$. Thus, targeted information is not strictly necessary. Instead, alignment shifts the information–escape frontier: task-aligned feedback requires less information to induce active verification and quality improvement, while the required information level remains task dependent.

7 Discussion

Scope and positioning. The framework gives sufficient conditions and operational diagnostics for saturation and escape, not a full classification of all possible closed-loop dynamics. It does not characterize oscillation, divergence, or adversarial feedback, and Section 5 analyzes a single external perturbation rather than a sequence of interacting perturbations. The information-theoretic bound also assumes a bounded representation space only when converting expected metric displacement into an escape-probability bound. The direct event-level KL threshold does not require bounded $\mathcal{X}$. Structural identifiability is relative to the declared probe distribution μ: no black-box experiment can distinguish parameter values that induce the same transition law on all probed states.

Extensions. The framework is instantiated in three distinct settings: LLM code repair, reinforcement learning, and Bayesian optimization. Future extensions include active learning, curriculum learning, and multi-agent feedback loops.

Empirical status of the structural and information quantities. The current experiments measure escape outcomes and feedback budgets, but do not yet estimate $\delta_{\mu,\nu}$ or the baseline-relative KL in Equation (6). Accordingly, they support the saturation–escape phenomenology but should not be read as direct validation of a governing-structure change or of the numerical information-divergence threshold. A direct test would pre-register probe states, resample transitions before and after intervention to estimate Equation (5), and estimate escape probabilities or density ratios relative to matched no-intervention continuations.

Conclusion. We introduced an operational framework for closed-loop knowledge dynamics based on representations, parameterized transition kernels, and a governing parameter identifiable up to observational equivalence. Classical Lyapunov drift conditions characterize fixed-parameter saturation, while pre-registered kernel discrepancies distinguish structural change from transient state movement. Escape is evaluated separately as a quality-aligned probability change relative to a matched no-intervention law, for which baseline-relative KL—rather than mutual information alone—provides the relevant information cost. The resulting contribution is a common measurement and intervention language for heterogeneous closed-loop systems, supported by cross-domain experiments and designed for direct empirical refinement.

References

Dimitri P. Bertsekas and John N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1996.

Vivek S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

Zhuo Chen, Xinzhe Yuan, Jianshu Zhang, Jinzong Dong, Ruichen Zhou, Yingchun Niu, Tianhang Zhou, Yu Yang Fredrik Liu, Yuqiang Li, Nanyang Ye, and Qinying Gu. Labo: Llm-accelerated bayesian optimization through broad exploration and selective experimentation, 2026.

Marquita Ellis and Paul Castro. Don't gamble, gamble: An analytical framework for ai-driven research systems, 2026.

Harold Kushner and G. George Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2nd edition, 2003.

Guangliang Liu, Haitao Mao, et al. On the convergence of moral self-correction in large language models. *arXiv preprint arXiv:2510.07290*, 2025.

Wei Liu, Siya Qi, Yali Du, and Yulan He. Self-play only evolves when self-synthetic pipeline ensures learnable information gain, 2026.

Aman Madaan, Niket Tandon, Prakhar Gupta, et al. Self-refine: Iterative refinement with self-feedback. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, 2nd edition, 2009.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, pages 400–407, 1951.

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*, volume 36, 2023.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning*, pages 1015–1022, 2010.

Xinzhe Yuan, Zhuo Chen, Jianshu Zhang, Huan Xiong, Nanyang Ye, Yuqiang Li, and Qinying Gu. Unleashing llms in bayesian optimization: Preference-guided framework for scientific discovery, 2026.

A Deferred Proofs

Proof of Proposition 1. Taking expectation in the assumed drift inequality gives $\mathbb{E}[d(x_{t+1}, x^\star)] \leq \rho\mathbb{E}[d(x_t, x^\star)] + \sigma$. Unrolling the recursion yields

$$\mathbb{E}[d(x_t, x^\star)] \leq \rho^t d(x_0, x^\star) + \sum_{s=0}^{t-1} \rho^s \sigma = \rho^t d(x_0, x^\star) + \frac{\sigma(1-\rho^t)}{1-\rho}.$$

For the step size, by the triangle inequality, $d(x_t, x_{t+1}) \leq d(x_t, x^\star) + d(x_{t+1}, x^\star)$. Taking expectations and applying the previous bound to both terms gives

$$\Delta_t \leq \mathbb{E}[d(x_t, x^\star)] + \mathbb{E}[d(x_{t+1}, x^\star)] \leq (1+\rho)\rho^t d(x_0, x^\star) + \frac{2\sigma}{1-\rho}.$$

Proof of Proposition 3. Since $x_{t_0} \in B(x^*_\theta, c)$, the triangle inequality gives $d(x_{t_0}, x^*_{\theta'}) \leq c + \Delta$. One step of the perturbed contraction then gives $\mathbb{E}[d(x_{t_0+1}, x^*_{\theta'})] \leq \rho'(c+\Delta) + \sigma_{\text{ext}}$. Applying the triangle inequality once more,

$$\mathbb{E}[d(x_{t_0+1}, x^*_\theta)] \geq \Delta - \mathbb{E}[d(x_{t_0+1}, x^*_{\theta'})] \geq \Delta - \rho'(c+\Delta) - \sigma_{\text{ext}}.$$

The right-hand side exceeds c exactly when $\Delta > \Delta^*$.

Remark 6 (Local linearization)**.** *When $\mathcal{X}$ is a vector space or smooth manifold and $\mathcal{T}_\omega$ is differentiable near a fixed point $x^\star$, local stability can be checked through the Jacobian*

$$J_\star = D_x \mathcal{T}_\omega(x^\star).$$

Writing $\delta x_{t+1} \approx J_\star \delta x_t$, the quadratic Lyapunov function $V(x) = \|x - x^\star\|^2$ gives local contraction whenever the spectral radius satisfies $\rho(J_\star) < 1$. The eigenvectors of $J_\star$ identify directions of fast and slow decay. This provides a differentiable analogue of the metric contraction condition used below for text-space self-refinement.

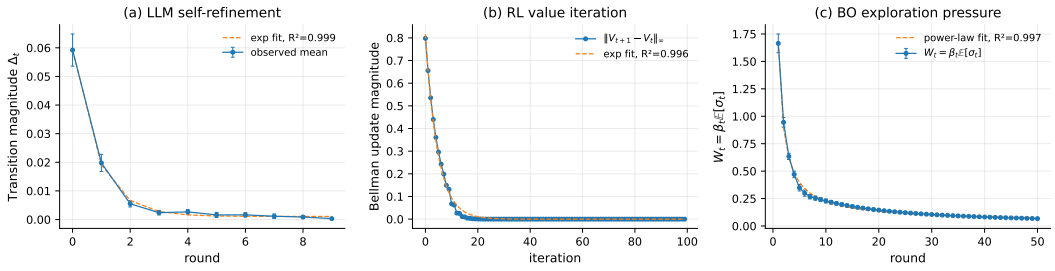

Figure 5: Saturation signatures across three closed-loop systems. (a) LLM self-refinement: transition magnitude decays toward a residual floor. (b) RL value iteration: Bellman update magnitude decays geometrically. (c) Bayesian optimization: exploration pressure $W_t = \beta_t \mathbb{E}[\sigma_t]$ decays as posterior uncertainty is resolved. Together, these examples show saturation across text refinement, value-function iteration, and belief-guided search.

A.1 Additional Validation A: Cross-System Saturation Signatures

We first test whether the saturation signature predicted by Theorem 1 appears across qualitatively different closed-loop systems. We consider three cases: recursive LLM self-refinement, tabular RL value iteration, and belief-guided Bayesian optimization. These systems differ in state space and feedback mechanism, but each repeatedly updates an internal knowledge representation under fixed governing dynamics. For LLM self-refinement, the state x_t is a draft and the transition magnitude is measured as $\Delta_t = d(x_t, x_{t+1})$. For RL, the state is the value function and we track the Bellman update magnitude $\|V_{t+1} - V_t\|_\infty$. For Bayesian optimization, the state is a Gaussian process posterior and we track the exploration pressure $W_t = \beta_t \mathbb{E}[\sigma_t]$. Figure 5 shows that all three systems exhibit a saturation pattern. LLM self-refinement follows an exponential-plus-floor curve, RL value iteration shows rapid geometric decay, and Bayesian optimization exhibits power-law attenuation of exploration pressure. These results support the claim that saturation is a general closed-loop phenomenon rather than an artifact of a single model or task.

A.2 Additional Validation B: Escape in Recursive Abstract Refinement

We next test whether external feedback can move a saturated self-refinement trajectory beyond its internal continuation baseline. After recursive self-refinement reaches a low-change regime at round t_0, we compare four conditions: baseline continuation, generic feedback, targeted feedback, and mismatched feedback. We define escape magnitude as

$$E_{\text{shift}} = d\left(x_{t_0+1}^{\text{condition}}, x_{t_0+1}^{\text{baseline}}\right),$$

where $x_{t_0+1}^{\text{baseline}}$ is the output produced by continuing internal self-refinement without external feedback. To distinguish mere displacement from useful escape, we also compute

$$E_{\text{useful}} = E_{\text{shift}} \max(\Delta Q, 0),$$

where ΔQ is the quality gain relative to the baseline continuation.

Figure 6 shows that targeted feedback produces the largest escape magnitude, approximately 0.54, compared with approximately 0.26 for generic feedback and 0.43 for mismatched feedback. The mismatched condition demonstrates that irrelevant feedback can still induce a large state displacement. However, useful escape is concentrated almost entirely in the targeted-feedback

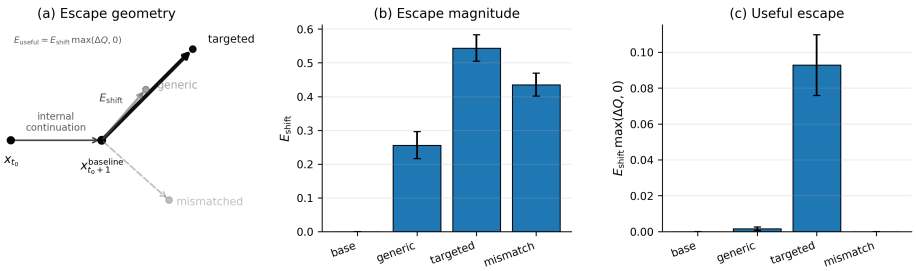

Figure 6: External feedback induces condition-dependent escape from saturated self-refinement trajectories. (a) Escape is measured relative to the internal continuation baseline $x_{t_0+1}^{\text{baseline}}$. Generic and targeted feedback move the trajectory in the same useful direction, while mismatched feedback moves it in a different direction. (b) Targeted feedback produces the largest escape magnitude $E_{\text{shift}} = d(x_{t_0+1}^{\text{condition}}, x_{t_0+1}^{\text{baseline}})$. (c) Useful escape, measured as $E_{\text{shift}} \max(\Delta Q, 0)$, is concentrated in the targeted-feedback condition, showing that successful escape requires feedback that is both strong enough to move the trajectory and relevant enough to improve it.

condition, with mean useful escape approximately 0.09, whereas generic and mismatched feedback produce near-zero useful escape. These results support the prediction that successful escape requires external information that is both strong enough to perturb the saturated trajectory and relevant enough to improve it.

A.3 Additional Validation C: Post-Perturbation Relaxation

We finally test whether successful external perturbation induces a renewed relaxation phase after saturation. Under fixed governing dynamics, recursive self-refinement should approach a low-change regime. If targeted external feedback shifts the effective transition regime, then the transition magnitude should exhibit a spike at the perturbation round followed by a second decay phase.

For each item, we run recursive self-refinement until round t_0, where the transition magnitude has reached a small residual floor. We then compare two continuations: a baseline trajectory that continues internal self-refinement without external feedback, and a targeted-perturbation trajectory that receives external feedback at round t_0 before continuing internal refinement. We track

$$\Delta_t = d(x_t, x_{t-1})$$

over the full trajectory. To quantify relaxation before and after the perturbation, we fit exponential-plus-floor curves:

$$\Delta_t^{\text{pre}} \approx A_1 e^{-k_1 t} + c_1, \qquad \Delta_t^{\text{post}} \approx A_2 e^{-k_2(t-t_0-1)} + c_2.$$

Table 4 shows the predicted double-saturation pattern. The pre-perturbation trajectory decays toward a small floor, indicating saturation under internal feedback. The targeted perturbation produces a spike in transition magnitude above the saturated floor, while the baseline continuation remains near the floor. After the perturbation, the transition magnitude decays again, producing a second relaxation phase. This provides an observable candidate signature for renewed relaxation: the external input moves the system away from its internal continuation, and the subsequent internal loop relaxes toward a new locally stable regime. Under Equation (5), however, a spike and second decay alone do not establish a governing-structure shift; that claim requires a reproducible post-intervention kernel discrepancy on pre-specified probes.

Together with Experiment 2, this result distinguishes useful escape from mere perturbation. Experiment 2 shows that targeted feedback produces the largest useful escape magnitude, while

Condition	Spike magnitude ↑	Spike rate ↑	Post-fit R^2	Final Δ_T ↓
Baseline continuation	0.004 ± 0.002	0.07	–	0.006 ± 0.001
Generic feedback	0.026 ± 0.006	0.38	0.841	0.010 ± 0.003
Targeted perturbation	0.071 ± 0.009	0.93	0.962	0.008 ± 0.002
Mismatched feedback	0.049 ± 0.008	0.71	0.624	0.021 ± 0.005

Table 4: Summary of double-saturation signatures after external perturbation. Targeted perturbation produces the largest spike rate and the clearest post-perturbation relaxation fit, while mismatched feedback induces displacement without a comparably stable second relaxation phase.

Experiment 3 shows that such perturbations can initiate a renewed relaxation phase rather than simply increasing noise.