ENVAUDIT: AUDITING THE TRAINING UTILITY OF SYNTHETIC AGENT ENVIRONMENTS UNDER A FIXED BUDGET

Anonymous authors
Paper under double-blind review

ABSTRACT

Can the target utility of synthetic training be bounded before running the training itself? We give an explicit end-to-end certificate for a finite learner and a class of state-dependent target tasks. The audited quantity is the improvement of the model returned under a resource budget, including base-model fallback on noncompletion. A static enclosure of the future policy combines with local target-transition constraints to give closed-form gain bounds, with attainable endpoints in a fixed-parameter subfamily. Preserving the same unknown environment across policies then yields completion-probability thresholds for choosing between training sources. For two fixed synthetic sources and a prescribed finite learner, the certificates establish opposite rankings at two budgets for every task horizon. At horizon 16, both rankings are determined even though independently bounded utilities overlap. The same task class also exposes a genuine information limit: two legal three-step targets satisfy the same supplied constraints yet give opposite rankings at one fixed budget for the same finite training endpoints. Together, these results specify when structural information certifies a useful decision and demonstrate why it cannot always do so. No future optimizer iterates or target samples are used. The guarantees concern the declared learner, progress-based success rule, and cost contract; they do not establish general neural fine-tuning prediction or empirical savings.

1 INTRODUCTION

Synthetic environments are useful only insofar as training on them improves the model returned for a target task. This question arises in instruction and rationale generation (Wang et al., 2022; Zelikman et al., 2022), executable software and command-line tasks (Yang et al., 2025; Lin et al., 2026), and interactive agent learning (Xi et al., 2024; Qi et al., 2024). Systems that generate executable worlds or jointly adapt environments and agents further expand the set of candidate training sources (Wang et al., 2026; Liu et al., 2026). Deciding which source to use can itself require costly training. We study what can be learned about that decision *before the prescribed training is executed*.

The relevant utility includes the entire resource allocation. Target benchmarks evaluate outcomes such as repository issue resolution, execution feedback, and functional web-task completion (Jimenez et al., 2023; Yang et al., 2023; Zhou et al., 2023). A pipeline may exhaust its budget during source preparation, collection, or updating before it returns an improved checkpoint. Cost-sensitive agent evaluation and uncertainty-aware reinforcement-learning evaluation motivate retaining these unsuccessful attempts (Kapoor et al., 2024; Agarwal et al., 2021). Under a base-model fallback rule, a deterministic completed endpoint with target gain g and completion probability $q(B)$ has complete utility $V(B) = q(B)g$. The better endpoint can therefore be the worse budgeted choice.

The question and the obstacle. *Can static information certify a range for $V(B)$, or a source ranking, without executing the future training updates?* A precise point prediction is unnecessary. A useful answer must nevertheless bridge two distinct unknowns: what policy the finite learner will return, and what that policy change will do in the target environment. Source optimization alone does not supply target success information. Moreover, comparing two policies through indepen-

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.

dent target-value intervals can discard their shared environmental uncertainty and leave a resolvable decision unanswered.

Main contribution. We give an explicit end-to-end certificate for a finite learner and a family of state-dependent progress tasks. Its central feature is a bound on the *change caused by training in one shared target world*, composed with a bound on the unexecuted training endpoint and the complete return rule. The resulting interval does not require either absolute target value to be identified. Three linked steps establish this one result:

1. A static contraction argument encloses the action-probability change produced by a prescribed finite scalar learner (Lemma 2).
2. Local transition constraints convert that change into closed-form target-gain bounds. Explicit kernels attain both bounds in a fixed-parameter subfamily (Theorem 1).
3. The shared-world comparison gives explicit completion-probability thresholds for a budget decision (Corollary 3). The same two finite source programs reverse their certified utility ordering as the budget increases, uniformly over every task horizon (Theorem 4).

These steps use the same learner interface and target class. The local proof permits later states to be harder and does not enumerate complete trajectories. At horizon 16, both budget rankings are certified even though independent utility intervals overlap. Within the same task class, two targets sharing all supplied information can nevertheless give opposite rankings at a fixed budget for the same finite learner (Proposition 5). This identifies an information boundary of the decision problem, without changing the learner or success metric.

What makes the result informative. The positive action advantage and policy shift supply the improvement direction; the result quantifies its magnitude, gives attainable gain endpoints, and shows how incomplete execution changes the source decision. Target transitions can remain unknown within the stated constraints. This is a constructive sufficient-information result, not a claim that arbitrary tasks admit these constraints. Established coupling and relational-verification principles supply the proof semantics (Avanzini et al., 2024); our contribution is their explicit source-to-target-to-budget instantiation. Section 6 separates this result from claims about general neural training, stronger verification baselines, or actual economic savings.

2 Related work

Environment design and synthetic training. PAIRED formulates environment design through regret; Prioritized Level Replay selects levels by estimated learning potential; and replay-guided adversarial design connects level curation to environment design (Dennis et al., 2020; Jiang et al., 2020; 2021). Self-Instruct, STaR, AgentGym, and WebRL construct training material or interactive learning procedures (Wang et al., 2022; Zelikman et al., 2022; Xi et al., 2024; Qi et al., 2024). SWE-smith, CLI-Gym, Agent World Model, and SPADE address task or environment supply (Yang et al., 2025; Lin et al., 2026; Wang et al., 2026; Liu et al., 2026). These works motivate the object of our audit. We fix the source procedure and certify its consequence under an explicit target contract. The scalar learner makes that connection provable; it is not a model of parameter-efficient neural adaptation such as LoRA (Hu et al., 2021).

Policy sensitivity and incomplete target evidence. Policy expansions and their off-policy-evaluation interpretation describe target return along policy curves (Tang et al., 2020; 2021). Their coefficients need target information, which source gradients do not generally determine. Partial-identification approaches bound what incomplete evidence identifies; Khan et al. (2023), for example, study smoothness beyond overlap. Finite-dimensional approximation duality relates incomplete linear information to indistinguishable bounded functions (Dachman-Soled et al., 2014). Our compatible-world formulation uses this established viewpoint. The additional object here is an unexecuted finite source program whose policy enclosure is propagated to a complete-budget decision. Target success criteria and evaluation costs remain separate inputs (Jimenez et al., 2023; Zhou et al., 2023; Kapoor et al., 2024).

Relational verification: the closest technical foundation. The eRHL logic reasons about paired probabilistic programs and provides soundness and completeness results for relevant terminating cases (Avanzini et al., 2024). Our gain bound is expressible in that logic using a nonnegative separable post-expectation; Appendix A states the reduction. Thus the generic paired proof rule is established. The concrete content here is the explicit potential, attaining target kernels, and composition with finite-source and complete-budget bounds. Logical completeness does not by itself construct these quantities or imply efficient general synthesis. A strong parameter-preserving verifier can exploit the same structure (Section 6).

Success structure. Reliability theory provides representations of monotone success systems through system signatures (Marichal et al., 2011; Marichal, 2015). Our quantitative result instead fixes the terminal success mean to normalized progress. This specific success rule is an input to the proof, and the formula does not extend unchanged to every monotone or terminal-goal criterion.

3 Complete utility and admissible static information

Fix a base policy π_0, a target return $J_\omega(\pi) \in [0,1]$, and an end-to-end budget B. A world ω fixes the target transition and reward mechanisms, the initial distribution, the source materials, and all parameters of the prescribed training and return contract. Method a attempts the entire pipeline and returns π_0 on noncompletion. Its utility is

$$V_a(B) = \mathbb{E}\big[J_\omega(\pi_a^{\mathrm{ret}}(B)) - J_\omega(\pi_0)\big] = \mathbb{E}[\mathbf{1}_{E_a(B)} g_a], \tag{1}$$

where g_a is the completed endpoint's gain on $E_a(B)$, extended by zero off that event. If $q_a = \mathbb{P}(E_a) > 0$, then $V_a = q_a \mathbb{E}[g_a \mid E_a]$. This identity does not assume independence. A missing evaluation after returning a model is missing evidence, not noncompletion.

Our constructive results use fixed source materials and a deterministic finite optimizer. Every completion returns the same endpoint $\pi_{a,K}$, and the cost event does not change that endpoint or the target world. Consequently,

$$g_a(\omega) = J_\omega(\pi_{a,K}) - J_\omega(\pi_0), \qquad V_a(\omega, B) = q_a(B) g_a(\omega). \tag{2}$$

Completion probabilities are uniform over the compatible target worlds in the example. If completion selects different endpoints, one must use (1) or retain the joint endpoint–completion distribution.

The information contract. Before certification, fix the program, policy interface, target state and success rules, shared parameter constraints, budget accounting, and fallback. A source enclosure must follow from the prescribed finite program. Target rules may be supplied explicitly or proved from a sound model of the actual task. Their verification, source inspection, interval computation, training stages, and return costs all enter the budget. Table 1 separates supplied information from quantities derived by the certificate. The target contract is not inferred from a successful synthetic training loss.

Let $\mathcal{W}(e)$ be the nonempty set of worlds compatible with the permitted information e. A certificate $[L_a, U_a]$ must contain $V_a(\omega, B)$ for every $\omega \in \mathcal{W}(e)$. Comparing A and B requires bounds on $V_A(\omega, B) - V_B(\omega, B)$ in the *same* world. Optimizing the two values separately and subtracting their intervals can discard useful dependence. Conversely, an interval crossing zero proves ambiguity only if compatible opposite-sign worlds actually exist. An empty compatibility set requires an inconsistency report.

4 A locally checkable progress certificate

For integer $H \geq 1$, the target state is progress $i \in \{0, \dots, H\}$. It starts at zero and runs for exactly H steps. At $i < H$, action 0 advances one state with probability λ_i, and action 1 advances with probability $\lambda_i + \eta$; otherwise the state stays unchanged. The parameters are fixed and shared by all policies, with

$$\lambda_i \geq 0, \quad \lambda_i + \eta \leq 1, \quad 0 < \eta_l \leq \eta \leq \eta_u. \tag{3}$$

Terminal success has conditional mean $\rho(i) = i/H$. Equivalently, draw an independent uniform threshold in $\{1, \dots, H\}$ and declare success when final progress reaches it; this threshold is not

Table 1: The certificate's information boundary. No trained target value or target trajectory is an input. Each supplied claim must describe the actual prescribed program or target.

Stage	Supplied or independently justified	Derived by this construction
Source	Fixed labels, finite update rule, policy interface	Enclosure of the future policy shift
Target	Legal shared kernel family, action advantage, local variation, success rule	Uniform target-gain interval
Budget	All-in cost and completion law, base fallback	Returned-model utility and source comparison

observed by the policies and does not influence transitions. The theorem does not replace an arbitrary task's success criterion with progress.

The base and candidate policies are

$$\beta_i = \pi_0(1 \mid i) = \frac{1}{4} + \frac{i}{4H}, \qquad \pi_d(1 \mid i) = \beta_i + d, \qquad 0 < d_l \le d \le d_u \le \frac{1}{4}. \tag{4}$$

They respond to visited states. Only the *difference* in their action probabilities is constant. Write $b_i = \lambda_i + \eta\beta_i$ for the base forward rate. The additional local contract is

$$|b_{i+1} - b_i| \le \kappa, \qquad 0 \le \kappa \le 1, \qquad i = 0, \dots, H-2. \tag{5}$$

It allows decreasing b_i, so subsequent states can be harder. Separate bounds on λ_i do not automatically imply a useful small κ.

Define $S_n(a) = \sum_{j=0}^{n-1} a^j$, with $S_0(a) = 0$, and

$$G_-(H,\kappa) = \frac{S_H(1-\kappa)}{H}, \qquad G_+(H,\kappa) = \frac{S_H(1+\kappa)}{H}. \tag{6}$$

Theorem 1 (Progress gain and attaining kernels) *Under* (3)–(5)*, the gain* $g = J_\omega(\pi_d) - J_\omega(\pi_0)$ *satisfies*

$$\eta_l d_l G_- \;\le\; g \;\le\; \min\{1, \eta_u d_u G_+\}. \tag{7}$$

For every $H \ge 1$*,* $\eta_l = \eta_u = \eta = 1/4$*, fixed* $d \in (0, 1/4]$*, and* $\kappa = 1/(4H)$*, there are legal shared target kernels attaining the lower and upper bounds with* $d_l = d_u = d$*. The fixed-parameter bounds are therefore sharp on this subfamily.*

Proof and explicit certificate. Use a fresh uniform random variable at every step to couple the two forward events, with candidate progress X_t and base progress Y_t. If the states agree, the candidate has the larger forward probability. If $X_t - Y_t \ge 1$, unit increments cannot reverse their order. Hence $D_t = X_t - Y_t \ge 0$ and

$$\mathbb{E}[D_{t+1} \mid X_t, Y_t] = D_t + b_{X_t} - b_{Y_t} + \eta d. \tag{8}$$

Telescoping (5) yields $|b_i - b_j| \le \kappa|i-j|$. Thus the conditional drift lies between $(1-\kappa)D_t + \eta_l d_l$ and $(1+\kappa)D_t + \eta_u d_u$. Iteration from $D_0 = 0$, followed by division by H, proves (7).

For a locally checkable proof, put $n = H - t$, $m_l = \eta_l d_l$, and $m_u = \eta_u d_u$. The paired potentials are

$$L_t(i,j) = \frac{(1-\kappa)^n(i-j) + m_l S_n(1-\kappa)}{H}, \qquad U_t(i,j) = \frac{(1+\kappa)^n(i-j) + m_u S_n(1+\kappa)}{H}. \tag{9}$$

They satisfy $L_t \le \mathbb{E}L_{t+1}$ and $U_t \ge \mathbb{E}U_{t+1}$ and agree with the reward difference at $t = H$. The domain is $0 \le j \le i \le t < H$; thus no positive drift is incorrectly required at an already absorbed (H, H). For attainment, choose $b_i^\pm = 3/8 \pm i/(4H)$ and $\lambda_i^\pm = b_i^\pm - \beta_i/4$. These are legal kernels and make the respective drift bound an equality. Appendix B supplies the remaining details and boundary cases.

What information does the work? Uniform action advantage, common positive policy shift, and unit-step order preservation already provide the improvement direction. The local variation bound controls its magnitude. No future advantage function or trained target value is an input. For $\kappa = 1/(4H)$, the gain bounds remain of constant order as H grows. This is not true uniformly at fixed positive κ: the lower bound can decay as $1/H$.

Verification cost. An explicit $H+1$-state contract has $O(H)$ legality and adjacent-variation conditions. The symbolic potentials avoid constructing an $O(H^3)$ time-indexed pair-state table. The geometric sums need $O(\log H)$ arithmetic operations using doubling, but exact rational bit lengths grow with H. If parameters form an explicitly given rational affine polytope, robust legality and variation checks can be reduced to linear optimization, whose encoding and solving costs must be charged. This is not an efficient general algorithm for extracting faithful contracts from arbitrary programs.

5 From unexecuted training to budgeted decisions

5.1 A static enclosure for the prescribed finite source

For $u \in [1/2, 3/4]$, let the source consist of fixed binary labels with positive fraction u. Prescribe the scalar loss and finite update rule

$$F_u(\theta) = \log(1 + e^\theta) - u\theta, \qquad \theta_0 = 0, \qquad \theta_{k+1} = \theta_k + u - \sigma(\theta_k), \tag{10}$$

where σ is the sigmoid. The target interface, fixed before training, is

$$\pi_\theta(1 \mid i) = \beta_i + \sigma(\theta) - \tfrac{1}{2}. \tag{11}$$

This interface is part of the prescribed program, not a property deduced from arbitrary neural-network optimization.

Lemma 2 (Finite source enclosure) *For every $K \geq 0$, the updates stay in $[0, \operatorname{logit}(u)]$ and*

$$0 \leq u - \sigma(\theta_K) \leq \delta_K, \qquad \delta_K = \tfrac{1}{2}(13/16)^K. \tag{12}$$

Thus the actual endpoint shift $d_K = \sigma(\theta_K) - 1/2$ lies in $[u - 1/2 - \delta_K, u - 1/2]$. This enclosure is calculated without generating the sequence of future updates.

The proof uses the invariant interval, the update map's $13/16$ contraction, and the sigmoid's $1/4$ Lipschitz constant (Appendix C). A negative lower endpoint is merely an uninformative enclosure; Theorem 1 requires a positive certified lower endpoint. The construction is a finite scalar learner, not an asserted surrogate for LoRA or another deployed fine-tuning procedure.

The end-to-end certificate. The source and target results now compose directly. For $1/2 < u \leq 3/4$, a prescribed K with $\delta_K < u - 1/2$, and a valid completion law $q(B)$, every compatible world satisfies

$$q(B)\eta_l\big(u - \tfrac{1}{2} - \delta_K\big)G_- \;\leq\; V(B) \;\leq\; q(B)\min\big\{1, \eta_u\big(u - \tfrac{1}{2}\big)G_+\big\}\,. \tag{13}$$

All quantities on the two sides are available from the declared inputs without computing $\theta_1, \ldots, \theta_K$. The formula is uniform over the target family; it requires no distribution over its unknown parameters. The budget in $q(B)$ includes the audit itself. This is the single-method output; the next step uses shared structure to make a tighter source comparison.

5.2 Comparing sources in the same unknown environment

Let A and B satisfy (4) and share the target rules and base policy, with fixed known $\eta > 0$ and endpoint enclosures $d_A \in [a_l, a_u]$, $d_B \in [b_l, b_u]$, where $0 < b_l \leq b_u < a_l \leq a_u \leq 1/4$. With B as the baseline, its forward rate is $b_i + \eta d_B$, so the adjacent differences and κ are unchanged. The same coupling therefore gives

$$\eta(d_A - d_B)G_- \leq J_\omega(\pi_{d_A}) - J_\omega(\pi_{d_B}) \leq \eta(d_A - d_B)G_+. \tag{14}$$

For fixed $0 \le q_A \le q_B \le 1$, write $\tau = V_A - V_B$. Using $\tau = q_A(g_A - g_B) - (q_B - q_A)g_B$ yields the valid enclosure

$$\begin{aligned} L_\tau &= \eta\{q_A(a_l - b_u)G_- - (q_B - q_A)b_u G_+\}, \\ U_\tau &= \eta\{q_A(a_u - b_l)G_+ - (q_B - q_A)b_l G_-\}. \end{aligned} \tag{15}$$

The two component extrema need not occur together, so this comparison enclosure is not claimed sharp. It still retains a useful same-world difference that separate single-method bounds discard.

Corollary 3 (Certified completion-probability thresholds) *Under the target and ordered-source assumptions above, define*

$$r_{\mathrm{B}} = \frac{b_l G_-}{b_l G_- + (a_u - b_l)G_+}, \qquad r_{\mathrm{A}} = \frac{b_u G_+}{b_u G_+ + (a_l - b_u)G_-}. \tag{16}$$

Then $0 < r_{\mathrm{B}} \le r_{\mathrm{A}} < 1$ *and, in every compatible world,*

$$r_{\mathrm{B}} \le r_*(\omega) := \frac{g_B(\omega)}{g_A(\omega)} \le r_{\mathrm{A}}. \tag{17}$$

For $q_B(B) > 0$, put $r = q_A(B)/q_B(B)$. If $r < r_{\mathrm{B}}$, B has strictly greater complete utility; if $r > r_{\mathrm{A}}$, A does. These implications hold even when $r > 1$. Inside $[r_{\mathrm{B}}, r_{\mathrm{A}}]$, these thresholds alone need not determine a strict ordering.

Proof and interpretation. Write $h = g_A - g_B > 0$. The ratio $r_* = g_B/(g_B + h)$ increases with g_B and decreases with h. Applying the single-method and paired gain bounds proves (17). Since $V_A - V_B = q_B g_A(r - r_*)$, its sign gives the stated decisions. The common positive factor η cancels. The thresholds enclose an unknown break-even ratio; the gap is an unresolved region of this certificate, not proof that compatible worlds have opposite rankings. Proposition 5 establishes such ambiguity for one explicit record in the same task class. Appendix D gives the endpoint and limiting cases.

Certificate procedure and decision rule. Verify the finite-source assumptions and ordered positive enclosures from (12); check robust transition legality and (5) while preserving parameter equalities; compute $G_\pm$ and the gain/ratio bounds; then insert the all-in completion probabilities. Equation (13) gives complete utilities, and (15) gives their difference when $q_A \le q_B$. Report numerical bounds with outward rounding. Invalid models, unverified inputs, and numerical failures are reported explicitly, rather than assigned zero utility.

For a prespecified margin $\gamma \ge 0$, $L_\tau > \gamma$ certifies A and $U_\tau < -\gamma$ certifies B. Execution additionally requires a positive single-method lower bound; otherwise abstain. These bounds hold uniformly over one compatible world set, so combining them adds no statistical error probability. For ordering without a positive margin, $q_B = 0 < q_A$ favors A, two zero probabilities return the base, and $q_A > q_B > 0$ favors A by Corollary 3. These cases do not use (15) outside its domain. Stronger information use may resolve an abstention.

5.3 A UNIFORM BUDGET REVERSAL FROM THE SAME CERTIFICATE

Theorem 4 (Budget reversal for the same finite learner and target class) *For every $H \ge 1$, use* (3)–(5) *with $\eta = 1/4$ and $\kappa = 1/(4H)$. Give A and B 100 fixed labels with respectively 75 and 70 positive labels, and prescribe $K = 100$ updates* (10) *and the shared interface* (11)*. Let*

$$C_H = C_{\mathrm{info}}(H) + C_{\mathrm{materials}} + C_{100} + C_{\mathrm{return}}, \tag{18}$$

where $C_{\mathrm{info}}(H)$ covers static checks and the two-method comparison. A requires one extra cost unit with probability $9/25$; B requires none. This event changes neither source endpoint nor target kernel, and noncompletion returns the base within the charged allowance. At separately allocated budgets $B_0 = C_H$ and $B_1 = C_H + 1$, both complete utilities are positive and

$$V_A(B_0) < V_B(B_0), \qquad V_A(B_1) > V_B(B_1), \qquad g_A > g_B > 0. \tag{19}$$

The claims hold for every compatible target kernel without executing either training program or sampling the target.

Table 2: Complete-utility difference $\tau = V_A - V_B$ at $H = 16$. All intervals include the static finite-source error and are rounded outward. Both individual utilities are positive.

Budget	Same-world certificate	Independent interval subtraction
C_H	$[-0.013144, -0.007027]$	$[-0.020672, 0.000501]$
$C_H + 1$	$[0.011136, 0.014078]$	$[-0.000626, 0.025840]$

Proof sketch. Put $\delta = \delta_{100}$, $[a_l, a_u] = [1/4 - \delta, 1/4]$, and $[b_l, b_u] = [1/5 - \delta, 1/5]$. At B_0, $(q_A, q_B) = (16/25, 1)$; at B_1, both equal one. Uniformly in H, $7/8 \le G_- \le 1$ and $1 \le G_+ \le 4/3$. Substitution into (15) gives

$$U_\tau(B_0) \le \tfrac{1}{4}\{-61/3000 + (91/75)\delta\} < 0, \qquad L_\tau(B_1) \ge \tfrac{7}{32}(1/20 - \delta) > 0. \tag{20}$$

Single-method positivity and completed-endpoint ordering follow from (7) and (14). Appendix D proves the uniform bounds and strict signs.

For $H = 16$, Table 2 reports outward-rounded enclosures including the finite-source error. The same-parameter comparison certifies opposite rankings at the two budgets, whereas independent subtraction does not. The same inputs give

$$r_{\mathrm{B}} \in [0.759872, 0.759873], \qquad r_{\mathrm{A}} \in [0.834878, 0.834879]. \tag{21}$$

Thus $q_A/q_B < 0.759872$ suffices to choose B, whereas $q_A/q_B > 0.834879$ suffices to choose A. The constructed budgets give ratios 0.64 and 1. These are sufficient conditions, not exact decision boundaries. These numbers are probability differences or dimensionless completion ratios in a declared mathematical model, not measured transfer performance. The cost unit is formal; C_H may grow with H, so the theorem does not fit arbitrarily long tasks into one unchanged numerical budget.

6 What a certificate can and cannot establish

When unresolved means genuinely missing information. An interval crossing zero alone proves little: Table 2 is resolved by preserving shared structure. The following complementary result shows that the same progress-task contract can also leave a ranking intrinsically undetermined.

Proposition 5 (Opposite rankings under the same static information) *Use the two fixed source programs and interface of Theorem 4. Supply only the progress rules* (3)–(5) *with $H = 3$, $\eta = 1/4$, $\kappa = 1/12$, and success mean $i/3$ as target information. At one fixed all-in budget, let $q_A = 4/5$ and $q_B = 1$, with the same endpoint-independent cost law in every target world. Two legal base-rate vectors are*

$$b^{\mathrm{v}} = (3/8, 7/24, 3/8), \qquad b^{\mathrm{p}} = (3/8, 11/24, 3/8). \tag{22}$$

For the same actual finite source endpoints, writing $\delta = \delta_{100}$, their complete-utility differences satisfy

$$\tau_{\mathrm{v}} \ge 1/28800 - \delta > 0, \qquad \tau_{\mathrm{p}} \le -1/28800 + \delta < 0. \tag{23}$$

Both sources improve on the base, and A has the larger completed gain in both worlds. No sound strict source ranking follows from this static record alone.

Proof sketch and scope. Writing $m = d/4$, exact three-step propagation gives gains $g_{\mathrm{v}}(d) = m(413/144 + m/6)/3$ and $g_{\mathrm{p}}(d) = m(449/144 - m/6)/3$. At the reference shifts $(1/4, 1/5)$, the linear terms in $(4/5)g_A - g_B$ cancel, giving $\pm 1/28800$. Each gain has derivative in $(0, 1/3)$ on $[0, 1/4]$; replacing reference shifts by the actual finite endpoints changes either difference by less than δ. Appendix E proves legality, the cost construction, and the finite-source bound. The reference shifts are used only for bounding this error, not substituted for the actual training output.

This ambiguity is caused by unobserved target structure, not by the source enclosure: the signs hold for every common pair of endpoints in that enclosure. It demonstrates an information limit at one completion ratio, rather than characterizing the entire unresolved threshold region. By contrast, Appendix E's separate routing example shows how discarding a *supplied* parameter equality can introduce avoidable ambiguity. These two failure modes require different remedies: additional target information, or a stronger use of information already supplied.

The contract's direction and metric matter. For $H = 2$, $\eta = 1/4$, $\lambda_0 = 1/2$, and $\lambda_1 = 1/8$, the gain is $(53/64)\eta d < \eta d$. Setting $\kappa = 0$ without justification would overstate its magnitude. The correct $\kappa = 11/32$ makes our lower bound exact. Gain remains positive, so this example does not show that missing variation information prevents all sign decisions. Likewise, unit progress, a fixed number of steps, and $\rho(i) = i/H$ are not interchangeable with arbitrary failure absorption or terminal-goal success. A different metric needs a different terminal certificate.

A strong comparator can exploit the same structure. On the affine subclass $b_i = c + i/(4H)$ with $c \in [1/4, 1/2]$ and fixed $m = \eta d$, direct parameterized evaluation gives $J_0 = cG_+$, $J_d = (c + m)G_+$, and $g = mG_+$. This strong baseline obtains the same useful gain without identifying c. Our certificate is not claimed superior to such shared-parameter verification. Directly deploying a reference policy is also a legitimate economic comparator when that policy is available, as it is in the scalar example. Direct training must receive any budget released by omitting the audit. None of these comparisons is settled by counting the optimizer steps that certification avoids.

Reproducible analytic checks. We check exact finite-state distributions, coupling marginals and local inequalities, the attaining affine kernels, finite-source allowances, complete-utility intervals, break-even ratios, budget decisions, and opposite-ranking witnesses with rational arithmetic and exact polynomial propagation. Forward propagation enumerates states, not complete trajectories, and is used only to check the formulas. It performs no source optimizer updates and draws no target samples. The universal claims follow from the proofs, not from this finite grid. These checks are analytic verification, not an empirical agent study or a confidence-interval coverage experiment.

7 Discussion and conclusion

Static utility assessment can answer a decision question without predicting an exact trained value. In our construction, a finite-source enclosure and local target rules produce a gain range; the same-world comparison then bounds the completion ratio at which the budgeted source ranking changes. Attaining kernels make the fixed-parameter gain bounds sharp, and the all-horizon budget reversal shows why endpoint improvement alone does not settle the allocation. The same task class contains targets that agree on all supplied information yet reverse the fixed-budget ranking for the same finite learner. Certification and this explicit information limit delimit one decision problem.

Its applicability depends on independently justified inputs: a faithful target model, a uniformly helpful action, a common positive policy shift, normalized-progress success, and an all-in completion law. The assumptions already determine the direction of single-method improvement. The work supplies quantitative bounds and their budget consequences, rather than discovering the action direction from source data. It does not establish these contracts for general language-model agents, sharpness of the finite-optimizer or comparison enclosures, or savings over strong verification and deployment alternatives. Deriving useful contracts for richer learners and tasks is the substantive next step; the present result specifies what such a derivation must preserve.

AI USE STATEMENT

Generative AI tools, including OpenAI Codex, assisted with the conceptual framework, mathematical formulation, proof development and checking, literature retrieval and comparison, code implementation, translation, manuscript drafting, and document preparation. Independent AI technical checks were used to examine assumptions, algebra, counterexamples, and claim scope; these are not human peer review. The work reports mathematical constructions and exact analytic checks, not an AI-generated empirical dataset or a completed target-agent experiment. The human authors remain responsible for the claims, citations, artifacts, and final submission.

REPRODUCIBILITY STATEMENT

Sections 3–5 specify the information, target class, finite learner, and complete cost contract. Appendices A–E provide full proofs, boundary cases, and counterexamples. The constructions and numerical bounds can be recomputed from the stated finite-state transition rules and exact formulas, including the finite-source allowance. No private target data, external model inference, or execution of the prescribed optimizer is needed for these checks.

REFERENCES

Rishabh Agarwal, Max Schwarzer, Pablo Samuel Castro, Aaron Courville, and Marc G. Bellemare. Deep Reinforcement Learning at the Edge of the Statistical Precipice, 2021. URL https://arxiv.org/abs/2108.13264v1. arXiv:2108.13264v1.

Martin Avanzini, Gilles Barthe, Davide Davoli, and Benjamin Grégoire. A quantitative probabilistic relational Hoare logic, 2024. URL https://arxiv.org/abs/2407.17127v2. arXiv:2407.17127v2.

Dana Dachman-Soled, Vitaly Feldman, Li-Yang Tan, Andrew Wan, and Karl Wimmer. Approximate resilience, monotonicity, and the complexity of agnostic learning, 2014. URL https://arxiv.org/abs/1405.5268v2. arXiv:1405.5268v2.

Michael Dennis, Natasha Jaques, Eugene Vinitsky, Alexandre Bayen, Stuart Russell, Andrew Critch, and Sergey Levine. Emergent Complexity and Zero-shot Transfer via Unsupervised Environment Design, 2020. URL https://arxiv.org/abs/2012.02096v2. arXiv:2012.02096v2.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models, 2021. URL https://arxiv.org/abs/2106.09685v2. arXiv:2106.09685v2.

Minqi Jiang, Edward Grefenstette, and Tim Rocktäschel. Prioritized Level Replay, 2020. URL https://arxiv.org/abs/2010.03934v4. arXiv:2010.03934v4.

Minqi Jiang, Michael Dennis, Jack Parker-Holder, Jakob Foerster, Edward Grefenstette, and Tim Rocktäschel. Replay-Guided Adversarial Environment Design, 2021. URL https://arxiv.org/abs/2110.02439v1. arXiv:2110.02439v1.

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?, 2023. URL https://arxiv.org/abs/2310.06770v3. arXiv:2310.06770v3.

Sayash Kapoor, Benedikt Stroebl, Zachary S. Siegel, Nitya Nadgir, and Arvind Narayanan. AI Agents That Matter, 2024. URL https://arxiv.org/abs/2407.01502v1. arXiv:2407.01502v1.

Samir Khan, Martin Saveski, and Johan Ugander. Off-policy evaluation beyond overlap: partial identification through smoothness, 2023. URL https://arxiv.org/abs/2305.11812v2. arXiv:2305.11812v2.

Yusong Lin, Haiyang Wang, Shuzhe Wu, Lue Fan, Feiyang Pan, Sanyuan Zhao, and Dandan Tu. CLI-Gym: Scalable CLI Task Generation via Agentic Environment Inversion, 2026. URL https://arxiv.org/abs/2602.10999v1. arXiv:2602.10999v1.

Bo Liu, Simon Yu, Yiding Jiang, Ao Qu, Andrew Zhao, Zichen Liu, Junsu Kim, Zijian Zhou, Seungone Kim, Tongzheng Ren, Mickel Liu, Hanfei Yu, Zhaorun Chen, Weiyan Shi, Paul Pu Liang, Luke Zettlemoyer, Yejin Choi, and Natasha Jaques. SPADE: Self-Play in Adaptive Synthetic Executable Environments, 2026. URL https://arxiv.org/abs/2608.19197v3. arXiv:2608.19197v3.

Jean-Luc Marichal. Algorithms and formulas for conversion between system signatures and reliability functions, 2015. URL https://arxiv.org/abs/1312.7456v2. arXiv:1312.7456v2.

Jean-Luc Marichal, Pierre Mathonet, and Tamás Waldhauser. On signature-based expressions of system reliability, 2011. URL https://arxiv.org/abs/1010.0162v2. arXiv:1010.0162v2.

Zehan Qi, Xiao Liu, Iat Long Iong, Hanyu Lai, Xueqiao Sun, Xinyue Yang, Jiadai Sun, Yu Yang, Shuntian Yao, Tianjie Zhang, Wei Xu, Jie Tang, and Yuxiao Dong. WebRL: Training LLM Web Agents via Self-Evolving Online Curriculum Reinforcement Learning, 2024. URL https://arxiv.org/abs/2411.02337v1. arXiv:2411.02337v1.

Yunhao Tang, Michal Valko, and Rémi Munos. Taylor Expansion Policy Optimization, 2020. URL https://arxiv.org/abs/2003.06259v1. arXiv:2003.06259v1.

Yunhao Tang, Tadashi Kozuno, Mark Rowland, Rémi Munos, and Michal Valko. Unifying Gradient Estimators for Meta-Reinforcement Learning via Off-Policy Evaluation, 2021. URL https://arxiv.org/abs/2106.13125v2. arXiv:2106.13125v2.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A. Smith, Daniel Khashabi, and Hannaneh Hajishirzi. Self-Instruct: Aligning Language Models with Self-Generated Instructions, 2022. URL https://arxiv.org/abs/2212.10560v2. arXiv:2212.10560v2.

Zhaoyang Wang, Canwen Xu, Boyi Liu, Yite Wang, Siwei Han, Zhewei Yao, Huaxiu Yao, and Yuxiong He. Agent World Model: Infinity Synthetic Environments for Agentic Reinforcement Learning, 2026. URL https://arxiv.org/abs/2602.10090v1. arXiv:2602.10090v1.

Zhiheng Xi, Yiwen Ding, Wenxiang Chen, Boyang Hong, Honglin Guo, Junzhe Wang, Dingwen Yang, Chenyang Liao, Xin Guo, Wei He, Songyang Gao, Lu Chen, Rui Zheng, Yicheng Zou, Tao Gui, Qi Zhang, Xipeng Qiu, Xuanjing Huang, Zuxuan Wu, and Yu-Gang Jiang. AgentGym: Evolving Large Language Model-based Agents across Diverse Environments, 2024. URL https://arxiv.org/abs/2406.04151v1. arXiv:2406.04151v1.

John Yang, Akshara Prabhakar, Karthik Narasimhan, and Shunyu Yao. InterCode: Standardizing and Benchmarking Interactive Coding with Execution Feedback, 2023. URL https://arxiv.org/abs/2306.14898v3. arXiv:2306.14898v3.

John Yang, Kilian Lieret, Carlos E. Jimenez, Alexander Wettig, Kabir Khandpur, Yanzhe Zhang, Binyuan Hui, Ofir Press, Ludwig Schmidt, and Diyi Yang. SWE-smith: Scaling Data for Software Engineering Agents, 2025. URL https://arxiv.org/abs/2504.21798v2. arXiv:2504.21798v2.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah D. Goodman. STaR: Bootstrapping Reasoning With Reasoning, 2022. URL https://arxiv.org/abs/2203.14465v2. arXiv:2203.14465v2.

Shuyan Zhou, Frank F. Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, Tianyue Ou, Yonatan Bisk, Daniel Fried, Uri Alon, and Graham Neubig. WebArena: A Realistic Web Environment for Building Autonomous Agents, 2023. URL https://arxiv.org/abs/2307.13854v4. arXiv:2307.13854v4.

A PAIRED CERTIFICATES AND THEIR ESTABLISHED SEMANTICS

This appendix states the general proof rule instantiated in Section 4. It is a standard relational-expectation argument, included to make the certificate self-contained. Consider two finite-state, finite-horizon Markov processes in a fixed compatible world; state may include any history needed to make the next-step distribution well-defined. Let R_t be a relation on their time-t states, and let L_t, U_t be finite real-valued potentials on R_t. Each supported initial state s must satisfy $(s, s) \in R_0$. Suppose

$$L_T(x, y) \le \rho_\omega(x) - \rho_\omega(y) \le U_T(x, y) \quad \text{for all } (x, y) \in R_T. \tag{24}$$

For every world, time $t < T$, and $(x, y) \in R_t$, require a coupling $C^-_{\omega,t,x,y}$ of the *correct* candidate and baseline next-state marginals, supported on R_{t+1}, such that

$$L_t(x, y) \le \mathbb{E}_{C^-} L_{t+1}(X', Y'). \tag{25}$$

An upper certificate may use a different coupling C^+ with the same required marginals and support, satisfying $U_t(x, y) \ge \mathbb{E}_{C^+} U_{t+1}(X', Y')$.

Fix a world before composing the lower couplings. Each marginal process then has its required path law, while the pair remains in the relations. Conditional expectation and induction imply

$$\mathbb{E}_{S\sim\mu_0} L_0(S, S) \le \mathbb{E} L_T(X_T, Y_T) \le J_\omega(\pi_d) - J_\omega(\pi_0). \tag{26}$$

Repeating the argument with the upper couplings proves the upper bound. This virtual common probability space does not require actual evaluations to share random seeds. Choosing a coupling separately in each fixed world is legitimate; replacing a fixed shared environment parameter by a fresh adversarial choice on every visit is a different model.

For bounded terminal reward, $g \ge L$ is equivalent to

$$\mathbb{E}[(1 - \rho(X_T)) + \rho(Y_T)] \le 1 - L. \tag{27}$$

The two post-expectation terms are nonnegative and separable. Finite-horizon termination and a shared initial-parameter relation place this statement within eRHL's semantics and separable-postexpectation completeness result (Avanzini et al., 2024, Definition 4.1 and Theorem 4.7). For an upper gain bound, use $\rho(X_T) + (1 - \rho(Y_T))$. The equality $\mathbb{E}[1] = 1$ relies on termination in this interpretation. Logical completeness is not an efficient algorithm for finding compact potentials or proving arbitrary parameter inequalities. Our particular potentials are given explicitly in (9).

B PROOFS FOR THE PROGRESS CLASS

B.1 LEGAL POLICIES AND ORDERED COUPLING

For $0 \le i \le H$, $\beta_i \in [1/4, 1/2]$ and $d \le 1/4$, so both action probabilities are legal. At a state $i < H$, the candidate forward rate is $b_i + \eta d = \lambda_i + \eta(\beta_i + d)$, which lies in $[0, 1]$ under (3). At H, either process can be treated as absorbing. Starting at zero with at most unit increments, however, it can reach H only at the final time, so absorption does not affect any preterminal calculation.

At each time draw a new $U_t \sim \mathrm{Uniform}[0, 1]$, independent of the past, and set

$$X_{t+1} = X_t + \mathbf{1}\{U_t \le b_{X_t} + \eta d\}, \qquad Y_{t+1} = Y_t + \mathbf{1}\{U_t \le b_{Y_t}\}. \tag{28}$$

These are the correct policy-averaged transition marginals. If $X_t = Y_t$, the candidate increment dominates. If $X_t > Y_t$, the smallest possible next difference is $X_t - Y_t - 1 \ge 0$. Induction gives $0 \le Y_t \le X_t \le t$ without assuming b_i increases with i. Taking conditional expectations gives (8). For $i \ge j$, the adjacent condition implies $|b_i - b_j| \le \sum_{k=j}^{i-1} |b_{k+1} - b_k| \le \kappa(i - j)$.

B.2 ITERATED DRIFT AND LOCAL POTENTIALS

Let $m = \eta d \in [m_l, m_u]$. Conditional drift satisfies

$$(1 - \kappa)D_t + m_l \le \mathbb{E}[D_{t+1} \mid X_t, Y_t] \le (1 + \kappa)D_t + m_u. \tag{29}$$

Since both multiplicative constants are nonnegative, induction from $D_0 = 0$ yields $m_l S_H(1-\kappa) \le \mathbb{E}D_H \le m_u S_H(1+\kappa)$. The gain equals $\mathbb{E}D_H/H$ by the specified success mean. Bounded returns permit clipping the final upper bound at one.

For the lower local potential with $a = 1-\kappa$ and $n = H - t \ge 1$,

$$\mathbb{E}L_{t+1}(X', Y') \ge \{a^{n-1}(a(i-j) + m_l) + m_l S_{n-1}(a)\}/H \tag{30}$$

$$= \{a^n(i-j) + m_l S_n(a)\}/H = L_t(i,j). \tag{31}$$

The upper argument uses $a = 1+\kappa$ and the reversed inequality. At $n = 0$, both potentials equal $(i-j)/H$. Powers of exponent zero mean the constant polynomial one, including when $a = 0$. At $\kappa = 0$, $G_- = G_+ = 1$. At $\kappa = 1$, $G_- = 1/H$. At $H = 1$, both G values are one and no adjacent condition is needed. These cases use finite sums and require no division by zero. We clip only the final global bound, not intermediate local potentials.

B.3 Attaining kernels and scaling

For the stated sharp subfamily, set

$$\lambda_i^+ = \frac{5}{16} + \frac{3i}{16H}, \qquad \lambda_i^- = \frac{5}{16} - \frac{5i}{16H}, \qquad 0 \le i < H. \tag{32}$$

Both are nonnegative, and adding $\eta = 1/4$ gives a number below one. They yield $b_i^{\pm} = 3/8 \pm i/(4H)$ and satisfy the adjacent condition with equality. In these worlds, $b^{\pm}_{X_t} - b^{\pm}_{Y_t} = \pm\kappa D_t$, so the corresponding conditional drift recurrence is exact. Consequently $g = \eta d G_-$ in the minus world and $g = \eta d G_+$ in the plus world. Neither endpoint is affected by upper clipping because $\eta d G_+ < 1$ here.

For $\kappa = 1/(4H)$,

$$G_- = 4\{1 - (1 - 1/(4H))^H\}, \qquad G_+ = 4\{(1 + 1/(4H))^H - 1\}. \tag{33}$$

They converge respectively to $4(1 - e^{-1/4})$ and $4(e^{1/4} - 1)$. In contrast, at a fixed $\kappa \in (0, 1]$, $G_- = [1 - (1-\kappa)^H]/(H\kappa)$, which is of order $1/H$. Sharpness here fixes d and η. Enlarging the actual finite optimizer's endpoint to an interval preserves validity but does not make all interval endpoints attainable by that optimizer.

C Static source proof and its interface

Let $\theta^* = \mathrm{logit}(u) \in [0, \log 3]$ and $T_u(\theta) = \theta + u - \sigma(\theta)$. For $\theta \in [0, \theta^*]$, $\sigma(\theta) \le u$, so $T_u(\theta) \ge \theta$. Also

$$T_u'(\theta) = 1 - \sigma(\theta)(1 - \sigma(\theta)) \in [3/4, 13/16]. \tag{34}$$

The map is increasing, $T_u(\theta^*) = \theta^*$, and therefore preserves $[0, \theta^*]$. The mean-value theorem gives $0 \le \theta^* - \theta_K \le (13/16)^K \theta^*$. Since $\theta^* < 2$ and the sigmoid is $1/4$-Lipschitz,

$$0 \le u - \sigma(\theta_K) \le \tfrac{1}{4}(13/16)^K \theta^* \le \tfrac{1}{2}(13/16)^K. \tag{35}$$

For $u = 1/2$ the sequence is identically zero and the same inequality holds. No value of $\theta_1, \ldots, \theta_K$ is evaluated in this calculation. The rule specifies exact full-gradient arithmetic; changing the update, adding finite-precision error, or introducing stochastic gradients requires a new valid enclosure.

For A, use a list of 75 ones and 25 zeros; for B, use 70 ones and 30 zeros. Label order has no effect on the full-gradient loss. At $K = 100$, $\delta < 1/1000$; for example $(13/16)^4 < 1/2$ gives $\delta < 2^{-26} < 1/1000$. Thus $a_l > b_u > 0$. The target mapping (11) is fixed for both sources and all target states. It is a designed routing interface, not a transfer property guaranteed by logistic regression on arbitrary inputs.

This construction supplies fixed synthetic source materials and a finite learner, rather than an adaptive environment generator. If source generation or collection can change the material, its possible outputs, costs, and correlation with completion must be incorporated into the compatibility set. The simple fixed-source example establishes existence of a complete static proof chain without claiming that these complications disappear in actual agent pipelines.

D SHARED-WORLD COMPARISON AND UNIFORM BUDGET REVERSAL

D.1 COMPARISON ENCLOSURE

For fixed actual endpoints, use B as the slower process in the ordered coupling. Its rate $b_i^B = b_i + \eta d_B$ has the same adjacent variation. A adds the common rate $\eta(d_A - d_B) > 0$, so the same proof gives (14). This reuse only requires legal transition rates, a common positive increment, and the adjacent condition; it does not require the new baseline action probability to equal the original β_i.

At fixed completion probabilities, write $h = g_A - g_B$ and $\Delta q = q_B - q_A \geq 0$. Then $\tau = q_A h - \Delta q\, g_B$. The lower bounds $h \geq \eta(a_l - b_u)G_-$ and $g_B \leq \eta b_u G_+$ imply the lower side of (15); the other sides give its upper bound. The reasoning never chooses different target worlds for h and g_B. Their separate bounding can introduce conservatism, so attainment of these comparison endpoints is not asserted.

D.2 PROOF OF THE COMPLETION-PROBABILITY THRESHOLDS

The source ordering and $G_-, G_+ > 0$ give $0 < g_B < g_A$. Define

$$B_l = \eta b_l G_-, \quad B_u = \eta b_u G_+, \quad D_l = \eta(a_l - b_u)G_-, \quad D_u = \eta(a_u - b_l)G_+. \tag{36}$$

The previous bounds imply $g_B \in [B_l, B_u]$ and $h \in [D_l, D_u]$. All four endpoints are strictly positive, $B_l \leq B_u$, and $D_l \leq D_u$. Monotonicity of $x/(x+y)$ for $x, y > 0$ gives

$$\frac{B_l}{B_l + D_u} \leq \frac{g_B}{g_B + h} \leq \frac{B_u}{B_u + D_l}. \tag{37}$$

Cancelling η gives (16) and proves $0 < r_{\mathrm{B}} \leq r_{\mathrm{A}} < 1$. This uses a box enclosure for (g_B, h); the box corners need not be jointly attainable, so neither threshold is asserted to be sharp. Factoring $V_A - V_B = q_B g_A (r - r_*)$ proves the sign assertions for any nonnegative ratio r, including $r > 1$.

At $r = r_{\mathrm{B}}$ we have $V_A - V_B \leq 0$, and at $r = r_{\mathrm{A}}$ we have $V_A - V_B \geq 0$. Neither boundary alone certifies a strict inequality. A ratio between the bounds leaves this sufficient test undecided; it does not establish intrinsic ambiguity. If $q_B = 0$, direct evaluation of $q_A g_A$ supplies the cases stated in the procedure. When $\kappa = 0$ and the source shifts are known exactly, $G_- = G_+ = 1$ and both thresholds reduce to d_B/d_A, the actual break-even ratio. Equation (21) uses the rational δ_{100} and $H = 16$; its endpoints are rounded outward before applying strict decision rules.

D.3 UNIFORM CONSTANTS AND SIGNS

Bernoulli's inequality gives $(1-\kappa)^j \geq 1 - j\kappa$ for $0 \leq \kappa \leq 1$, and hence, when $\kappa = 1/(4H)$,

$$G_- \geq 1 - \frac{\kappa(H-1)}{2} = \frac{7}{8} + \frac{1}{8H} \geq \frac{7}{8}, \qquad G_- \leq 1. \tag{38}$$

Also $G_+ \geq 1$. To bound it above, note that $(1 + 1/(4H))^j \leq e^{j/(4H)} \leq e^{1/4} < 4/3$ for $j \leq H - 1$. The last strict inequality follows from $e^{1/4} = \sum_{k\geq 0}(1/4)^k/k! < \sum_{k\geq 0}(1/4)^k = 4/3$.

For the low budget, substituting $q_A = 16/25$, $q_B = 1$, and the source intervals yields

$$U_\tau \leq \frac{1}{4}\left\{\frac{16}{25}(1/20 + \delta)\frac{4}{3} - \frac{9}{25}(1/5 - \delta)\frac{7}{8}\right\} \tag{39}$$

$$= \frac{1}{4}\left\{-\frac{61}{3000} + \frac{701}{600}\delta\right\} \leq \frac{1}{4}\left\{-\frac{61}{3000} + \frac{91}{75}\delta\right\} < 0. \tag{40}$$

The final inequality uses $\delta < 1/1000$. At the high budget, $\Delta q = 0$, and $L_\tau \geq (1/4)(1/20 - \delta)(7/8) > 0$. Each single-method gain is positive by Theorem 1; all completion probabilities here are positive. Equation (14) also proves A's larger completed gain independently of the budget.

The all-in cost contract is a constructive assumption. C_{info} includes both-source inspection and the full comparison, so shared information is not obtained free of charge. The extra-unit event can be placed at a gate with a reserved fallback allowance, ensuring that the noncompletion return fits B_0. For each method, every completion returns that method's fixed deterministic endpoint. The example does not claim measured failure probabilities or a distribution-free prediction of runtime.

D.4 Exact numbers and strong baseline

For $H = 16$, let $\kappa = 1/64$ and form $G_\pm$ as finite sums of rational powers. Substitute the exact rational $\delta = \frac{1}{2}(13/16)^{100}$ into (15). Independent single-method subtraction instead uses

$$[\eta\{q_A a_l G_- - q_B b_u G_+\},\ \eta\{q_A a_u G_+ - q_B b_l G_-\}]\,. \tag{41}$$

The checker computes these bounds exactly before outward rounding.

For the affine comparator $b_i = c + i/(4H)$, a progress process with fixed increment $m \geq 0$ satisfies $\mathbb{E}Z_{t+1} = (1+1/(4H))\mathbb{E}Z_t + c + m$ before the final time. Hence $J_m = (c+m)G_+$, and subtracting the $m = 0$ baseline removes c. For $c \in [1/4, 1/2]$, $\lambda_i = b_i - \beta_i/4 \geq 0$ and $\lambda_i + 1/4 < 1$. This is a legal strong comparator that preserves the same parameter dependence.

E Failure diagnostics and information limits

E.1 Opposite rankings in the same progress class

For Proposition 5, the policies and score are exactly those of the main progress class. With $\beta_i = 1/4 + i/12$, the rates in (22) correspond to

$$\lambda^{\mathrm{v}} = (5/16, 5/24, 13/48), \qquad \lambda^{\mathrm{p}} = (5/16, 3/8, 13/48). \tag{42}$$

All entries lie in $[0, 3/4]$, so adding $\eta = 1/4$ gives legal action rates. Both adjacent differences of either b vector have magnitude $1/12$. The two worlds therefore satisfy the same supplied bounds, base policy, and success rule. The actual rate vector is not part of this information record. For the cost law, use Theorem 4's all-in construction with extra-unit probability $1/5$ instead of $9/25$. At $B = C_3$, $q_A = 4/5$ and $q_B = 1$ in both worlds, with the same charged audit and base fallback. This probability change defines the fixed-budget witness; the all-horizon reversal theorem keeps its original cost law.

Here is an exact return calculation. Put $p_i = b_i + m$, $\alpha = b_1 - b_0$, and $\zeta = b_2 - b_1$. After two steps the progress probabilities are $(1-p_0)^2$, $p_0(2-p_0-p_1)$, and $p_0 p_1$. Using these for the final expected increment yields

$$3J(m) = (3 + 3\alpha + \alpha\zeta)p_0 + (\zeta - \alpha)p_0^2. \tag{43}$$

Subtracting $J(0)$ gives

$$J(m) - J(0) = \frac{m}{3}\{3 + 3\alpha + \alpha\zeta + 2(\zeta - \alpha)b_0 + (\zeta - \alpha)m\}. \tag{44}$$

For the valley vector, $(\alpha, \zeta) = (-1/12, 1/12)$; for the peak vector they are $(1/12, -1/12)$. These substitutions give the two gain polynomials in the main text. With $m_A = 1/16$ and $m_B = 1/20$, $(4/5)m_A - m_B = 0$ and $(4/5)m_A^2 - m_B^2 = 1/1600$. Hence the reference differences are exactly $\pm 1/28800$.

For either polynomial, $0 < g'(d) < 1/3$ on $[0, 1/4]$. Lemma 2 places the actual shifts within δ below their respective reference values. Thus changing to the same actual source endpoints in both worlds perturbs either utility difference by at most $(4/5 + 1)\delta/3 = 3\delta/5 < \delta$. The inequality $(13/16)^4 < 1/2$ gives $\delta < 2^{-26} < 1/28800$, proving (23). Positive individual gains and A's larger completed gain follow from (7) and (14).

The input record is identical across these worlds, including the actual source programs and cost law. A strict ranking valid in every compatible world would have to agree with both opposite signs, which is impossible. This does not rule out useful interval output, abstention, or a decision based on richer target evidence. In particular, it neither proves that all ratios between r_{B} and r_{A} are ambiguous nor makes these sufficient thresholds sharp.

E.2 Missing variation information

At $H = 2$, $\eta = 1/4$, and $(\lambda_0, \lambda_1) = (1/2, 1/8)$, the base rates are $b_0 = 9/16$ and $b_1 = 7/32$. For a policy increment $m = \eta d$, direct two-step expectation gives

$$\mathbb{E}X_2 - \mathbb{E}Y_2 = m\{2 + b_1 - b_0\} = m\frac{53}{32}. \tag{45}$$

The success-mean difference is half this quantity: $g = 53m/64$. With $\kappa = 11/32$, $G_- = (2 - \kappa)/2 = 53/64$, so the lower certificate is exact. An unsupported choice $\kappa = 0$ would assert $g \geq m$ and fail. The correct information still certifies positive gain.

E.3 Discarded parameter equalities versus genuinely absent information

This diagnostic uses a separate routing task. A root decision selects R with probability p and S otherwise. S succeeds with probability $5/16$. R succeeds if two independent coins with the same fixed parameter $\theta \in [1/4, 3/4]$ differ:

$$G(\theta) = 2\theta(1 - \theta) \in [3/8, 1/2]. \tag{46}$$

The base uses $p_0 = 1/2$ and the completed policy $p_K = 1/2 + d_K$. Appendix C's source with $u = 3/4$ and $K = 100$ supplies $d_K > 0$, through a *different target interface* controlling this root decision. The fixed-d_K gain range is exactly $[d_K/16, 3d_K/16]$; its source enclosure gives $[d_l/16, 3d_u/16]$.

Dropping the equality between the three local parameter uses permits a second-coin probability of $1/4$ after zero and $3/4$ after one. For any first-coin probability $x \in [1/4, 3/4]$,

$$G_{\text{relaxed}} = (1 - x)/4 + x(1 - 3/4) = 1/4, \qquad g_{\text{relaxed}} = -d_K/16. \tag{47}$$

Both policies can share this enlarged kernel: the relaxation loses a cross-state equality. Two separately fixed coin parameters do not suffice for this negative example, since their minimum disagreement probability on $[1/4, 3/4]^2$ is still $3/8$.

When equality is supplied, the negative world is incompatible. With only local ranges and no equality, however, the fixed-$\theta = 1/2$ and outcome-dependent worlds are both compatible. They yield $3d_K/16$ and $-d_K/16$ for the same source, base, cost law, and visible information. Their opposite signs prove ambiguity for that weaker interface, without implying impossibility for richer static evidence.

E.4 Changing the success mechanism

For terminal-goal reward $\mathbf{1}\{i = H\}$, gain is a tail-probability difference, so $g = \mathbb{E}D_H/H$ no longer holds. Ordered coupling proves nonnegative gain for nondecreasing scores, but a constant score has zero gain. Thus (7) requires the specified success mean $\rho(i) = i/H$.