WHEN PREDICTIONS BECOME SUPERVISION: SHARP MINIMAX LAWS FOR STOCHASTIC REPLAY

Anonymous authors
Paper under double-blind review

ABSTRACT

When past predictors are reused to generate new training labels, each prediction shapes both current accuracy and future supervision. We study this feedback in online learning over K thresholds with a fixed but unknown target. After an initial truthful round, an independent audit reveals the true label with probability ρ, and otherwise an adversary-selected past predictor generates the feedback label. For horizons $T \le K^{1/2-\varepsilon}$ with fixed $\varepsilon \in (0, 1/2)$, we establish sharp minimax rates for every ρ. At $\rho = 1/2$, the minimax cumulative loss is $\Theta_\varepsilon(\min\{T, T^{2/3}(\log K)^{1/3}\})$. Achieving this rate requires sacrificing some current accuracy to preserve future information, and this tradeoff is unavoidable even for improper learners. Restricting replay to sufficiently recent predictors on average reduces the loss to $\Theta(\min\{T, (\log K)/\rho\})$. For proper learners, revealing audit status or the selected replay source achieves the same rate. We further derive sharp rates for partial or noisy information about label generation and show that equally frequent disclosures can yield different minimax rates. Finally, with the replay source revealed but audit status hidden, fixing the reusable first predictor to $p_1 \equiv 1/2$ raises the long-horizon loss from $\Theta((\log K)/\rho)$ to $\Theta((\log K)/\rho^2)$.

1 INTRODUCTION

Model-generated supervision creates a feedback loop between what a system predicts and what it subsequently learns. In self-training, a classifier labels unannotated examples that are then used to update the model (Xie et al., 2020; Sohn et al., 2020). Recursive synthetic-data pipelines extend this reuse across successive model generations. Studies of model collapse have shown that repeatedly training on generated data can degrade model quality or diversity (Shumailov et al., 2024; Alemohammad et al., 2024). Retaining real data and accumulating synthetic samples can prevent collapse in some settings, showing that the data-reuse protocol matters (Gerstgrasser et al., 2024; Kazdan et al., 2025). These findings raise a question beyond the behavior of any particular training procedure. *What can a learner guarantee when its own earlier predictors become sources of future supervision?*

The difficulty is not simply that model-generated labels may be incorrect. Data feedback loops can amplify biases depending on how faithfully model outputs sample the training distribution (Taori & Hashimoto, 2023). In replay, the label distribution depends on predictors that the learner itself produced, and replacing a predictor for deployment need not prevent it from generating labels later. An early prediction can therefore affect learning long after the learner has changed its behavior. Each deployed predictor determines current prediction error and helps determine the feedback available for subsequent learning. A strategy that considers only current accuracy may create a history from which future labels are difficult to interpret. How should a learner balance current accuracy against the informativeness of its future supervision?

In their SODA 2026 paper, Dmitriev et al. (2026) established a foundation for this question through *online learning with a replay adversary*. Their learner deploys binary-valued predictors and receives either a true label or an earlier predictor's output, without observing its origin. They characterize when learning is possible in this replay model and identify structural conditions governing proper learning. In their discussion of open problems, they explicitly propose stochastic predictions, with replay labels sampled from earlier predictors, and an audited fraction of ground-truth feedback as extensions of this framework. We address these two directions jointly for finite thresholds under independent audits, asking how their interaction determines cumulative prediction error.

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

We study online learning over $K \geq 3$ thresholds with a fixed unknown target and no initial history. The learner commits to prediction probabilities for every input before an adversary that knows the target selects the input. The first feedback label is truthful. Thereafter, an independent audit returns the true label with probability ρ, and otherwise feedback is freshly sampled from an earlier predictor selected by the adversary. The learner observes the input and feedback label, but neither audit status nor source identity. Our objective is *cumulative deployment loss*, which sums prediction-error probabilities against the target over all rounds.

Our results give a sharp answer under the horizon condition stated below. A learner can attain the minimax rate by balancing current accuracy against the information preserved in future feedback. One way to achieve this balance is to keep prediction probabilities away from zero and one when needed, accepting some current error to preserve the information supplied by later audits. At audit rate $\rho = 1/2$, replaying prediction probability zero when the truth is one produces the same feedback distribution as replaying probability one when the truth is zero. Avoiding overly extreme predictions prevents this cancellation. The misleading predictors are not supplied to the adversary in advance. They must arise from the learner's actual actions, starting from empty history. Our matching lower bound accounts for both the loss incurred while producing these predictors and the difficulty of learning when they are repeatedly reused. With fully truthful feedback, the classical online-learning benchmark is $\Theta(\min\{T, \log K\})$ (Littlestone, 1988). Our three main contributions quantify the additional cost of replay and identify when it can be reduced.

1. **Sharp minimax rates when replay sources are hidden.** For fixed $\varepsilon \in (0, 1/2)$ and $1 \leq T \leq K^{1/2-\varepsilon}$, we establish sharp minimax rates for every audit rate ρ (Theorem 3.1). At the critical rate $\rho = 1/2$, the minimax loss is $\Theta_\varepsilon(\min\{T, T^{2/3}(\log K)^{1/3}\})$. The upper bound captures the tradeoff between current accuracy and future learnability: clipping with a probability floor α incurs loss of order αT, while the feedback signal it preserves yields a learning cost of order $(\log K)/\alpha^2$. The matching lower bound shows that this tradeoff is unavoidable even for improper learners that may choose arbitrary probability-valued prediction functions before the query is selected. For proper learners, we further obtain sharp rates for all horizons when the positive audit rate ρ stays below $1/3$ by a fixed margin (Theorem 3.2).
2. **Recent predictors and additional feedback information reduce the loss.** We study two ways to reduce the loss in Section 4. If the expected delay between deployment and reuse is bounded by a fixed $D \geq 1$, proper learners achieve $\Theta_D(\min\{T, (\log K)/\rho\})$ for all horizons and $\rho > 0$, even when the selected predictor and audit status remain hidden (Theorem 4.1). With unrestricted replay, revealing audit status or the selected predictor's index or score after feedback yields the same rate (Theorem 4.2). Under Theorem 3.1's horizon condition, we further characterize partial disclosure and noisy observations of audit status, and show that two protocols revealing the same expected number of source indices can differ in loss by a factor $1/\rho$ (Section 4.2 and Corollary 4.5).
3. **Revealing the source does not eliminate the cost of fresh replay sampling.** We consider the setting in which the selected predictor's index or exact score is revealed, audit status remains hidden, and the first predictor is fixed as $p_1 \equiv c$. All later proper actions remain unrestricted, and the first predictor remains available for replay. We establish sharp minimax rates for every $c \in [0, 1]$, $\rho > 0$, and horizon T (Theorem 5.1). When $c \in \{0, 1\}$, the loss is $\Theta(\min\{T, (\log K)/\rho\})$, whereas $c = 1/2$ yields $\Theta(\min\{T, (\log K)/\rho^2\})$. Thus, even with exact source information, fresh sampling from a stochastic replay source can incur an additional factor $1/\rho$. Revealing audit status eliminates this separation.

2 Model and Learning Objective

We formalize online threshold learning in which past prediction functions serve as sources of future feedback. The objective is to minimize cumulative prediction error against a fixed target under independent audits and adversarial replay.

Predictions and replay. Let $\Delta(A)$ denote distributions on a finite set A, with $\Delta_n = \Delta([n])$. For an integer $K \geq 2$, the input space is $\mathcal{X} = [K-1] = \{1, \ldots, K-1\}$ and the hypothesis class is $\mathcal{H}_K = \{h_a : a \in [K]\}$, where $h_a(x) = \mathbf{1}\{x \geq a\}$. On round t, the learner submits a score function $p_t : \mathcal{X} \to [0, 1]$, where $p_t(x)$ is the probability of predicting one at x. A proper randomized learner chooses $Q_t \in \Delta(\mathcal{H}_K)$, giving $p_t(x) = \mathbb{E}_{h \sim Q_t} h(x) = \sum_{a \leq x} Q_t(h_a)$. Proper score functions are

exactly the nondecreasing maps $\mathcal{X} \to [0,1]$, whereas improper learners may submit arbitrary maps $\mathcal{X} \to [0,1]$. Each function is retained unchanged for replay. Replaying source s at input x draws a fresh label from $\mathrm{Bern}(p_s(x))$, reusing prediction probabilities rather than a threshold sampled once from Q_s.

Interaction and feedback. The class size K, integer horizon $T \geq 1$, and audit rate $\rho \in [0,1]$ are known. Before interaction, a target cutoff $a^* \in [K]$ is fixed independently of the learner's private randomness. The target $h^* := h_{a^*}$ is hidden from the learner and known to the adversary. On every round, the learner submits its full function p_t before the adversary chooses the query x_t. For $t \geq 2$, the adversary also chooses a probability vector $\boldsymbol{\lambda}_t = (\lambda_t(1), \ldots, \lambda_t(t-1)) \in \Delta_{t-1}$ over previous rounds and draws a replay source $S_t \sim \boldsymbol{\lambda}_t$. These choices may depend on the target, observed history, p_t, and the adversary's randomness, but not on the learner's private seed. Write $y_t^* := h^*(x_t)$ for the true label. After source selection, a fresh audit indicator $A_t \sim \mathrm{Bern}(\rho)$ is drawn independently of the target, private seeds, and preceding randomness. If $A_t = 1$, the learner receives y_t^*. Otherwise, the selected predictor generates a fresh replay label. Given the choices already made, audit and replay sampling use independent randomness. Thus, conditional on past functions and observations, the target, current action, query, and selected source,

$$Y_t \sim \mathrm{Bern}(\rho y_t^* + (1-\rho) p_{S_t}(x_t)), \qquad t \geq 2. \tag{1}$$

The learner observes only (x_t, Y_t). The audit indicator A_t, source distribution $\boldsymbol{\lambda}_t$, and realized source S_t remain hidden. The audit indicator records how the label was obtained, not whether it happens to be correct. Since the source is chosen before the audit, it is well defined on audited rounds too. History starts empty, so round one has no replay source. After the learner's first prediction, $Y_1 = y_1^*$ is revealed truthfully, even when $\rho = 0$, and the first predictor becomes available for later replay.

Learning objective. For a prediction probability p and binary label y, define $\ell(p,y) := |p - y|$. The round-t *deployment loss* $\ell_t := \ell(p_t(x_t), y_t^*)$ is the learner's probability of predicting the wrong label against the fixed target. It is charged on every round, regardless of how feedback is generated, giving cumulative loss $L_T := \sum_{t=1}^{T} \ell_t$. These losses are used only for evaluation and are not separately revealed. For a finite binary class $\mathcal{H}$ on a finite input space, let Alg range over proper randomized learners using mixtures of $\mathcal{H}$, and let $\mathfrak{A}_\rho$ be the adversaries under the same protocol. The minimax cumulative deployment loss is

$$\mathcal{R}_T(\mathcal{H}, \rho) := \inf_{\mathsf{Alg}} \sup_{h^* \in \mathcal{H},\, \mathsf{Adv} \in \mathfrak{A}_\rho} \mathbb{E}_{\mathsf{Alg},\mathsf{Adv},h^*} L_T. \tag{2}$$

For thresholds, abbreviate $\mathcal{R}_T(K,\rho) := \mathcal{R}_T(\mathcal{H}_K, \rho)$. The expectation covers all randomness in the interaction. Allowing improper learners gives $\mathcal{R}_T^{\mathrm{imp}}(K,\rho)$, with the same prediction-before-query timing, so $\mathcal{R}_T^{\mathrm{imp}}(K,\rho) \leq \mathcal{R}_T(K,\rho)$. Unless stated otherwise, learners are proper and the model above remains unchanged. Appendix C gives the formal information structure and randomized-strategy model. All logarithms are natural unless a base is specified.

3 The Cost of Hidden Replay

We first study the original model, in which any past predictor may be replayed and neither the selected predictor nor audit status is revealed. Its difficulty depends on both the feedback generated by past predictions and the loss incurred while producing those predictions.

3.1 Sharp Minimax Law

Theorem 3.1 (Sharp minimax rate under hidden replay)**.** *Fix $\varepsilon \in (0, 1/2)$. Suppose the learner submits its entire prediction function before the query is chosen, and neither the selected replay source nor the audit status is revealed. For integers $K \geq 2$, $T \geq 1$, and $\rho \in [0,1]$, define*

$$r_\rho(T,K) := \begin{cases} \min\left\{T,\ T^{2/3}(\log K)^{1/3},\ \dfrac{\log K}{(2\rho-1)^2}\right\}, & \rho > \frac{1}{2}, \\ \min\left\{T,\ T^{2/3}(\log K)^{1/3}\right\}, & \rho = \frac{1}{2}, \\ \min\left\{T,\ (1-2\rho)T + T^{2/3}(\log K)^{1/3}\right\}, & \rho < \frac{1}{2}. \end{cases} \tag{3}$$

Then, uniformly over these parameters with $1 \le T \le K^{1/2-\varepsilon}$,

$$\mathcal{R}_T(K,\rho) = \Theta_\varepsilon\big(r_\rho(T,K)\big), \qquad \mathcal{R}_T^{\mathrm{imp}}(K,\rho) = \Theta_\varepsilon\big(r_\rho(T,K)\big). \tag{4}$$

The constants hidden by Θ_ε *depend only on* ε. *Moreover, for every integer* $K \ge 2$, $1 \le T \le \lfloor \log_2 K \rfloor$, *and* $\rho \in [0,1]$, *the exact identity* $\mathcal{R}_T(K,\rho) = \mathcal{R}_T^{\mathrm{imp}}(K,\rho) = T/2$ *holds without the preceding horizon restriction.*

Remark 3.1 (**Rate regimes and scope**). For $\log K \ll T$ within the theorem's horizon range, the critical audit rate $\rho = 1/2$ gives $T^{2/3}(\log K)^{1/3}$, reflecting the tradeoff between current accuracy and preserving future information. When $\rho > 1/2$, replay cannot eliminate the excess truthful signal, so loss is capped at order $\log K/(2\rho-1)^2$. This cap becomes active at $T \asymp \log K/(2\rho-1)^3$ whenever that scale lies in the theorem's range. When $\rho < 1/2$, the audit deficit adds the linear term $(1-2\rho)T$, so any fixed gap below $1/2$ gives linear loss. The transition occurs at $|2\rho-1| \asymp (\log K/T)^{1/3}$. These comparisons concern joint growth of K and T. For fixed finite K and any $\rho > 0$, the minimax cumulative loss is uniformly bounded in T (Appendices T and S.1).

The sharp rate in Theorem 3.1 is restricted to $T \le K^{1/2-\varepsilon}$. In the low-audit regime, however, the minimax rate can be characterized for every horizon.

Theorem 3.2 (Low-audit rate for all horizons). *Fix* $\delta \in (0, 1/3)$. *Uniformly over integers* $K \ge 3$, $T \ge 1$, *and* $0 < \rho \le 1/3 - \delta$, *the proper minimax value satisfies* $\mathcal{R}_T(K,\rho) = \Theta_\delta(\min\{T, K/\rho\})$, *where the constants depend only on* δ.

The upper bound adapts the conservative threshold strategy of Dmitriev et al. (2026). Starting from the constant-zero predictor, it moves the cutoff left to the current query only when a zero prediction is followed by feedback one. All earlier predictors also output zero there, so such feedback certifies a true positive (Appendix T). The cutoff $1/3$ marks the range of the separate lower-bound construction, not a claimed transition (Appendix W).

3.2 Why This Law Appears

The upper bound sacrifices current accuracy to preserve future information. The lower bound first induces predictors whose replay can obscure unresolved target labels.

Clipping preserves information in future replay. For $\alpha \in [0, 1/2]$, keeping deployed scores in $[\alpha, 1-\alpha]$ preserves a signed separation of at least $\gamma = 2\rho - 1 + 2(1-\rho)\alpha$ between possible feedback means under the two target labels. Set $p_1 \equiv 1/2$ and $w_2(a) = 1/K$, ignoring the first truthful label. For $t \ge 2$, form $r_t(x) = \sum_{a \le x} w_t(a)$ and apply affine clipping, $p_t(x) = \alpha + (1-2\alpha) r_t(x)$. Mixing with the two endpoint thresholds keeps this predictor proper. After observing (x_t, Y_t), update $w_{t+1}(a) \propto w_t(a) \exp\{\frac{\gamma}{2}(2h_a(x_t) - 1)(Y_t - \frac{1}{2})\}$ and normalize over $a \in [K]$. This constant-α rule uses only observed feedback and current weights (Appendix N, Eq. (100)).

Proposition 3.3 (Clipped upper bound). *For* $T \ge 2$, *constant* $\alpha \in [0, 1/2]$, *and* $\gamma = 2\rho - 1 + 2(1-\rho)\alpha > 0$, *the proper learner above satisfies*

$$\mathbb{E}[L_T] \le \frac{1}{2} + (T-1)\alpha + \frac{4 \log K}{\gamma^2}. \tag{5}$$

Remark 3.2 (**How the three audit regimes emerge**). Clipping costs at most α per round, while the preserved feedback signal contributes a learning cost of order $(\log K)/\gamma^2$. At $\rho = 1/2$, we have $\gamma = \alpha$, so optimizing $\alpha T + (\log K)/\alpha^2$ gives the critical rate $T^{2/3}(\log K)^{1/3}$. When $\rho > 1/2$, the positive offset $2\rho - 1$ allows $\alpha = 0$, giving the additional cap $(\log K)/(2\rho-1)^2$. When $\rho < 1/2$, clipping must first compensate for the negative offset $2\rho - 1$, which produces the additional linear term in Theorem 3.1. Together with the constant-$1/2$ predictor, whose loss is $T/2$, optimizing the clipping level gives Theorem 3.1's upper bound at every horizon, including $\rho = 0$.

The potential argument uses $r_t(x_t)(1 - r_t(x_t)) \le |r_t(x_t) - y_t^*|$ to control the second-order term by the internal prediction error. This argument extends to any finite binary class admitting a proper predictor that outputs $1/2$ everywhere, with $\log|\mathcal{H}|$ replacing $\log K$. Thresholds attain the worst-class order under the horizon condition of Theorem 3.1, while some classes admit smaller bounds through the closure method of Dmitriev et al. (2026) (Appendix U). Appendices M and N give the full analysis and an anytime implementation.

Why the tradeoff is unavoidable. The adversary induces two actual predictors and reuses them to test d unresolved branch bits along a balanced search-tree path, with m setting the time scale of construction and reuse. The construction and these bits encode one cutoff fixed before interaction.

Proposition 3.4 (Two-predictor lower bound)**.** *Let $m \geq 4$ and $1 \leq d \leq m/2$ be integers satisfying $3m \leq T$ and $(m+1)^2 2^d \leq K$, and put $R_{m,d} = (dm^2)^{1/3}$. There is a universal constant $c > 0$ such that, under the prediction-before-query timing,*

$$\mathcal{R}_T^{\mathrm{imp}}(K, \rho) \geq c \begin{cases} \min\left\{ R_{m,d},\ \dfrac{d}{(2\rho-1)^2} \right\}, & \rho \geq \frac{1}{2}, \\ \min\{m,\ (1-2\rho)m + R_{m,d}\}, & \rho < \frac{1}{2}, \end{cases} \tag{6}$$

for every $\rho \in [0,1]$, with division by zero interpreted as $+\infty$.

Remark 3.3 (**Why the lower bound matches the upper bound**). Let L_{build} be the expected loss while producing two replay sources in the lower-bound experiment. Large construction loss already suffices. Otherwise, the selected scores approach opposite extremes at enough later test points. At $\rho = 1/2$, feedback gaps of at most $O(L_{\mathrm{build}}/m)$ make unresolved target bits difficult to distinguish. Reusing the pair across d bits yields

$$\mathbb{E}[L_T] \geq L_{\mathrm{build}} + c \min\left\{ m,\ \frac{d}{(L_{\mathrm{build}}/m)^2} \right\}. \tag{7}$$

Balancing construction loss against the second term gives $(dm^2)^{1/3}$, matching the upper-bound tradeoff. For proper learners, monotonicity extends score control from the source-selection points to the test interval. For improper learners, independent potential queries drawn after submission of the full function give control over a constant fraction of tests in expectation. For $\rho > 1/2$, the gap also contains $2\rho - 1$. For $\rho < 1/2$, sufficiently separated scores let a mixture of the two sources make the feedback laws coincide.

Both expectations refer to an enhanced experiment: the current construction node is revealed before prediction, a completed branch after loss, and a testing bit only after its level's last loss and feedback. The target is independent of the learner's seed. Each level starts with a conditionally fair bit. Appendices O and V give the full timing and constructions. The construction uses $(m+1)^2 2^d$ cutoff leaves. With $m = \lfloor T/6 \rfloor$, the restriction $T \leq K^{1/2-\varepsilon}$ leaves room for $d = \Theta_\varepsilon(\log K)$ branch bits. For $T \geq C_\varepsilon \log K$ with sufficiently large C_ε, these choices satisfy Proposition 3.4 and give Theorem 3.1's lower bound. Shorter horizons use the balanced-tree bound (Appendices P and V).

Robustness to revealing the query. The lower bound does not rely on hiding the current query from a proper learner. Let $\mathcal{R}_T^{\mathrm{qf}}(K, \rho)$ denote the proper minimax value when the query is revealed before the learner submits its full prediction function, while the adversary still selects a past predictor after submission and before the audit.

Corollary 3.5 (Query revealed before prediction)**.** *The proper minimax value $\mathcal{R}_T^{\mathrm{qf}}(K, \rho)$ satisfies the same rate and the same exact short-horizon identity as in Theorem 3.1.*

Appendix L gives the corresponding proper lower-bound construction. This robustness need not extend to improper learners. If the query is revealed before prediction and arbitrary probability-valued prediction functions are also allowed, then for $\rho \geq 1/2$ the minimax loss admits the upper bound $O(\min\{T, \log^2(2K)\})$ from Appendix X. Thus either relaxation alone preserves the rate of Theorem 3.1, while allowing both can reduce the minimax loss.

4 When Does Replay Become Easier?

We study two ways to reduce replay's cost: limit how far back replay can reach on average while keeping sources and audit status hidden, or allow unrestricted replay and reveal how feedback was generated.

4.1 Limiting How Far Replay Reaches

We examine how the delay between deployment and reuse affects the minimax loss. To isolate this effect, we bound the conditional expected reuse delay while keeping the selected predictor and audit status hidden.

Bounding how far back replay can reach. For known $D \geq 1$, let $\mathfrak{A}_\rho(D) \subseteq \mathfrak{A}_\rho$ consist of adversaries satisfying $\sum_{s<t} \lambda_t(s)(t-s) \leq D$ almost surely on every round $t \geq 2$. The selection distribution $\boldsymbol{\lambda}_t$ is conditional on the target and legal history, including the current predictor and query, and the bound holds before the source is sampled. It therefore bounds the conditional expected time since deployment rather than restricting replay to the most recent D rounds. Older predictors remain eligible, and neither the selected predictor nor audit status is revealed. Let $\mathcal{R}_T(K, \rho; D)$ denote the corresponding proper minimax value.

Theorem 4.1 (Sharp rate under a bound on expected replay age)**.** *Fix $D \geq 1$. Under the restriction above, with the selected predictor and audit status hidden, uniformly over integers $K \geq 3$, $T \geq 1$, and $0 < \rho \leq 1$,*

$$\mathcal{R}_T(K, \rho; D) = \Theta_D\left(\min\left\{T, \frac{\log K}{\rho}\right\}\right).$$

The constants depend only on D.

Remark 4.1 (**Why recent replay is easier**). Bounding how far back replay can reach removes much of the cost of unrestricted replay. At $\rho = 1/2$ and $\log K \ll T \leq K^{1/2-\varepsilon}$, Theorem 4.1 improves the rate in Theorem 3.1 by a factor of order $(T/\log K)^{2/3}$. In the low-audit regime of Theorem 3.2, once $T \geq K/\rho$, the corresponding rates are K/ρ under unrestricted replay and $(\log K)/\rho$ under the restriction above. Comparison constants may depend on the fixed parameters ε, δ, D.

The upper bound holds a binary predictor fixed through a waiting period and subsequent internal-weight updates. Once older sources are unlikely, centered feedback has conditional signed mean of order ρ and second moment $O(\rho)$, so a potential bound controls errors. The predictor is reselected only after successive weight updates accumulate total variation $1/4$. This measures the total magnitude of weight changes, summing half the ℓ_1 distance of each update. Its expected total is $O(\log K)$, bounding reselections and their waiting-period loss. This yields the $(\log K)/\rho$ rate for fixed D. The lower bound already holds for $S_t = t - 1$. Appendix J gives the proof. Dependence on growing D remains unmatched.

4.2 Revealing How Feedback Was Generated

We now return to unrestricted replay and instead reveal information about how feedback was generated. This information arrives after the current prediction, loss, and feedback label, while deployment loss itself remains unobserved.

Revealing audit status or source information. For $t \geq 2$, write $z_t = p_{S_t}(x_t)$ for the selected predictor's score, which its index also determines. Because the source is selected before the audit, both its index and score are defined even on audited rounds. An audit flag identifies feedback from the true-label mechanism, while the selected predictor's score allows the learner to subtract the mean contribution of replay. We first consider revealing such information on every round.

Theorem 4.2 (Information revealed on every round)**.** *For integers $K \geq 3$, $T \geq 1$, and $\rho \in (0, 1]$, let I be a fixed choice of the audit flag, the selected predictor's exact score, its index, the bit indicating whether $z_t \geq 1/2$, or all these quantities together. If I is revealed after feedback on every round $t \geq 2$, then, with universal implicit constants, the proper minimax value satisfies*

$$\mathcal{R}_T^{\mathsf{I}}(K, \rho) = \Theta\left(\min\left\{T, \frac{\log K}{\rho}\right\}\right). \tag{8}$$

Remark 4.2 (**Why different observations give the same rate**). Audit flags identify truthful feedback directly. With source scores, the learner instead maintains normalized internal weights $\boldsymbol{w}_t$ over cutoffs and deploys the binary weighted-median threshold $p_t = h_{a_t}$ from round one, where $a_t = \min\{a : \sum_{b \leq a} w_t(b) \geq 1/2\}$. The weights determine the cutoff, while the deployed scores are binary. Hence the bit indicating whether $z_t \geq 1/2$ recovers the selected score. Conditional on the target and pre-feedback information, including the selected source, the centered feedback $s_t = Y_t - (1-\rho)z_t - \rho/2$ satisfies $\mathbb{E}s_t = \rho(y_t^* - 1/2)$ and $\mathbb{E}s_t^2 \leq \rho$. These moment bounds give the $1/\rho$ dependence. The matching lower bound allows the audit flag, index, and score to be revealed together, so combining these observations cannot improve the order (Appendices D and E). Equal rates do not make the observations equivalent for arbitrary past predictors, a distinction that becomes important in Section 5.

Partial disclosure. We now reveal the same information only occasionally. After loss and feedback on each $t \geq 2$, draw $E_t \sim \mathrm{Bern}(q)$ independently, with known q. The learner observes E_t and, when $E_t = 1$, also observes a fixed choice I from Theorem 4.2. Let $\mathcal{R}_T^{\mathsf{I},q}(K,\rho)$ denote the corresponding proper minimax value.

Theorem 4.3 (Sharp rates with partial disclosure)**.** *Fix $\varepsilon \in (0, 1/2)$. Uniformly over integers $K \geq 3$, $1 \leq T \leq K^{1/2-\varepsilon}$, $\rho, q \in [0,1]$, and choices I from Theorem 4.2 revealed without error,*

$$\mathcal{R}_T^{\mathsf{I},q}(K,\rho) = \Theta_\varepsilon\left(\min\left\{r_\rho(T,K), \frac{\log K}{q\rho}\right\}\right), \tag{9}$$

where zero denominators are interpreted as $+\infty$.

Remark 4.3 (**When partial disclosure improves the rate**). The learner can ignore the additional information and retain the rate $r_\rho(T,K)$ from Theorem 3.1, or use the revealed observations and obtain the scale $(\log K)/(q\rho)$. The lower bound shows that combining the two cannot improve the order. At $\rho = 1/2$ and $\log K \ll T \leq K^{1/2-\varepsilon}$, the two terms balance when $q \asymp (\log K/T)^{2/3}$. A strict asymptotic improvement occurs when q is asymptotically larger than this scale. Appendix G gives the proof.

Noisy audit status. We next replace the exact audit flag by a noisy observation through a known channel $Q = (Q_0, Q_1)$ on a finite alphabet. Conditional on $A_t = i$, draw $W_t \sim Q_i$ using fresh independent randomness. The learner still observes (x_t, Y_t, E_t) and receives W_t as its only additional observation when $E_t = 1$. The scale κ_Q quantifies effective information from one noisy audit observation, replacing ρ in the disclosure term. For $0 < \rho < 1$, write $\bar{Q} = (1-\rho)Q_0 + \rho Q_1$ and define

$$J_Q = \sum_{w:\bar{Q}(w)>0} \frac{(Q_1(w) - Q_0(w))^2}{\bar{Q}(w)}, \qquad \kappa_Q = \frac{\rho^2 J_Q}{1 + \rho^2 J_Q}. \tag{10}$$

Let $\mathcal{R}_T^{Q,q}(K,\rho)$ denote the corresponding proper minimax value.

Theorem 4.4 (Sharp rates with noisy audit-status observations)**.** *Fix $\varepsilon \in (0, 1/2)$. Uniformly over integers $K \geq 3$, $1 \leq T \leq K^{1/2-\varepsilon}$, $0 < \rho < 1$, $q \in [0,1]$, and all such Q_0, Q_1,*

$$\mathcal{R}_T^{Q,q}(K,\rho) = \Theta_\varepsilon\left(\min\left\{r_\rho(T,K), \frac{\log K}{q\kappa_Q}\right\}\right), \tag{11}$$

with the same zero-denominator convention.

Remark 4.4 (**When noisy disclosure improves the rate**). Theorem 4.4 has the same form as the audit-flag case of Theorem 4.3, with ρ replaced by κ_Q in the disclosure term. A perfect audit flag has $\kappa_Q = \rho$, while $Q_0 = Q_1$ gives $\kappa_Q = 0$ and no improvement from disclosure. If the audit flag is independently flipped with known probability $\xi \leq 1/4$, then $\kappa_Q \asymp \rho^2/(\rho + \xi)$. For fixed $q > 0$, when the disclosure term controls the rate, noise of order ρ or smaller preserves the $1/\rho$ dependence of exact disclosure, whereas a fixed positive noise level gives $1/\rho^2$ dependence as audits become rare. Appendix H gives the proof.

Disclosure only on replay rounds. Theorem 4.3 reveals information independently of whether the round is audited or replayed. We now reveal information only after replay. For $t \geq 2$, let $Z_t = (1 - A_t)E_t$, where $E_t \sim \mathrm{Bern}(q)$ is drawn independently after loss and feedback but is not separately observed. The learner sees Z_t and may additionally observe (S_t, z_t) when $Z_t = 1$. The known parameter q is thus the disclosure probability conditional on replay. In general, $Z_t = 0$ leaves open whether the round was audited or replay occurred without disclosure.

Corollary 4.5 (Disclosure only on replay rounds)**.** *Fix $\varepsilon \in (0, 1/2)$. Uniformly over integers $K \geq 3$, $1 \leq T \leq K^{1/2-\varepsilon}$, and $\rho, q \in [0,1]$, let $\kappa_{\mathrm{rep}} = q\rho^2/[q\rho + (1-q)(1-\rho)]$, with $\kappa_{\mathrm{rep}} = 0$ when $q = 0$ or $\rho = 0$. Whether the learner observes Z_t alone or also (S_t, z_t) when $Z_t = 1$, its proper minimax value satisfies, with constants depending only on ε and zero denominators meaning $+\infty$,*

$$\mathcal{R}_T^{\mathrm{rep}}(K,\rho,q) = \Theta_\varepsilon\left(\min\left\{r_\rho(T,K), \frac{\log K}{\kappa_{\mathrm{rep}}}\right\}\right). \tag{12}$$

Moreover, compare source-index disclosure with $q_{\mathrm{rep}} = 1/2$ on replay rounds and $q_{\mathrm{ind}} = (1-\rho)/2$ independently on every round. Both reveal $(1-\rho)(T-1)/2$ source indices in expectation. Let $\mathcal{R}_T^{\mathrm{rep}}$ and $\mathcal{R}_T^{\mathrm{ind}}$ denote these values. If $0 < \rho \le 1/4$ and $2\log K/\rho^2 \le T \le K^{1/2-\varepsilon}$, then

$$\mathcal{R}_T^{\mathrm{rep}} = \Theta_\varepsilon\left(\frac{\log K}{\rho^2}\right), \qquad \mathcal{R}_T^{\mathrm{ind}} = \Theta_\varepsilon\left(\frac{\log K}{\rho}\right). \tag{13}$$

Remark 4.5 (**What absence reveals**). With binary stored predictors, a disclosed source score certifies an audit whenever $Y_t \neq z_t$, since replay returns z_t. Independent disclosure can expose these audited rounds, whereas replay-only disclosure excludes them. An absent replay-only disclosure can still indicate an audit and identifies one exactly at $q = 1$. This information is captured by κ_{rep}, which includes the disclosure probability q. For $q, \rho > 0$,

$$\frac{\log K}{\kappa_{\mathrm{rep}}} = \frac{\log K}{\rho} + \frac{(1-\rho)(1-q)\log K}{q\rho^2},$$

so the additional $1/\rho^2$ cost disappears when every replay round is disclosed. The equal-count comparison therefore shows that the value of disclosure depends on when it occurs, not only on how often it occurs.

The upper bound uses only Z_t as additional information, while the matching lower bound allows the source index and score when $Z_t = 1$ (Appendix H). Thus hiding the replay source does not explain the separation. This distinction motivates Section 5, which asks whether source information still suffices when the learner does not control its first predictor.

5 The Lasting Cost of Stochastic Sources

The score- and index-based upper bounds in Section 4.2 use binary predictors from the first round. We now test whether this initial choice is essential. Fix the first predictor to $p_1 \equiv c$ for known $c \in [0,1]$, while leaving all later proper actions unrestricted. After feedback on every $t \ge 2$, the learner observes either the selected predictor's exact score or its index, including on audited rounds, while audit status remains hidden. The first round remains truthful, and the first predictor remains available for replay throughout. Write the proper minimax value as $\mathcal{R}_T^{\mathsf{I},\mathrm{init}(c)}(K,\rho)$, where I denotes either disclosure protocol.

Theorem 5.1 (Sharp minimax rate with a fixed first predictor)**.** *For either disclosure protocol, define $v_c = \rho + (1-\rho)c(1-c)$. With universal implicit constants, uniformly over integers $K \ge 3$, $T \ge 1$, $\rho \in (0,1]$, and $c \in [0,1]$,*

$$\mathcal{R}_T^{\mathsf{I},\mathrm{init}(c)}(K,\rho) = \Theta\left(\min\left\{T, \frac{v_c \log K}{\rho^2}\right\}\right). \tag{14}$$

If the audit flag is revealed instead, the rate is $\Theta(\min\{T, (\log K)/\rho\})$ for every c under the same first-predictor constraint, with or without the selected predictor's score or index.

Remark 5.1 (**Why knowing the source is not enough**). When the horizon cap is inactive, the rate is of order $(\log K)/\rho + (1-\rho)c(1-c)(\log K)/\rho^2$. Thus a constant-zero or constant-one first predictor gives the $1/\rho$ rate from Theorem 4.2, while $c = 1/2$ gives $\Theta((\log K)/\rho^2)$. The extra factor does not come from uncertainty about which predictor was replayed, since its score or index is revealed. It comes from repeatedly drawing fresh labels from a stochastic predictor. In particular, at $c = 1/2$ the first-round loss is only $1/2$, while later replay creates the additional cost. More generally, as $\rho \to 0$ with the horizon cap inactive and c allowed to vary, the rate remains $\Theta((\log K)/\rho)$ if and only if $c(1-c) = O(\rho)$.

Lower bound by replaying the first predictor. The adversary can choose $S_t = 1$ on every round $t \ge 2$. The disclosed index and score are then always 1 and c, so they reveal nothing about the current target label. The feedback law is $Y_t \mid y_t^* \sim \mathrm{Bern}((1-\rho)c + \rho y_t^*)$, and later actions cannot change it. Testing an initially equally likely label for h rounds incurs expected loss $\Omega(\min\{h, v_c/\rho^2\})$. A balanced threshold construction fixes the target before interaction, independently of the learner's seed, and supplies $\Theta(\log K)$ such labels. Each remains conditionally fair when its test begins, given earlier observations and that seed, and the forced first label reveals none of them. Reusing the first predictor for order v_c/ρ^2 rounds per label and truncating at the horizon gives the lower bound. The separate first-round minimax loss $\max\{c, 1-c\} \ge 1/2$ covers $T = 1$ and the initial constant.

Matching upper bound. From round two onward, deploy binary weighted-median thresholds. The selected score then always lies in $\{0, c, 1\}$, even under adaptive source selection. Either disclosure determines z_t, so the learner can subtract the mean contribution of replay using $s_t = Y_t - (1 - \rho)z_t - \rho/2$. Conditional on the target and pre-feedback history, including the selected source, this statistic has mean $\rho(y_t^* - 1/2)$ and second moment at most v_c. The weighted-potential argument therefore gives $O(v_c \log K/\rho^2)$ expected cumulative loss, including round one, while $L_T \le T$ gives the horizon cap. Knowing the source removes uncertainty about its mean contribution, but not the randomness of a fresh replay sample. Later binary predictors add no further replay sampling noise, but the first predictor remains available for replay and continues to determine the rate. Appendix E.2 gives the complete proof.

6 Related Work

Self-training and recursive synthetic data. Self-training admits guarantees under expansion assumptions (Wei et al., 2021), Gaussian-mixture structure (Oymak & Gulcu, 2021), and gradual domain shift (Kumar et al., 2020). Analyses of synthetic-data training identify changes in scaling laws (Dohmatob et al., 2024) and persistent regression error (Dohmatob et al., 2025). With hidden sample origins, Amin et al. (2026) show that empirical risk minimization can fail in recursive classification even when other procedures succeed. These results highlight the roles of distributional structure and the training rule. We instead optimize cumulative deployment loss over all admissible learners facing target-aware reuse of their actual earlier predictors, and quantify the effects of audits and feedback information.

Online learnability and replay. Classical realizable online learning is governed by Littlestone dimension (Littlestone, 1988), with additional constraints when predictors must belong to a prescribed family (Hanneke et al., 2021). The closest predecessor, Dmitriev et al. (2026), characterizes binary replay learning through extended threshold dimension and closure structure. We study its proposed stochastic-prediction and truthful-feedback extensions jointly for finite thresholds under independent audits. Their loss counts errors when the observed label agrees with the target, whereas ours charges deployment error on every round. Fresh replay sampling creates a statistical tradeoff between current accuracy and future information. Replay also appears in language generation (Racca et al., 2026), where the objective is eventual valid generation rather than cumulative supervised prediction error.

Feedback information and deployment effects. Partial monitoring (Bartók et al., 2014) and Apple Tasting (Helmbold et al., 2000; Raman et al., 2024) study how actions affect both loss and available observations. Performative prediction studies distributions that respond to deployment, including state-dependent responses (Perdomo et al., 2020; Brown et al., 2022). Our target stays fixed, while earlier deployments determine the sources of later supervision. Origin information also changes minimax rates in recursive distribution estimation (Kanabar & Gastpar, 2026a). We distinguish audit status from source identity or score, quantify partial and noisy disclosure, and show that identifying a source need not remove its sampling cost. Appendix Y develops these connections and discusses label noise, active learning, watermarking, and further recursive-estimation results.

7 Conclusion

In this work, we establish sharp minimax characterizations of online threshold learning under stochastic replay and independent audits. The analysis identifies a fundamental tradeoff between prediction accuracy and the informativeness of future supervision, with matching lower bounds accounting for the cost of generating replay sources through the learner's own actions. We further characterize how bounded expected reuse delay and additional feedback information reduce cumulative loss, and quantify the effects of partial and noisy disclosure. Finally, a fixed stochastic first predictor that remains reusable can incur an additional sampling cost despite exact source information, whereas audit-status disclosure removes this penalty. These results demonstrate that the statistical difficulty of replay learning depends jointly on the construction of prediction histories, their subsequent reuse, and the information accompanying feedback.

AI USE STATEMENT

Generative AI tools were used to assist with language revision, manuscript organization, presentation refinement, and LaTeX preparation. The research questions, learning model, minimax formulations, algorithmic ideas, lower-bound constructions, theorem statements, proof strategies, and proofs were formulated and developed by the authors. Only after the proofs had been completed by the authors were generative AI tools used as supplementary verification aids to stress-test the completed proofs and to flag possible logical gaps, overly compressed steps, unstated assumptions, edge cases, inconsistencies in quantifiers or parameter regimes, and issues concerning the timing and information structure of the learning protocols. All such issues and any resulting revisions were independently assessed, carried out, and verified by the authors. The authors take full responsibility for the correctness, originality, and content of the final manuscript.

ETHICS STATEMENT

This theoretical study uses no personal data or human subjects and makes no empirical claims about deployed systems.

REPRODUCIBILITY STATEMENT

The model, observation protocol, and learning objective are specified in Section 2 and Appendix C. Complete proofs, explicit learner updates, and lower-bound constructions are provided in the appendices. Appendix B summarizes the assumptions, parameter ranges, and rates of the main results.

REFERENCES

Sina Alemohammad, Josue Casco-Rodriguez, Lorenzo Luzi, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkoohi, and Richard G. Baraniuk. Self-consuming generative models go MAD. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=ShjMHfmPs0.

Kareem Amin, Alex Bie, Weiwei Kong, Umar Syed, and Sergei Vassilvitskii. Learning from synthetic data: Limitations of ERM. In *Proceedings of the 37th International Conference on Algorithmic Learning Theory*, volume 313 of *Proceedings of Machine Learning Research*, pp. 1–23. PMLR, 2026. URL https://proceedings.mlr.press/v313/amin26a.html.

Dana Angluin and Philip Laird. Learning from noisy examples. *Machine Learning*, 2(4):343–370, 1988. doi: 10.1023/A:1022873112823.

Raman Arora, Ofer Dekel, and Ambuj Tewari. Online bandit learning against an adaptive adversary: From regret to policy regret. In *Proceedings of the 29th International Conference on Machine Learning*, 2012. URL https://icml.cc/2012/papers/749.pdf.

Peter Auer and Nicolò Cesa-Bianchi. On-line learning with malicious noise and the closure algorithm. *Annals of Mathematics and Artificial Intelligence*, 23(1–2):83–99, 1998. URL https://homes.di.unimi.it/~cesabian/Pubblicazioni/amai-98.pdf.

Peter Auer and Ronald Ortner. A new PAC bound for intersection-closed concept classes. *Machine Learning*, 66:151–163, 2007. doi: 10.1007/s10994-006-8638-3. URL https://link.springer.com/article/10.1007/s10994-006-8638-3.

Gábor Bartók, Dean P. Foster, Dávid Pál, Alexander Rakhlin, and Csaba Szepesvári. Partial monitoring—classification, regret bounds, and algorithms. *Mathematics of Operations Research*, 39(4):967–997, 2014. doi: 10.1287/moor.2014.0663.

Shai Ben-David, Dávid Pál, and Shai Shalev-Shwartz. Agnostic online learning. In *Proceedings of the 22nd Annual Conference on Learning Theory*, 2009. URL https://www.cs.mcgill.ca/~colt2009/papers/032.pdf.

Quentin Bertrand, Avishek Joey Bose, Alexandre Duplessis, Marco Jiralerspong, and Gauthier Gidel. On the stability of iterative retraining of generative models on their own data. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=JORAfH2xFd.

Olivier Bousquet, Steve Hanneke, Shay Moran, and Nikita Zhivotovskiy. Proper learning, Helly number, and an optimal SVM bound. In *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 582–609. PMLR, 2020. URL https://proceedings.mlr.press/v125/bousquet20a.html.

Martin Briesch, Dominik Sobania, and Franz Rothlauf. Large language models suffer from their own output: An analysis of the self-consuming training loop. *arXiv preprint arXiv:2311.16822*, 2023. URL https://arxiv.org/abs/2311.16822.

Gavin Brown, Shlomi Hod, and Iden Kalemaj. Performative prediction in a stateful world. In *Proceedings of the Twenty-Fifth International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 6045–6061. PMLR, 2022. URL https://proceedings.mlr.press/v151/brown22a.html.

Nicolò Cesa-Bianchi, Ofer Dekel, and Ohad Shamir. Online learning with switching costs and other adaptive adversaries. In *Advances in Neural Information Processing Systems 26*, pp. 1160–1168, 2013.

Moses Charikar and Chirag Pabbaraju. Exploring facets of language generation in the limit. In *Proceedings of the Thirty-Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 854–887, 2025. URL https://proceedings.mlr.press/v291/charikar25a.html.

David Cohn, Les Atlas, and Richard Ladner. Improving generalization with active learning. *Machine Learning*, 15:201–221, 1994. doi: 10.1007/BF00993277. URL https://link.springer.com/article/10.1007/BF00993277.

Malte Darnstädt. The optimal PAC bound for intersection-closed concept classes. *Information Processing Letters*, 115(4):458–461, 2015. doi: 10.1016/j.ipl.2014.12.001. URL https://www.sciencedirect.com/science/article/pii/S0020019014002658.

Sumanth Dathathri, Abigail See, Sumedh Ghaisas, Po-Sen Huang, Rob McAdam, Johannes Welbl, Vandana Bachani, Alex Kaskasoli, Robert Stanforth, Tatiana Matejovicova, Jamie Hayes, Nidhi Vyas, Majd Al Merey, Jonah Brown-Cohen, Rudy Bunel, Borja Balle, Taylan Cemgil, Zahra Ahmed, Kitty Stacpoole, Ilia Shumailov, Ciprian Baetu, Sven Gowal, Demis Hassabis, and Pushmeet Kohli. Scalable watermarking for identifying large language model outputs. *Nature*, 634:818–823, 2024. doi: 10.1038/s41586-024-08025-4. URL https://www.nature.com/articles/s41586-024-08025-4.

Daniil Dmitriev, Harald Eskelund Franck, Carolin Heinzler, and Amartya Sanyal. Learning in an echo chamber: Online learning with replay adversary. In *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 6674–6695. SIAM, 2026. doi: 10.1137/1.9781611978971.239. arXiv:2509.25135.

Elvis Dohmatob, Yunzhen Feng, Pu Yang, François Charton, and Julia Kempe. A tale of tails: Model collapse as a change of scaling laws. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 11165–11197, 2024. URL https://proceedings.mlr.press/v235/dohmatob24b.html.

Elvis Dohmatob, Yunzhen Feng, Arjun Subramonian, and Julia Kempe. Strong model collapse. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/284afdc2309f9667d2d4fb9290235b0c-Abstract-Conference.html.

Ran El-Yaniv and Yair Wiener. On the foundations of noise-free selective classification. *Journal of Machine Learning Research*, 11(53):1605–1641, 2010. URL https://jmlr.org/papers/v11/el-yaniv10a.html.

Shi Fu, Yingjie Wang, Yuzhu Chen, Li Shen, and Dacheng Tao. Self-verification provably prevents model collapse in recursive synthetic training. In *Advances in Neural Information Processing Systems*, volume 38, pp. 40550–40603, 2025. doi: 10.52202/085713-1213. URL https://papers.neurips.cc/paper_files/paper/2025/hash/3380e8116452e0efbf36f35d95e88c94-Abstract-Conference.html.

Roei Gelbhart and Ran El-Yaniv. The relationship between agnostic selective classification, active learning and the disagreement coefficient. *Journal of Machine Learning Research*, 20(33):1–38, 2019. URL https://jmlr.org/papers/v20/17-147.html.

Matthias Gerstgrasser, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Tomasz Korbak, Henry Sleight, Rajashree Agrawal, John Hughes, Dhruv Bhandarkar Pai, Andrey Gromov, Daniel A. Roberts, Diyi Yang, David L. Donoho, and Sanmi Koyejo. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=5B2K4LRgmz.

Surbhi Goel, Steve Hanneke, Shay Moran, and Abhishek Shetty. Adversarial resilience in sequential prediction via abstention. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/1967f962c7c2083618236d80eeb9d1ac-Abstract-Conference.html.

Steve Hanneke. Refined error bounds for several learning algorithms. *Journal of Machine Learning Research*, 17(135):1–55, 2016. URL https://jmlr.org/papers/v17/15-655.html.

Steve Hanneke and Liu Yang. Minimax analysis of active learning. *Journal of Machine Learning Research*, 16(109):3487–3602, 2015. URL https://jmlr.org/papers/v16/hanneke15a.html.

Steve Hanneke, Roi Livni, and Shay Moran. Online learning with simple predictors and a combinatorial characterization of minimax in 0/1 games. In *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pp. 2289–2314. PMLR, 2021. URL https://proceedings.mlr.press/v134/hanneke21a.html.

David P. Helmbold, Nicholas Littlestone, and Philip M. Long. Apple tasting. *Information and Computation*, 161(2):85–139, 2000. doi: 10.1006/inco.2000.2870.

Meena Jagadeesan, Tijana Zrnic, and Celestine Mendler-Dünner. Regret minimization with performative feedback. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9760–9785. PMLR, 2022. URL https://proceedings.mlr.press/v162/jagadeesan22a.html.

Pooria Joulani, András György, and Csaba Szepesvári. Delay-tolerant online convex optimization: Unified analysis and adaptive-gradient algorithms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, pp. 1744–1750, 2016. doi: 10.1609/aaai.v30i1.10320. URL https://ojs.aaai.org/index.php/AAAI/article/view/10320.

Alkis Kalavasis, Anay Mehrotra, and Grigoris Velegkas. On the limits of language generation: Trade-offs between hallucination and mode-collapse. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, 2025. URL https://arxiv.org/abs/2411.09642.

Millen Kanabar and Michael Gastpar. Model non-collapse: Minimax bounds for recursive discrete distribution estimation. *IEEE Transactions on Information Theory*, 72(3):1817–1830, 2026a. doi: 10.1109/TIT.2025.3649611. URL https://arxiv.org/abs/2501.19273. First posted in 2025. A shorter version appeared at ISIT 2025.

Millen Kanabar and Michael Gastpar. Minimax bounds for watermarked and masked recursive discrete distribution estimation. *arXiv preprint arXiv:2608.31091*, 2026b. URL https://arxiv.org/abs/2608.31091. First posted August 31, 2026. A shorter version to appear at ITW 2026.

Joshua Kazdan, Rylan Schaeffer, Apratim Dey, Matthias Gerstgrasser, Rafael Rafailov, David L. Donoho, and Sanmi Koyejo. Collapse or thrive: Perils and promises of synthetic data in a self-generating world. In *Proceedings of the 42nd International Conference on Machine Learning*,

volume 267 of *Proceedings of Machine Learning Research*, pp. 29469–29494. PMLR, 2025. URL https://proceedings.mlr.press/v267/kazdan25a.html.

John Kirchenbauer, Jonas Geiping, Yuxin Wen, Jonathan Katz, Ian Miers, and Tom Goldstein. A watermark for large language models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 17061–17084. PMLR, 2023. URL https://proceedings.mlr.press/v202/kirchenbauer23a.html.

John Kirchenbauer, Jonas Geiping, Yuxin Wen, Manli Shu, Khalid Saifullah, Kezhi Kong, Kasun Fernando, Aniruddha Saha, Micah Goldblum, and Tom Goldstein. On the reliability of watermarks for large language models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=DEJIDCmWOz.

Jon Kleinberg and Sendhil Mullainathan. Language generation in the limit. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/7988e9b3876ad689e921ce05d711442f-Abstract-Conference.html.

Ananya Kumar, Tengyu Ma, and Percy Liang. Understanding self-training for gradual domain adaptation. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 5468–5479. PMLR, 2020. URL https://proceedings.mlr.press/v119/kumar20c.html.

Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, 2:285–318, 1988. doi: 10.1023/A:1022869011914. URL https://link.springer.com/article/10.1023/A:1022869011914.

Matteo Marchi, Stefano Soatto, Pratik Chaudhari, and Paulo Tabuada. Heat death of generative models in closed-loop learning. In *2024 IEEE 63rd Conference on Decision and Control*, 2024. URL https://arxiv.org/abs/2404.02325.

Pascal Massart and Élodie Nédélec. Risk bounds for statistical learning. *The Annals of Statistics*, 34 (5):2326–2366, 2006. doi: 10.1214/009053606000000786.

Anay Mehrotra, Grigoris Velegkas, Xifan Yu, and Felix Zhou. Language generation with infinite contamination. In *Proceedings of the Thirty-Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 5053–5112, 2026. URL https://proceedings.mlr.press/v336/mehrotra26a.html.

John P. Miller, Juan C. Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7710–7720. PMLR, 2021. URL https://proceedings.mlr.press/v139/miller21a.html.

Nagarajan Natarajan, Inderjit S. Dhillon, Pradeep K. Ravikumar, and Ambuj Tewari. Learning with noisy labels. In *Advances in Neural Information Processing Systems*, volume 26, pp. 1196–1204, 2013. URL https://papers.nips.cc/paper/5073-learning-with-noisy-labels.

Samet Oymak and Talha Cihad Gulcu. A theoretical characterization of semi-supervised learning with self-training for Gaussian mixture models. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 3601–3609. PMLR, 2021. URL https://proceedings.mlr.press/v130/oymak21a.html.

Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7599–7609. PMLR, 2020. URL https://proceedings.mlr.press/v119/perdomo20a.html.

Giorgio Racca, Michal Valko, and Amartya Sanyal. Language generation with replay: A learning-theoretic view of model collapse. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2603.11784. arXiv:2603.11784.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Stochastic, constrained, and smoothed adversaries. In *Advances in Neural Information Processing Systems*, volume 24, 2011. URL https://papers.nips.cc/paper_files/paper/2011/hash/692f93be8c7a41525c0baf2076aecfb4-Abstract.html.

Ananth Raman and Vinod Raman. Generation from noisy examples. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 51079–51093, 2025. URL https://proceedings.mlr.press/v267/raman25a.html.

Vinod Raman, Unique Subedi, Ananth Raman, and Ambuj Tewari. Apple tasting: Combinatorial dimensions and minimax rates. In *Proceedings of the Thirty-Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pp. 4358–4380. PMLR, 2024. URL https://proceedings.mlr.press/v247/raman24a.html.

Vinod Raman, Jiaxun Li, and Ambuj Tewari. Generation through the lens of learning theory. In *Proceedings of the Thirty-Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 4740–4776, 2025. URL https://proceedings.mlr.press/v291/raman25a.html.

J. Hyam Rubinstein and Benjamin I. P. Rubinstein. Unlabelled sample compression schemes for intersection-closed classes and extremal classes. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/54d6a55225cebbdc16fbb0e45c5bdf2b-Abstract-Conference.html.

Clayton Scott, Gilles Blanchard, and Gregory Handy. Classification with asymmetric label noise: Consistency and maximal denoising. In *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pp. 489–511, 2013. URL https://proceedings.mlr.press/v30/Scott13.html.

Mohamed El Amine Seddik, Suei-Wen Chen, Soufiane Hayou, Pierre Youssef, and Merouane Debbah. How bad is training on synthetic data? a statistical analysis of language model collapse. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=t3z6UlV09o.

Lianghe Shi, Meng Wu, Huijie Zhang, Zekai Zhang, Molei Tao, and Qing Qu. A closer look at model collapse: From a generalization-to-memorization perspective. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://arxiv.org/abs/2509.16499.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. AI models collapse when trained on recursively generated data. *Nature*, 631:755–759, 2024. doi: 10.1038/s41586-024-07566-y.

Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. FixMatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems*, volume 33, pp. 596–608, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/06964dce9addb1c5cb5d6e3d9838f733-Abstract.html.

Ananda Theertha Suresh, Andrew Thangaraj, and Aditya Nanda Kishore Khandavally. Rate of model collapse in recursive training. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 1396–1404, 2025. URL https://proceedings.mlr.press/v258/suresh25a.html.

Rohan Taori and Tatsunori B. Hashimoto. Data feedback loops: Model-driven amplification of dataset biases. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 33883–33920. PMLR, 2023. URL https://proceedings.mlr.press/v202/taori23a.html.

Colin Wei, Kendrick Shen, Yining Chen, and Tengyu Ma. Theoretical analysis of self-training with deep networks on unlabeled data. In *The Ninth International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=rC8sJ4i6kaH.

Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V. Le. Self-training with Noisy Student improves ImageNet classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. doi: 10.1109/CVPR42600.2020.01070. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Xie_Self-Training_With_Noisy_Student_Improves_ImageNet_Classification_CVPR_2020_paper.html.

Bingji Yi, Qiyuan Liu, Yuwei Cheng, and Haifeng Xu. Escaping model collapse via synthetic data verification: Near-term improvements and long-term convergence. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/29806b2e4dcb35e42abed21d43e2b7d8-Abstract-Conference.html.

Jinghui Zhang, Dandan Qiao, Mochen Yang, and Qiang Wei. Regurgitative training: The value of real data in training large language models. *arXiv preprint arXiv:2407.12835*, 2024. URL https://arxiv.org/abs/2407.12835.

Contents of Appendix

A Notation Reference **19**

B Summary of settings, assumptions, and rates **21**

C Formal model and preliminaries **22**

C.1 Proper randomized threshold predictors . . . 22

C.2 Probability space and filtrations . . . 22

C.3 Feedback distributions from past predictors . . . 23

C.4 Bayes lower bounds for randomized learners . . . 24

C.5 A sequential Bernoulli testing lemma . . . 25

C.6 A balanced-tree horizon lower bound . . . 25

D Proofs with additional feedback information **26**

D.1 Revealing whether feedback came from an audit . . . 26

D.2 Upper bound . . . 27

D.3 Lower bound . . . 27

D.4 Revealing the selected predictor's score or index . . . 29

D.5 Lower bound when audits and replay are equally likely . . . 30

E Binary predictions and the effect of the first predictor **30**

E.1 Information on every round at each positive audit rate . . . 31

E.2 Proof of Theorem 5.1 . . . 32

F Approximate selected-predictor scores or known selection probabilities **33**

F.1 Approximate scores and coarse quantization . . . 34

F.2 Known conditional selection probabilities . . . 34

G Sharp rates when information is revealed independently **35**

G.1 What is revealed and when . . . 35

G.2 Upper bounds, including noisy audit flags . . . 35

G.3 Testing with all observed feedback information . . . 36

G.4 The first phase produces predictors for later replay . . . 37

G.5 Audit rates at or above one-half . . . 38

G.6 Audit rates between one-quarter and one-half . . . 39

G.7 Audit rates at most one-quarter . . . 39

G.8 Completing the lower bound . . . 40

H Noisy audit status and information revealed only on replay rounds **41**

H.1 Minimum-second-moment unbiased weight for audit status . . . 41

H.2 A proper upper bound for arbitrary earlier predictors . . . 42

H.3 Information in the disclosed audit-status observation . . . 42

H.4 Information revealed only on replay rounds . . . 43

I Interpreting noisy audit flags **45**

J An inverse-linear audit bound for recent predictors **45**

J.1 A query-tree lower bound using the previous predictor 46

J.2 Learning while holding predictions fixed 46

K An upper bound for every proper finite class **49**

L Revealing the query before prediction **51**

M Clipped upper bounds for finite hypothesis classes **52**

M.1 Proper affine clipping 52

M.2 An upper bound without observing the selected predictor 52

M.3 All audit rates and the worst case over finite classes 54

N Complete proof of the clipped upper bounds **54**

N.1 The learner and bounds on all earlier scores 55

N.2 Applying the finite-class upper bound 55

N.3 Efficient threshold implementation 55

N.4 Optimization for $\rho \geq 1/2$ 56

N.5 Optimization for $\rho < 1/2$ 56

O Lower bound by producing and repeatedly reusing two predictors **56**

O.1 A target prior with two stages for selecting the sources 56

O.2 Additional information given to the learner 57

O.3 Relating earlier prediction loss to the selected predictors' scores 58

O.4 Reusing the sources across unresolved target bits 59

O.5 Audit rates at or above one half 59

O.6 Audit rates below one half 60

O.7 Checking the horizon, target support, and protocol 61

P Derivation of the sharp minimax rates **61**

P.1 The case $\rho = 1/2$ 63

P.2 The case $\rho > 1/2$ 63

P.3 The case $\rho < 1/2$ 63

P.4 Uniformity and scope 63

Q Learning under a fixed breadth-first query order **63**

Q.1 Mixing the endpoints of the candidate interval 64

Q.2 Estimating the change point at each depth 64

Q.3 Earlier scores equal one half at unresolved queries 64

Q.4 Total loss 65

Q.5 Extension to every positive audit rate 65

R Small classes and extreme audit rates **66**
R.1 One query point . . . 67
R.2 Every label is truthful . . . 67
R.3 No audits after the first round . . . 67

S Bounds for finite threshold classes at every horizon **68**
S.1 Comparison of the available bounds . . . 68
S.2 Proof of the lower bound at audit rate one half . . . 71
S.3 Comparing the critical upper and lower bounds . . . 71
S.4 A lower bound for every audit rate . . . 72

T A cumulative loss bound for every horizon **73**

U Replay in classes closed under intersection **74**

V Lower bound for arbitrary probability-valued predictors **75**
V.1 The target prior and independent query distributions . . . 75
V.2 How earlier loss controls the average score at future queries . . . 76
V.3 Testing unresolved target bits with the same two sources . . . 77
V.4 Optimizing the parameters and endpoints . . . 78

W Low-audit bounds through extended threshold dimension **79**
W.1 Extended threshold dimension and its witnesses . . . 79
W.2 Sharp minimax rates at low audit probabilities . . . 81
W.3 The result for thresholds . . . 83

X Improper learning when the input is revealed first **84**
X.1 Learning at one coordinate when all earlier scores equal one half . . . 84
X.2 A finite-class upper bound . . . 85

Y Additional Related Work **86**
Y.1 Self-Training and Recursive Synthetic Data . . . 86
Y.2 Replay and Language Generation . . . 87
Y.3 Online Learnability, Proper Prediction, and Closure . . . 88
Y.4 Selective Feedback and Noisy Supervision . . . 88
Y.5 Provenance and Recursive Estimation . . . 88
Y.6 Action-Dependent Feedback and Deployment Effects . . . 89

A Notation Reference

This table collects the main notation used throughout the paper. The final column points to the first formal definition or statement fixing the indicated meaning. Symbols with different local meanings are distinguished by context.

Table 1: Notation reference.

Notation	Meaning	First introduced
Model and feedback		
K, T	Number of threshold hypotheses and interaction horizon.	Section 2
ρ	Audit probability on rounds $t \geq 2$. Round one is always truthful.	Section 2
$\mathcal{X}$	Input space $[K-1] = \{1, \ldots, K-1\}$.	Section 2
$\mathcal{H}_K, h_a$	Threshold class $\{h_a : a \in [K]\}$, with $h_a(x) = \mathbf{1}\{x \geq a\}$.	Section 2
a^*, h^*	Fixed target cutoff and target hypothesis $h^* = h_{a^*}$.	Section 2
p_t, Q_t	Deployed prediction probabilities and, for proper learning, the distribution over thresholds inducing them.	Section 2
x_t, y_t^*	Current input and its true label $y_t^* = h^*(x_t)$.	Section 2
$\boldsymbol{\lambda}_t, S_t$	Distribution over earlier rounds and selected replay-source index, with $S_t \sim \boldsymbol{\lambda}_t$.	Section 2
A_t, Y_t	Audit indicator and observed feedback label.	Section 2, Eq. (1)
Losses and minimax rates		
ℓ_t, L_T	Deployment loss $\lvert p_t(x_t) - y_t^* \rvert$ and cumulative loss $\sum_{t=1}^{T} \ell_t$.	Section 2
$\mathcal{R}_T(\mathcal{H}, \rho), \mathcal{R}_T(K, \rho)$	Proper minimax cumulative loss for a finite binary class and its threshold specialization.	Eq. (2)
$\mathcal{R}_T^{\mathrm{imp}}, \mathcal{R}_T^{\mathrm{qf}}$	Values with arbitrary prediction functions, and with proper predictions after the query is revealed, respectively.	Section 2, Corollary 3.5
$r_\rho(T, K)$	Piecewise rate governing hidden replay in the stated horizon range.	Eq. (3)
ε, δ	Fixed margins in $T \leq K^{1/2-\varepsilon}$ and $\rho \leq 1/3 - \delta$. Subscripts on Θ indicate permitted constant dependence.	Theorems 3.1 and 3.2
Replay restrictions and additional observations		
$D, \mathfrak{A}_\rho(D)$, $\mathcal{R}_T(K, \rho; D)$	Conditional expected replay-age bound, admissible adversaries, and corresponding proper minimax value.	Section 4.1
z_t, I	Selected source score $z_t = p_{S_t}(x_t)$ and the fixed type of information disclosed.	Section 4.2, Theorem 4.2
$\mathcal{R}_T^{\mathsf{I}}(K, \rho)$	Proper minimax value when I is revealed after every round $t \geq 2$.	Eq. (8)
q, E_t	Disclosure probability and independent coin $E_t \sim \mathrm{Bern}(q)$. In the replay-only protocol, q is conditional on replay.	Section 4.2
$\mathcal{R}_T^{\mathsf{I},q}(K, \rho)$	Proper minimax value with independent partial disclosure of I.	Eq. (9)
$Q = (Q_0, Q_1), W_t$	Known audit-observation channel and its output, with $W_t \mid A_t = i \sim Q_i$. This Q is distinct from the action distribution Q_t.	Section 4.2, Theorem 4.4
$\bar{Q}$	Marginal channel-output law $(1-\rho)Q_0 + \rho Q_1$.	Eq. (10)
J_Q, κ_Q	Channel separation and information scale $\kappa_Q = \rho^2 J_Q / (1 + \rho^2 J_Q)$ for $0 < \rho < 1$. Fixed-channel endpoint limits are given in Appendix H.	Eq. (10)
$\mathcal{R}_T^{Q,q}(K, \rho)$	Proper minimax value with independent disclosure of noisy audit status through Q with probability q.	Eq. (11)
ξ	Independent flip probability for a noisy observation of audit status.	Section 4.2

Continued on the next page.

Table 1: Notation reference (continued).

Notation	Meaning	First introduced
Z_t	Replay-only disclosure indicator $(1 - A_t)E_t$. Its independent coin E_t is not separately observed.	Section 4.2
$\kappa_{\mathrm{rep}}, \mathcal{R}_T^{\mathrm{rep}}$	Information scale and proper minimax value for disclosure only on replay rounds.	Corollary 4.5
$c, \mathcal{R}_T^{\mathrm{I,init}(c)}$	Prescribed first score $p_1 \equiv c$ and the corresponding value under exact source-score or source-index disclosure.	Section 5
v_c	Second-moment scale $\rho + (1 - \rho)c(1 - c)$ for a fixed first predictor.	Theorem 5.1
Quantities in the main proofs		
α, γ	Clipping floor and preserved signed feedback separation $\gamma = 2\rho - 1 + 2(1 - \rho)\alpha$.	Section 3.2
$r_t(x)$	Internal threshold-mixture prediction before affine clipping.	Section 3.2
$m, d, R_{m,d}$	Block-size and depth parameters of the lower-bound construction, and scale $R_{m,d} = (dm^2)^{1/3}$.	Proposition 3.4
L_{build}	Expected first-phase loss in the enhanced lower-bound experiment.	Remark 3.3
s_t	Feedback statistic defined locally. It is $Y_t - (1 - \rho)z_t - \rho/2$ under source disclosure and $Y_t - 1/2$ in the clipped update.	Remark 4.2, Appendix N
$\underline{p}, \overline{p}, I_0, I_1$	Smallest and largest historical scores at a fixed input, and attainable feedback-mean intervals under target labels zero and one.	Proposition C.3
$U, \mathcal{F}_{t-1}^{\mathrm{pub}}, \mathcal{G}_t$	Learner's private seed, public past, and enlarged pre-feedback information used for conditional analysis.	Appendix C
Additional appendix notation		
$u(w), v_Q$	Minimum-second-moment unbiased audit weight and its second moment $v_Q = \rho^2 + 1/J_Q$ for $0 < \rho < 1$ and $J_Q > 0$.	Eq. (60)
$\mathrm{ExThD}(\mathcal{H})$	Extended threshold dimension, locally denoted by d. It is distinct from the block-depth parameter above.	Eq. (193)

Other symbols are defined locally where they are used.

B SUMMARY OF SETTINGS, ASSUMPTIONS, AND RATES

Appendix reading guide. Appendix C formalizes the model. Table 2 summarizes the assumptions and rates of the main results. Bounds involving $T \le K^{1/2-\varepsilon}$ take $\varepsilon \in (0, 1/2)$ fixed.

Table 2: **Sharp minimax rates for threshold replay learning.** All displayed rates are Θ laws.

Feedback / replay structure	Minimax rate
Hidden sources	
Unrestricted replay	r
Low audit, $\rho \le \frac{1}{3} - \delta$	$\min\{T, K/\rho\}$
Expected reuse delay $\le D$	$\min\{T, s/\rho\}$
Additional information	
Audit flag or source information on every round	$\min\{T, s/\rho\}$
Independent disclosure with probability q	$\min\{r, s/(q\kappa)\}$
Disclosure only on replay rounds	$\min\{r, s/\kappa_{\mathrm{rep}}\}$
A stochastic source remains reusable	
Source score or index revealed, $p_1 \equiv c$	$\min\{T, v_c s/\rho^2\}$
Audit flag revealed, $p_1 \equiv c$	$\min\{T, s/\rho\}$

Here $s = \log K$, $r = r_\rho(T, K)$, and $v_c = \rho + (1-\rho)c(1-c)$. For independent disclosure, $\kappa = \rho$ for exact information and $\kappa = \kappa_Q$ for a known noisy audit-status channel (Theorem 4.4 with endpoints in Appendix H). The unrestricted, independent-disclosure, and replay-only rates are stated for $T \le K^{1/2-\varepsilon}$. The low-audit and bounded-delay results hold for all T. The latter assumes fixed D. Hidden-source results allow proper or improper learners as specified in the corresponding theorems. The information-revelation results concern proper learners.

Appendices H–I study noisy observations of audit status. Appendix K gives the complementary upper bound for general proper finite classes. When the query is revealed before prediction, the proper and improper results appear in Appendices L and X. Appendix F studies known conditional selection probabilities and approximate estimates of the selected predictor's score. Appendix U explains how class structure can reduce the loss. Appendices S and T give finite-domain bounds for arbitrary horizons, Appendix Q studies a fixed breadth-first query order, and Appendix R treats boundary cases.

Table 2 distinguishes what the learner may predict, which past predictors may be replayed, and what the learner observes. History starts empty and the first feedback label is truthful. The learner chooses its first predictor except in the rows that fix $p_1 \equiv c$. When information is revealed independently after each round, the learner sees whether that round was selected for disclosure. When information is revealed only on replay rounds, the learner sees $Z_t = (1 - A_t)E_t$ but not the independent coin E_t. Neither rule imposes a fixed total count. Known parameters determine the learner before round one. Changing its updates later does not remove earlier predictors.

Table 2 assumes $K \ge 3$. The low-audit, bounded-delay, and every-round disclosure rows take $\rho > 0$. Implicit constants may depend only on fixed $\varepsilon \in (0, 1/2)$ for restricted horizons, fixed $\delta \in (0, 1/3)$ for low audits, and fixed $D \ge 1$ for bounded delay. Otherwise, they are universal. Corollary 4.5 defines κ_{rep}. Zero denominators mean $+\infty$. In the score-or-index row with $p_1 \equiv c$, $c \in [0, 1]$ and audit status remains hidden.

The low-audit and expected-delay rows retain the broader appendix statements of Corollaries W.8 and J.4. Their proper, prediction-before-query cases are Theorems 3.2 and 4.1. The expected-delay row bounds the number of rounds between deployment and reuse in conditional expectation at every round, given the target and current history. It supplies no additional observation.

Original timing means that the learner chooses prediction probabilities for all inputs before the query is revealed. Corollary 3.5 allows proper learners to see the query first. Allowing both arbitrary prediction functions and earlier knowledge of the query instead gives the upper bound $C \min\{T, \log^2(2K)\}$ for $\rho \ge 1/2$ (Appendix X). This upper bound is not a sharp rate in the table. Theorems 4.2 and 4.3 concern information revealed on every round or independently on some rounds. Theorem 4.4 treats independently disclosed noisy audit status. Corollary 4.5 instead reveals information only on replay

rounds. Theorem 5.1 fixes the first predictor while revealing the selected predictor. All of these results concern proper learners.

C Formal model and preliminaries

This appendix formalizes who observes what and when. The target is fixed before interaction and independently of the learner's private randomness. Replay uses only predictors deployed in earlier rounds, and the predictor to be replayed is selected before the current audit coin.

For probability laws P, Q, write $\mathrm{Aff}(P,Q) = \int \sqrt{dP\,dQ}$ for their Hellinger affinity and use the squared Hellinger distance $H^2(P,Q) = \int(\sqrt{dP} - \sqrt{dQ})^2 = 2 - 2\,\mathrm{Aff}(P,Q)$. These integrals are taken with respect to a common dominating measure, so $0 \le H^2(P,Q) \le 2$.

C.1 Proper randomized threshold predictors

For $a \in [K]$, recall $h_a(x) = \mathbf{1}\{x \ge a\}$. A distribution $Q \in \Delta(\mathcal{H}_K)$ gives the probability of predicting one at each input:

$$p_Q(x) = \sum_{a=1}^{K} Q(h_a) h_a(x) = \sum_{a \le x} Q(h_a).$$

Lemma C.1 (CDF representation)**.** *A function $p : \mathcal{X} \to [0,1]$ is induced by a proper randomized threshold predictor if and only if it is nondecreasing.*

Proof. Every p_Q is nondecreasing because the sets $\{a : a \le x\}$ increase with x. Conversely, let p be nondecreasing and put $p(0) = 0$. Define

$$Q(h_1) = p(1), \qquad Q(h_a) = p(a) - p(a-1) \quad (2 \le a \le K-1), \qquad Q(h_K) = 1 - p(K-1).$$

All coordinates are nonnegative and sum to one. For every $x \in [K-1]$, $\sum_{a \le x} Q(h_a) = p(x)$. Thus $p = p_Q$. □

The two endpoint cutoffs $a = 1$ and $a = K$ are the constant-one and constant-zero hypotheses. In particular, affine clipping with $\alpha \in [0, 1/2]$ preserves properness, where δ_h is a point mass at h:

$$\alpha + (1 - 2\alpha) p_Q = p_{\alpha\delta_{h_1} + (1-2\alpha)Q + \alpha\delta_{h_K}}.$$

For improper learning, the action space is $[0,1]^{K-1}$. The learner still submits a complete function before the query, and that function remains available for replay. The probability space and order of observations below are unchanged. For a finite binary class $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$, proper prediction functions belong to $\mathrm{conv}(\mathcal{H})$ and improper ones to $[0,1]^{\mathcal{X}}$. Later appendices state any additional restrictions explicitly.

C.2 Probability space and filtrations

The class size K, horizon T, and audit rate ρ are common knowledge before round one. When additional feedback information is disclosed, its disclosure probability and noise parameters are also known at that time. A finite-horizon upper bound may select a learner using these parameters before any action is deployed. Changing learning rules later does not remove earlier predictors from the history. The target $a^* \in [K]$ is selected before round one. A randomized learner uses a private random seed U independent of the target and the adversary's randomness. The adversary cannot observe this seed. This representation is equivalent to a behavioral strategy. The public filtration before round t is

$$\mathcal{F}_{t-1}^{\mathrm{pub}} = \sigma\big(p_s, x_s, Y_s : 1 \le s < t\big).$$

The learner's action p_t is measurable with respect to $\mathcal{F}_{t-1}^{\mathrm{pub}} \vee \sigma(U)$. On round one, after observing p_1, the adversary chooses x_1 and the feedback is forced to be truthful, $Y_1 = y_1^*$. For every $t \ge 2$, after observing the full action, the adversary chooses a public query x_t and a probability vector $\boldsymbol{\lambda}_t \in \Delta_{t-1}$

on $[t-1]$. These choices may depend on the fixed target and the adversary's private random seed. A source index $S_t \sim \boldsymbol{\lambda}_t$ is then sampled. Define the enlarged pre-feedback sigma-field, for $t \geq 2$,

$$\mathcal{G}_t = \sigma\big(\mathcal{F}_{t-1}^{\mathrm{pub}}, U, a^*, p_t, x_t, \boldsymbol{\lambda}_t, S_t, \text{adversary randomness used so far}\big).$$

For $t \geq 2$, conditional on $\mathcal{G}_t$, draw $A_t \sim \mathrm{Bern}(\rho)$ and $\widetilde{Y}_t \sim \mathrm{Bern}(p_{S_t}(x_t))$ independently. The audit coins are i.i.d. and independent of the target, all private seeds, and all prior randomness. Set

$$Y_t = A_t y_t^* + (1 - A_t)\widetilde{Y}_t.$$

The learner observes (x_t, Y_t) but not $A_t, \boldsymbol{\lambda}_t, S_t$, or $\widetilde{Y}_t$. Here A_t identifies the mechanism that generated the label, not whether the label happens to be correct. The selected predictor is defined even when $A_t = 1$. Our analysis conditions on information the learner need not observe: drift inequalities fix S_t, whereas the next proposition averages over its draw to compare the feedback distributions obtainable from different source choices.

Lemma C.2 (Counting errors on every round or only on audits)**.** *Recall $\ell_t = |p_t(x_t) - y_t^*|$ and define the audit-only loss, including the first round with forced truthful feedback, under the same independent audits, with their outcomes hidden, by*

$$L_T^{\mathrm{aud}} = \ell_1 + \sum_{t=2}^{T} A_t \ell_t. \tag{15}$$

For every fixed learner, adversary, and target,

$$\mathbb{E}L_T^{\mathrm{aud}} = \rho\,\mathbb{E}L_T + (1-\rho)\mathbb{E}\ell_1. \tag{16}$$

If $\mathcal{R}_T^{\mathrm{aud}}(K,\rho)$ denotes the minimax value with this objective and the same strategy classes, then

$$\rho\mathcal{R}_T(K,\rho) \leq \mathcal{R}_T^{\mathrm{aud}}(K,\rho) \leq \rho\mathcal{R}_T(K,\rho) + 1 - \rho, \qquad \mathcal{R}_T^{\mathrm{aud}}(K,\rho) = \Theta(1 + \rho\mathcal{R}_T(K,\rho)). \tag{17}$$

This comparison keeps the model fixed and changes only which rounds count toward the loss.

Proof. For $t \geq 2$, the deployment loss ℓ_t is $\mathcal{G}_t$-measurable and A_t is conditionally independent with mean ρ. Hence $\mathbb{E}[A_t\ell_t] = \rho\mathbb{E}\ell_t$, which gives Eq. (16). Since $0 \leq \ell_1 \leq 1$, taking the same infimum and supremum on the pointwise inequalities proves the first two bounds in Eq. (17). For the remaining lower bound, put a uniform prior on the two endpoint targets h_1 and h_K and fix any legal first query. Before its forced truthful observation, every first-round score has average loss $1/2$, so $\mathcal{R}_T^{\mathrm{aud}}(K,\rho) \geq 1/2$. Therefore

$$\mathcal{R}_T^{\mathrm{aud}}(K,\rho) \geq \max\{1/2, \rho\mathcal{R}_T(K,\rho)\} \geq \tfrac{1}{4}(1 + \rho\mathcal{R}_T(K,\rho)),$$

while the preceding upper bound is at most $1 + \rho\mathcal{R}_T(K,\rho)$. □

C.3 Feedback distributions from past predictors

Proposition C.3 (Attainable feedback distributions given the history)**.** *Fix a round $t \geq 2$, realized earlier prediction functions $p_1, \ldots, p_{t-1}$, and a current query x. Write*

$$\underline{p} = \min_{s<t} p_s(x), \qquad \overline{p} = \max_{s<t} p_s(x).$$

When the true label is $B = b \in \{0,1\}$, let the adversary choose any distribution $\boldsymbol{\lambda}_b$ over the actual earlier indices $[t-1]$. It draws the source $S \sim \boldsymbol{\lambda}_b$ before the independent audit coin. The learner does not observe S, so its feedback law averages over this draw. The attainable conditional means of the observed label are exactly

$$I_0 = (1-\rho)[\underline{p}, \overline{p}] \quad \textit{under } B = 0, \qquad I_1 = \rho + (1-\rho)[\underline{p}, \overline{p}] \quad \textit{under } B = 1.$$

Consequently,

$$\begin{aligned} \inf_{a \in I_0,\, b \in I_1} \mathrm{TV}(\mathrm{Bern}(a), \mathrm{Bern}(b)) &= \big[\rho - (1-\rho)(\overline{p} - \underline{p})\big]_+ \\ &= \big[2\rho - 1 + (1-\rho)(1 - \overline{p} + \underline{p})\big]_+ . \end{aligned} \tag{18}$$

If I_0 and I_1 are disjoint, the minimum is attained by an actual historical maximizer under $B = 0$ and an actual historical minimizer under $B = 1$. If they overlap, label-dependent randomization between those actual endpoint sources makes the two visible laws identical.

Moreover, there is a legal minimizing channel W^* that is Blackwell least informative within this fixed-history, fixed-query one-step family. For every other legal channel W, there is a binary stochastic kernel K_W, independent of B, such that $W^* = WK_W$.

Proof. For $b \in \{0, 1\}$, let

$$z_b = \sum_{s<t} \lambda_b(s) p_s(x).$$

After integrating out the hidden source draw, replay sample, and audit coin,

$$Y \mid B = b \sim \mathrm{Bern}(\rho b + (1 - \rho) z_b)\,.$$

Every z_b lies in $[\underline{p}, \overline{p}]$. Conversely, choose actual indices $s_{\min}, s_{\max}$ attaining the finite minimum and maximum. Randomization between these two indices realizes every point of $[\underline{p}, \overline{p}]$. This proves the two attainable intervals without introducing an exogenous source.

For Bernoulli laws, total variation is the absolute difference of the means. Since the lower endpoint of I_1 is $\rho + (1 - \rho)\underline{p}$ and the upper endpoint of I_0 is $(1 - \rho)\overline{p}$, their distance gives Eq. (18). When the distance is positive, the nearest endpoints are attained by $s_{\max}$ under $B = 0$ and $s_{\min}$ under $B = 1$.

Suppose the intervals overlap. Then $\rho < 1$ and $\overline{p} - \underline{p} \geq \rho/(1 - \rho)$. Under $B = 1$, select $s_{\min}$. Under $B = 0$, use an endpoint mixture whose mean score is $z_0 = \underline{p} + \rho/(1 - \rho) \leq \overline{p}$. Both visible means equal $c = \rho + (1 - \rho)\underline{p}$. If $\overline{p} = \underline{p}$, overlap forces $\rho = 0$ and the mixture is degenerate. Thus the legal minimizing channel is independent of B.

In the overlap case, the kernel that ignores the output of any W and returns $\mathrm{Bern}(c)$ proves Blackwell minimality. Now suppose the intervals are disjoint and set

$$\theta_0^* = (1 - \rho)\overline{p}, \qquad \theta_1^* = \rho + (1 - \rho)\underline{p}, \qquad \theta_0^* < \theta_1^*.$$

Write any other legal channel as $W(1 \mid 0) = a$ and $W(1 \mid 1) = b$. The interval constraints give $0 \leq a \leq \theta_0^* < \theta_1^* \leq b \leq 1$. Define

$$k = \frac{\theta_1^* - \theta_0^*}{b - a}, \qquad r = \theta_0^* - ka.$$

Then $0 < k \leq 1$, $r \geq \theta_0^* - a \geq 0$, and

$$1 - r - k = \frac{(b - \theta_1^*)(1 - \theta_0^*) + (\theta_0^* - a)(1 - \theta_1^*)}{b - a} \geq 0.$$

Therefore $K_W(1 \mid 0) = r$ and $K_W(1 \mid 1) = r + k$ define a stochastic kernel. Its output-one probabilities are $r + ka = \theta_0^*$ under $B = 0$ and $r + kb = \theta_1^*$ under $B = 1$. Hence $W^* = WK_W$. □

Remark C.1 (Scope). For a fixed history and query, the smallest and largest earlier scores determine this one-step comparison after the hidden source is integrated out. A sequential lower bound must additionally show how the learner produces these scores and how their later reuse affects future observations.

C.4 Bayes lower bounds for randomized learners

A lower-bound construction samples the target and all adversary randomness at time zero, independently of the learner's private seed. Suppose this same distribution gives expected loss at least R for every deterministic learner. Conditional on a randomized learner's private seed $U = u$, the learner is deterministic. Averaging the same lower bound over u gives Bayes risk at least R for every randomized learner. Since the average over targets is at least R, some fixed target and corresponding legal adversary have expected loss at least R. This is the form of Yao's principle used below: the target prior is the same for every learner seed.

C.5 A SEQUENTIAL BERNOULLI TESTING LEMMA

Lemma C.4 (Testing an initially unbiased binary label)**.** *Let $B \sim \mathrm{Bern}(1/2)$ be fixed before an h-round experiment. Before round r, an arbitrary randomized rule observes the previous bits and chooses a prediction probability $p_r \in [0,1]$. Under $B = b$, the next observation is Bernoulli with a conditional parameter $\theta_{b,r}$ that may depend on the common visible history. Suppose either*

(i) $|\theta_{1,r} - \theta_{0,r}| \le G \le 1/4$ and both parameters belong to $[1/4, 3/4]$ on every common history, or

(ii) the first action is taken before any observation and $G > 1/4$.

Then the expected deployment loss against the hidden bit satisfies

$$\frac{1}{2}\mathbb{E}_1 \sum_{r=1}^{h}(1-p_r) + \frac{1}{2}\mathbb{E}_0 \sum_{r=1}^{h} p_r \ge c_0 \min\{h, G^{-2}\} \tag{19}$$

for a universal $c_0 > 0$.

Proof. First assume case (i). Let P_b^{r-1} be the law of the visible history before round r under $B = b$. For Bernoulli parameters in $[1/4, 3/4]$,

$$D_{\mathrm{KL}}(\mathrm{Bern}(p) \| \mathrm{Bern}(q)) \le \frac{(p-q)^2}{q(1-q)} \le \frac{16}{3}(p-q)^2 \le 6G^2.$$

The sequential chain rule, including the learner's conditional action distributions, gives

$$D_{\mathrm{KL}}(\mathsf{P}_1^{r-1} \| \mathsf{P}_0^{r-1}) \le 6(r-1)G^2.$$

Pinsker's inequality implies

$$\mathrm{TV}(\mathsf{P}_1^{r-1}, \mathsf{P}_0^{r-1}) \le \sqrt{3(r-1)}G.$$

For any history-measurable $p_r \in [0,1]$,

$$\begin{aligned}\frac{1}{2}\mathbb{E}_1(1-p_r) + \frac{1}{2}\mathbb{E}_0 p_r &= \frac{1}{2}\left(1 - \mathbb{E}_1 p_r + \mathbb{E}_0 p_r\right) \\ &\ge \frac{1}{2}\left(1 - \mathrm{TV}(\mathsf{P}_1^{r-1}, \mathsf{P}_0^{r-1})\right).\end{aligned}$$

Thus every $r \le (48G^2)^{-1} + 1$ contributes at least $1/4$. Summing proves Eq. (19) after reducing the universal constant. In case (ii), averaging over the two equally likely labels gives first-round loss exactly $1/2$, while $G^{-2} < 16$. Again a smaller universal constant proves the result. □

Lemma C.5 (Identical feedback distributions)**.** *If the complete observation history has the same law under both values of a fixed, initially equally likely bit throughout h rounds, then every randomized learner has expected deployment loss at least $h/2$.*

Proof. The visible history and hence the current score have the same law in both worlds. On each round, $\frac{1}{2}\mathbb{E}_1(1-p_r) + \frac{1}{2}\mathbb{E}_0 p_r = 1/2$. □

C.6 A BALANCED-TREE HORIZON LOWER BOUND

Lemma C.6 (An unresolved target decision on each round)**.** *For every K, T, and ρ,*

$$\mathcal{R}_T(K, \rho) \ge \frac{1}{2}\min\{T, \lfloor \log_2 K \rfloor\}. \tag{20}$$

Proof. Let $d = \min\{T, \lfloor \log_2 K \rfloor\}$ and restrict the target prior to 2^d consecutive cutoff leaves, uniformly. Arrange them as a balanced binary tree. Before round $j \le d$, tell the learner which node contains the target and query the split between its two children. Revealing this extra information strengthens the learner, so a lower bound remains valid. For every round after the first, choose the fixed actual source $S_t = 1$. Given earlier child bits, all past queries, labels, selected predictors' scores, and feedback kernels are independent of the next child bit. The first forced truthful observation

likewise depends only on the first completed bit. Conditional on the preceding $j-1$ child bits and the complete visible history, the current child is therefore fair and is not revealed until after the action. The two child blocks lie on opposite sides of the query, so the target labels are one and zero. Every score $p \in [0,1]$ has fair average deployment loss $\frac{1}{2}(1-p)+\frac{1}{2}p = 1/2$. Feedback may reveal the bit after the loss is paid. Summing the d rounds proves the Bayes lower bound. The target and all d bits were sampled at time zero. □

D PROOFS WITH ADDITIONAL FEEDBACK INFORMATION

This appendix proves the audit-flag case of Theorem 4.2 and a finite-class upper bound when the selected predictor's score is revealed. We treat these separately: an audit flag identifies labels generated by the true target, whereas the selected predictor's score allows the learner to subtract the mean contribution of replay. Appendix E improves the latter bound for thresholds by deploying binary predictors. All additional information arrives after the current action, loss, and feedback label.

D.1 REVEALING WHETHER FEEDBACK CAME FROM AN AUDIT

For notational convenience set $A_1 = 1$, since the first label is forced to be truthful. On every round $t \geq 2$, the action, query, source distribution, realized source, audit coin, and replay sample are generated in the same order as in Appendix C. Only after the current loss has been incurred and Y_t has been formed is the learner additionally shown the audit indicator A_t. The indicator reports whether the label came from an audit. A replayed label may also be correct. A randomized learner may choose p_t using an independent private seed and the filtration

$$\mathcal{F}_{t-1}^{\mathrm{flag}} = \sigma\big(p_s, x_s, Y_s, A_s : 1 \leq s < t\big),$$

but it cannot use A_t to choose its current action. The distribution $\boldsymbol{\lambda}_t$ and the realized source S_t remain hidden. When $A_t = 0$, the realized replay bit equals the observed label Y_t, but the predictor that generated it is not disclosed. The adversary retains its original information filtration. The same bounds remain valid if past audit indicators are also disclosed to the adversary.

Let $\mathcal{R}_T^{\mathrm{flag}}(K,\rho)$ denote the minimax deployment loss under this information structure:

$$\mathcal{R}_T^{\mathrm{flag}}(K,\rho) = \inf_{\mathsf{Alg}\ \text{proper,randomized}} \sup_{a^*\in[K],\,\mathsf{Adv}\in\mathfrak{A}_\rho} \mathbb{E}_{\mathsf{Alg},\mathsf{Adv},a^*} L_T. \tag{21}$$

Here $\mathfrak{A}_\rho$ is the original adversary class. The learner submits its prediction function before the query, and the adversary selects the source before the current audit outcome is drawn.

We first give the potential calculation behind the upper bound.

Lemma D.1 (Audited version-space shrinkage)**.** *Let $\mathcal{V}_t$ be a finite version space containing the fixed target. Suppose the learner deploys the uniform mixture over $\mathcal{V}_t$. On a revealed audit it keeps the hypotheses agreeing with the observed true label, and on a replay round it leaves $\mathcal{V}_t$ unchanged. If ℓ_t is the deployment loss at the current query, then, for every $t \geq 2$,*

$$\rho\,\mathbb{E}\ell_t \leq \mathbb{E}\big[\log|\mathcal{V}_t| - \log|\mathcal{V}_{t+1}|\big]. \tag{22}$$

The inequality remains valid when the adversary adapts its queries and source choices using the target and past observations.

Proof. Because the action is the uniform mixture, exactly an ℓ_t fraction of $\mathcal{V}_t$ disagrees with the target at x_t. An audit therefore leaves $(1-\ell_t)|\mathcal{V}_t|$ hypotheses, while a replay round leaves the version space unchanged. Conditional on the pre-feedback sigma-field $\mathcal{G}_t$ from Appendix C, the value ℓ_t is fixed and the fresh audit coin is independent with mean ρ. Hence

$$\mathbb{E}[\log|\mathcal{V}_t| - \log|\mathcal{V}_{t+1}| \,|\, \mathcal{G}_t] = \rho[-\log(1-\ell_t)] \geq \rho\ell_t.$$

The target is always retained, so $\ell_t < 1$ and the logarithm is finite. Taking expectations proves the claim. The proof uses only the identified audits and makes no assumption about the replay labels. □

For the lower bound, we arrange that replayed labels reveal nothing about the current target bit. The learner must therefore incur loss while waiting for an audit, even though it sees all audit flags.

Lemma D.2 (Loss before an audit is observed)**.** *Let $B \sim \mathrm{Bern}(1/2)$ be fixed before a block, and let the independent audit coins have mean $\rho \in (0, 1]$. At offset r of the block, let $p_r \in [0, 1]$ be the learner's already committed score at the loss query. Suppose that, conditional on no audit in the preceding r block rounds, p_r is independent of B. Then every prefix of $s \le \lceil 1/\rho \rceil$ block rounds has expected deployment loss at least $s/4$.*

Proof. On the event that no preceding block round was audited, averaging over the two equally likely target labels gives

$$\frac{1}{2}p_r + \frac{1}{2}(1 - p_r) = \frac{1}{2}.$$

That event has probability $(1 - \rho)^r$. Loss on its complement is nonnegative, so the expected loss over the prefix is at least

$$\frac{1}{2}\sum_{r=0}^{s-1}(1 - \rho)^r.$$

Bernoulli's inequality gives $(1 - \rho)^r \ge 1 - r\rho$. Since $s \le \lceil 1/\rho \rceil$, we have $\rho(s - 1) < 1$, and consequently

$$\sum_{r=0}^{s-1}(1 - \rho)^r \ge s - \frac{\rho s(s - 1)}{2} \ge \frac{s}{2}.$$

This proves the lemma. The loss on the first audited round is included because the action and loss both precede disclosure of its audit indicator. □

D.2 Upper bound

We first give a strategy that uses the revealed truthful labels. On round one, deploy the constant score $p_1 \equiv 1/2$. This is a proper stochastic threshold, realized by placing mass $1/2$ on each of the endpoint cutoffs 1 and K. Its first loss is exactly $1/2$. After the forced true label at the adversarial query $x_1 \in [K - 1]$, initialize

$$\mathcal{V}_2 = \{a \in [K] : h_a(x_1) = Y_1\}.$$

Both labels split off a nonempty proper subset of $[K]$, so $1 \le |\mathcal{V}_2| \le K - 1$.

For every $t \ge 2$, deploy the uniform distribution over $\mathcal{V}_t$. This is one distribution over the cutoff class and hence one legal proper action chosen before the query. If $A_t = 1$, retain the cutoffs agreeing with Y_t. If $A_t = 0$, ignore Y_t. The rule uses only the transcript available to the learner and is therefore a legal deterministic learner, which is also allowed in the randomized class. Summing Lemma D.1 and using $|\mathcal{V}_{T+1}| \ge 1$ gives

$$\rho\,\mathbb{E}\sum_{t=2}^{T}\ell_t \le \mathbb{E}\log\frac{|\mathcal{V}_2|}{|\mathcal{V}_{T+1}|} \le \log(K - 1). \tag{23}$$

Thus this strategy satisfies

$$\mathbb{E}L_T \le \frac{1}{2} + \frac{\log(K - 1)}{\rho}.$$

A separate legal learner predicts one with probability $1/2$ at every input on every round. This is proper and incurs loss $T/2$ on every path. We have proved

$$\mathcal{R}_T^{\mathrm{flag}}(K, \rho) \le \min\left\{\frac{T}{2}, \frac{1}{2} + \frac{\log(K - 1)}{\rho}\right\}. \tag{24}$$

D.3 Lower bound

Fix $K \ge 3$ and write

$$d = \lfloor \log_2(K - 1) \rfloor, \qquad M = 2^d, \qquad m = \left\lceil \frac{1}{\rho} \right\rceil.$$

At time zero, draw the target cutoff uniformly from $\{2, 3, \ldots, M + 1\}$. These M consecutive cutoffs are the leaves of a complete balanced binary interval tree. The target is equivalently described by independent fair branch bits $B_1, \ldots, B_d$.

On round one, the adversary fixes $x_1 = 1$. Every target in the support has label zero there, so the forced truthful observation reveals none of the branch bits. The learner's first action was chosen without target information and is now an actual historical predictor.

For each level $j = 1, \dots, d$, allocate a block of m rounds. Conditional on the earlier branch bits, the target-containing tree node is a consecutive interval $[a_j^{\mathrm{lo}}, a_j^{\mathrm{hi}}]$ of cutoff indices of even size. Put

$$x^{(j)} = \frac{a_j^{\mathrm{lo}} + a_j^{\mathrm{hi}} - 1}{2}.$$

The left child contains cutoffs at most $x^{(j)}$ and therefore has label one at $x^{(j)}$. The right child has label zero. Thus this label is the fair bit B_j. The pivot belongs to $[K-1]$ and depends only on earlier branch bits, not on B_j. On every round in block j, the adversary queries $x^{(j)}$ and sets $\boldsymbol{\lambda}_t = \delta_1$. Hence the realized source is the actual round-one predictor, $S_t = 1$, selected before the fresh audit coin.

We verify the conditional independence needed for Lemma D.2. Fix the learner's private seed and the earlier branch bits. Before block j, every query, true label, replay draw, audit flag, and learner action depends only on those earlier bits and on randomness independent of B_j. Within the block, on the event of no preceding audit, all previously disclosed flags in the block equal zero and all feedback bits are replay samples from

$$\mathrm{Bern}\big(p_1(x^{(j)})\big).$$

Since p_1 was written before any target information and $x^{(j)}$ does not depend on B_j, this entire replay sequence is jointly independent of the current branch bit. Inductively, every subsequent randomized action based on that common transcript is also independent of B_j. The score extracted from the committed action at $x^{(j)}$ therefore satisfies the hypothesis of Lemma D.2.

This argument is unaffected if the first query of the next block discloses an earlier branch through its location. The learner has already written its action on that round, and the new query depends only on earlier bits. After conditioning on all of them, the next branch bit remains independent and fair. This gives the required induction across blocks.

Apply Lemma D.2 to every complete block and to the final block prefix, if the horizon truncates one. After all d blocks, the adversary may continue legally with query $x = 1$ and actual source $S_t = 1$. The nonnegative loss on those remaining rounds is ignored. The $T-1$ rounds after the first therefore contribute at least

$$\mathbb{E}L_T \geq \frac{1}{4}\min\{T-1, dm\} \geq \frac{1}{4}\min\left\{T-1, \frac{d}{\rho}\right\}. \tag{25}$$

The construction is a valid distribution for a lower bound against randomized learners. The target and all branch bits are drawn at time zero. Conditional on a learner's independent private seed, its strategy is deterministic and the argument above is unchanged. Averaging over the seed preserves Eq. (25). Equivalently, the prior is chosen at time zero over fixed targets and legal adversaries that know those targets. Every such adversary uses only actual source $S_t = 1$. Since the average risk has the displayed lower bound, some fixed target and its corresponding legal adversary attain at least that risk.

There is also a universal first-round lower bound. Draw the target uniformly from the two endpoint cutoffs 1 and K, and query $x_1 = 1$. The target label is a fair bit before the learner acts, so every first score p has average loss

$$\frac{1}{2}p + \frac{1}{2}(1-p) = \frac{1}{2}.$$

Combining the two valid Bayes lower bounds yields

$$\mathcal{R}_T^{\mathrm{flag}}(K, \rho) \geq \max\left\{\frac{1}{2}, \frac{1}{4}\min\left\{T-1, \frac{d}{\rho}\right\}\right\}. \tag{26}$$

For $K \geq 3$, one has $d = \Theta(\log(K-1)) = \Theta(\log K)$. The independent $1/2$ term absorbs the $T = 1$ endpoint and the additive constant lost in replacing $T-1$ by T. Equations (24) and (26) therefore give

$$\mathcal{R}_T^{\mathrm{flag}}(K, \rho) = \Theta\left(\min\left\{T, 1 + \frac{\log(K-1)}{\rho}\right\}\right) = \Theta\left(\min\left\{T, \frac{\log K}{\rho}\right\}\right).$$

All constants here are universal.

It remains to verify the exceptional one-point domain. When $K = 2$, the only query is $x = 1$. The learner deploys $p_1(1) = 1/2$, incurs loss $1/2$, and the forced truthful label identifies the target cutoff exactly. It deploys that target thereafter and incurs no further loss. The endpoint prior above forces first-round loss $1/2$, so

$$\mathcal{R}_T^{\mathrm{flag}}(2, \rho) = \frac{1}{2} \qquad (K = 2). \tag{27}$$

Together with Eqs. (24) and (26), this proves the audit-status part of Theorem 4.2 for every $K \geq 3$, $T \geq 1$, and $\rho \in (0, 1]$. The $K = 2$ argument does not use the audit coin and therefore also covers $\rho = 0$.

At $\rho = 1/2$ and $K \geq 3$, the theorem specializes to $\mathcal{R}_T^{\mathrm{flag}}(K, 1/2) = \Theta(\min\{T, \log K\})$. For $\log K \leq T$ in the horizon range of Theorem 3.1, this is $\Theta(\log K)$, while the value when neither the source nor audit status is revealed is $\Theta_\varepsilon(T^{2/3}(\log K)^{1/3})$. Revealing the audit indicator therefore removes the critical polynomial penalty without revealing the historical source.

D.4 Revealing the selected predictor's score or index

We now hide the audit flag and instead reveal information about the predictor selected for replay. This information arrives after the loss and label on every round after the first, including audited rounds. If only the score is disclosed, the learner receives

$$z_t = p_{S_t}(x_t).$$

If the index is disclosed, it receives S_t and can therefore reconstruct z_t from its own past actions. The source distribution $\boldsymbol{\lambda}_t$ remains hidden. Write $\mathcal{R}_T^{\mathrm{src}}(\mathcal{H}, \rho)$ and $\mathcal{R}_T^{\mathrm{score}}(\mathcal{H}, \rho)$ for the proper minimax values with index and score disclosure, respectively. A target-dependent index can carry information beyond the score, so

$$\mathcal{R}_T^{\mathrm{src}}(\mathcal{H}, \rho) \leq \mathcal{R}_T^{\mathrm{score}}(\mathcal{H}, \rho).$$

The learner below uses only z_t, the selected predictor's probability of predicting one. It therefore applies to both variants.

Theorem D.3 (Finite-class upper bound with the selected predictor's score)**.** *Let $\mathcal{H} \subseteq \{0, 1\}^{\mathcal{X}}$ be finite, and suppose a distribution ν over $\mathcal{H}$ gives prediction probability $1/2$ at every input. Fix $\rho \in (0, 1]$ and a prior $\pi \in \Delta(\mathcal{H})$. There is a proper learner such that, for every target $h^* \in \mathrm{supp}(\pi)$ and every legal adversary,*

$$\mathbb{E}L_T \leq \frac{1}{2} + \frac{16 \log(1/\pi(h^*))}{\rho^2}. \tag{28}$$

Consequently,

$$\mathcal{R}_T^{\mathrm{src}}(\mathcal{H}, \rho) \leq \mathcal{R}_T^{\mathrm{score}}(\mathcal{H}, \rho) \leq \min\left\{\frac{T}{2}, \frac{1}{2} + \frac{16 \log |\mathcal{H}|}{\rho^2}\right\}. \tag{29}$$

Proof. Deploy the probability-$1/2$ predictor given by ν on round one and initialize $\boldsymbol{w}_2 = \pi$. For $t \geq 2$, deploy the mixture with weights $\boldsymbol{w}_t$, with score $r_t(x) = \mathbb{E}_{h \sim \boldsymbol{w}_t} h(x)$. After observing (x_t, Y_t, z_t), set

$$c_t = (1 - \rho) z_t + \frac{\rho}{2}, \qquad s_t = Y_t - c_t, \qquad \eta = \frac{\rho}{8}. \tag{30}$$

With $v_{t,h} = 2h(x_t) - 1$, update

$$w_{t+1}(h) = \frac{w_t(h) e^{\eta v_{t,h} s_t}}{\mathbb{E}_{g \sim \boldsymbol{w}_t} e^{\eta v_{t,g} s_t}}. \tag{31}$$

Fix a target h^* and abbreviate $y_t^* = h^*(x_t)$, $r_t = r_t(x_t)$, $e_t = |r_t - y_t^*|$, and $\bar{v}_t = 2r_t - 1$. Conditional on $\mathcal{G}_t$, the selected predictor's score z_t is fixed and

$$\mathbb{E}[s_t \mid \mathcal{G}_t] = \rho\left(y_t^* - \frac{1}{2}\right).$$

If $y_t^* = 1$, then $v_{t,h^*} - \bar{v}_t = 2e_t$. If $y_t^* = 0$, the difference is $-2e_t$. Therefore

$$\mathbb{E}[(v_{t,h^*} - \bar{v}_t) s_t \mid \mathcal{G}_t] = \rho e_t. \tag{32}$$

The center c_t belongs to $[\rho/2, 1-\rho/2]$, so $|s_t| \le 1$. For an auxiliary $h \sim \boldsymbol{w}_t$, let $X_t = (v_{t,h} - \bar{v}_t)s_t$. Conditional on the realized s_t, one has $\mathbb{E}_{\boldsymbol{w}_t} X_t = 0$, $|X_t| \le 2$, and

$$\mathbb{E}_{\boldsymbol{w}_t} X_t^2 \le \mathrm{Var}_{\boldsymbol{w}_t}(v_{t,h}) = 4r_t(1-r_t) \le 4e_t.$$

Since $|\eta X_t| \le 1/4$, the same exponential inequality used in Appendix M gives

$$\log \mathbb{E}_{\boldsymbol{w}_t} e^{\eta X_t} \le 4\eta^2 e_t.$$

For $\Phi_t = \log(1/w_t(h^*))$, Eqs. (32) and (31) imply

$$\mathbb{E}[\Phi_t - \Phi_{t+1} \mid \mathcal{G}_t] \ge (\eta\rho - 4\eta^2)e_t = \frac{\rho^2}{16} e_t.$$

Summing gives $\sum_{t=2}^{T} \mathbb{E}e_t \le 16\log(1/\pi(h^*))/\rho^2$. The first predictor has loss exactly $1/2$, proving Eq. (28). Taking π uniform and comparing with the separate strategy that always predicts with probability $1/2$ proves Eq. (29). □

D.5 Lower bound when audits and replay are equally likely

Lemma D.4 (Unresolved target bits with the selected predictor's score or index)**.** *For both disclosure choices and every K, T, ρ,*

$$\mathcal{R}_T^{\mathrm{score}}(K,\rho) \ge \frac{1}{2}\min\{T, \lfloor \log_2 K \rfloor\}, \qquad \mathcal{R}_T^{\mathrm{src}}(K,\rho) \ge \frac{1}{2}\min\{T, \lfloor \log_2 K \rfloor\}. \tag{33}$$

Proof. Use the balanced-tree target prior from Lemma C.6. At each of the first $d = \min\{T, \lfloor \log_2 K \rfloor\}$ rounds, announce the split of the current target-containing block. Conditional on all earlier branch bits and on the complete visible transcript, the current branch bit remains fair before the learner acts. For $t \ge 2$, choose the fixed actual source $S_t = 1$. Its disclosed identity is constant, while its score $p_1(x_t)$ depends only on the already revealed pivot and the round-one action, not on the current branch bit. Thus neither disclosed quantity changes the equal probabilities of the two labels before the action, so the expected deployment loss is $1/2$. Feedback may reveal the bit after the loss is paid. Summing the d rounds proves the Bayes lower bound. The target and all branch bits are sampled at time zero. □

At $\rho = 1/2$, Theorem D.3 and Lemma D.4 give, for every $K \ge 3$,

$$\mathcal{R}_T^{\mathrm{score}}(K, 1/2) = \Theta(\min\{T, \log K\}), \qquad \mathcal{R}_T^{\mathrm{src}}(K, 1/2) = \Theta(\min\{T, \log K\}).$$

For $K = 2$, the round-one truthful label identifies the only unknown cutoff after an unavoidable loss of $1/2$, so both exact values are $1/2$. This gives Theorem 4.2 at audit rate $1/2$. Appendix E proves the sharper dependence on the audit rate.

E Binary predictions and the effect of the first predictor

Knowing the selected predictor's score permits a stronger bound than Theorem D.3 when the learner controls its predictions from round one. For thresholds, the learner can maintain internal weights while deploying their deterministic weighted majority over the entire domain. Every predictor available for later replay then has binary scores. Replaying one of these predictors adds no sampling noise at a fixed input, which improves the dependence on the audit rate.

Theorem E.1 (Upper bound from deploying binary predictors)**.** *Let $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ be a finite binary class closed under the pointwise weighted-majority operation $h_{\boldsymbol{w}}(x) = \mathbf{1}\{\mathbb{E}_{h\sim\boldsymbol{w}} h(x) \ge 1/2\}$ for $\boldsymbol{w} \in \Delta(\mathcal{H})$. Let π be a full-support prior, $\rho > 0$, and $q > 0$. If, independently with probability q, the selected predictor's score, its index, or the bit indicating whether its score is at least $1/2$ is disclosed after the loss, a proper learner satisfies, for every fixed target h^*,*

$$\mathbb{E}L_T \le 1 + \frac{32\log(1/\pi(h^*))}{q\rho}. \tag{34}$$

The same bound holds for an arbitrary finite binary class if improper deterministic actions $h_{\boldsymbol{w}}$ are allowed and replay samples the learner's actual earlier actions. A disclosed score with pointwise error strictly less than $1/2$ also suffices in this algorithm.

Proof. Keep weights $\boldsymbol{w}_t$, but deploy $p_t = h_{\boldsymbol{w}_t}$, rather than their mixture. Set $\boldsymbol{w}_1 = \boldsymbol{w}_2 = \pi$, ignoring the first forced true label for this bound. The first loss is at most one. Since deployment is binary from round one, every earlier predictor's score $z_t = p_{S_t}(x_t)$ belongs to $\{0, 1\}$. Thus the indicator $\mathbf{1}\{z_t \geq 1/2\}$ equals z_t, and any score report with error less than $1/2$ recovers it by rounding. An index also determines it from the learner's past actions.

For $t \geq 2$, update only when source information is disclosed, using

$$s_t = Y_t - (1-\rho)z_t - \rho/2, \qquad \eta = 1/8, \qquad w_{t+1}(h) \propto w_t(h) e^{\eta(2h(x_t)-1)s_t}.$$

Condition on the fixed target and the pre-feedback history, including the realized source but excluding the independent audit and disclosure coins. Write $y_t^* = h^*(x_t)$, $r = \mathbb{E}_{\boldsymbol{w}_t} h(x_t)$, $e = |r - y_t^*|$, $v_h = 2h(x_t) - 1$, and $\bar{v} = 2r - 1$. The deployed loss is $\ell_t = \mathbf{1}\{h_{\boldsymbol{w}_t}(x_t) \neq y_t^*\} \leq 2e$. The mean and second moment of the correction satisfy

$$\mathbb{E}s_t = \rho(y_t^* - 1/2), \qquad \mathbb{E}s_t^2 = \begin{cases} \rho^2/4, & y_t^* = z_t, \\ \rho - 3\rho^2/4, & y_t^* \neq z_t, \end{cases} \leq \rho. \tag{35}$$

Indeed, if $y_t^* = z_t$, the observed label is that same binary value. If $y_t^* \neq z_t$, it equals y_t^* with probability ρ and z_t otherwise. This calculation remains valid when source selection uses the target, because selection precedes the fresh audit coin.

For $X_h = (v_h - \bar{v})s_t$, one has $\mathbb{E}_{\boldsymbol{w}_t} X_h = 0$, $|\eta X_h| \leq 1/4$, and $\mathbb{E}_{\boldsymbol{w}_t} X_h^2 = 4r(1-r)s_t^2 \leq 4es_t^2$. The inequality $e^u \leq 1 + u + u^2$ for $|u| \leq 1/4$ and $\log(1+v) \leq v$ give

$$\log \mathbb{E}_{\boldsymbol{w}_t} e^{\eta X_h} \leq 4\eta^2 e s_t^2.$$

For $\Phi_t = \log(1/w_t(h^*))$, the target term has mean $\mathbb{E}[(v_{h^*} - \bar{v})s_t] = \rho e$. Consequently an update with disclosed source information has expected potential decrement at least $(\eta - 4\eta^2)\rho e = \rho e/16$. Independence of disclosure multiplies this by q. Telescoping and using $\Phi_{T+1} \geq 0$ gives $\mathbb{E}\sum_{t=2}^{T} \ell_t \leq 32\Phi_2/(q\rho)$. This proves the bound. The argument uses closure only to ensure that the deployed weighted-majority predictor belongs to the class. □

For threshold weights, $r_t(x) = \sum_{a \leq x} w_t(a)$ is a CDF and $h_{\boldsymbol{w}_t}$ is exactly the threshold at its lower weighted median. Thus Theorem E.1 is a proper threshold algorithm committed before the adversarial query. For uniform weights, comparison with a separate learner that always predicts one with probability $1/2$ gives

$$\mathcal{R}_T^{\mathsf{I},q}(K, \rho) \leq \min\{T/2,\, 1 + 32 \log K/(q\rho)\} \tag{36}$$

for disclosure of the score, index, or whether the selected predictor's score is at least $1/2$. This minimax comparison chooses one learner at time zero, before any predictions have been made. For a general class and a nonuniform prior, Eq. (34) itself is the targetwise statement, without an additional targetwise $T/2$ cap.

E.1 Information on every round at each positive audit rate

The lower-bound construction in Appendix D, which always selects the first predictor, also permits joint disclosure of (A_t, S_t, z_t) on every round after the first. The adversary always selects the first predictor, so its index is one and its score is determined by the first action and the current query. Neither depends on the next fair branch bit. Before the first audit at that level, all labels are replays of this same first action. The independence argument of Lemma D.2 is therefore unchanged by joint disclosure. In particular, for $K \geq 3$,

$$\mathcal{R}_T^{\mathrm{joint},1}(K, \rho) \geq c \min\{T, \log K/\rho\}.$$

Joint disclosure supplies each of these quantities. Combining this lower bound with Eq. (36) and the flag upper bound proves Theorem 4.2 for every $\rho > 0$, including disclosure of whether the selected predictor's score is at least $1/2$. The one-point domain $K = 2$ has value $1/2$ by the forced truthful first label. None of these results needs the horizon restriction in Theorem 3.1.

E.2 Proof of Theorem 5.1

We prove Theorem 5.1 with

$$v_c = \rho + (1-\rho)c(1-c), \qquad \kappa_c = \rho^2/v_c.$$

Only the complete first predictor $p_1 \equiv c$ is fixed. All later proper actions remain unrestricted. On every round after the first, the learner observes the exact score or index of the predictor selected for replay, including when the label is audited. The audit flag stays hidden, and the first predictor remains available for replay throughout.

Upper bound. Deploy $p_1 \equiv c$, with loss at most one, and use the binary weighted-majority learner from round two, starting from a uniform internal prior. Now the selected predictor's score always belongs to $\{0, c, 1\}$. Given the target label y_t^* and the selected predictor's score z, put $s = Y - (1-\rho)z - \rho/2$. Then

$$\begin{aligned} \mathbb{E}s &= \rho(y_t^* - 1/2), \\ \mathbb{E}s^2 &= (1-\rho)z(1-z) + \rho(1-\rho)(y_t^* - z)^2 + \rho^2/4 \le v_c. \end{aligned} \tag{37}$$

The identity follows by conditioning on the audit coin and applying the law of total variance, then adding $(\mathbb{E}s)^2 = \rho^2/4$. Thus v_c is a uniform second-moment bound, not a variance identity. For these three possible values of the selected predictor's score, $z(1-z) \le c(1-c)$, and $\rho(1-\rho) + \rho^2/4 \le \rho$. Also $|s| \le 1$ and $v_c \ge \rho$. Use the same centered update as in Theorem E.1 with $\eta = \rho/(8v_c) \le 1/8$. Its expected remainder is at most $4\eta^2 v_c e_t$, where e_t is internal error, and the target term is $\eta\rho e_t$. Hence the expected potential decrement is at least $\kappa_c e_t/16$. The deployed median loses at most $2e_t$, yielding

$$\mathbb{E}L_T \le 1 + 32 v_c \log K/\rho^2. \tag{38}$$

The same learner has $L_T \le T$ on every path. Combining this with Eq. (38) gives $\mathbb{E}L_T \le \min\{T, 1 + 32 v_c \log K/\rho^2\}$. Since $K \ge 3$ and $v_c \ge \rho$, the additive constant is absorbed into $v_c \log K/\rho^2$, proving the claimed upper order. The first constraint need not permit an exact $T/2$ cap. Thus the free-first-action identity $\mathcal{R}_T(K,\rho) = T/2$ for $T \le \lfloor \log_2 K \rfloor$ in Lemma P.1 does not in general transfer to the constrained game.

Lower bound by replaying the first predictor. At time zero choose the target uniformly among cutoffs $2, \dots, 2^d + 1$, where $d = \lfloor \log_2(K-1) \rfloor$, independently of the learner's private seed. The first query is $x_1 = 1$, so its forced true label is zero and reveals no descendant bit. Use actual source $S_t = 1$ thereafter. At each balanced descendant pivot, the current fair bit B has channel

$$Y_t \mid B \sim \mathrm{Bern}((1-\rho)c + \rho B). \tag{39}$$

The revealed index and score are always 1 and c and provide no additional information. Before each new level is queried, its bit is fair conditional on all completed bits, the previous observations, and the learner's private seed.

Set $a = (1-\rho)c$, $b = a + \rho$. Directly, with the continuous convention at zero numerator and denominator,

$$\begin{aligned} H^2(\mathrm{Bern}(a), \mathrm{Bern}(b)) &= \rho^2 \left\{ \frac{1}{(\sqrt{a} + \sqrt{b})^2} + \frac{1}{(\sqrt{1-a} + \sqrt{1-b})^2} \right\} \\ &\le \frac{2\rho^2}{(a+b)(2-a-b)}. \end{aligned} \tag{40}$$

For $0 < \rho \le 3/4$, the denominator is

$$4(1-\rho)^2 c(1-c) + \rho(2-\rho) \ge v_c.$$

For $\rho > 3/4$, use $H^2 \le 2$ and $v_c \le 1$ instead. Thus in all cases the squared distance is at most $4\kappa_c$. The Hellinger bound also covers $c = 0, 1$, where a KL calculation requiring probabilities strictly between zero and one would not apply.

Repeat each pivot for $m = \lceil \kappa_c^{-1} \rceil$ rounds, truncating the last block to the remaining horizon, and reveal the completed bit before proceeding to its child. Lemma G.1 gives $c_0 \min\{s, \kappa_c^{-1}\} \ge c_1 s$

expected loss on every prefix of length $s \le m$, since $\kappa_c \le 1$. After the d levels, fill any remaining rounds with $x = 1, S = 1$. The resulting post-start lower bound is $c_2 \min\{T-1, dv_c/\rho^2\}$. Separately, the minimax first-round loss for the prescribed constant is $\max\{c, 1-c\} \ge 1/2$. Taking the larger of these two valid minimax lower bounds handles $T = 1$ and absorbs the initial additive constant. Since $d \asymp \log K$ for $K \ge 3$, this proves Eq. (14) at every horizon.

Flags and the one-point domain. With a disclosed audit flag, the version-space learner of Appendix D can use any fixed first action. Only its first-loss constant changes to at most one. The lower bound obtained by replaying the first predictor while waiting for an audit allows joint disclosure and remains valid under this restriction. It gives the order $\min\{T, \log K/\rho\}$ for every c. For $K = 2$, the forced first label identifies the whole target and the exact constrained value is $\max\{c, 1-c\}$. For $c = 1/2$, this equals $1/2$.

A first predictor with probability one-half. At $c = 1/2$, $v_c = (1+3\rho)/4 \asymp 1$, so the law becomes

$$\mathcal{R}_T^{\mathrm{I,init}(1/2)}(K, \rho) = \Theta(\min\{T, \log K/\rho^2\}). \tag{41}$$

For the lower-bound choice $S_t = 1$ on $t \ge 2$, the feedback law is $Y_t \mid y_t^* \sim \mathrm{Bern}((1-\rho)/2 + \rho y_t^*)$. At $c = 0, 1$, $v_c = \rho$, recovering inverse-linear audit dependence. More generally, when the time cap is inactive, the ratio to the inverse-linear scale is exactly of order $1+(1-\rho)c(1-c)/\rho$. The first predictor can be sampled repeatedly at later queries. This ratio measures the resulting cumulative cost, rather than its one-round prediction loss. The indicator $\mathbf{1}\{z_t \ge 1/2\}$ recovers z_t when all earlier predictions are binary, as in Theorem E.1. Here it need not distinguish c from zero or one. The theorem therefore explicitly assumes disclosure of the selected predictor's exact score or index. Output masking to improve future information has a direct precedent in recursive distribution estimation (Kanabar & Gastpar, 2026b, Sections III-B and IV-B). The contribution here is the proper cumulative-loss upper bound and matching all-horizon dependence on an initial predictor that remains legally reusable forever.

F Approximate selected-predictor scores or known selection probabilities

We prove Proposition F.2 by correcting the mean contribution of replay in two ways. One uses an approximate score of the selected predictor. The other averages the scores of past predictors under their actual conditional selection probabilities. Both yield upper bounds for any finite binary class admitting a proper predictor that assigns probability $1/2$ at every input. Neither argument uses the ordering of thresholds.

Lemma F.1 (A generic finite-class update bound)**.** *Let $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ be finite and admit a proper predictor with probability $1/2$ at every input. Let π be a prior positive at the fixed target h^*, and let $\gamma \in (0, 1]$. Suppose that, after each prediction from round two onward, the learner can compute $s_t \in [-1, 1]$ satisfying*

$$(2y_t^* - 1)\mathbb{E}[s_t \mid \mathcal{J}_t] \ge \gamma/2,$$

where $\mathcal{J}_t$ contains the target, current action, and query, but no fresh randomness used to form s_t. Then there is a proper learner with

$$\mathbb{E}L_T \le \frac{1}{2} + \frac{16\log(1/\pi(h^*))}{\gamma^2}.$$

If such a signal is available only on independent Bernoulli-q disclosure rounds and the condition holds there, the denominator becomes $q\gamma^2$. The disclosure coin is drawn after the action, query, and source selection, and is independent of all their randomness and the feedback randomness.

Proof. Deploy the probability-$1/2$ predictor on round one and initialize $\boldsymbol{w}_2 = \pi$. Thereafter deploy the mixture with weights $\boldsymbol{w}_t$, put $v_{t,h} = 2h(x_t) - 1$, and update by

$$w_{t+1}(h) \propto w_t(h)\exp\{\eta v_{t,h} s_t\}, \qquad \eta = \gamma/8.$$

Write $r_t = \mathbb{E}_{\boldsymbol{w}_t} h(x_t)$, $e_t = |r_t - y_t^*|$, $\bar{v}_t = 2r_t - 1$, and $\Phi_t = \log(1/w_t(h^*))$. The signal condition gives

$$\mathbb{E}[(v_{t,h^*} - \bar{v}_t)s_t \mid \mathcal{J}_t] \ge \gamma e_t.$$

For fixed realized s_t, the centered variable $X = (v_{t,h} - \bar{v}_t)s_t$, $h \sim \boldsymbol{w}_t$, has mean zero, magnitude at most two, and second moment at most $4r_t(1-r_t) \le 4e_t$. Since $|\eta X| \le 1/4$, $e^u \le 1 + u + u^2$ yields

$$\log \mathbb{E}_{\boldsymbol{w}_t} e^{\eta X} \le 4\eta^2 e_t.$$

Thus $\mathbb{E}[\Phi_t - \Phi_{t+1} \mid \mathcal{J}_t] \ge (\eta\gamma - 4\eta^2)e_t = \gamma^2 e_t/16$. Telescope and use $\Phi_{T+1} \ge 0$. When information is disclosed only on some rounds, leave $\boldsymbol{w}_t$ unchanged whenever the signal is unavailable. Independence multiplies both the drift and the remainder by q, so the same calculation gives $q\gamma^2 e_t/16$. □

Proposition F.2 (Approximate selected-predictor scores and known selection probabilities)**.** *For $\rho > 0$, the following guarantees hold for every horizon. If a reported score $\widetilde{z}_t \in [0,1]$ has error at most ε_z and $\gamma = \rho - 2(1-\rho)\varepsilon_z > 0$, then*

$$\mathbb{E}L_T \le \frac{1}{2} + 16\log K/\gamma^2. \tag{42}$$

If the actual conditional source distribution $\boldsymbol{\lambda}_t$ is known and its mean score $\bar{z}_t = \sum_{s<t} \lambda_t(s)p_s(x_t)$ can be computed, then the same bound holds with $\gamma = \rho$, even though A_t and S_t remain hidden. Both guarantees extend to finite binary classes admitting the same probability-$1/2$ predictor, with $\log|\mathcal{H}|$ in place of $\log K$.

The score report is formed before the audit and feedback coins and revealed after prediction. Its error guarantee includes audited rounds, since it describes the predictor selected for replay. Earlier predictors may have any scores in $[0,1]$. If they are binary, score error strictly below $1/2$ instead permits exact rounding at every $\rho > 0$. The second guarantee requires the actual conditional probabilities of selecting each source, as specified below. Marginal frequencies alone do not suffice.

F.1 Approximate scores and coarse quantization

The selected predictor's score $z_t = p_{S_t}(x_t)$ is fixed before the audit and replay coins. A score report $\widetilde{z}_t \in [0,1]$ is also fixed then, possibly using its own randomness, and is disclosed only after prediction. Assume $|z_t - \widetilde{z}_t| \le \varepsilon_z$ almost surely. This requirement concerns the selected source even on audited rounds. It excludes a report chosen after inspecting the realized feedback. With

$$s_t = Y_t - (1-\rho)\widetilde{z}_t - \rho/2, \qquad \gamma = \rho - 2(1-\rho)\varepsilon_z > 0,$$

one has $|s_t| \le 1$ and, conditional on the target, source, and report,

$$\mathbb{E}[s_t \mid \mathcal{J}_t] = \rho(y_t^* - 1/2) + (1-\rho)(z_t - \widetilde{z}_t).$$

Multiplication by $2y_t^* - 1$ lower-bounds this by $\gamma/2$. Lemma F.1 proves

$$\mathbb{E}L_T \le \frac{1}{2} + \frac{16\log K}{[\rho - 2(1-\rho)\varepsilon_z]^2}. \tag{43}$$

For $b \ge 1$ bits, partition $[0,1]$ into 2^b equal-width bins (assign the endpoint one to the last bin) and report the bin index. Midpoint reconstruction has error at most 2^{-b-1}, so $\gamma = \rho - (1-\rho)2^{-b}$. This is a sufficient signal condition, not a converse for arbitrary quantization protocols. At $\rho = 1/2$, one bit $\mathbf{1}\{z_t \ge 1/2\}$ gives $\varepsilon_z = 1/4$, $\gamma = 1/4$, and $\mathbb{E}L_T \le 1/2 + 256\log K$. An independent disclosure probability $q > 0$ changes this to $1/2 + 256\log K/q$. The lower bound in Lemma D.4, based on unresolved target decisions, also holds for this coarser report.

F.2 Known conditional selection probabilities

Let $\boldsymbol{\lambda}_t$ be the actual conditional distribution of S_t given the target, history, current action, and query, before the source is sampled. Assume the learner receives this vector with feedback and can use it before updating. The vector need not be disclosed before the current action or query. In particular, the rule may be a known uniform, recency-weighted, or last-predictor kernel. A public marginal distribution is insufficient if conditioning on the target changes the actual source law. The learner computes

$$\bar{z}_t = \sum_{s<t} \lambda_t(s)p_s(x_t).$$

Here neither S_t nor A_t is disclosed. Conditioning before the source draw, the audit remains independent and

$$\mathbb{E}[Y_t \mid \mathcal{J}_t] = \rho y_t^* + (1-\rho)\bar{z}_t.$$

Consequently $s_t = Y_t - (1-\rho)\bar{z}_t - \rho/2$ satisfies the condition in Lemma F.1 with $\gamma = \rho$, proving

$$\mathbb{E}L_T \le \frac{1}{2} + \frac{16 \log K}{\rho^2}. \tag{44}$$

This upper bound assumes the learner can compute the conditional mean. Direct computation evaluates all $t-1$ earlier predictors on round t. Faster evaluation requires additional structure in the source distribution and in how the earlier predictors are represented.

G SHARP RATES WHEN INFORMATION IS REVEALED INDEPENDENTLY

This appendix proves Theorem 4.3 when additional feedback information is revealed with probability q independently on each round after the first. The information can identify an audit, give the selected predictor's index or score, or report whether that score is at least one-half. We also allow independently flipped audit flags. Appendix H treats general finite channels for observing audit status and information revealed only on replay rounds. The upper bounds use the additional observations to correct updates. The lower bounds account for everything the learner sees, including observations that identify the current target bit exactly. Targets are fixed at time zero, and the adversary selects actual earlier predictors before the independent audit coin.

G.1 WHAT IS REVEALED AND WHEN

Fix a known $q \in [0,1]$. The first round is unchanged and has no replay source. For $t \ge 2$, first generate the current action p_t, query x_t, source distribution $\boldsymbol{\lambda}_t$, realized source S_t, audit coin A_t, replay sample, and visible label Y_t, in the order of Appendix C. After these variables and the deployment loss have been determined, generate a fresh $E_t \sim \mathrm{Bern}(q)$, independent of the target, all private seeds, and all other randomness. The learner receives (x_t, Y_t, E_t) and, if $E_t = 1$, the additional quantity O_t. Equivalently, the independent coins E_t can be sampled at time zero and kept inaccessible until their respective feedback times. They are not available when selecting the current action, query, or source.

The choices of O_t considered here are A_t, $z_t = p_{S_t}(x_t)$, S_t, $b_t = \mathbf{1}\{z_t \ge 1/2\}$, and the joint observation (A_t, S_t, z_t). Scores are defined and can be disclosed also on audited rounds, because a source was already selected before the audit coin. The learner's filtration contains its past actions and all past revealed feedback information. We may also give these observations to the adversary. This strengthening does not affect the proof: the next audit and disclosure coins remain independent of the enlarged past. No present or future disclosure decision is available to the adversary in advance.

With noisy audit flags, a disclosure instead supplies only $W_t = A_t \oplus N_t$, where $N_t \sim \mathrm{Bern}(\xi)$ is fresh and independent, with known $\xi \in [0, 1/2]$. Its noise coins, like the disclosure coins, are unavailable to the learner and adversary before feedback.

Revealing the audit flag together with the selected predictor's index and score supplies at least as much information as revealing any one quantity alone. We prove the lower bound with all three quantities visible and an upper bound for each individual choice. The noisy-flag game has its own matching lower bound. For a fixed information type I, the notation $\mathcal{R}_T^{\mathsf{I},q}(K,\rho)$ means that O_t is revealed independently with probability q on each round after the first. The total number of these observations is not fixed.

G.2 UPPER BOUNDS, INCLUDING NOISY AUDIT FLAGS

For perfect flags, shrink a uniform version space only when the round is audited and its flag is disclosed. Lemma D.1 gives

$$\mathcal{R}_T^{\mathrm{flag},q}(K,\rho) \le \min\{T/2,\ 1/2 + \log(K-1)/(q\rho)\}. \tag{45}$$

For the selected predictor's score, index, or the indicator that its score is at least one-half, the learner that deploys binary predictors in Theorem E.1 gives $1 + 32 \log K/(q\rho)$. If all three quantities are revealed, it can use the audit flag. These guarantees apply at every positive audit rate.

We also allow the disclosed audit flag to be corrupted: $W_t = A_t \oplus N_t$, where $N_t \sim \mathrm{Bern}(\xi)$ is fresh, independent of all other variables, and $\xi \in [0, 1/2]$ is known. Only W_t is disclosed in this variant. The learner receives no separate exact flag or selected predictor's index or score. For $\xi < 1/2$ put

$$a = 1 - 2\xi, \qquad \nu = \frac{\xi(1-\xi)}{a^2}, \qquad \kappa = \frac{\rho^2}{\rho + \nu}. \tag{46}$$

Set $\kappa = 0$ if $\rho = 0$ or $\xi = 1/2$. For any finite binary class with prior π, deploy its weighted mixture and, when W_t is revealed, use

$$s_t = \frac{W_t - \xi}{a}(Y_t - 1/2), \qquad \eta = \frac{\rho}{4(\rho + \nu)}, \qquad w_{t+1}(h) \propto w_t(h) e^{\eta(2h(x_t)-1)s_t}. \tag{47}$$

Do not update when W_t is withheld. The cases $q\kappa = 0$ will use the learner for hidden replay instead. The first action can be the prior mixture, with loss at most one, and $\boldsymbol{w}_2 = \pi$. If the class admits a proper predictor with probability one-half at every input, that predictor instead gives first loss exactly $1/2$.

Condition on the target and pre-feedback history, including the selected source. Since $\mathbb{E}[(W_t - \xi)/a \mid A_t] = A_t$, and the flag noise is independent of the label given A_t, the weighted centered label satisfies

$$\mathbb{E}s_t = \rho(y_t^* - 1/2), \qquad \mathbb{E}s_t^2 = (\rho + \nu)/4. \tag{48}$$

The second equality uses $(Y_t - 1/2)^2 = 1/4$ and $\mathbb{E}[((W_t - \xi)/a)^2 \mid A_t] = A_t + \nu$. Let r_t be the mixture score and $e_t = |r_t - y_t^*|$. For $v_h = 2h(x_t) - 1$, $\bar{v} = 2r_t - 1$, and $X_h = (v_h - \bar{v})s_t$, the target term has mean ρe_t, and $\mathbb{E}_{\boldsymbol{w}_t} X_h^2 \le 4e_t s_t^2$. Moreover $|s_t| \le M = (1-\xi)/(2a)$ and $2M \le 1 + \nu$, while $\rho + \nu \ge \rho(1 + \nu)$. Consequently $|\eta X_h| \le 1/4$, and the same centered exponential inequality as before bounds the expected remainder by $\eta^2(\rho + \nu)e_t$. The expected potential decrement when W_t is revealed is at least

$$(\eta\rho - \eta^2(\rho + \nu))e_t = 3\kappa e_t/16.$$

Telescoping, with disclosure probability q, proves the finite-class upper bound

$$\mathbb{E}L_T \le 1 + \frac{16 \log(1/\pi(h^*))}{3q\kappa}. \tag{49}$$

The first constant is $1/2$ if the first predictor has probability one-half at every input. No majority-closure assumption is needed for this noisy-flag learner.

The learner from Proposition 3.3 can instead ignore all additional feedback information. Its proof uses the bounds placed on every deployed prediction probability and the independence of the current audit coin. It remains valid when the adversary sees past disclosures. Choosing a learner at time zero using the known parameters therefore gives, for every $K \ge 3$ and every horizon,

$$\mathcal{R}_T^{\mathsf{I},q}(K, \rho) \le C \min\{r_\rho(T, K), \log K/(q\kappa)\}, \tag{50}$$

where $\kappa = \rho$ for any of the five exact quantities listed above and is given by Eq. (46) for a noisy flag. Zero denominators mean infinity. The additive first-round constants are absorbed because $K \ge 3$. Predicting with probability one-half on every round supplies the time cap. Each learner is selected before its first action, with all subsequent actions remaining available for replay.

G.3 Testing with all observed feedback information

We use the squared Hellinger distance defined in Appendix C because it stays finite when a disclosure identifies the target bit exactly. In that case one of the directional KL divergences may be infinite, so an interior-Bernoulli KL argument alone would not cover the complete observation.

Lemma G.1 (Sequential Hellinger testing)**.** *Let a fair bit be fixed before an h-round interaction. Suppose the conditional observation kernels under its two values have squared Hellinger distance at most J on every common supported history. Actions are selected before the next observation. Then every randomized learner has expected cumulative loss at least $c \min\{h, J^{-1}\}$, with the usual convention at $J = 0$.*

Proof. For $J \geq 1/4$ the fair first action suffices, after adjusting c. Otherwise, let P_b^r be the law of the complete transcript after r observations under bit b. Include the independent learner seed in the initial transcript and the chosen action in each transition. The learner's conditional action distribution is identical in the two bit worlds. Consequently the transition kernels $K_{b,r}(\cdot \mid h)$ have affinity at least $1 - J/2$ on every common supported history h. Write $a_r(h) = \mathrm{Aff}(K_{0,r}(\cdot \mid h), K_{1,r}(\cdot \mid h))$. For densities p_b^{r-1} with respect to a common dominating measure μ_{r-1}, the sequential factorization gives

$$\begin{aligned}\mathrm{Aff}(P_0^r, P_1^r) &= \int \sqrt{p_0^{r-1}(h) p_1^{r-1}(h)}\, a_r(h)\, d\mu_{r-1}(h) \\ &\geq (1 - J/2)\, \mathrm{Aff}(P_0^{r-1}, P_1^{r-1}).\end{aligned}$$

Histories outside common support contribute zero to this integral. The initial seed has the same law in both worlds, so the initial affinity is one and iteration gives affinity at least $(1 - J/2)^r$. Thus transcript squared Hellinger distance is at most rJ, and its total variation is at most $\sqrt{rJ}$. Before action $r + 1$, the expected loss under the uniform prior on the bit is at least $(1 - \sqrt{rJ})/2$. Charge the first c'/J actions, or all h if shorter. If $J = 0$, all observation laws agree and every action costs $1/2$ on average. □

We next consider a block that repeats the same query to test one fixed target bit. The observation includes both the feedback label and any additional information revealed on that round.

Lemma G.2 (Information for interior Bernoulli means)**.** *Suppose $\rho \geq 1/4$. In an h-round repeated-query block with a fair fixed target bit, let the two ordinary feedback means lie in $[1/4, 3/4]$ and differ by at most $G \leq 1/4$ on every common history. Reveal additional information independently with probability q. Set $\kappa = \rho$ if the audit flag and the selected predictor's index and score are revealed together. For a noisy audit flag use Eq.* (46)*. Then the conditional squared Hellinger distance of the complete observation is at most $C(G^2 + q\kappa)$. Consequently the block costs at least*

$$c \min\{h, (G^2 + q\kappa)^{-1}\}. \tag{51}$$

If $G > 1/4$, the same loss bound follows from the first action without any interior-parameter assumption.

Proof. Interior Bernoulli laws have squared Hellinger distance at most $6G^2$. Because the disclosure indicator is itself observed and has the same independent law in both worlds, the total squared distance is exactly $(1 - q)$ times the distance for the label alone plus q times the distance when the extra information is included. If the exact audit flag, index, and score are all included, the latter is at most $2 \leq 8\rho$, regardless of what its source index discloses.

For a noisy flag with $\xi \leq 1/4$, one has $\nu \leq 3/4$ and $\kappa \geq 1/16$, so the same bound $H^2 \leq 2 \leq 32\kappa$ suffices. For $\xi > 1/4$, put $a = 1 - 2\xi$ and let $c_i(w) = \mathbb{P}(W = w \mid A = i)$. Each joint cell $\mathbb{P}_b(Y = y, W = w)$ is at least $\xi \mathbb{P}_b(Y = y) \geq 1/16$. If z_b is the conditional replay mean and $\theta_b = \rho b + (1 - \rho) z_b$, the difference in the $Y = 1$ cell is

$$c_0(w)(\theta_1 - \theta_0) + \rho(c_1(w) - c_0(w)).$$

Its absolute value is at most $G + a$, and the $Y = 0$ difference is its negative. Summing the squared differences divided by a lower cell probability bounds squared Hellinger distance by $C(G^2 + a^2)$. For $\rho \geq 1/4$,

$$\kappa = \frac{\rho^2 a^2}{\rho a^2 + \xi(1 - \xi)} \geq a^2/20.$$

This proves the displayed information bound, including $\xi = 1/2$ by the same calculation with $a = 0$. Apply Lemma G.1. If $G > 1/4$, the inverse denominator is less than 16, and the fair first action again suffices. □

G.4 The first phase produces predictors for later replay

Use the target prior and the first two stages of Appendix O, with integers

$$m \geq 4, \qquad 1 \leq d \leq m/2, \qquad 3m \leq T, \qquad (m + 1)^2 2^d \leq K.$$

The two geometric stopping times I, J, the d descendant bits, and all independent random seeds are sampled at time zero. During this first phase, give the learner the same extra information as before: the current node before its action and the completed child after its loss. On every round after the first, replay the actual first predictor, $S_t = 1$. Each stage ends at its sampled stopping time or after m rounds. Run the second stage even if the first reaches its limit without stopping. Generate the independent reveal coins and any flag-noise coins throughout these stages. For the noisy-flag lower bound, we may additionally reveal the exact flag and the selected predictor's index and score during this phase. This only strengthens the learner.

Here $L_{\mathrm{build}} = L_{\mathrm{lo}} + L_{\mathrm{hi}}$ is the learner's actual expected loss during these two stages with the extra information just described. Apply the stopping-time score calculation to this experiment and this L_{build}. Conditional on a stage not having stopped, all its possible remaining stopping times produce the same previous target labels, queries, source rule, and feedback kernels. The revealed source index is always 1. The revealed score depends on the same first action and query. Audit, reveal, and flag-noise coins are independent of the remaining stopping time. Thus, even after observing this information, the conditional probability of stopping on the current round is exactly $1/m$.

Let δ_{lo} be the prediction probability at the first stopping query, and let δ_{hi} be one minus the prediction probability at the second stopping query. The stopping-time calculation gives

$$\mathbb{E}[(\delta_{\mathrm{lo}} + \delta_{\mathrm{hi}})\mathbf{1}_E] \le 2L_{\mathrm{build}}/m, \quad E = \{I \le m, J \le m\}, \quad \mathbb{P}(E) \ge \pi_0 = (1 - e^{-1})^2.$$

Consequently the event

$$F = E \cap \{\delta_{\mathrm{lo}} + \delta_{\mathrm{hi}} \le 4L_{\mathrm{build}}/(\pi_0 m)\}$$

has probability at least $\pi_0/2$, including $L_{\mathrm{build}} = 0$ by nonnegativity. On F, the predictors $p^{\mathrm{lo}}, p^{\mathrm{hi}}$ deployed at these stopping rounds satisfy at every internal pivot x of the selected target block

$$1 - p^{\mathrm{hi}}(x) + p^{\mathrm{lo}}(x) \le 4L_{\mathrm{build}}/(\pi_0 m). \tag{52}$$

Conditional on the two stopping times, the complete observations from this first phase are independent of the descendant bits. Every query in the phase separates superblocks, so all descendants of the selected block have the same true label there. Replaying the first predictor, revealing the extra node information, and revealing audit or source information preserve this independence inductively.

The conditional independence used at every level. Let $B_1, \dots, B_d$ be the time-zero descendant bits. Let $\mathcal{F}_{\mathrm{build}}$ be the sigma-field generated by the stopping times, the independent learner seed, and all observations from the first phase. Let $\mathcal{F}_{j-1}$ be generated by all actions, feedback, additional information, and completed-bit revelations through level $j-1$. When both stages stopped within their limits,

$$\mathbb{P}\big((B_j, \dots, B_d) = \beta \mid \mathcal{F}_{\mathrm{build}} \vee \mathcal{F}_{j-1}\big) = 2^{-(d-j+1)} \quad \text{for every } \beta \in \{0,1\}^{d-j+1}.$$

The first-phase observations are independent of all descendant bits. Each subsequent level uses only its current bit, the two predictors already created, and fresh randomness. Its disclosures therefore carry no information about deeper bits. Induction gives the displayed conditional law at the start of every new level. Observations at that level may then reveal its current bit. The event F is $\mathcal{F}_{\mathrm{build}}$-measurable, so conditioning further on F leaves this equality unchanged. This argument covers both exact audit and source information revealed together and noisy flags, including their observations during the first phase.

After the first phase, if either stopping time equals ∞, fill the rest of the horizon with the legal query $x = 1$ and actual source $S = 1$. Otherwise, run the repeated-query blocks described next. They use the two predictors deployed at the stopping rounds, including when F fails. The lower bound only counts their losses on F.

For the audit-rate cases below, set $b = 4/\pi_0$ and, when $\rho < 1/2$, write $\delta_\rho = 1 - 2\rho$.

G.5 Audit rates at or above one-half

At each descendant level, repeat its midpoint for $h = \lfloor m/d \rfloor$ rounds and disclose the completed branch only after these losses. Under its current bit $B = 1$, replay p^{lo}. Under $B = 0$, replay p^{hi}. Both source indices were fixed in the first phase. The additional feedback information can reveal the

current bit, but it depends on no deeper bit. Each next level therefore starts with a fair bit even after conditioning on the previous disclosures and the learner's independent private seed.

For $\rho \geq 1/2$, put $\Delta_\rho = 2\rho - 1$. Conditional on a first-phase history in F, the two ordinary feedback means are

$$\theta_1 = \rho + (1-\rho)p^{\mathrm{lo}}(x), \qquad \theta_0 = (1-\rho)p^{\mathrm{hi}}(x).$$

Their separation is at most $G = \Delta_\rho + bL_{\mathrm{build}}/m$. If $G \leq 1/4$, both means belong to $[1/4, 3/4]$. Otherwise the first-action case applies. Lemma G.2, $dh \geq m/2$, and $\mathbb{P}(F) \geq \pi_0/2$ yield

$$\mathbb{E}L_T \geq L_{\mathrm{build}} + c_0 \min\left\{m, \frac{d}{(\Delta_\rho + bL_{\mathrm{build}}/m)^2 + q\kappa}\right\}. \tag{53}$$

Set $R_0 = \min\{(dm^2)^{1/3}, d/(\Delta_\rho^2 + q\kappa)\}$. Then $R_0 \leq m$, $R_0^3 \leq dm^2$, and $\Delta_\rho^2 + q\kappa \leq d/R_0$. If $L_{\mathrm{build}} \geq R_0$, the first phase already incurs loss R_0. Otherwise

$$(\Delta_\rho + bL_{\mathrm{build}}/m)^2 + q\kappa \leq 2\Delta_\rho^2 + 2b^2R_0^2/m^2 + q\kappa \leq (2 + 2b^2)d/R_0.$$

The resulting lower bound is therefore $\Omega(R_0)$.

G.6 Audit rates between one-quarter and one-half

Let $1/4 < \rho < 1/2$. If $L_{\mathrm{build}} \geq m/(4b)$, the first phase already incurs loss proportional to m. Otherwise on F put $\underline{p} = p^{\mathrm{lo}}(x)$ and $\overline{p} = p^{\mathrm{hi}}(x)$. Then $\underline{p} \leq 1/4$, $\overline{p} \geq 3/4$, and $1 - \overline{p} + \underline{p} \leq bL_{\mathrm{build}}/m$. Use the negative-correlation rule from Appendix O: if $\overline{p} - \underline{p} \geq \rho/(1-\rho)$, replay the low source under $B = 1$ and mix the endpoints under $B = 0$ to obtain replay mean $\underline{p} + \rho/(1-\rho)$. The ordinary observation laws then agree. If this overlap is infeasible, use the low source under $B = 1$ and the high source under $B = 0$. Their positive feedback gap is at most

$$G = [bL_{\mathrm{build}}/m - \delta_\rho]_+.$$

In both branches the means lie in $[1/4, 3/4]$ and have separation at most $G \leq 1/4$. Lemma G.2 gives

$$\mathbb{E}L_T \geq L_{\mathrm{build}} + c_1 \min\left\{m, \frac{d}{[bL_{\mathrm{build}}/m - \delta_\rho]_+^2 + q\kappa}\right\}. \tag{54}$$

Write $R = (dm^2)^{1/3}$ and $R_0 = \min\{m, m\delta_\rho + R, d/(q\kappa)\}$. If $L_{\mathrm{build}} \geq \varepsilon_0 R_0$, the first phase gives the required loss. Choose a universal $\varepsilon_0 > 0$ small enough below and otherwise assume $L_{\mathrm{build}} < \varepsilon_0 R_0$. If $m\delta_\rho \geq R$, then $R_0 \leq 2m\delta_\rho$ and $bL_{\mathrm{build}}/m \leq 2b\varepsilon_0\delta_\rho \leq \delta_\rho$. The positive part in Eq. (54) vanishes and its repeated-query term is at least a constant times R_0. If $m\delta_\rho < R$, then $R_0 \leq 2R$ and $R_0^3 \leq 8dm^2$. Hence

$$[bL_{\mathrm{build}}/m - \delta_\rho]_+^2 + q\kappa \leq b^2\varepsilon_0^2 R_0^2/m^2 + d/R_0 \leq (1 + 8b^2\varepsilon_0^2)d/R_0.$$

The repeated-query blocks again incur $\Omega(R_0)$ loss. This also covers $q\kappa = 0$.

G.7 Audit rates at most one-quarter

For $0 \leq \rho \leq 1/4$, counting loss only until information is first revealed loses the dependence on ρ. Instead, we bound the information in the full observation on every round. As above, either $L_{\mathrm{build}} \geq m/(4b)$ already pays $\Omega(m)$, or on F the two selected predictors' scores obey $\underline{p} \leq 1/4$, $\overline{p} \geq 3/4$. Choose source mixtures whose means under the current bit are

$$z_1 = \frac{1/2 - \rho}{1 - \rho}, \qquad z_0 = \frac{1}{2(1-\rho)}. \tag{55}$$

Both belong to $[1/3, 2/3] \subseteq [\underline{p}, \overline{p}]$. They are realized by mixing the two predictors from the first phase before the audit coin is drawn. Both ordinary feedback laws are exactly $\mathrm{Bern}(1/2)$. The mixing probabilities for the high source differ by

$$w = \frac{\rho}{(1-\rho)(\overline{p} - \underline{p})}.$$

When the exact audit flag, index, and score are all revealed, couple the two source choices with disagreement probability w and use the same independent audit coin. When it is a replay and the sources agree, couple their replay samples identically. The joint observations can differ only on an audit or a source disagreement during replay. Their total variation is thus at most

$$\rho + (1-\rho)w = \rho + \frac{\rho}{\overline{p} - \underline{p}} \le 3\rho.$$

Its squared Hellinger distance is at most $6\rho = 6\kappa$. This coupling bounds observation laws. The adversary still cannot see the independent reveal coin before selecting the source. The same source rule generates observations whether or not the extra information is revealed.

For a noisy audit flag, the selected predictor's index and score are not disclosed. Put $a = 1 - 2\xi$, $v = \rho a$, $u_0 = \xi/2$, and $u_1 = (1-\xi)/2$. In the cell order $(Y, W) = (1,1), (0,1), (1,0), (0,0)$, the two laws are

$$P_1 = (u_0 + v, u_0, u_1 - v, u_1), \qquad P_0 = (u_0, u_0 + v, u_1, u_1 - v). \tag{56}$$

Here $u_1 - v \ge 1/4$. Their squared Hellinger distance is

$$2(\sqrt{u_0 + v} - \sqrt{u_0})^2 + 2(\sqrt{u_1} - \sqrt{u_1 - v})^2 \le \frac{2v^2}{u_0 + v} + 8v^2 \le 8\kappa.$$

For the last inequality let $v_0 = \rho a^2 + \xi(1-\xi)$, so that $v^2 = \kappa v_0$, $v_0 \le 2(u_0 + v)$, and $v_0 \le 1/2$. The zero-information cases $\rho = 0$ and $\xi = 1/2$ follow directly from equal cell probabilities, avoiding $0/0$ expressions.

Since the label alone has the same law in both worlds, the complete observation has squared Hellinger distance at most $8q\kappa$ for either disclosure choice. Lemma G.1 therefore charges $c\min\{h, (q\kappa)^{-1}\}$ at each level. Adding the first-phase loss and averaging over F proves

$$\mathbb{E}L_T \ge c\min\{m, d/(q\kappa)\}. \tag{57}$$

Because $\delta_\rho \ge 1/2$ in this range, this is equivalent up to universal constants to $\min\{m, m\delta_\rho + (dm^2)^{1/3}, d/(q\kappa)\}$.

G.8 Completing the lower bound

The adversary must also be defined when the two stopping rounds occurred but F fails. The source rules remain legal as follows. For the positive-rate rule the two predictors from those rounds always suffice. For the subcritical overlap rule, use the mixture only when the endpoints are ordered and its mean is attainable. For Eq. (55), use its mixtures only if both means belong to the actual interval $[\underline{p}, \overline{p}]$ with $\underline{p} < \overline{p}$. Otherwise use the low predictor under $B = 1$ and the high predictor under $B = 0$. No adversary needs to test F. All these rules depend only on the first-phase history and the current descendant bit, so deeper bits remain independent. If a stage reaches its limit without stopping, use the already specified continuation with source one. Fill unused rounds with $x = 1, S = 1$.

The construction uses at most $2m + dh \le 3m$ rounds and $(m+1)^2 2^d$ target cutoffs. For randomized learners, the conditional calculations include their independent private seed. The target prior at time zero is the same for every seed. Thus some fixed target and legal target-aware adversary attain the resulting Bayes lower bound. Writing $R = (dm^2)^{1/3}$, we have proved

$$\mathcal{R}_T^{\mathsf{I},q}(K,\rho) \ge c\begin{cases}\min\{R, d/((2\rho-1)^2 + q\kappa)\}, & \rho \ge 1/2,\\ \min\{m, (1-2\rho)m + R, d/(q\kappa)\}, & \rho < 1/2,\end{cases} \tag{58}$$

Here I is either the exact audit flag, index, and score together or the noisy flag. The lower bound with all three exact quantities revealed also applies to each quantity individually.

For $C_\varepsilon \log K \le T \le K^{1/2-\varepsilon}$, choose m, d as in Lemma P.2, so $m = \Theta(T)$ and $d = \Theta_\varepsilon(\log K)$ with all feasibility conditions valid. Equation (58) gives the desired lower bound. For $\rho \ge 1/2$, use $\max\{(2\rho-1)^2, q\kappa\} \le (2\rho-1)^2 + q\kappa \le 2\max\{(2\rho-1)^2, q\kappa\}$ to compare the two displayed forms of the rate.

For shorter horizons, use one fresh balanced-tree bit per round and fixed source $S_t = 1$. Every disclosed quantity, including the full joint observation, depends only on the current or earlier bits, and is disclosed after its current loss. Thus

$$\mathcal{R}_T^{\mathsf{I},q}(K,\rho) \geq \tfrac{1}{2}\min\{T, \lfloor \log_2 K \rfloor\}.$$

When $T < C_\varepsilon \log K$, this is $\Omega_\varepsilon(T)$. Every entry in the claimed rate is also $\Omega_\varepsilon(T)$: $\kappa \leq \rho \leq 1$ and $(2\rho-1)^2 \leq 1$, while $T^{2/3}(\log K)^{1/3} = \Omega_\varepsilon(T)$. This covers all short horizons. If $T \leq \lfloor \log_2 K \rfloor$, the value is exactly $T/2$ by always predicting with probability one-half. For $K = 2$ the forced first true label identifies the target after loss $1/2$, giving that exact value for all horizons and choices of additional information.

This proper lower-bound construction also allows the learner to see the query before submitting its complete action. In the first phase it already knows the current node, and later queries depend only on completed bits. The upper-bound learners can ignore an early query. Thus the same minimax rate holds with this order of moves. The cases $q = 0, 1$, $\rho = 0, 1$, and $\xi = 0, 1/2$ are all included.

H Noisy audit status and information revealed only on replay rounds

We prove Theorem 4.4 and extend its noisy-disclosure law to audit-rate endpoints and to queries revealed before prediction. The noise affects the observed audit status, while audited labels remain truthful. A known finite channel lets us construct an unbiased correction. A separate lower bound proves its minimax order when the adversary replays actual earlier predictors. We then prove Corollary 4.5, where information is disclosed only on replay rounds. In that protocol the learner does not see the independent coin used to decide disclosure, so the meaning of an absent disclosure changes.

For a known finite channel $Q = (Q_0, Q_1)$ and $0 < \rho < 1$, recall $\bar{Q}$, J_Q, and κ_Q from Eq. (10).

Theorem H.1 (Endpoint and query-first extensions of the noisy disclosure law)**.** *Let $Q = (Q_0, Q_1)$ be a known channel on a finite alphabet. Conditional on $A_t = i$, draw $W_t \sim Q_i$, independently of the source, replay sample, and other current randomness. The learner observes the independent disclosure indicator $E_t \sim \mathrm{Bern}(q)$ after feedback and, only when $E_t = 1$, additionally receives W_t, with no exact flag or source information. Write $\mathcal{R}_T^{Q,q}(K,\rho)$ for this proper threshold value. Fix $\varepsilon \in (0, 1/2)$. For integers $K \geq 3$, $1 \leq T \leq K^{1/2-\varepsilon}$ and $\rho, q \in [0,1]$,*

$$\mathcal{R}_T^{Q,q}(K,\rho) = \Theta_\varepsilon\left(\min\left\{r_\rho(T,K), \frac{\log K}{q\kappa_Q}\right\}\right), \tag{59}$$

where κ_Q is defined by Eq. (10) for $0 < \rho < 1$ and by fixed-channel one-sided limits at the audit-rate endpoints. Zero denominators mean $+\infty$. The constants depend only on ε, uniformly in K, T, ρ, q, Q, and its alphabet size. The law also holds if the query is shown before the learner submits its complete action. Independently of the stated horizon restriction, the value is exactly $T/2$ for $K \geq 2$ and $1 \leq T \leq \lfloor \log_2 K \rfloor$, and exactly $1/2$ for $K = 2$ at every horizon.

H.1 Minimum-second-moment unbiased weight for audit status

Let Q_0, Q_1 be probability distributions on a finite alphabet $\mathcal{W}$. Conditional on $A_t = i$, draw $W_t \sim Q_i$, independently of the source, replay sample, and other current randomness. The channel is fixed and known. The learner sees W_t only when $E_t = 1$. We seek a weight whose conditional mean is zero on replay rounds and one on audited rounds. Multiplying the centered label by this weight will cancel the mean contribution of replay. For $0 < \rho < 1$, write

$$\bar{Q} = (1-\rho)Q_0 + \rho Q_1, \qquad d(w) = \frac{Q_1(w) - Q_0(w)}{\bar{Q}(w)}, \qquad J_Q = \mathbb{E}_{\bar{Q}} d^2.$$

Discard symbols of zero $\bar{Q}$ mass. If $J_Q > 0$, define

$$u(w) = \rho + \frac{d(w)}{J_Q}, \qquad v_Q = \rho^2 + \frac{1}{J_Q}, \qquad \kappa_Q = \frac{\rho^2}{v_Q}. \tag{60}$$

If $J_Q = 0$, set $\kappa_Q = 0$ and do not use this update. At $\rho = 0, 1$, define κ_Q by its one-sided limit with the channel held fixed. In particular it is zero at $\rho = 0$.

Lemma H.2 (Minimum-second-moment unbiased weight for audit status)**.** *For $0 < \rho < 1$ and $J_Q > 0$, the weight in Eq.* (60) *satisfies*

$$\mathbb{E}_{Q_0} u = 0, \qquad \mathbb{E}_{Q_1} u = 1, \qquad \mathbb{E}_{\bar{Q}} u^2 = v_Q.$$

It minimizes $\mathbb{E}_{\bar{Q}} v^2$ over all weights with these two conditional means. Moreover $v_Q \geq \rho$, so $0 < \kappa_Q \leq \rho$.

Proof. The identities $\mathbb{E}_{\bar{Q}} d = 0$, $\mathbb{E}_{Q_1} d = (1-\rho) J_Q$, and $\mathbb{E}_{Q_0} d = -\rho J_Q$ verify the displayed formulas. Any feasible v has $\mathbb{E}_{\bar{Q}} v = \rho$ and $\mathbb{E}_{\bar{Q}}[(v-\rho)d] = 1$. Cauchy–Schwarz gives $\mathbb{E}_{\bar{Q}}(v-\rho)^2 \geq 1/J_Q$, with equality for u. Conditional Jensen gives $\mathbb{E} u(W)^2 \geq \mathbb{E}(\mathbb{E}[u(W) \mid A])^2 = \rho$. This minimizes the second moment among unbiased scalar weights for audit status. The minimax statement follows separately from the sequential lower bound. □

H.2 A PROPER UPPER BOUND FOR ARBITRARY EARLIER PREDICTORS

Let $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ be any finite binary class with full-support prior π. For $0 < \rho \leq 3/4$, $q > 0$, and $J_Q > 0$, deploy the internal weighted mixture, start with weights $\boldsymbol{w}_2 = \pi$, and when W_t is disclosed use

$$s_t = u(W_t)(Y_t - 1/2), \qquad \eta = \frac{\rho}{16 v_Q}, \qquad w_{t+1}(h) \propto w_t(h) e^{\eta(2h(x_t)-1)s_t}. \tag{61}$$

Ignore the first forced true label for this upper bound. Its action may be the prior mixture, or a proper probability-one-half predictor if available. Conditional on the target and the selected source, independence of W_t and Y_t given A_t gives, for every value of the selected predictor's score in $[0,1]$,

$$\mathbb{E} s_t = \rho(y_t^* - 1/2), \qquad \mathbb{E} s_t^2 = v_Q/4. \tag{62}$$

The correction works for any value of the selected predictor's score in $[0,1]$. Unlike the score-based strategy, this learner need not deploy binary predictors or assume closure under weighted majority.

The exponential update is well defined even for rare channel symbols. Indeed $-1/(1-\rho) \leq d(w) \leq 1/\rho$, so

$$|\eta u(w)| \leq \frac{\rho^2 J_Q + \max\{1, \rho/(1-\rho)\}}{16(1+\rho^2 J_Q)} \leq \frac{3}{16} < \frac{1}{4}.$$

Writing e_t for the internal mixture error, the centered potential calculation used in Eq. (49) has target term $\eta\rho e_t$ and expected remainder at most $\eta^2 v_Q e_t$. Its expected decrement is therefore at least $15\kappa_Q e_t/256$ per update with an observed W_t. Revealing W_t independently with probability q multiplies this by q, and telescoping gives

$$\mathbb{E} L_T \leq 1 + \frac{256 \log(1/\pi(h^*))}{15 q \kappa_Q}. \tag{63}$$

If the first predictor has probability one-half at every input, replace the first constant by $1/2$.

For $\rho > 3/4$, one may instead ignore the channel and use the unclipped midpoint update. All historical scores lie in $[0,1]$, so the smallest separation between the two target-label feedback means is $2\rho - 1 > 1/2$. The same potential proof gives $1 + 4\log(1/\pi(h^*))/(2\rho-1)^2$, at most the right side of Eq. (63) because $q\kappa_Q \leq 1$. This argument permits any first proper action, whether or not the class admits a probability-one-half predictor. It also handles $\rho = 1$. If $q\kappa_Q = 0$, simply use the learner for hidden replay. Selecting among these learners at time zero proves the upper bound of Theorem H.1 at every horizon.

H.3 INFORMATION IN THE DISCLOSED AUDIT-STATUS OBSERVATION

Use the first phase from Appendix G, including its two stopping times and its legal continuation when either stage fails to stop. During this phase, we may additionally reveal the exact audit flag and then draw W_t from Q_{A_t}. Given the specified label and audit status, this operation is independent of the remaining stopping time. It therefore preserves the stopping probability calculation, the actual first-phase loss L_{build}, and the conditional fairness of the remaining descendant bits.

We now bound how much the disclosed W_t helps distinguish the two target bits. These bounds allow us to use the preceding appendix's lower-bound optimization. First suppose the two ordinary label means θ_b lie in $[1/4, 3/4]$ and differ by at most G. Conditional on $W = w$, let

$$a_w = \frac{\rho Q_1(w)}{\bar{Q}(w)}, \qquad \theta_b(w) = a_w b + (1 - a_w) z_b, \qquad \theta_b = \rho b + (1 - \rho) z_b.$$

Here z_b is the actual conditional replay mean in bit world b. The channel marginal $\bar{Q}$ is the same in both worlds, and $|\theta_b(w) - \theta_b| \le |a_w - \rho|$. An interior Bernoulli parameter $v \in [1/4, 3/4]$ satisfies $H^2(\mathrm{Bern}(u), \mathrm{Bern}(v)) \le 8(u - v)^2$ for every $u \in [0, 1]$. Therefore the squared Hellinger distance between the joint law in world b and $\bar{Q} \otimes \mathrm{Bern}(\theta_b)$ is at most $8\rho^2(1 - \rho)^2 J_Q$. The triangle inequality for H, followed by squaring, bounds the distance between the two joint laws by

$$C\{G^2 + \rho^2(1 - \rho)^2 J_Q\} \le C(G^2 + \kappa_Q). \tag{64}$$

For the last inequality, Lemma H.2 implies $\rho^2 J_Q \le \rho/(1 - \rho)$ and hence $(1 - \rho)^2(1 + \rho^2 J_Q) \le 1$. The case $J_Q = 0$ follows directly since $Q_0 = Q_1$. The learner also observes the independent reveal coin, which gives a bound for the complete observation of $C(G^2 + q\kappa_Q)$, as in Lemma G.2. If $G > 1/4$, use that lemma's fair-first-action argument instead.

For $0 < \rho \le 1/4$, use the legal source mixtures in Eq. (55). Both ordinary label means are exactly $1/2$. Conditional on a common history, let P_b be the joint law of (Y, W) under bit b. Put $\delta_w = Q_1(w) - Q_0(w)$. For each symbol w, the two Y cells are swapped between the bit worlds and have masses $Q_0(w)/2$ and $Q_0(w)/2 + \rho\delta_w$. These masses are nonnegative, including when $Q_0(w) = 0$. Their sum is $\bar{Q}(w)$, so

$$\begin{aligned} H^2(P_1, P_0) &= 2\rho^2 \sum_w \frac{\delta_w^2}{\left(\sqrt{Q_0(w)/2 + \rho\delta_w} + \sqrt{Q_0(w)/2}\right)^2} \\ &\le 2\rho^2 J_Q \le \tfrac{8}{3}\kappa_Q. \end{aligned} \tag{65}$$

The last step uses $\rho^2 J_Q \le 1/3$. Terms with both masses zero are zero. This calculation allows disjoint support and does not assume finite directional KL. Revealing W_t independently with probability q therefore gives squared Hellinger distance $O(q\kappa_Q)$ for the complete observation. At $\rho = 0$ the two laws coincide directly.

Apply Lemma G.1 to each descendant level, and substitute κ_Q into the positive, subcritical, and small-audit lower bounds in Appendix G. The same optimization and choice of target support apply. The adversary selects between the two earlier predictors using only the current bit. Observing the noisy audit status does not disclose later bits. This proves Theorem H.1, including queries shown before complete actions, the exact short-horizon value, and $K = 2$. At $\rho = 1$, the lower bound using one unresolved label per round and the truthful-label upper bound already give the claimed order, irrespective of the channel.

H.4 Information revealed only on replay rounds

Fix $q \in [0, 1]$. Generate an unobserved, independent $E_t \sim \mathrm{Bern}(q)$ after the current loss and label. Instead of observing E_t, the learner sees the indicator

$$Z_t = (1 - A_t)E_t.$$

If $Z_t = 1$, it may additionally receive the selected predictor's index and score (S_t, z_t). If $Z_t = 0$, either the round was audited or the index and prediction probability were withheld on a replay round. At $q = 1$, absence identifies an audit. The coin E_t is not separately visible. The expected number of disclosures is $q(1 - \rho)(T - 1)$. Here q is the probability of revealing information conditional on a replay round. Sources are still selected before the audit.

The conditional laws of this observed indicator satisfy $Q_1(Z = 1) = 0$ and $Q_0(Z = 1) = q$. When $q > 0$ its unbiased weight and second moment are explicitly

$$u(Z) = 1 - Z/q, \qquad v_{\mathrm{rep}} = \rho + (1 - \rho)\frac{1 - q}{q}, \qquad \kappa_{\mathrm{rep}} = \frac{q\rho^2}{q\rho + (1 - q)(1 - \rho)}. \tag{66}$$

Set $\kappa_{\mathrm{rep}} = 0$ at $q = 0$ or $\rho = 0$. For $q > 0, \rho = 1$ it equals one. The correction $s_t = (1 - Z_t/q)(Y_t - 1/2)$ uses only Z_t and the label, has mean $\rho(y_t^* - 1/2)$ and second moment $v_{\mathrm{rep}}/4$, and is valid for

arbitrary earlier predictors with scores in $[0, 1]$. Apply the preceding upper bound with channel output Z_t and outer disclosure probability one. This proves the upper bound of Corollary 4.5. The selected predictor's index and score are unnecessary for this bound.

The lower bound must still allow both the index and score to be revealed when $Z_t = 1$. A lower bound with Z_t as the only extra observation would not suffice. For $\rho \geq 1/4$, one has $\kappa_{\mathrm{rep}} \geq q/16$ because the denominator in Eq. (66) is at most one. Enrich the observation by revealing the independent E_t and, whenever $E_t = 1$, the joint observation (A_t, S_t, z_t). The actual observation is a deterministic function of this enriched observation. In a repeated-query block whose two label means lie in the specified interior interval, its squared Hellinger distance is at most $C(G^2 + q)$ by Lemma G.2, hence at most $C(G^2 + \kappa_{\mathrm{rep}})$. For larger G the first action suffices.

For $0 \leq \rho \leq 1/4$, retain the equal-mean source mixtures in Eq. (55). On the event F from the first phase, let $\underline{p} \leq 1/4$, $\overline{p} \geq 3/4$ be the two selected predictors' scores. The probabilities $a_b = (z_b - \underline{p})/(\overline{p} - \underline{p})$ of selecting the high source belong to $[1/9, 8/9]$, and

$$|a_1 - a_0| = \frac{\rho}{(1-\rho)(\overline{p} - \underline{p})}.$$

When $Z_t = 1$, the round uses replay, so $A_t = 0$. Conditional on the selected source, the label and its reported score have exactly the same law in both bit worlds. Thus the contribution to squared distance from all observations with $Z_t = 1$ is exactly

$$q(1-\rho)H^2(\mathrm{Bern}(a_1), \mathrm{Bern}(a_0)) \leq Cq\rho^2.$$

This statement includes the full source index, not merely its score. For an observation with $Z_t = 0$, the two Y masses are swapped and equal $(1-q)/2$ and $(1-q)/2 + q\rho$. Its contribution is bounded by

$$2\left(\sqrt{(1-q)/2 + q\rho} - \sqrt{(1-q)/2}\right)^2 \leq \frac{4q^2\rho^2}{1-q+q\rho}.$$

Here P_b denotes the conditional law of (Y_t, Z_t) together with (S_t, z_t) when $Z_t = 1$, given a common history and bit b. Adding both contributions gives

$$H^2(P_1, P_0) \leq \frac{Cq\rho^2}{1-q+q\rho} \leq C\kappa_{\mathrm{rep}}, \tag{67}$$

since $q\rho + (1-q)(1-\rho) \leq 1 - q + q\rho$. At zero-information endpoints the laws agree. At $q = 1$ this bound has order ρ, accounting for the fact that $Z_t = 0$ identifies an audit.

During the first phase, replaying the fixed source $S = 1$ and optionally revealing the exact audit flag preserve the stopping probabilities and descendant independence. Outside F, use the mixture when its endpoints allow it and otherwise the designated endpoint sources. If a stage reaches its limit without stopping, continue with $x = 1, S = 1$. These are the complete continuation rules of Appendix G. The lower-bound inequalities and their optimization therefore hold with $q\kappa$ replaced by κ_{rep}. This proves the minimax law in Corollary 4.5 with the selected predictor's index and score disclosed when $Z_t = 1$, and hence also when Z_t is the only extra observation. The same arguments using one unresolved label per round and the forced first label give both variants the exact value $T/2$ for $K \geq 2$ and $1 \leq T \leq \lfloor \log_2 K \rfloor$, independently of the stated horizon restriction, and $1/2$ for $K = 2$ at every horizon. These identities and the matching law hold when the query is shown before the complete action, including all endpoints.

Comparison at the same expected number of disclosures. Take $q_{\mathrm{rep}} = 1/2$ for disclosure only on replay rounds and $q_{\mathrm{ind}} = (1-\rho)/2$ for revealing the source index independently on each round. Both protocols reveal the source index $(1-\rho)(T-1)/2$ times in expectation. The quantities entering their respective minimax denominators are ρ^2 and $\rho(1-\rho)/2$. If $0 < \rho \leq 1/4$ and $2\log K/\rho^2 \leq T \leq K^{1/2-\varepsilon}$, then $r_\rho(T, K) \geq T/2$, so the terms from the revealed information determine both minima. Their minimax losses are therefore respectively $\Theta_\varepsilon(\log K/\rho^2)$ and $\Theta_\varepsilon(\log K/\rho)$, a ratio of order $1/\rho$. This proves the equal-count comparison in Corollary 4.5, including when both protocols disclose the source index and recover its score from the earlier actions. This comparison fixes an expected count, not a hard budget or an adaptive rule for deciding when to reveal information.

Relation to existing detector information. For $0 < \rho < 1$, the channel divergence used in recursive estimation by Kanabar & Gastpar (2026b) obeys $\chi^2(Q_1\|\bar{Q}) = (1-\rho)^2 J_Q$. Consequently our index can also be written $\rho^2\chi^2/((1-\rho)^2 + \rho^2\chi^2)$. To establish the minimax conclusion here, we also correct an unknown replay mean that can depend on the target. The lower bound pays the learner's loss while creating the predictors used for replay, then controls every later observation. Equation (67) also permits the actual source to be revealed on replay rounds with $Z_t = 1$, a stronger observation than the detector alone.

I Interpreting noisy audit flags

For a symmetrically flipped audit flag, $W = A \oplus N$ with known independent $N \sim \mathrm{Bern}(\xi)$, Eq. (10) becomes

$$\kappa = \frac{\rho^2}{\rho + \xi(1-\xi)/(1-2\xi)^2}, \qquad 0 \le \xi < 1/2, \tag{68}$$

with $\kappa = 0$ at $\xi = 1/2$ or $\rho = 0$. At $\rho = 1/2$ and $\log K \ll T$ with $T \le K^{1/2-\varepsilon}$, the rate from exact audit flags equals the order obtainable from labels alone when the reveal probability is $q_c \asymp (\log K/T)^{2/3}$. With noisy flags, the relevant quantity is instead $q(1-2\xi)^2$, up to universal constants. Values comparable to q_c give the same order as learning from the labels alone. A strict asymptotic improvement requires this quantity to be larger by a diverging factor. For $\xi \le 1/4$, $\kappa \asymp \rho^2/(\rho+\xi)$, so inverse-linear audit dependence as $\rho \to 0$, when time truncation is inactive, requires $\xi = O(\rho)$. The example shows how audit-flag noise increases the second moment of the weighted centered label, while audited labels remain truthful. The proof and update for this channel remain in Appendix G.

J An inverse-linear audit bound for recent predictors

How does limiting how far back replay reaches help when the selected predictors remain hidden? This appendix proves Theorem 4.1 and extends the upper and lower bounds to finite classes. Corollary J.4 gives the threshold result for both proper and improper learners in either timing. The restriction bounds the expected time elapsed since the selected predictor was deployed, conditional on the target and history, as in Appendix K. The adversary may still choose sources using the target and history without disclosing its choice. We obtain an upper bound logarithmic in class size and inverse-linear in the audit rate for unrestricted improper learners and for proper learners whose class is closed under weighted majority. Appendix K gives a separate proper guarantee without this closure assumption.

Throughout, $\mathcal{H}$ is a finite binary class of $N \ge 2$ distinct hypotheses on a finite domain $\mathcal{X}$. The target is fixed before interaction and is independent of the learner's private seed. Proper scores lie in $\mathrm{conv}(\mathcal{H})$, and unrestricted improper scores may be any full functions into $[0,1]$. Each complete deployed score function remains available for replay. The source is drawn before the independent audit coin, and the first round gives a forced true label after prediction. We first prove lower bounds using only the immediately preceding predictor. We then construct an upper-bound strategy that holds one binary predictor fixed while learning from the resulting feedback.

Lemma J.1 (A two-target lower bound using the previous predictor)**.** *For every finite binary class of size $N \ge 3$, $T \ge 1$, and $\rho \in (0,1]$, a fixed-target prior and the actual-source policy $S_t = t-1$ force expected deployment loss at least $c\min\{T, 1/\rho\}$ against every unrestricted improper learner, for a universal $c > 0$. The statement holds in both timings, even if audit flags are disclosed. At $\rho = 0$, the corresponding lower bound is at least $T/4$.*

Proof. Among three distinct hypotheses, some pair agrees at a coordinate x_0 and differs at another coordinate x. Sample this pair fairly at time zero. Query x_0 on the forced-truth first round and x on every subsequent round. The first label reveals no information about the bit at x. Use $S_t = t-1$ throughout the post-start interaction.

Fix the learner's independent private seed. On the event that no previous query at x was audited, all preceding predictors and replayed labels are generated recursively without using the fair target bit. In particular, the previous predictor's score at x is determined by a transcript independent of that

bit. Hence the current score has average loss $1/2$ on that event, even when the current query and all previous audit flags are visible. The post-start loss is at least

$$\frac{1}{2}\sum_{r=0}^{T-2}(1-\rho)^r \geq c_0 \min\{T-1, 1/\rho\} \qquad (T \geq 2).$$

The absence of previous audits is an event for the analysis, not information used by the policy. The policy uses the same query and source rule on every round. Separately, a fair prior on two hypotheses and a distinguishing first query gives first-round minimax loss $1/2$. Taking the larger of these valid lower bounds proves the positive-audit statement, including $T = 1$. At zero audit the first construction gives $(T-1)/2$, whose maximum with $1/2$ is at least $T/4$. □

J.1 A QUERY-TREE LOWER BOUND USING THE PREVIOUS PREDICTOR

To obtain a class-dependent lower bound, repeat this argument at successive queries whose labels are initially unresolved. Let $\ell = \mathrm{Ldim}(\mathcal{H})$ be the maximum depth of a complete binary query tree whose every root-to-leaf label path is realized by a target.

Lemma J.2 (Testing cost when the previous predictor is replayed)**.** *For $N \geq 3$, $T \geq 1$, and $\rho \in (0, 1]$, the policy $S_t = t - 1$ permits a lower bound $c \min\{T, \ell/\rho\}$ against every unrestricted improper learner, for a universal $c > 0$. It holds in either timing and remains valid if all audit flags are revealed. Because every selected predictor was deployed one round earlier, the result satisfies any per-round conditional bound $D \geq 1$ on the expected elapsed time.*

Proof. Suppose first that $\ell \geq 2$. Choose a complete depth-ℓ query tree, sample its independent fair path bits at time zero, and select a realizing target for each leaf before interaction. The resulting fixed-target prior is independent of the learner's private seed. Spend the forced truthful first round at the root. It exposes only the root bit and incurs expected loss $1/2$.

At each of the remaining $\ell - 1$ levels, repeat the query at that node for $m = \lceil 1/\rho \rceil$ rounds, truncating the last block if the horizon ends. At the end of a complete block, a genie reveals its true bit before the next pivot is chosen. This strengthens the learner. Pivots and block endpoints depend only on completed bits and the fixed schedule, never on whether an unobserved audit has occurred. On every post-start round choose the actual immediately preceding action as source.

At a new level, its bit is fair conditional on the learner's private seed and the complete preceding transcript, including the completed genies and audit flags. On the event that the first $r - 1$ audit coins in its block are zero, those observations are recursively generated from previous predictions and fresh replay samples independent of the current bit. The previous full action evaluated at the current pivot is fixed by this bit-independent transcript. Thus the round-r prediction has mean loss $1/2$ on that event. The event is used only in the analysis. The source policy never tests it.

For any block prefix of length $s \leq m$, its expected loss is therefore at least

$$\frac{1}{2}\sum_{r=1}^{s}(1-\rho)^{r-1} = \frac{1-(1-\rho)^s}{2\rho} \geq \frac{1-e^{-2}}{4}s. \tag{69}$$

Indeed $\rho s \leq \rho\lceil 1/\rho \rceil \leq 2$, and $1 - e^{-u} \geq (1 - e^{-2})u/2$ for $0 \leq u \leq 2$. This also covers $\rho = 1$, when $m = 1$. Summing over complete blocks and the possible final prefix gives at least

$$\frac{1}{2} + c_0 \min\{T-1, (\ell-1)m\} \geq c \min\{T, \ell/\rho\},$$

using $\ell - 1 \geq \ell/2$. After all levels, any remaining rounds can use the root query and source $t - 1$. Their losses are nonnegative. If $\ell = 1$, Lemma J.1 gives the same bound. The construction reveals each current pivot before prediction if desired, so both timings are covered. It always replays the predictor deployed one round earlier. □

J.2 LEARNING WHILE HOLDING PREDICTIONS FIXED

We now bound the expected time since the replayed predictor was deployed, as in Appendix K:

$$\sum_{s<t}\lambda_t(s)(t-s) \leq D \quad \text{almost surely,} \qquad D \geq 1. \tag{70}$$

This is a per-round conditional bound: the target, public history, current predictor, query, and source distribution are fixed before the source is sampled. The learner knows D and ρ, but does not observe the source distribution, which may depend on the target. The distribution may still assign positive probability to arbitrarily old predictors. Write $\mathcal{R}_T^{\mathrm{imp}}(\mathcal{H}, \rho; D)$ for the improper minimax loss under this restriction, with predictions submitted before queries.

Theorem J.3 (Upper bound by holding binary predictions fixed)**.** *There is a universal C such that, for every finite binary class of size $N \geq 2$, all $\rho \in (0, 1]$, and all $T \geq 1$, unrestricted improper learning under* (70) *satisfies*

$$\mathcal{R}_T^{\mathrm{imp}}(\mathcal{H}, \rho; D) \leq C \min\left\{T, \frac{[1 + (1-\rho)D]\log(2N)}{\rho}\right\}. \tag{71}$$

The same bound is proper when the pointwise binary weighted majority of every distribution on $\mathcal{H}$ belongs to $\mathcal{H}$. The learner commits before the query and keeps all earlier predictors available for replay.

J.2.1 The learner

Fix $\eta = 1/64$, initialize strictly positive uniform internal weights $\boldsymbol{w} \in \Delta(\mathcal{H})$, and set

$$b = \max\{1, \lceil 16(1-\rho)D/\rho \rceil\}. \tag{72}$$

The learner divides time into epochs. At the beginning of each epoch, save $\boldsymbol{w}^0 = \boldsymbol{w}$ and hold fixed the full binary function

$$f(x) = \mathbf{1}\{\mathbb{E}_{h\sim\boldsymbol{w}^0} h(x) \geq 1/2\}.$$

First deploy this same f for a waiting period of b rounds, called burn-in. Ignore these observations and leave $\boldsymbol{w}$ fixed. The period may include the first forced truthful round. The remaining rounds of the epoch are called stable rounds. Continue deploying the same f on those rounds, but now update the internal weights. For each stable observation at x, compute

$$s = Y - (1-\rho)f(x) - \rho/2, \qquad w'(h) = \frac{w(h)\exp(\eta s h(x))}{\sum_g w(g)\exp(\eta s g(x))}. \tag{73}$$

Accumulate the total variation distances $\mathrm{TV}(\boldsymbol{w}', \boldsymbol{w})$ of these internal updates. End the epoch after the first update that makes this sum at least $1/4$. Begin the next epoch with the current weights and their binary majority as the new predictor. All earlier predictors remain available for replay, including during burn-in.

J.2.2 Controlling replay from older predictors after burn-in

In a stable round, the previous b complete actions all equal f. With $z = \sum_{s<t} \lambda_t(s) p_s(x)$, Markov's inequality applied to (70) gives

$$(1-\rho)|z - f(x)| \leq (1-\rho)\mathbb{P}(t - S_t > b) \leq \frac{(1-\rho)D}{b+1} \leq \frac{\rho}{16}. \tag{74}$$

After burn-in, older predictors contribute little to the feedback mean. A label that contradicts the current prediction can still come from replay, so the learner uses the quantitative bound above rather than interpreting a contradiction as an audit. Write $\mathcal{J}_t$ for the enlarged pre-source information, which includes the fixed target, public history, current internal weights, deployed function, query, and actual source distribution, but not the current source draw or fresh feedback randomness. Writing $y^* = h^*(x)$ and conditioning on $\mathcal{J}_t$ while integrating out the source,

$$\begin{aligned} (2y^* - 1)\mathbb{E}s &\geq 7\rho/16, \\ \mathbb{E}|s| \leq 25\rho/16, &\qquad \mathbb{E}s^2 \leq 3\rho. \end{aligned} \tag{75}$$

Indeed $\mathbb{P}(Y \neq f(x)) \leq \rho + (1-\rho)|z - f(x)| \leq 17\rho/16$ and $s = (Y - f(x)) + \rho(f(x) - 1/2)$. The absolute and quadratic bounds follow from this representation. Neither the audit flag nor the selected predictor's score is observed.

J.2.3 Bounding prediction loss after burn-in

Fix the true target and put $r = \mathbb{E}_{h\sim \boldsymbol{w}} h(x)$, $e = |r - y^*|$, and $\Phi = -\log w(h^*)$. The centered exponential inequality, with $|s| \le 1$, gives

$$\Phi' - \Phi \le \eta s(r - y^*) + \eta^2 s^2 r(1 - r).$$

Since $r(1 - r) \le e$, Eq. (75) implies

$$\mathbb{E}[\Phi' - \Phi \mid \mathcal{J}_t] \le -\eta(7\rho/16)e + 3\eta^2\rho e \le -\eta\rho e/4. \tag{76}$$

Internal weights are unchanged during burn-in and are not reset at an epoch boundary. Stable-round indicators are determined before the current observation. Telescoping the same nonnegative potential over all stable updates up to the horizon, including any unfinished final epoch, yields

$$\mathbb{E} \sum_{\text{stable}} e \le \frac{4 \log N}{\eta\rho}. \tag{77}$$

We next relate the internal prediction error to the loss of the predictor actually deployed. Before every stable update, the accumulated variation since the epoch began is less than $1/4$, so $\mathrm{TV}(\boldsymbol{w}, \boldsymbol{w}^0) < 1/4$. If the fixed binary majority $f(x)$ is wrong, the hypotheses with the wrong label have mass at least $1/2$ under $\boldsymbol{w}^0$ and at least $1/4$ under $\boldsymbol{w}$. Thus the deployment loss of f is at most $4e$, and its total stable loss is bounded by $16 \log N/(\eta\rho)$.

J.2.4 Bounding how often the predictor changes

The exponential update preserves the conditional distributions inside the two groups $h(x) = 0$ and $h(x) = 1$. Hence its total variation is exactly the change in their mass. A direct logistic calculation gives

$$\mathrm{TV}(\boldsymbol{w}', \boldsymbol{w}) = |r' - r| \le 2\eta|s|r(1 - r) \le 2\eta|s|e. \tag{78}$$

Using (75) and (77),

$$\mathbb{E} \sum_{\text{stable}} \mathrm{TV}(\boldsymbol{w}', \boldsymbol{w}) \le 4\eta\rho\mathbb{E} \sum_{\text{stable}} e \le 16 \log N.$$

This variation bound also pays for the waiting periods. Every completed epoch contributes at least $1/4$ to the accumulated variation, so the expected number of epochs begun is at most $1 + 64 \log N$. Each contributes at most b burn-in rounds, including a possibly truncated final burn-in. The loss during these rounds is at most their number, giving

$$\mathbb{E}L_T \le \frac{16 \log N}{\eta\rho} + b(1 + 64 \log N) \le C\frac{[1 + (1 - \rho)D] \log(2N)}{\rho}.$$

The same learner has the pathwise bound $L_T \le T$, which completes (71) without switching learners or resetting history. The argument also covers $\rho = 1$, when $b = 1$ and replay vanishes. Every deployment is binary and is proper under the stated majority-closure condition. This proves Theorem J.3.

Corollary J.4 (Threshold bounds with recent predictors)**.** *For $K \ge 3$, $D \ge 1$, $T \ge 1$, and $\rho \in (0, 1]$, thresholds under* (70) *satisfy*

$$\begin{aligned} c \min\{T, \log K/\rho\} \le \mathcal{R}_T^{\mathrm{imp}}(K, \rho; D) \le \mathcal{R}_T(K, \rho; D) \\ \le C \min\{T, [1 + (1 - \rho)D] \log K/\rho\}. \end{aligned} \tag{79}$$

In particular, both values are $\Theta_D(\min\{T, \log K/\rho\})$, with constants independent of K, T, ρ. Both timings are covered. At zero audit the value is $\Theta(T)$, and $K = 2$ retains its forced-first-label value $1/2$.

Proof. Threshold weighted majority is a threshold. Apply Theorem J.3. For the lower bound use the repeated balanced query-tree blocks of Lemma J.2 with source $S_t = t - 1$. The selected predictor was always deployed one round earlier. The threshold Littlestone dimension is $\lfloor \log_2 K \rfloor$ and is comparable to $\log K$, including the small-class cases handled there. Lemma J.1 gives the zero-audit lower bound with source $t - 1$. The constant-half threshold supplies the upper bound $T/2$. For $K = 2$, the forced first truthful label identifies the entire target, giving the exact value $1/2$. The first-round lower bound absorbs small horizons. □

Comparison with the general proper-class guarantee. Theorem K.1 in Appendix K gives a proper upper bound $4[4 + (1 - \rho)D] \log N/\rho^2$ for every finite class, together with a target-dependent nonuniform-prior refinement. Equation (71) improves the audit dependence by one factor of $1/\rho$ for unrestricted improper actions and for proper majority-closed classes, including thresholds. These assumptions differ, so both guarantees are useful. The upper and lower bounds in (79) match for fixed D. Their dependence on growing D remains unmatched.

J.2.5 Every proper finite class: a maximum-weight alternative

Theorem J.5 (No majority-closure assumption)**.** *Under the same per-round conditional bound on the expected time since the selected predictor was deployed, every finite binary class of $N \geq 2$ targets admits a proper learner satisfying*

$$\mathbb{E}L_T \leq C \min\left\{T, \frac{N[1 + (1 - \rho)D]}{\rho}\right\}. \tag{80}$$

Thus for every fixed finite class with $N \geq 3$ and every fixed D, the proper and improper minimax values under this restriction are $\Theta_{\mathcal{H},D}(\min\{T, 1/\rho\})$, uniformly in T and $0 < \rho \leq 1$. This last comparison allows class-dependent constants, unlike the logarithmic-size upper bound for majority-closed classes.

Proof. Initialize uniform positive weights and use the same step $\eta = 1/64$, burn-in length (72), and update (73). At an epoch start choose a maximum-weight hypothesis h_0 and freeze its full binary prediction. After burn-in, continue the same exponential update, ending the epoch as soon as some hypothesis has weight at least $2w(h_0)$. This stopping condition is observable and does not use the true target. Before each stable update, $w(h_0) > \max_h w(h)/2$.

For analysis set $v_h = \sqrt{w(h)/w(h^*)}$ and $\Xi = \sum_{h \neq h^*} v_h$, initially $N - 1$. On an agreement with the true label, v_h does not change. On a disagreement it is multiplied by $\exp\{-\eta(2y^* - 1)s/2\}$. Equation (75) gives

$$\mathbb{E}[v'_h \mid \mathcal{J}_t] \leq (1 - \eta\rho/8)v_h, \qquad \mathbb{E}[|v'_h - v_h| \mid \mathcal{J}_t] \leq 2\eta\rho v_h$$

on a disagreement. Consequently

$$\mathbb{E}\sum_{\text{stable}} \sum_{h:h(x)\neq y^*} v_h \leq \frac{8(N-1)}{\eta\rho}, \qquad \mathbb{E}\sum_{\text{stable}} \sum_{h} |v'_h - v_h| \leq 16(N-1).$$

If the fixed predictor h_0 is wrong, its pre-update ratio satisfies $v_{h_0} \geq 1/\sqrt{2}$. The first display therefore bounds its loss during stable rounds, up to a factor $\sqrt{2}$.

It remains to count the waiting periods. Each completed epoch contributes at least $1/4$ to the ℓ_1 variation in the vector $\boldsymbol{v} = (v_h)_h$. To check this, at its start let $v_0 = v_{h_0} \geq 1$, with every $v_h \leq v_0$. At its end some $v_g \geq \sqrt{2}v_{h_0}$. If $\Delta\boldsymbol{v}$ denotes the net change over the epoch, then $\sqrt{2}|\Delta v_{h_0}| + |\Delta v_g| \geq (\sqrt{2} - 1)v_0$. The total coordinatewise variation is at least $|\Delta v_{h_0}| + |\Delta v_g|$, and hence at least $(1 - 1/\sqrt{2})v_0 > 1/4$. The case $g = h^*$ satisfies the same inequality since that coordinate is constantly one. Hence the expected number of epochs begun is at most $1 + 64(N - 1)$. Adding burn-in loss proves (80).

For the fixed-class matching claim, Lemma J.1 already uses $S_t = t - 1$. Prior to its first audit, successive predictions and replayed labels are recursively independent of the fair bit. This rule always selects the predictor deployed one round earlier. It gives a universal $c \min\{T, 1/\rho\}$ lower bound even against improper learners. Combine it with (80) and fix the class and D. □

K An upper bound for every proper finite class

This appendix gives an upper bound that applies to every finite proper class, without the majority-closure assumption used in Appendix J. The learner changes its full prediction function slowly enough that its current score approximates the unknown mean of replayed feedback. Consider a finite binary class $\mathcal{H} \subseteq \{0, 1\}^{\mathcal{X}}$ of size $N \geq 2$, with proper scores $p_t(x) = \mathbb{E}_{h\sim Q_t} h(x)$. For the actual source

distribution conditional on the target and history, the expected time since the selected predictor was deployed is bounded by known $D \geq 1$:

$$\sum_{s<t} \lambda_t(s)(t-s) \leq D \tag{81}$$

almost surely on every $t \geq 2$, before the source and feedback draws. Sources and their distribution remain hidden. The first label is still forced truthful.

Theorem K.1 (Proper upper bound using recent predictors)**.** *For $\rho > 0$, there is a proper learner satisfying*

$$\mathbb{E}L_T \leq \frac{4[4 + (1-\rho)D] \log N}{\rho^2} \tag{82}$$

for every horizon and every legal adversary obeying Eq. (81)*. More generally, initializing at a full-support prior π replaces $\log N$ by $\log(1/\pi(h^*))$. Every such guarantee is also bounded by T. For the uniform-prior minimax bound, a separate learner that always predicts one with probability one half gives an additional $T/2$ cap when $\mathcal{H}$ admits a proper constant-half score. Selecting between these two learners uses only known parameters, not the target.*

Proof. Initialize internal weights $\boldsymbol{w}_1$ uniformly on $\mathcal{H}$ (or at a full-support prior π for the target-dependent version) and set

$$\eta = \frac{\rho}{4 + (1-\rho)D} \leq \frac{1}{4}.$$

On round t, deploy the proper distribution Q_t with masses $Q_t(h) = w_t(h)$, put $r_t = p_t(x_t)$ and, after the observed label, set

$$s_t = Y_t - (1-\rho)r_t - \rho/2, \qquad w_{t+1}(h) = \frac{w_t(h) \exp\{\eta s_t h(x_t)\}}{\sum_{g \in \mathcal{H}} w_t(g) \exp\{\eta s_t g(x_t)\}}. \tag{83}$$

The update uses only the current prediction and observed label. In particular, $|s_t| \leq 1$, and it does not require the unknown source distribution.

We first bound how much the prediction function changes in one round. Put $z = \eta s_t$, so $|z| \leq \eta$. Multiplying hypothesis weights by $\exp(zh(x_t))$ preserves their conditional distributions within the two groups $h(x_t) = 0$ and $h(x_t) = 1$. Hence its total variation distance from the untilted distribution is exactly the change of mass of one group. After the update, the mass of the group predicting one is $\mathrm{logistic}(\mathrm{logit}(r_t) + z)$. The logistic derivative is at most $1/4$, and endpoint masses $r_t \in \{0, 1\}$ are unchanged. Consequently

$$\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_{\mathrm{TV}} \leq \eta/4, \qquad |p_s(x) - p_t(x)| \leq \frac{\eta}{4}(t-s) \quad (s < t)$$

for every query x on every path. Because this bound holds for the complete function, it also holds at a query selected after the adversary observes the current predictor.

Condition on the enlarged pre-source information that includes the target, history, $\boldsymbol{w}_t$, x_t, and $\boldsymbol{\lambda}_t$, and integrate out the source draw. Writing $y_t^* = h^*(x_t)$ and $\bar{z}_t = \sum_{s<t} \lambda_t(s) p_s(x_t)$ gives

$$|\bar{z}_t - r_t| \leq \eta D/4, \qquad \mathbb{E}[s_t \mid \text{pre-source information}] = \rho(y_t^* - 1/2) + (1-\rho)(\bar{z}_t - r_t).$$

Therefore

$$(2y_t^* - 1)\mathbb{E}[s_t \mid \text{pre-source information}] \geq \rho/2 - (1-\rho)\eta D/4. \tag{84}$$

On round one $Y_1 = y_1^*$ deterministically, and directly $(2y_1^* - 1)s_1 \geq \rho/2$, so the same lower estimate holds.

Let $\Phi_t = -\log w_t(h^*)$ and $e_t = |r_t - y_t^*|$. For a Bernoulli random variable H of mean r, $|s| \leq 1$, and $\eta \leq 1/4$, the elementary bound $e^u \leq 1 + u + u^2$ for $|u| \leq 1$ gives

$$\log \mathbb{E} e^{\eta s(H-r)} \leq \eta^2 s^2 r(1-r) \leq \eta^2 r(1-r).$$

The update thus satisfies

$$\Phi_{t+1} - \Phi_t \leq \eta s_t (r_t - y_t^*) + \eta^2 r_t(1-r_t).$$

Since $r_t - y_t^* = -(2y_t^* - 1)e_t$ and $r_t(1 - r_t) \le e_t$, Eq. (84) yields

$$\begin{aligned}\mathbb{E}[\Phi_{t+1} - \Phi_t \mid \text{pre-source information}] &\le -\eta e_t \left[\frac{\rho}{2} - \eta\left(1 + \frac{(1-\rho)D}{4}\right)\right] \\ &= -\frac{\eta\rho}{4} e_t.\end{aligned}$$

Telescoping, using $\Phi_1 = \log N$ (respectively $\Phi_1 = \log(1/\pi(h^*))$) and $\Phi_{T+1} \ge 0$, proves Eq. (82).

□

Remark K.1 (Comparison with holding binary predictions fixed). The proof uses the current prediction probability to approximate the unknown replay mean, by controlling changes in the full function. Related control of changing predictions appears in delayed-feedback learning (Joulani et al., 2016). Appendix J instead holds a binary predictor fixed and explicitly pays for the waiting period before updating. That strategy gives ρ^{-1} audit dependence with logarithmic class-size dependence for improper actions or proper majority-closed classes, including thresholds. The present ρ^{-2} guarantee applies to every finite proper class.

Corollary K.2 (How far replay must reach for the critical lower bound)**.** *Suppose a family of adversaries bounds the expected time since the selected predictor was deployed by D, given the target, history, current predictor, and input on every round, and forces at $\rho = 1/2$ the critical threshold lower bound $cT^{2/3}(\log K)^{1/3}$ against every proper learner. Then*

$$D + 1 \ge c'\left(\frac{T}{\log K}\right)^{2/3},$$

where $c' > 0$ depends only on c. Thus this polynomial lower bound requires the per-round conditional bound on the expected elapsed time to grow when $T/\log K \to \infty$.

Proof. At $\rho = 1/2$, Eq. (82) is $16(4 + D/2)\log K$. Compare the assumed lower bound with this valid upper bound for the restricted family and rearrange. □

L Revealing the query before prediction

This appendix proves Corollary 3.5 for proper learners that see x_t before submitting the complete predictor p_t. The adversary then chooses the source distribution and draws the source before the audit coin. The target remains fixed at time zero. Write $\mathcal{R}_T^{\mathrm{qf}}(K, \rho)$ for this game's minimax value. We show that knowing the query leaves the uncertainty needed for the lower bound.

Proof of Corollary 3.5. A learner can ignore the query until after submitting its predictor. Every upper-bound strategy in Theorem 3.1 is therefore still legal, and

$$\mathcal{R}_T^{\mathrm{qf}}(K, \rho) \le \mathcal{R}_T(K, \rho),$$

so all upper bounds for actions submitted before queries transfer directly.

For the reverse inequalities, consider the lower-bound distributions in Appendices C, O, and P. In the short-horizon construction of Lemma C.6, the query splits the current interval containing the target. Previous branch bits determine that interval, but not the next fair bit that determines its label. Announcing the split before prediction therefore leaves the Bayes loss at $1/2$.

The two-predictor construction already gives the learner more information. During both first-phase clocks, the genie in Appendix O announces the tree node before prediction. The node determines the query, but whether the clock stops is revealed only after loss. During later tests, completed branch bits determine the next midpoint. By Lemma O.3, the new branch bit is fair conditional on the complete visible history at the level's start. Announcing the query early preserves this testing bound. Both sources are actual earlier predictors, selected after the action and before the independent audit.

Thus Proposition O.1 and the lower bound using unrevealed target labels hold with $\mathcal{R}_T^{\mathrm{qf}}(K, \rho)$ in place of $\mathcal{R}_T(K, \rho)$. The parameter choice and optimization in Appendix P give the proper lower bound asserted in Eq. (4), including both transition windows. Querying unrevealed target labels

also gives the exact short-horizon identity. The same reasoning applies to the block optimization in Appendix S, so its global critical lower bound transfers as well. Combining these lower bounds with the earlier upper-bound strategies proves the corollary. □

M CLIPPED UPPER BOUNDS FOR FINITE HYPOTHESIS CLASSES

The clipped upper bound needs only a proper predictor that outputs one with probability one half at every input. This appendix proves the bound for any finite binary class with that property. Let $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ be finite. A proper stochastic action is a distribution $Q \in \Delta(\mathcal{H})$ whose score $p_Q(x) = \mathbb{E}_{h\sim Q} h(x)$ is its probability of predicting one. The target $h^* \in \mathcal{H}$ is fixed before round one, and all source, audit, and replay randomness follow Appendix C. Write $\mathcal{R}_T(\mathcal{H}, \rho)$ for the minimax deployment loss.

Formally, assume there is a distribution $\nu \in \Delta(\mathcal{H})$ such that

$$\mathbb{E}_{h\sim\nu} h(x) = \frac{1}{2} \qquad \text{for every } x \in \mathcal{X}. \tag{85}$$

M.1 PROPER AFFINE CLIPPING

Lemma M.1 (Clipping by mixing predictions)**.** *For every $\boldsymbol{w} \in \Delta(\mathcal{H})$ and $\alpha \in [0, 1/2]$, the distribution*

$$Q = (1-2\alpha)\boldsymbol{w} + 2\alpha\nu \tag{86}$$

is proper. If $r(x) = \mathbb{E}_{h\sim\boldsymbol{w}} h(x)$, then

$$p_Q(x) = \alpha + (1-2\alpha) r(x) \in [\alpha, 1-\alpha]. \tag{87}$$

Proof. The coefficients in Eq. (86) are nonnegative and sum to one. Linearity and Eq. (85) give Eq. (87). □

Proposition M.2 (When affine clipping preserves properness)**.** *Let $\mathcal{P} = \mathrm{conv}(\mathcal{H})$ be the set of proper prediction functions, and let $u(x) = 1/2$ for every $x \in \mathcal{X}$. Define $C_\alpha(r)(x) = \alpha + (1-2\alpha)r(x)$. If $u \in \mathcal{P}$, then $C_\alpha(\mathcal{P}) \subseteq \mathcal{P}$ for every $\alpha \in [0, 1/2]$. Conversely, legality of $C_{1/2}$ on any element of $\mathcal{P}$ implies $u \in \mathcal{P}$.*

Proof. If $u \in \mathcal{P}$, then $C_\alpha(r) = (1-2\alpha)r + 2\alpha u$ is a convex combination of two functions in $\mathcal{P}$. Conversely, $C_{1/2}(r) = u$ for every r. □

M.2 AN UPPER BOUND WITHOUT OBSERVING THE SELECTED PREDICTOR

Theorem M.3 (Clipped upper bound for a finite class)**.** *Assume Eq. (85). Fix $T \geq 2$, $\rho \in [0, 1]$, a prior $\pi \in \Delta(\mathcal{H})$, and a deterministic nonincreasing sequence $\alpha_2, \ldots, \alpha_T \in [0, 1/2]$. Define*

$$\gamma_t = 2\rho - 1 + 2(1-\rho)\alpha_t \tag{88}$$

and suppose $\gamma_t > 0$. There is one proper learner, depending only on $(\rho, \pi, \nu, (\alpha_t))$, such that for every $h^* \in \mathrm{supp}(\pi)$ and every legal adversary that knows the target,

$$\mathbb{E}L_T \leq \frac{1}{2} + \sum_{t=2}^{T} \alpha_t + \frac{4\log(1/\pi(h^*))}{\gamma_T^2}. \tag{89}$$

For $T = 1$, predicting one with probability one half has loss exactly $1/2$.

Proof. On round one deploy $Q_1 = \nu$, so $p_1 \equiv 1/2$, and initialize $\boldsymbol{w}_2 = \pi$. At round $t \geq 2$, before the query, form

$$r_t(x) = \mathbb{E}_{h\sim\boldsymbol{w}_t} h(x), \qquad Q_t = (1-2\alpha_t)\boldsymbol{w}_t + 2\alpha_t\nu.$$

Lemma M.1 gives the full deployed score

$$p_t(x) = \alpha_t + (1-2\alpha_t) r_t(x). \tag{90}$$

After observing (x_t, Y_t), set

$$s_t = Y_t - \frac{1}{2}, \qquad v_{t,h} = 2h(x_t) - 1, \qquad \eta_t = \frac{\gamma_t}{2}, \tag{91}$$

and update

$$w_{t+1}(h) = \frac{w_t(h)\exp\{\eta_t v_{t,h} s_t\}}{\mathbb{E}_{g\sim \boldsymbol{w}_t}\exp\{\eta_t v_{t,g} s_t\}}. \tag{92}$$

The learner implements this update from the observed label and its own weights. It does not need the audit status, the selected source, or the smallest and largest earlier scores.

To analyze the update, fix $h^* \in \mathrm{supp}(\pi)$. Abbreviate $y_t^* = h^*(x_t)$, $r_t = r_t(x_t)$, $e_t = |r_t - y_t^*|$, and $\bar{v}_t = \mathbb{E}_{h\sim \boldsymbol{w}_t} v_{t,h} = 2r_t - 1$. Condition on the enlarged pre-feedback field $\mathcal{G}_t$ from Appendix C, with h^* replacing a^*. This field includes the realized source. The nonincreasing clipping floors ensure that every earlier predictor has score in $[\alpha_t, 1-\alpha_t]$ at x_t. This controls the feedback mean regardless of which source the adversary selects.

If $y_t^* = 1$, the selected predictor's score is at least α_t, so

$$\mathbb{E}[s_t \mid \mathcal{G}_t] \geq \rho + (1-\rho)\alpha_t - \frac{1}{2} = \frac{\gamma_t}{2}, \qquad v_{t,h^*} - \bar{v}_t = 2e_t.$$

If $y_t^* = 0$, the selected predictor's score is at most $1 - \alpha_t$, so

$$\mathbb{E}[s_t \mid \mathcal{G}_t] \leq (1-\rho)(1-\alpha_t) - \frac{1}{2} = -\frac{\gamma_t}{2}, \qquad v_{t,h^*} - \bar{v}_t = -2e_t.$$

Hence

$$\mathbb{E}[(v_{t,h^*} - \bar{v}_t)s_t \mid \mathcal{G}_t] \geq \gamma_t e_t. \tag{93}$$

The same prediction error also bounds the second-order term in the potential calculation. Condition additionally on Y_t and draw an auxiliary $h \sim \boldsymbol{w}_t$. For $X_t = (v_{t,h} - \bar{v}_t)s_t$, one has $\mathbb{E}_{\boldsymbol{w}_t} X_t = 0$, $|X_t| \leq 1$, and

$$\mathbb{E}_{\boldsymbol{w}_t} X_t^2 = s_t^2 \,\mathrm{Var}_{\boldsymbol{w}_t}(v_{t,h}) = r_t(1-r_t) \leq e_t.$$

Since $0 < \gamma_t \leq 1$, we have $|\eta_t X_t| \leq 1/2$. The inequality $e^z \leq 1 + z + z^2$ for $|z| \leq 1$ gives

$$\log \mathbb{E}_{\boldsymbol{w}_t} e^{\eta_t X_t} \leq \eta_t^2 e_t. \tag{94}$$

Let $\Phi_t = \log(1/w_t(h^*))$. Centering the normalizer in Eq. (92) gives

$$\Phi_t - \Phi_{t+1} = \eta_t(v_{t,h^*} - \bar{v}_t)s_t - \log \mathbb{E}_{h\sim \boldsymbol{w}_t} e^{\eta_t (v_{t,h} - \bar{v}_t)s_t}.$$

Taking conditional expectation and using Eqs. (93) and (94),

$$\mathbb{E}[\Phi_t - \Phi_{t+1} \mid \mathcal{G}_t] \geq (\eta_t\gamma_t - \eta_t^2)e_t = \frac{\gamma_t^2}{4} e_t. \tag{95}$$

Summing, using $\Phi_{T+1} \geq 0$, and recalling $\Phi_2 = \log(1/\pi(h^*))$ yields

$$\sum_{t=2}^{T} \gamma_t^2 \mathbb{E} e_t \leq 4 \log \frac{1}{\pi(h^*)}. \tag{96}$$

Because (γ_t) is nonincreasing, the weighted error bound also gives

$$\sum_{t=2}^{T} \mathbb{E} e_t \leq \frac{4\log(1/\pi(h^*))}{\gamma_T^2}. \tag{97}$$

Finally, Eq. (90) gives $\ell_t \leq \alpha_t + e_t$ for either target label. Adding the exact first-round loss $1/2$ proves Eq. (89). □

M.3 All audit rates and the worst case over finite classes

Corollary M.4 (All-ρ finite-class upper bound)**.** *There is a universal constant C such that every finite binary class satisfying Eq.* (85) *satisfies*

$$\mathcal{R}_T(\mathcal{H}, \rho) \le C r_\rho(T, |\mathcal{H}|) \tag{98}$$

for every $T \ge 1$ and $\rho \in [0, 1]$.

Proof. Let $s = \log |\mathcal{H}|$ and take π uniform. The constant-half predictor gives the pathwise cap $T/2$. If $T = 1$, Eq. (85) implies $|\mathcal{H}| \ge 2$, so $s \ge \log 2$ and $r_\rho(1, |\mathcal{H}|)$ is bounded below by a positive universal constant for every ρ. The cap proves the claim in this case. Henceforth assume $T \ge 2$, so every expression involving $T - 1$ below is well-defined.

At $\rho = 1/2$, Eq. (88) gives $\gamma_t = \alpha_t$. If $T - 1 \ge 64s$, use the constant floor $\alpha_t = 2(s/(T-1))^{1/3} \le 1/2$. Theorem M.3 gives

$$\mathbb{E}L_T \le \frac{1}{2} + 3(T-1)^{2/3} s^{1/3}.$$

If $T - 1 < 64s$, the constant-half cap is $O(T^{2/3} s^{1/3})$. This proves the critical line.

For $\rho > 1/2$, write $\Delta_\rho = 2\rho - 1$. The critical floor remains valid because $\gamma_t \ge \alpha_t$, while the zero floor gives $\gamma_t = \Delta_\rho$ and loss at most $1/2 + 4s/\Delta_\rho^2$. Intersecting with the horizon cap proves the supercritical line.

For $\rho < 1/2$, write $\delta_\rho = 1 - 2\rho$ and $\alpha_0 = \delta_\rho/(1 + \delta_\rho)$. Use a constant floor $\alpha = \alpha_0 + b \le 1/2$. Then $\gamma = (1 + \delta_\rho)b$, and Theorem M.3 gives

$$\mathbb{E}L_T \le \frac{1}{2} + (T-1)(\alpha_0 + b) + \frac{4s}{(1+\delta_\rho)^2 b^2}.$$

Set

$$b_{\max} = \frac{1 - \delta_\rho}{2(1 + \delta_\rho)}, \qquad b^* = \left(\frac{8s}{(T-1)(1+\delta_\rho)^2} \right)^{1/3}.$$

If $b^* \le b_{\max}$, substitution gives $O(\delta_\rho T + T^{2/3} s^{1/3})$, using $\delta_\rho/2 \le \alpha_0 \le \delta_\rho$. If $b^* > b_{\max}$ and $\delta_\rho \ge 1/2$, then $\delta_\rho T \ge T/2$. If $b^* > b_{\max}$ and $\delta_\rho < 1/2$, then $b_{\max} \ge 1/6$, so $s/(T-1) \ge c$ and $T^{2/3} s^{1/3} \ge c' T$. Thus the constant-half cap proves the claimed horizon-capped order in every infeasible case. The endpoint $\rho = 0$ is also covered by the constant-half cap. □

The upper bound is uniform over class structure, but individual classes can be easier. To express the worst-case comparison, for an integer $M \ge 2$ let $\mathfrak{N}_M$ be the collection of all pairs $(\mathcal{X}, \mathcal{H})$ with $|\mathcal{H}| = M$ and the constant function $1/2$ in $\mathrm{conv}(\mathcal{H})$. Define

$$\mathfrak{W}_{M,T,\rho} = \sup_{(\mathcal{X},\mathcal{H}) \in \mathfrak{N}_M} \mathcal{R}_T(\mathcal{H}, \rho).$$

Corollary M.5 (Thresholds attain the worst-case rate over finite classes)**.** *Fix $\varepsilon \in (0, 1/2)$. Uniformly over $\rho \in [0, 1]$ and $1 \le T \le M^{1/2-\varepsilon}$,*

$$\mathfrak{W}_{M,T,\rho} = \Theta_\varepsilon\big(r_\rho(T, M)\big). \tag{99}$$

If $T \le \lfloor \log_2 M \rfloor$, then $\mathfrak{W}_{M,T,\rho} = T/2$ exactly.

Proof. Corollary M.4 gives the upper bound uniformly over $\mathfrak{N}_M$. For the lower bound, take the class of M thresholds on the ordered domain $[M - 1]$. It contains the constant-zero and constant-one hypotheses, so it admits the constant-half score. Theorem 3.1, with $K = M$, gives the matching lower bound. In the short-horizon range, every class satisfying Eq. (85) has value at most $T/2$ by the constant-half action, while the threshold class has value exactly $T/2$. □

N Complete proof of the clipped upper bounds

We prove Proposition 3.3 by specializing the finite-class argument to thresholds. We then describe how to implement its updates and choose the clipping floor at each audit rate.

N.1 THE LEARNER AND BOUNDS ON ALL EARLIER SCORES

Set $p_1(x) = 1/2$ for every x, which incurs loss exactly $1/2$, and initialize $\boldsymbol{w}_2$ uniformly on $[K]$. Let $(\alpha_t)_{t\geq 2}$ be deterministic, nonincreasing, and contained in $[0, 1/2]$. At round $t \geq 2$, define

$$r_t(x) = \sum_{a\leq x} w_t(a), \qquad p_t(x) = \alpha_t + (1 - 2\alpha_t)r_t(x).$$

The predictor is submitted before the query. After observing (x_t, Y_t), put

$$\gamma_t = 2\rho - 1 + 2(1 - \rho)\alpha_t, \qquad s_t = Y_t - \frac{1}{2}, \qquad \eta_t = \frac{\gamma_t}{2}.$$

Assume $\gamma_t > 0$. With $v_{t,a} = 2h_a(x_t) - 1$, update

$$w_{t+1}(a) = \frac{w_t(a)e^{\eta_t v_{t,a} s_t}}{\sum_b w_t(b)e^{\eta_t v_{t,b} s_t}}. \tag{100}$$

Every coordinate remains positive, so the learner keeps a positive weight on the target throughout.

Lemma N.1 (Properness and bounds on earlier scores)**.** *Every p_t is a full proper stochastic threshold. Moreover, every historical score at the current query lies in $[\alpha_t, 1 - \alpha_t]$.*

$$\alpha_t \leq p_s(x_t) \leq 1 - \alpha_t \qquad \textit{for every } s < t. \tag{101}$$

Proof. By Lemma C.1, affine clipping is proper: it mixes $\boldsymbol{w}_t$ with equal mass on the two constant endpoint hypotheses. The first score $1/2$ belongs to every interval $[\alpha_t, 1 - \alpha_t]$. For $2 \leq s < t$, the floors are nonincreasing, so $\alpha_s \geq \alpha_t$ and $p_s(x_t) \in [\alpha_s, 1 - \alpha_s] \subseteq [\alpha_t, 1 - \alpha_t]$. □

N.2 APPLYING THE FINITE-CLASS UPPER BOUND

Apply Theorem M.3 with $\mathcal{H}_K = \{h_a : a \in [K]\}$, the uniform prior, and $\nu = (\delta_{h_1} + \delta_{h_K})/2$. Its mixture and update are exactly those above, as are its bounds on earlier prediction probabilities. Hence

$$\mathbb{E}L_T \leq \frac{1}{2} + \sum_{t=2}^{T} \alpha_t + \frac{4\log K}{\gamma_T^2}. \tag{102}$$

Setting $\alpha_t \equiv \alpha$ proves Proposition 3.3, while Eq. (102) also allows nonincreasing floors. Equations (93)–(96) show how the feedback signal and the error-dependent second-order term give this bound. That potential argument applies to every class satisfying the stated assumption. The ordered structure of thresholds is used in the implementation below and in the proper lower bound.

N.3 EFFICIENT THRESHOLD IMPLEMENTATION

Proposition N.2 (Segment-tree implementation)**.** *On a real-RAM, the threshold learner's current state has an implicit representation using $O(K)$ space. It supports evaluation of $p_t(x)$ at any x and the update after (x_t, Y_t) in $O(\log K)$ time. If the protocol implementation must retain queryable copies of all emitted actions, a persistent version uses $O(\log K)$ new nodes per round and supports each historical score query in $O(\log K)$ time.*

Proof. Store unnormalized positive weights $\widetilde{w}_t(a)$ at the leaves of a lazy-propagation segment tree. Each node stores the sum of its interval and a pending multiplicative factor. A prefix-sum query returns $r_t(x) = \sum_{a\leq x} \widetilde{w}_t(a)/\sum_a \widetilde{w}_t(a)$ in $O(\log K)$ time, hence also $p_t(x)$.

At the observed query x_t, one has $v_{t,a} = 1$ for $a \leq x_t$ and $v_{t,a} = -1$ for $a > x_t$. Equation (100) therefore multiplies the prefix $[1, x_t]$ by $e^{\eta_t s_t}$ and the suffix $[x_t + 1, K]$ by $e^{-\eta_t s_t}$. Two lazy range-multiplication operations take $O(\log K)$ time. Global normalization is unnecessary because all score queries divide by the root sum.

To evaluate earlier predictors, use the standard path-copying persistent variant. Each range multiplication copies only the $O(\log K)$ nodes visited by the update and leaves every prior root unchanged. Store the contemporaneous floor α_t with the root of version t. A prefix-sum query at an old root then recovers $r_t(x)$ and hence $p_t(x)$ in $O(\log K)$ time. The learner's live working state remains $O(K)$, while storing T deployed functions requires $O(K + T \log K)$ total nodes. In the information-theoretic protocol, the environment retains earlier functions once they are deployed. The learner's update rule itself does not need to query them. □

N.4 Optimization for $\rho \geq 1/2$

Let $s = \log K$. At $\rho = 1/2$, $\gamma_t = \alpha_t$. Choose

$$\alpha_t = \min\left\{\frac{1}{4}, \left(\frac{s}{t}\right)^{1/3}\right\}. \tag{103}$$

If $T \geq 64s$, then $\alpha_T = (s/T)^{1/3}$ and

$$\sum_{t=2}^{T} \alpha_t \leq 16s + s^{1/3} \sum_{t=\lceil 64s \rceil}^{T} t^{-1/3} \leq 16s + \frac{3}{2} T^{2/3} s^{1/3}.$$

The inverse-square term in Eq. (102) is $4T^{2/3}s^{1/3}$. Enlarging a universal constant gives

$$\mathcal{R}_T(K, 1/2) \leq C T^{2/3} s^{1/3}. \tag{104}$$

For $T < 64s$, the constant-half action gives the required horizon cap.

Now suppose $\rho > 1/2$ and write $\Delta_\rho = 2\rho - 1$. Since $\gamma_t \geq \alpha_t$, the same schedule gives the critical upper bound. The zero floor gives $\gamma_t = \Delta_\rho$ and

$$\mathbb{E}L_T \leq \frac{1}{2} + \frac{4s}{\Delta_\rho^2}.$$

Intersecting these bounds with $T/2$ proves the supercritical line of Theorem 3.1.

N.5 Optimization for $\rho < 1/2$

The subcritical floor and its infeasible cases were already optimized in the proof of Corollary M.4. Taking $|\mathcal{H}_K| = K$ there gives the subcritical upper half of Theorem 3.1, including the constant-half cap at $\rho = 0$. The decreasing schedule above gives an anytime implementation at the critical rate. For the other choices, the finite-class optimization already supplies the required threshold bound.

O Lower bound by producing and repeatedly reusing two predictors

Proposition O.1 (Lower bound from repeatedly reusing two earlier predictors)**.** *There are universal constants $c > 0, m_0$ such that, if integers $m \geq m_0$, $1 \leq d \leq m/2$ satisfy $3m \leq T$ and $(m+1)^2 2^d \leq K$, then, with $R_{m,d} = (dm^2)^{1/3}$,*

$$\mathcal{R}_T(K, \rho) \geq c \begin{cases} \min\{R_{m,d}, d/(2\rho-1)^2\}, & \rho \geq 1/2, \\ \min\{m, m(1-2\rho) + R_{m,d}\}, & \rho < 1/2. \end{cases} \tag{105}$$

Division by zero is $+\infty$. The target is fixed at time zero, sources are actual past actions, and charged queries may be announced before prediction.

The proof has two phases. The first phase produces two predictors through the learner's actual actions. Avoiding predictions that could later hide the target incurs loss in this phase. The second phase reuses the selected pair while testing unresolved target labels. Both phases use one target distribution fixed before interaction. The adversary receives no external predictors.

O.1 A target prior with two stages for selecting the sources

Fix integers m, d satisfying the feasibility conditions of Proposition O.1 and set $\alpha = 1/m$. Let

$$\mathcal{C}_m = \{1, 2, \ldots, m, \infty\}.$$

The two clocks determine the stages at which we select the predictors. Sample $I, J \in \mathcal{C}_m$ independently, with

$$\mathbb{P}(I = s) = \mathbb{P}(J = s) = \alpha(1-\alpha)^{s-1} \quad (1 \leq s \leq m), \qquad \mathbb{P}(I = \infty) = \mathbb{P}(J = \infty) = (1-\alpha)^m. \tag{106}$$

We arrange $(m+1)^2$ consecutive superblocks in an ordered binary tree. At stage s of the first chain, the child $\{I = s\}$ lies left of the child $\{I > s\}$. The resulting fixed-I subtrees are ordered $1, 2, \ldots, m, \infty$ from left to right. Inside each such subtree the order is reversed: $\{J > s\}$ lies left of $\{J = s\}$. Hence the fixed-J superblocks are ordered $\infty, m, \ldots, 1$ from left to right. Each node has two contiguous sets of cutoff leaves.

Replace each superblock by a contiguous block of 2^d cutoff leaves. The condition $(m+1)^2 2^d \le K$ embeds the used leaves in $[K]$. Conditional on (I, J), select the target cutoff uniformly in the corresponding block. Equivalently, sample d independent fair branch bits for a balanced binary tree inside that block. The clocks and all branch bits are drawn at time zero and determine one fixed cutoff a^*.

O.2 Additional information given to the learner

During the first phase, a genie announces the current tree node before each action. After the prediction loss, it reveals which child contains the target. The learner therefore does not know whether the clock stops before making its prediction. Each within-block branch bit also remains hidden until its testing level is complete. A lower bound for learners receiving these extra messages holds in the original protocol as well.

The adversary queries the announced split after seeing the action. On each first-phase round after round one, it selects the actual first predictor for replay. This choice does not depend on the target, and the predictor's values are not needed in the loss calculation. Conditional on reaching a clock stage, all remaining clock values give the same distribution of previous queries, labels, and feedback. Conditioning on what the learner has seen therefore preserves the truncated geometric hazard:

$$\mathbb{P}(I = s \mid I \ge s, \text{visible history}) = \alpha, \qquad \mathbb{P}(J = s \mid J \ge s, \text{visible history}) = \alpha. \tag{107}$$

Complete source and continuation rules. On each first-phase round after round one, use $S_t = 1$. The first label is forced truthful. Each clock runs until its sampled finite value is reached or all m stages have passed. Run the second clock even if $I = \infty$. Write $i_0 = \min\{I, m\}$ and $j_0 = \min\{J, m\}$, with $\min\{\infty, m\} = m$. The first phase ends at $\tau = i_0 + j_0$. If $I = \infty$ or $J = \infty$, use $x_t = 1$ and $S_t = 1$ on all remaining rounds and ignore their nonnegative losses. Otherwise begin testing the within-block target bits at round $\tau + 1$. The two selected sources then have indices $\tau_L = I < \tau_U = I + J = \tau$, so both are strictly historical. Denote their prediction functions by $p^{\mathrm{lo}} = p_{\tau_L}$ and $p^{\mathrm{hi}} = p_{\tau_U}$. For each internal pivot x of the selected block, let

$$z_x^{\mathrm{lo}} = p^{\mathrm{lo}}(x), \qquad z_x^{\mathrm{hi}} = p^{\mathrm{hi}}(x).$$

On every successful-clock history use the testing rule, whether or not the analysis event F defined below occurs. For audit rates below one half, use the endpoint mixture only when $z_x^{\mathrm{lo}} \le z_x^{\mathrm{hi}}$ and $z_x^{\mathrm{hi}} - z_x^{\mathrm{lo}} \ge \rho/(1-\rho)$. Otherwise replay the predictor selected by the first clock under label one and that selected by the second clock under label zero. Thus the mixture is used only when its probabilities are valid, and the rule remains legal on histories outside F as well. After these tests, fill any remaining rounds with $x_t = 1, S_t = 1$. The adversary never needs to test F or use a predictor that was not actually produced.

Conditional hazard. The histories in Eq. (107) are pre-action, learner-visible histories, including genie signals and the learner's independent private seed. They do not include the fixed target itself. On $I \ge s$, every possible remaining clock value $i \in \{s, \ldots, m, \infty\}$ induces the same preceding nodes where the clock has not stopped, queries, true labels, and source-one feedback kernels. Induction over these rounds shows that the visible transcript has the same likelihood under every such i. Its conditional law therefore does not change the geometric hazard. The same argument applies to $J \ge s$ after conditioning on the completed first-clock value. These equalities hold almost surely on supported histories. Current actions and scores can be included because the learner's conditional action distribution is the same in all surviving worlds. Independently revealing source or audit information after loss preserves this likelihood identity: during the first phase an index report is the constant one, a score report is $p_1(x_t)$, and an audit report has the same independent coin law in every surviving world.

O.3 Relating earlier prediction loss to the selected predictors' scores

At a first-clock node with $I \geq s$, let q_s be the learner's realized score at the announced pivot. The child $I = s$ is left and has label one. The child $I > s$ is right and has label zero. Conditional on the pre-action history and q_s, the deployment loss is

$$\alpha(1 - q_s) + (1 - \alpha)q_s = \alpha + (1 - 2\alpha)q_s. \tag{108}$$

If $I = s$, the final target block and all its internal pivots lie left of this pivot. The predictor submitted at round I is nondecreasing, so

$$p_I(x) \leq q_I \quad \text{for every internal pivot } x \text{ of the selected block.} \tag{109}$$

Define $\delta_{\mathrm{lo}} = q_I$ on $\{I \leq m\}$. This selects the predictor from actual round $\tau_L = I$.

For the reversed second clock, let r_s be the score at its current pivot. The child $J > s$ is left and has label one. The child $J = s$ is right and has label zero. The conditional loss is

$$(1 - \alpha)(1 - r_s) + \alpha r_s = \alpha + (1 - 2\alpha)(1 - r_s). \tag{110}$$

If $J = s$, the corresponding actual round is $\tau_U = \min\{I, m\} + J$, where $\min\{\infty, m\} = m$. The final block lies wholly to the right of the pivot, so

$$p_{\tau_U}(x) \geq r_J \quad \text{for every internal pivot } x \text{ of the selected block.} \tag{111}$$

Define $\delta_{\mathrm{hi}} = 1 - r_J$ on $\{J \leq m\}$. Set each defect to zero outside its defining event.

Let $L_{\mathrm{lo}}, L_{\mathrm{hi}}$ be the learner's expected losses during the two clocks, and set $L_{\mathrm{build}} = L_{\mathrm{lo}} + L_{\mathrm{hi}}$. Thus L_{build} is its actual expected loss in the first phase. The next lemma relates this loss to the selected predictors' scores.

Lemma O.2 (Early loss controls the selected predictors)**.** *Let $E = \{I \leq m, J \leq m\}$. Then*

$$\mathbb{E}[(\delta_{\mathrm{lo}} + \delta_{\mathrm{hi}})\mathbf{1}_E] \leq \frac{2L_{\mathrm{build}}}{m}. \tag{112}$$

Moreover,

$$\mathbb{P}(E) = \left(1 - (1 - 1/m)^m\right)^2 \geq \pi_0 := (1 - e^{-1})^2. \tag{113}$$

Proof. By predictability of q_s and Eq. (107),

$$\begin{aligned}\mathbb{E}[\delta_{\mathrm{lo}}\mathbf{1}_{\{I \leq m\}}] &= \sum_{s=1}^{m} \mathbb{E}[q_s \mathbf{1}_{\{I=s\}}] \\ &= \alpha \mathbb{E} \sum_{s \leq I \wedge m} q_s.\end{aligned}$$

Equation (108) and $1 - 2\alpha \geq 1/2$ for $m \geq 4$ imply

$$\mathbb{E}[\delta_{\mathrm{lo}}\mathbf{1}_{\{I \leq m\}}] \leq \frac{\alpha L_{\mathrm{lo}}}{1 - 2\alpha}.$$

The same argument conditionally inside every fixed-I subtree gives

$$\mathbb{E}[\delta_{\mathrm{hi}}\mathbf{1}_{\{J \leq m\}}] \leq \frac{\alpha L_{\mathrm{hi}}}{1 - 2\alpha}.$$

Restricting both nonnegative terms to E only decreases them. Since $\alpha/(1 - 2\alpha) \leq 2/m$, Eq. (112) follows. Equation (113) follows from independence and $(1 - 1/m)^m \leq e^{-1}$. □

We now choose a single event on which this control holds throughout the block:

$$F = E \cap \left\{ \delta_{\mathrm{lo}} + \delta_{\mathrm{hi}} \leq \frac{4L_{\mathrm{build}}}{\pi_0 m} \right\}. \tag{114}$$

If $L_{\mathrm{build}} = 0$, Lemma O.2 and nonnegativity imply $(\delta_{\mathrm{lo}} + \delta_{\mathrm{hi}})\mathbf{1}_E = 0$ almost surely, so $F = E$ almost surely. If $L_{\mathrm{build}} > 0$, apply Markov's inequality and Lemma O.2. Both cases give

$$\mathbb{P}(F) \geq \pi_0/2. \tag{115}$$

On F, both selected predictors were actually deployed. Equations (109), (111), and (114) yield

$$1 - z_x^{\mathrm{hi}} + z_x^{\mathrm{lo}} \le \delta_{\mathrm{lo}} + \delta_{\mathrm{hi}} \le \frac{4L_{\mathrm{build}}}{\pi_0 m}. \tag{116}$$

The same event F controls all internal pivots because proper prediction probabilities are nondecreasing. The first phase has revealed none of the d within-block branch bits. Its queries lie between superblocks, where all 2^d cutoffs in the selected block have the same label. Source choices and genie messages are also the same across those cutoffs.

Lemma O.3 (Unresolved target bits remain conditionally fair)**.** *Conditional on the complete first-phase transcript and any completed descendant levels, the next within-block branch bit is fair and independent of the learner's visible history before that level.*

Proof. The descendant bits are sampled independently at time zero. First-phase queries lie between superblocks, so their labels, source rules, feedback kernels, and genie signals do not depend on any descendant bit. At a completed level only its current bit is revealed, while its feedback depends on that bit and the fixed scores of earlier predictors. Induction over the levels proves the claim. □

O.4 Reusing the sources across unresolved target bits

Set

$$h = \left\lfloor \frac{m}{d} \right\rfloor. \tag{117}$$

Since $d \le m/2$,

$$dh \ge m - d \ge m/2. \tag{118}$$

In the second phase, traverse the balanced tree in the selected block. Query each level's midpoint for exactly h rounds, and only then reveal the child containing the target. Each level starts with a conditionally fair branch bit. Its feedback can reveal that bit, but depends on no deeper bit: it uses only the current bit and the two selected predictors' scores at the midpoint. This lets us add the losses from all d levels while paying for the two predictors only once.

If the target lies in the left child of a midpoint, its threshold cutoff is smaller than the separating domain point, so the true label is one. The adversary replays p^{lo}. If the target lies in the right child, the true label is zero and the adversary replays p^{hi}. These are actual earlier predictors. Their indices are selected before each audit coin.

O.5 Audit rates at or above one half

Assume $\rho \ge 1/2$ and put $\Delta_\rho = 2\rho - 1$. At a current internal pivot x, the two feedback parameters under the fair branch bit are

$$\theta_1 = \rho + (1-\rho) z_x^{\mathrm{lo}}, \qquad \theta_0 = (1-\rho) z_x^{\mathrm{hi}}. \tag{119}$$

Their separation is

$$\theta_1 - \theta_0 = \Delta_\rho + (1-\rho)(1 - z_x^{\mathrm{hi}} + z_x^{\mathrm{lo}}) \le G,$$

where

$$G = \Delta_\rho + \frac{4(1-\rho)L_{\mathrm{build}}}{\pi_0 m}. \tag{120}$$

If $G \le 1/4$, then $\theta_0 \le 1/2 \le \theta_1$ and their distance is at most $1/4$. Hence both parameters belong to $[1/4, 3/4]$. Lemma C.4 gives at least $c_0 \min\{h, G^{-2}\}$ expected loss at this level. If $G > 1/4$, the same conclusion follows from the first-action case of that lemma after reducing c_0.

Condition on the first-phase history in F and apply the lemma successively to the d unresolved branch bits. Using Eq. (118), the tests cost at least

$$c_1 d \min\{h, G^{-2}\} \ge c_2 \min\{m, dG^{-2}\}. \tag{121}$$

Multiply by $\mathbb{P}(F) \ge \pi_0/2$ and add the first-phase loss. The resulting lower bound on Bayes risk is

$$L_{\mathrm{build}} + c_3 \min\left\{ m, \frac{d}{(\Delta_\rho + bL_{\mathrm{build}}/m)^2} \right\} \tag{122}$$

for a universal $b > 0$.

Let

$$r_0 = \min\left\{(dm^2)^{1/3}, \frac{d}{\Delta_\rho^2}\right\}. \tag{123}$$

Because $d \le m$, $r_0 \le m$. Also, from the two inequalities defining the minimum,

$$\Delta_\rho \le \sqrt{d/r_0}, \qquad r_0/m \le \sqrt{d/r_0}. \tag{124}$$

The second relation is equivalent to $r_0^3 \le dm^2$. These inequalities separate two possibilities. If $L_{\mathrm{build}} \ge \varepsilon r_0$ for a sufficiently small universal ε, the learner has already incurred $\Omega(r_0)$ loss. Otherwise,

$$\Delta_\rho + bL_{\mathrm{build}}/m \le (1 + b\varepsilon)\sqrt{d/r_0},$$

so the second term of Eq. (122) is $\Omega(r_0)$. This proves Eq. (105), including the critical case $\Delta_\rho = 0$.

O.6 Audit rates below one half

Assume $\rho < 1/2$ and write $\delta_\rho = 1 - 2\rho$. Recall that $\delta_{\mathrm{lo}}, \delta_{\mathrm{hi}}, z_x^{\mathrm{lo}}$, and $1 - z_x^{\mathrm{hi}}$ are nonnegative on the common event F, with

$$0 \le z_x^{\mathrm{lo}} \le \delta_{\mathrm{lo}}, \qquad 0 \le 1 - z_x^{\mathrm{hi}} \le \delta_{\mathrm{hi}}, \qquad \delta_{\mathrm{lo}} + \delta_{\mathrm{hi}} \le \frac{4L_{\mathrm{build}}}{\pi_0 m}.$$

Put $b_0 = 4/\pi_0$, $R = (dm^2)^{1/3}$, and

$$r_- := \min\{m, m\delta_\rho + R\}.$$

Since $d \le m/2$, we have $R \le m$. Choose a universal $c^* \le 1/(4b_0)$. If $L_{\mathrm{build}} \ge c^* m$, the first phase alone contributes $\Omega(m) = \Omega(r_-)$, so assume henceforth that

$$L_{\mathrm{build}} < c^* m. \tag{125}$$

If $L_{\mathrm{build}} = 0$, then the bounds below hold with zero defects directly. Otherwise the definition of F follows from the Markov argument above.

On F, the individual defect bounds now give

$$z_x^{\mathrm{lo}} \le b_0 L_{\mathrm{build}}/m \le \frac{1}{4}, \qquad 1 - z_x^{\mathrm{hi}} \le b_0 L_{\mathrm{build}}/m \le \frac{1}{4}. \tag{126}$$

Thus $z_x^{\mathrm{lo}} \le z_x^{\mathrm{hi}}$, so the selected predictors' scores are ordered at every internal pivot. The event F and complete first-phase transcript are independent of the descendant bits sampled at time zero. After earlier completed bits are revealed, the next bit is still fair at the start of its test.

For this pair, Proposition C.3 shows that the two feedback distributions can be made identical if and only if

$$z_x^{\mathrm{hi}} - z_x^{\mathrm{lo}} \ge \frac{\rho}{1-\rho}, \qquad \text{equivalently} \qquad 1 - z_x^{\mathrm{hi}} + z_x^{\mathrm{lo}} \le \frac{\delta_\rho}{1-\rho}. \tag{127}$$

When the intervals overlap, select the low source under label one. Under label zero, randomize between the two selected predictors to obtain mean score $z_x^{\mathrm{lo}} + \rho/(1-\rho)$. Draw this hidden source before the independent audit coin. Averaging over the source choice makes the two visible Bernoulli laws agree, so Lemma C.5 gives $h/2$ expected loss.

When Eq. (127) fails, replaying the low source under label one and the high source under label zero gives the positive separation

$$\Gamma_x = (1-\rho)(1 - z_x^{\mathrm{hi}} + z_x^{\mathrm{lo}}) - \delta_\rho, \qquad 0 < \Gamma_x \le [b_0 L_{\mathrm{build}}/m - \delta_\rho]_+. \tag{128}$$

First consider $1/4 \le \rho < 1/2$. The two feedback means in the positive-separation branch are

$$\theta_1 = \rho + (1-\rho)z_x^{\mathrm{lo}}, \qquad \theta_0 = (1-\rho)z_x^{\mathrm{hi}}.$$

Equation (126) gives

$$\frac{1}{4} \le \theta_1 \le \frac{1}{4} + \frac{3\rho}{4} \le \frac{5}{8}, \qquad \frac{3}{8} \le \theta_0 \le 1 - \rho \le \frac{3}{4}. \tag{129}$$

Both parameters lie in $[1/4, 3/4]$. Set $G = [b_0 L_{\mathrm{build}}/m - \delta_\rho]_+$. Equation (125) gives $G \le 1/4$. At each level, identical feedback distributions give $h/2$ loss. Otherwise Lemma C.4, Eq. (128), and Eq. (129) give at least $c_0 \min\{h, G^{-2}\}$ loss. With $1/0^2 = +\infty$, both branches have this common lower bound.

Condition on the first-phase history in F and apply the bound successively to all d unresolved branch bits, revealing each completed bit before the next level. Using $dh \ge m/2$, $\mathbb{P}(F) \ge \pi_0/2$, and $b = b_0$ gives

$$L_{\mathrm{build}} + c_4 \min\left\{m, \frac{d}{[bL_{\mathrm{build}}/m - \delta_\rho]_+^2}\right\}. \tag{130}$$

We optimize this bound. If $m\delta_\rho \ge R$, then $r_- = \Theta(\min\{m, m\delta_\rho\})$. Unless $L_{\mathrm{build}} \ge \varepsilon r_-$, the inequality $r_- \le m\delta_\rho + R \le 2m\delta_\rho$ gives

$$bL_{\mathrm{build}}/m \le 2b\varepsilon\delta_\rho < \delta_\rho$$

for sufficiently small universal ε. The positive part vanishes. Equation (116) then satisfies the overlap condition, and the block pays $\Omega(m) \ge \Omega(r_-)$. If $m\delta_\rho < R$, then $r_- = \Theta(R)$. Unless $L_{\mathrm{build}} \ge \varepsilon R$,

$$[bL_{\mathrm{build}}/m - \delta_\rho]_+ \le b\varepsilon R/m,$$

so

$$\frac{d}{[bL_{\mathrm{build}}/m - \delta_\rho]_+^2} \ge \frac{dm^2}{b^2\varepsilon^2 R^2} = \frac{R}{b^2\varepsilon^2}.$$

Here $R^3 = dm^2$. The tests cost $\Omega(R) = \Omega(r_-)$.

It remains to treat $0 \le \rho < 1/4$. In this range

$$\frac{\delta_\rho}{1-\rho} = \frac{1-2\rho}{1-\rho} \ge \frac{2}{3} > \frac{1}{4}$$

and $r_- = \Theta(m)$. Under the standing assumption on L_{build}, Eq. (116) is at most $b_0 L_{\mathrm{build}}/m < 1/4$, so the feedback distributions can be made identical at every internal pivot on F. The tests therefore cost $\Omega(m)$. The case of larger first-phase loss was handled at the start. This proves Eq. (105) for all $\rho < 1/2$.

O.7 Checking the horizon, target support, and protocol

The two stages that select the sources use at most $2m$ rounds. Testing uses $dh \le m$ more rounds, so the entire construction fits within $3m \le T$. Any remaining rounds follow the legal continuation specified above. The target support contains exactly $(m+1)^2 2^d$ cutoff leaves and therefore embeds in $[K]$.

The target (I, J, internal bits) is sampled before round one, independently of the learner's private randomness. The first feedback is forced truthful. Every testing source is one of the two actual selected predictors, chosen before the independent audit coin. Revealing first-phase queries before prediction gives the learner extra information. During each later test, the query is the same under both values of the current branch bit until all its charged predictions are complete. Conditioning on the learner's independent private seed completes the Yao reduction and proves Proposition O.1 with actions submitted before queries.

P Derivation of the sharp minimax rates

This appendix derives the proper part of Theorem 3.1 from Proposition 3.3 and the proper two-predictor construction in Proposition O.1. Short horizons use labels that have not yet been revealed. For longer horizons, we choose parameters so that the first phase produces two predictors while leaving $\Theta_\varepsilon(\log K)$ unresolved target decisions. We then compare the rates for each audit regime. Appendix V uses the same parameters in Proposition 3.4 to obtain the improper lower bound. The proper construction remains separate because Corollary 3.5 permits the query to be revealed before prediction.

Lemma P.1 (Short-horizon law)**.** *Fix $\varepsilon \in (0, 1/2)$ and the sufficiently large constant C_ε used below. If $1 \le T < C_\varepsilon \log K$, then uniformly over $\rho \in [0, 1]$,*

$$\mathcal{R}_T(K, \rho) = \Theta_\varepsilon(T), \tag{131}$$

and the rate $r_\rho(T, K)$ in Eq. (3) is $\Theta_\varepsilon(T)$. Moreover, if $T \le \lfloor \log_2 K \rfloor$, then $\mathcal{R}_T(K, \rho) = T/2$ exactly.

Proof. The constant proper CDF $p_t \equiv 1/2$ gives the pathwise upper bound $T/2$. Lemma C.6 gives

$$\mathcal{R}_T(K, \rho) \ge \frac{1}{2} \min\{T, \lfloor \log_2 K \rfloor\}.$$

For $K \ge 2$, $\lfloor \log_2 K \rfloor \ge (\log K)/(2 \log 2)$. Combining this with $T < C_\varepsilon \log K$ proves Eq. (131), and the two bounds coincide when $T \le \lfloor \log_2 K \rfloor$.

It remains to compare with Eq. (3). Put $s_K = \log K$ and $R = T^{2/3} s_K^{1/3}$. The short-horizon condition gives $R \ge C_\varepsilon^{-1/3} T$. For $\rho > 1/2$, also $s_K/(2\rho - 1)^2 \ge s_K > T/C_\varepsilon$. For $\rho < 1/2$, the second entry in the minimum is at least R. Every displayed rate is therefore bounded above by T and below by a positive ε-dependent multiple of T. □

We now treat $C_\varepsilon \log K \le T \le K^{1/2-\varepsilon}$. The next lemma checks that the first phase fits the available time and finite set of thresholds, while leaving enough target uncertainty for later tests.

Lemma P.2 (Unresolved target decisions after the first phase)**.** *Fix $\varepsilon \in (0, 1/2)$. Let $s_K = \log K$ and suppose*

$$C_\varepsilon s_K \le T \le K^{1/2-\varepsilon}.$$

With

$$m = \lfloor T/6 \rfloor, \qquad d = \min\left\{ \lfloor m/2 \rfloor, \left\lfloor \log_2 \frac{K}{(m+1)^2} \right\rfloor \right\},$$

there are constants $c_\varepsilon, C'_\varepsilon > 0$ such that, for C_ε sufficiently large,

$$c_\varepsilon s_K \le d \le C'_\varepsilon s_K, \qquad m = \Theta(T). \tag{132}$$

The feasibility conditions of Proposition O.1 hold.

Proof. The upper bound $d \le \log_2 K = s_K / \log 2$ is immediate. Since $T \ge C_\varepsilon s_K$, choosing C_ε large gives

$$\lfloor m/2 \rfloor \ge c C_\varepsilon s_K.$$

For the number of branch bits allowed by the target support, use $m + 1 \le T$ for large parameters:

$$\begin{aligned} \log_2 \frac{K}{(m+1)^2} &\ge \log_2 K - 2 \log_2 T \\ &\ge \log_2 K - 2(1/2 - \varepsilon) \log_2 K = 2\varepsilon \log_2 K. \end{aligned}$$

Rounding changes this by at most one, so both entries in the minimum are at least $c_\varepsilon s_K$. Enlarging C_ε also ensures $m \ge m_0$, $d \le m/2$, and $3m \le T$. Finally, the definition of d gives $(m+1)^2 2^d \le K$, as required for a target prior supported on the given thresholds. □

By Lemma P.2,

$$(dm^2)^{1/3} = \Theta_\varepsilon(T^{2/3} s_K^{1/3}). \tag{133}$$

Because $T \ge C_\varepsilon s_K$, the critical scale is at most a constant multiple of T:

$$T^{2/3} s_K^{1/3} = T(s_K/T)^{1/3} \le C_\varepsilon^{-1/3} T.$$

Thus the lower bound from the block construction is consistent with the explicit horizon cap throughout this range.

P.1 THE CASE $\rho = 1/2$

At $\rho = 1/2$, Eq. (105) has no plateau term and gives

$$\mathcal{R}_T(K, 1/2) \geq c(dm^2)^{1/3} \geq c_\varepsilon T^{2/3} s_K^{1/3}.$$

Proposition 3.3 gives the matching upper bound. Taking its minimum with the constant-half strategy's horizon bound proves the proper critical case.

P.2 THE CASE $\rho > 1/2$

Let $\Delta_\rho = 2\rho - 1 > 0$. The block lower gives

$$\begin{aligned}\mathcal{R}_T(K, \rho) &\geq c \min\left\{(dm^2)^{1/3}, \frac{d}{\Delta_\rho^2}\right\} \\ &\geq c_\varepsilon \min\left\{T^{2/3} s_K^{1/3}, \frac{s_K}{\Delta_\rho^2}\right\}.\end{aligned}$$

The upper bound contains the same two terms and the horizon cap, proving the proper case above one half. The growing term reaches the T-independent bound when

$$T^{2/3} s_K^{1/3} = s_K/\Delta_\rho^2,$$

which is equivalent to

$$T = s_K/\Delta_\rho^3 \tag{134}$$

up to constants. At $\Delta_\rho = 1$, the plateau is s_K, consistent with the perfect-audit endpoint.

P.3 THE CASE $\rho < 1/2$

Let $\delta_\rho = 1 - 2\rho > 0$. The block lower and Eq. (133) give

$$\begin{aligned}\mathcal{R}_T(K, \rho) &\geq c \min\{m, m\delta_\rho + (dm^2)^{1/3}\} \\ &\geq c_\varepsilon \min\{T, \delta_\rho T + T^{2/3} s_K^{1/3}\}.\end{aligned}$$

The upper bound in the third line of Eq. (3) matches it. The linear and critical terms meet when

$$\delta_\rho T = T^{2/3} s_K^{1/3}, \qquad \text{or} \qquad \delta_\rho = (s_K/T)^{1/3}.$$

For every fixed $\rho < 1/2$, the value is therefore linear in T throughout the stated range once the horizon is sufficiently large. At $\rho = 0$, the order statement agrees with the exact endpoint in Appendix R.

P.4 UNIFORMITY AND SCOPE

The upper-bound constants are universal. The lower bound depends on ε only through $d = \Theta_\varepsilon(s_K)$ and the short-horizon comparison. All constants used in the first phase, the label tests, and the optimization are uniform in ρ. Together with Lemma P.1, this proves the proper part of Theorem 3.1.

The restriction $T \leq K^{1/2-\varepsilon}$ leaves enough thresholds inside each of the $(m+1)^2$ first-phase blocks to preserve a constant fraction of the original $\log K$ target uncertainty. When T is of order $\sqrt{K}$, the same choice can leave only $O(1)$ branch bits and no longer matches the clipped upper bound's factor $s_K^{1/3}$. Appendix S optimizes the finite-domain construction separately to prove Corollary S.1.

Q LEARNING UNDER A FIXED BREADTH-FIRST QUERY ORDER

Proposition Q.1 (Proper upper bound for the restricted query order)**.** *For $T \geq 2$ in the restricted breadth-first game described below, there is a deterministic proper learner satisfying*

$$\sup_{a^* \in [K]} \mathbb{E}_{a^*} L_T = O(\log^2(T+1)). \tag{135}$$

We prove the proposition for the following restricted adversary. The case $T = 1$ is immediate, so assume $T \geq 2$ and let $K = T^2$. Arrange the K consecutive cutoff cells in an ordered, height-balanced binary partition tree. Split each nonsingleton node into two consecutive child intervals whose sizes differ by at most one. This uses only actual cutoffs, so every query remains in the legal domain without requiring K to be a power of two. The public query sequence consists of the first T internal pivots in breadth-first order. At $\rho = 1/2$, the adversary selects an earlier predictor with minimum score at the query when the true label is one, and one with maximum score when the true label is zero. The guarantee below uses both restrictions on the adversary.

Let $d = \lfloor \log_2 T \rfloor$. The queried prefix contains every node at depths $0, \ldots, d-1$ and possibly a prefix of depth d. Since $K = T^2$, all of these nodes contain at least two cutoff cells.

Q.1 Mixing the endpoints of the candidate interval

At the beginning of depth j, the learner keeps a consecutive block I_j of depth-j cutoff cells as candidates for the target. Their union is the current candidate interval $[\underline{a}_j, \overline{a}_j]$. Throughout depth j, the learner deploys

$$Q_j = \tfrac{1}{2}\delta_{\underline{a}_j} + \tfrac{1}{2}\delta_{\overline{a}_j}.$$

Its full CDF is

$$p^{I_j}(x) = \begin{cases} 0, & x < \underline{a}_j, \\ 1/2, & \underline{a}_j \leq x < \overline{a}_j, \\ 1, & x \geq \overline{a}_j. \end{cases} \tag{136}$$

If the target cutoff belongs to $[\underline{a}_j, \overline{a}_j]$, every query below $\underline{a}_j$ has target label zero and score zero, every query at or above $\overline{a}_j$ has target label one and score one, and every pivot inside the interval has score $1/2$. Thus loss is zero outside the candidate interval and $1/2$ inside it.

On round one, use the root endpoint mixture. The forced true label at the root identifies the target-containing child exactly. Let this singleton depth-one cell block be I_1.

Q.2 Estimating the change point at each depth

Suppose all queries at depth j have been completed and the candidate interval contains m consecutive cells. Order their pivots from left to right, with feedback bits $Y_1, \ldots, Y_m$. If the interval contains the target, the true labels change once from zero to one. Write k^* for the number of label-zero pivots.

Define

$$f_0(y) = (1/4)^y (3/4)^{1-y}, \qquad f_1(y) = (3/4)^y (1/4)^{1-y},$$

and let $\widehat{k} \in \{0, \ldots, m\}$ maximize

$$\mathcal{L}(k) = \prod_{i \leq k} f_0(Y_i) \prod_{i > k} f_1(Y_i), \tag{137}$$

with deterministic tie breaking. Set

$$r = \left\lceil 8 \log\big(18(d+1)T\big) \right\rceil. \tag{138}$$

Among the $2m$ children of the current cells, indexed from left to right, define the next block by

$$\alpha = \max\{1, 2(\widehat{k} - r)\}, \qquad \beta = \min\{2m, 2(\widehat{k} + r) + 1\}. \tag{139}$$

The next candidate interval contains child cells $\alpha, \ldots, \beta$. No update is needed after the final partially queried depth.

Q.3 Earlier scores equal one half at unresolved queries

Lemma Q.2 (Constant-half scores at unresolved queries)**.** *Assume the target belongs to every preceding candidate interval. Then every earlier endpoint-mixture CDF equals $1/2$ at every pivot in the current interval. Consequently, the restricted replay rule produces independent observations*

$$Y \sim \mathrm{Bern}(1/4) \quad \textit{under label zero}, \qquad Y \sim \mathrm{Bern}(3/4) \quad \textit{under label one}. \tag{140}$$

Proof. The successive cutoff intervals are nested. Every unresolved current pivot lies between the endpoints of all preceding intervals, where each earlier endpoint mixture predicts one with probability $1/2$. Thus the minimum and maximum earlier scores are both $1/2$. With $\rho = 1/2$, a label-zero audit/replay mixture has mean $(1/2)0 + (1/2)(1/2) = 1/4$, and a label-one mixture has mean $(1/2)1 + (1/2)(1/2) = 3/4$. Fresh audit and replay coins give independence across rounds. □

Lemma Q.3 (Change-point localization)**.** *Under Eq.* (140)*,*

$$\mathbb{P}\{|\widehat{k} - k^*| > r\} \le 18e^{-r/8}. \tag{141}$$

Proof. Suppose $\widehat{k} \ge k^* + s$. Then some candidate that treats s true-one observations as zeros has likelihood at least the true change point. For a single observation,

$$\frac{f_0(0)}{f_1(0)} = 3, \qquad \frac{f_0(1)}{f_1(1)} = 1/3.$$

The candidate can beat the truth only if at least half of those s observations are zero. Under a true-one law, their count is $\mathrm{Binomial}(s, 1/4)$. Hoeffding's inequality gives probability at most $e^{-s/8}$. A union bound over all $s > r$ gives

$$\sum_{s=r+1}^{\infty} e^{-s/8} = \frac{e^{-(r+1)/8}}{1 - e^{-1/8}}.$$

The left tail is identical after interchanging zeros and ones. The sum of the two geometric tails is at most $18e^{-r/8}$. □

If $|\widehat{k} - k^*| \le r$, the target-containing child has relative index in $\{2k^*, 2k^* + 1\} \cap [2m]$ and therefore lies in Eq. (139)*.* Also,

$$|I_{j+1}| \le \min\{2|I_j|, 4r + 2\}. \tag{142}$$

Q.4 Total loss

Let G be the event that the target remains in the candidate interval after every completed-level update. At the first failed update, all preceding intervals still contain the target, so Lemma Q.3 applies. There are at most d completed updates. By Eq. (138)*,*

$$\mathbb{P}(G^c) \le 18de^{-r/8} \le 1/T. \tag{143}$$

On G, Eq. (142) and $|I_1| = 1$ imply $|I_j| \le 4r + 2$ at every relevant depth. Every queried pivot outside the candidate interval has zero loss, while every pivot inside has loss $1/2$. Including the root and the final partial depth,

$$L_T \mathbf{1}_G \le \frac{1}{2} + \frac{d}{2}(4r + 2). \tag{144}$$

On G^c, use $L_T \le T$. Combining Eqs. (143) and (144)*,*

$$\sup_{a^* \in [K]} \mathbb{E}_{a^*} L_T \le \frac{3}{2} + d(2r + 1) = O(\log^2(T + 1)).$$

Each deployed prediction function is the CDF of a distribution on actual thresholds, so every action is proper. This completes the proof of Proposition Q.1.

Q.5 Extension to every positive audit rate

Proposition Q.4 (The restricted query order at general audit rates)**.** *Let $T \ge 2$ and $K = T^2$. Consider the restricted game with the first T publicly specified breadth-first queries. Knowing the target, the adversary replays an earlier predictor with minimum score under target label one and maximum score under target label zero. For every $\rho \in (0, 1]$, there is a deterministic proper learner with*

$$\sup_{a^* \in [K]} \mathbb{E}_{a^*} L_T \le C \min\left\{T, \rho^{-2} \log^2(T + 1)\right\}. \tag{145}$$

This is a bound for the restricted query and source rule, not for the unrestricted minimax game.

Proof. Use the same learner, which mixes the endpoints of nested intervals. Condition on the transcript at the start of a level and on the target belonging to every preceding candidate interval. Each earlier prediction function then equals $1/2$ at every currently unresolved pivot. The minimum and maximum earlier scores are therefore both $1/2$. Fresh audit and replay coins give conditionally independent observations within the completed level with laws

$$Y \sim \mathrm{Bern}\left(\frac{1-\rho}{2}\right) \quad \text{under label zero,} \qquad Y \sim \mathrm{Bern}\left(\frac{1+\rho}{2}\right) \quad \text{under label one.} \tag{146}$$

The source is always an actual earlier predictor, selected before the fresh audit coin.

At $\rho = 1$, the completed level reveals the exact step vector. Set the change-point radius to zero. Each update retains the two child cells adjacent to the true change point, so the target is never discarded and the candidate interval has at most two cells. The calculation below becomes the deterministic bound $L_T \le 1/2 + d$, where $d = \lfloor \log_2 T \rfloor$.

Fix $0 < \rho < 1$ and put

$$a = \frac{1-\rho}{2}, \qquad b = \frac{1+\rho}{2} = 1 - a.$$

After querying the m consecutive cells of the current interval, define the maximum-likelihood change point as in Eq. (137), now with

$$f_0(y) = a^y(1-a)^{1-y}, \qquad f_1(y) = b^y(1-b)^{1-y}.$$

Let k^* be the true change point. A candidate shifted s positions to the right assigns the label-zero law to s observations drawn from the label-one law. Because

$$\frac{f_0(0)}{f_1(0)} = \frac{b}{a}, \qquad \frac{f_0(1)}{f_1(1)} = \frac{a}{b},$$

the candidate can beat the truth only if at least half of those observations are zero. Their count is $\mathrm{Binomial}(s, a)$ and $1/2 - a = \rho/2$. Hoeffding's inequality gives

$$\mathbb{P}\{\text{a fixed shift by } s \text{ beats } k^*\} \le e^{-s\rho^2/2}. \tag{147}$$

The left tail is symmetric. There are at most T candidate shifts in either direction, so for every integer $r \ge 0$,

$$\mathbb{P}\{|\widehat{k} - k^*| > r\} \le 2Te^{-(r+1)\rho^2/2}. \tag{148}$$

This bound conditions on the transcript at the start of the level and on the target belonging to all preceding candidate intervals.

Choose

$$b_\rho = \left\lceil \frac{2}{\rho^2} \log\bigl(2T^2(d+1)\bigr) \right\rceil. \tag{149}$$

Use the child interval in Eq. (139) with $r = b_\rho$. If $|\widehat{k} - k^*| \le b_\rho$, the next candidate interval contains the target and has at most $4b_\rho + 2$ cells. At the first failed update, the conditional assumptions for Eq. (148) hold. A union bound over at most d updates gives failure probability at most $1/T$.

On target retention, including the root and the final partial depth,

$$L_T \le \frac{1}{2} + \frac{d}{2}(4b_\rho + 2).$$

On failure, use $L_T \le T$. Hence

$$\sup_{a^* \in [K]} \mathbb{E}_{a^*} L_T \le C\rho^{-2} \log^2(T+1).$$

The constant proper CDF $1/2$ has pathwise loss $T/2$. Intersecting the two bounds proves Eq. (145). □

R SMALL CLASSES AND EXTREME AUDIT RATES

The first round always reveals the true label after prediction. This matters when the class has only two hypotheses or no later labels are audited. We give these cases together with the standard bound when every label is truthful.

R.1 One query point

Proposition R.1. *For every $T \geq 1$ and $\rho \in [0, 1]$,*

$$\mathcal{R}_T(2, \rho) = 1/2.$$

Proof. The two targets are the constant-one and constant-zero thresholds. The learner predicts one with probability $p_1(1) = 1/2$ and incurs loss $1/2$. The forced true label identifies the target, which the learner uses on all later rounds. This gives the upper bound $1/2$. Under the uniform prior on the two targets, every first score has average loss $\frac{1}{2}(1-p) + \frac{1}{2}p = 1/2$, which proves the lower bound. □

R.2 Every label is truthful

Proposition R.2. *At $\rho = 1$,*

$$\mathcal{R}_T(K, 1) = \Theta(\min\{T, \log K\}).$$

Proof. For the upper bound, maintain the version space $\mathcal{V}_t \subseteq [K]$ of cutoffs consistent with every observed true label. Deploy the uniform distribution on this set. At query x_t, let e_t be the fraction of surviving cutoffs that predict the wrong label. This fraction is exactly the learner's prediction error probability. After observing the label,

$$|\mathcal{V}_{t+1}| = (1 - e_t)|\mathcal{V}_t|.$$

Since $e \leq -\log(1-e)$ for $e \in [0, 1)$,

$$\sum_t e_t \leq \sum_t -\log(1 - e_t) = \log \frac{|\mathcal{V}_1|}{|\mathcal{V}_{T+1}|} \leq \log K.$$

If $e_t = 1$, the true target would be eliminated, which is impossible, so the logarithm is well-defined. The horizon cap gives $O(\min\{T, \log K\})$. Lemma C.6 gives the matching lower bound up to universal constants. □

R.3 No audits after the first round

Proposition R.3. *For $\rho = 0$ and $K \geq 3$,*

$$\frac{T-1}{2} \leq \mathcal{R}_T(K, 0) \leq \frac{T}{2}. \tag{150}$$

For $K \geq 4$,

$$\mathcal{R}_T(K, 0) = T/2. \tag{151}$$

Proof. Predicting one with probability $p_t(x) = 1/2$ everywhere is proper and has pathwise loss $T/2$, proving both upper bounds.

For Eq. (150), choose adjacent targets h_1, h_2. They disagree at $x = 1$ and agree at the distinct point $z = 2$, which exists because $K \geq 3$. Sample the two targets fairly at time zero. Query z in round one. Its forced label is one under either target, so the first predictor and visible transcript have the same distribution in both cases. On every later round, query $x = 1$ and replay the actual round-one predictor. Since $\rho = 0$, the feedback is sampled from the same prediction probability $p_1(1)$ under either target. By induction, the complete visible transcripts have the same law. Conditional on any such history, the target label at $x = 1$ is equally likely to be zero or one, so every prediction probability has average loss $1/2$. The last $T - 1$ rounds therefore cost $(T-1)/2$.

For Eq. (151), use a uniform prior on four consecutive cutoffs $1, 2, 3, 4$, which are available when $K \geq 4$. Query $x = 2$ on round one. The true label is one for targets $1, 2$ and zero for targets $3, 4$, so every first action has average loss $1/2$. The forced label reveals one of two adjacent pairs. If the surviving pair is $\{1, 2\}$, query their disagreement point $x = 1$ thereafter. If it is $\{3, 4\}$, query $x = 3$. In either case replay the actual round-one predictor. Within the selected pair, replay has the same law under either target, and the visible histories remain indistinguishable. Every remaining round therefore has average loss $1/2$ under the conditional fair prior, giving total Bayes loss $T/2$. □

S BOUNDS FOR FINITE THRESHOLD CLASSES AT EVERY HORIZON

S.1 COMPARISON OF THE AVAILABLE BOUNDS

The following corollary describes what the construction using two earlier predictors proves when the number of thresholds is fixed and the horizon is unrestricted. Appendix S proves it by optimizing the construction's integer parameters, and also gives a lower bound for every audit rate.

Corollary S.1 (Bounds at audit rate one half for every horizon)**.** *Define* $R(K,T) = \min\{T, [T^2\log(1+K/T^2)]^{1/3}\}$*. There are universal constants* $c, C > 0$ *such that, for all* $K \geq 2$ *and* $T \geq 1$*,*

$$\mathcal{R}_T(K, 1/2) \geq cR(K,T). \tag{152}$$

For $T \leq \sqrt{K}$*, the upper bound satisfies*

$$\mathcal{R}_T(K, 1/2) \leq C\left(\frac{\log(2K)}{\log(1+K/T^2)}\right)^{1/3} R(K,T). \tag{153}$$

For $T \geq \sqrt{K}$*, the lower bound gives* $\mathcal{R}_T(K, 1/2) = \Omega(K^{1/3})$*. It also holds when the learner sees the query before choosing its predictor.*

For $T \leq K^{1/2-\varepsilon}$, the comparison factor in Eq. (153) is bounded by an ε-dependent constant. This agrees with the sharp rates for proper and improper learners that choose the predictor before seeing the query in Theorem 3.1 and Appendix V. At $T = \sqrt{K}$, the lower and clipped upper bounds differ by $(\log K)^{1/3}$. For longer horizons, the lower bound $\Omega(K^{1/3})$ and the conservative upper bound $K-1$ from Eq. (182) leave both the eventual cumulative loss and the time needed to reach it undetermined up to constants. The sharp rate at audit probability one half therefore remains open beyond the main theorem's horizon range.

Theorem 3.2 gives a separate all-horizon result for proper learners when $0 < \rho \leq 1/3 - \delta$, with fixed $\delta \in (0, 1/3)$. The rate is $\min\{T, K/\rho\}$. Corollary W.8 in Appendix W proves the same order for improper learners and in either order of prediction and query. These conclusions use a different lower-bound construction from Proposition 3.4.

We now optimize the lower bound using two earlier predictors without assuming that K grows sufficiently quickly with T. The number of available targets limits how many unresolved target labels remain after those predictors have been produced. The first result expresses that constraint as an explicit rate. The later arguments compare it with the clipped upper bound and extend the lower bound to other audit rates.

Definition S.2 (Optimizing the lower-bound parameters)**.** *For integers* $K, T \geq 1$*, define*

$$\mathfrak{R}(K,T) := \sup_{\substack{m,d\in\mathbb{Z}_{\geq 1}\\ 3m\leq T,\ d\leq m/2\\ (m+1)^2 2^d \leq K}} (dm^2)^{1/3}, \tag{154}$$

where the supremum over the empty feasible set is zero. Set

$$\Lambda(K,T) := \log_2\left(1+\frac{K}{T^2}\right), \qquad R_2(K,T) := \min\left\{T,\ [T^2\Lambda(K,T)]^{1/3}\right\}. \tag{155}$$

Theorem S.3 (An explicit form of the optimized lower bound)**.** *There are universal constants* $0 < c < C < \infty$ *such that, for every pair of integers* $K, T \geq 1$*,*

$$c\,[1 + R_2(K,T)] \leq 1 + \mathfrak{R}(K,T) \leq C\,[1 + R_2(K,T)]. \tag{156}$$

Replacing $\log_2$ *by the natural logarithm changes only the universal constants.*

Proof. We first bound what any feasible choice of parameters can achieve, then construct choices attaining that order. Throughout, c, C are positive universal constants and may change from line to line.

Step 1: the horizon upper bound. Fix a feasible pair (m, d). Since $d \le m/2$ and $m \le T/3$,

$$dm^2 \le \frac{m^3}{2} \le \frac{T^3}{54}. \tag{157}$$

Consequently, $(dm^2)^{1/3} \le CT$.

Step 2: the support upper bound. The support constraint implies

$$2^d \le \frac{K}{(m+1)^2} \le \frac{K}{m^2}, \qquad \text{hence} \qquad d \le \log_2 \frac{K}{m^2}.$$

Put

$$\chi := \frac{K}{T^2}, \qquad x := \frac{m}{T} \le \frac{1}{3}.$$

Then

$$dm^2 \le T^2 x^2 \log_2 \frac{\chi}{x^2}. \tag{158}$$

We show that the right-hand side is at most $CT^2 \log_2(1+\chi)$.

If $\chi \ge 1$, then

$$x^2 \log_2 \frac{\chi}{x^2} = x^2 \log_2 \chi + 2x^2 \log_2 \frac{1}{x}.$$

The first term is at most $\log_2 \chi$. The function $x \mapsto x^2 \log_2(1/x)$ is bounded on $(0, 1/3]$, while $\log_2(1+\chi) \ge 1$. Therefore

$$x^2 \log_2 \frac{\chi}{x^2} \le C \log_2(1+\chi) \qquad (\chi \ge 1). \tag{159}$$

If $0 < \chi < 1$, feasibility and $d \ge 1$ give $2(m+1)^2 \le K$, and in particular

$$u := \frac{m^2}{K} \le \frac{1}{2}.$$

Using $d \le \log_2(K/m^2)$,

$$dm^2 \le m^2 \log_2 \frac{K}{m^2} = Ku \log_2 \frac{1}{u} \le CK. \tag{160}$$

For $0 < \chi < 1$, $\chi \asymp \log_2(1+\chi)$, and $K = \chi T^2$. Hence Eq. (160) is at most $CT^2 \log_2(1+\chi)$.

Combining the two cases with Eq. (157),

$$(dm^2)^{1/3} \le C \min\Big\{T,\ [T^2 \Lambda(K,T)]^{1/3}\Big\} = C R_2(K,T).$$

Taking the supremum over feasible pairs proves the upper half of Eq. (156).

Step 3: reduction of bounded exceptions. Small values of the proposed rate are already covered by the additive one. To make this explicit, fix $S_0 \ge 4096$. If $R_2(K,T) \le S_0$, then

$$1 + \mathfrak{R}(K,T) \ge 1 \ge \frac{1}{1+S_0}[1 + R_2(K,T)].$$

We may therefore assume $R_2(K,T) > S_0$. If $\chi \ge 1$, then $R_2(K,T) \le T$, so $T > S_0$. If $\chi \le 1$, the elementary bound $\log_2(1+\chi) \le \chi / \log 2$ gives

$$R_2(K,T)^3 \le T^2 \log_2(1+\chi) \le K / \log 2.$$

Hence $K \ge (\log 2) S_0^3$. This also ensures that the integer choices in Step 4 satisfy all required floor estimates. Thus the remaining construction needs only treat parameters outside a fixed bounded range.

Step 4: lower construction when $\chi \le 1$. Take

$$d = 1, \qquad m = \left\lfloor \frac{\sqrt{K}}{8} \right\rfloor. \tag{161}$$

The nonexceptional assumption ensures $m \ge 2$, so $d \le m/2$. Since $\chi = K/T^2 \le 1$,

$$3m \le \frac{3\sqrt{K}}{8} \le \frac{3T}{8} \le T.$$

For K above a universal constant,

$$m + 1 \le \frac{\sqrt{K}}{4},$$

and therefore

$$(m+1)^2 2^d \le 2\left(\frac{\sqrt{K}}{4}\right)^2 = \frac{K}{8} \le K.$$

Thus (m, d) is feasible. Also, $m \ge \sqrt{K}/16$ after enlarging the exceptional range, so

$$dm^2 = m^2 \ge \frac{K}{256}. \tag{162}$$

For $0 < \chi \le 1$, $\Lambda(K,T) \asymp \chi$ and $\Lambda(K,T) \le 1$. Hence

$$R_2(K,T)^3 = T^2\Lambda(K,T) \asymp T^2\chi = K.$$

Equation (162) gives

$$\mathfrak{R}(K,T) \ge cR_2(K,T).$$

Step 5: lower construction when $\chi \ge 1$. Let

$$s_0 := \min\{T, \Lambda(K,T)\},$$

and choose

$$m = \left\lfloor \frac{T}{64} \right\rfloor, \qquad d = \max\left\{1, \left\lfloor \frac{s_0}{2048} \right\rfloor\right\}. \tag{163}$$

For T outside the bounded exceptional range,

$$3m \le T, \qquad m \ge \frac{T}{128}, \qquad m + 1 \le \frac{T}{32}.$$

Moreover,

$$d \le 1 + \frac{T}{2048} \le \frac{T}{1024} \le \frac{m}{2},$$

again after increasing the fixed threshold on T.

For the support constraint, $s_0 \le \Lambda(K,T)$ gives

$$2^d \le 2\,2^{s_0/2048} \le 2\,2^{\Lambda(K,T)/2048} = 2(1+\chi)^{1/2048}.$$

Since $\chi \ge 1$,

$$(1+\chi)^{1/2048} \le (2\chi)^{1/2048} \le 2\chi.$$

Therefore

$$(m+1)^2 2^d \le \frac{T^2}{1024}\,4\chi \le \chi T^2 = K. \tag{164}$$

Thus (m, d) is feasible.

Finally, $d \ge s_0/4096$: if $s_0 < 4096$ this follows from $d \ge 1$, while if $s_0 \ge 4096$ it follows from $\lfloor s_0/2048 \rfloor \ge s_0/4096$. Consequently,

$$m^2 d \ge cT^2 s_0 = cT^2 \min\{T, \Lambda(K,T)\} = cR_2(K,T)^3. \tag{165}$$

Taking cube roots yields $\mathfrak{R}(K,T) \ge cR_2(K,T)$.

The two constructions cover both ways the constraint can matter: limited target support when $\chi \le 1$, and limited time when $\chi \ge 1$. Together with Step 3 for the bounded exceptions, they prove the lower half of Eq. (156). □

Remark S.1 (Why the additive one is useful). The feasible set in Eq. (154) can be empty for small horizons because $3m \le T$ and $d \le m/2$ require $m \ge 2$ and hence $T \ge 6$. The additive one records this bounded discrete effect without changing any asymptotic regime.

S.2 Proof of the lower bound at audit rate one half

Proof of the lower bound in Corollary S.1. The preceding optimization allowed all feasible integers, whereas Proposition O.1 requires $m \geq m_0$. We first check that this distinction changes the bound by only a universal constant. Let m_0 and c_0 be the constants in that proposition. Define $\mathfrak{R}_{\geq m_0}(K,T)$ by the feasible maximum in Eq. (154), restricted to $m \geq m_0$, and set it to zero if the restricted set is empty. At $\rho = 1/2$, the proposition gives

$$\mathcal{R}_T(K, 1/2) \geq c_0 \mathfrak{R}_{\geq m_0}(K,T). \tag{166}$$

If a feasible pair has $m < m_0$, then $d \leq m/2$ and

$$(dm^2)^{1/3} \leq (m^3/2)^{1/3} \leq B_0, \qquad B_0 := m_0/2^{1/3}.$$

Consequently,

$$\mathfrak{R}(K,T) \leq \mathfrak{R}_{\geq m_0}(K,T) + B_0. \tag{167}$$

Lemma C.6 also gives, for $K \geq 2$ and $T \geq 1$,

$$\mathcal{R}_T(K, 1/2) \geq \frac{1}{2} \min\{T, \lfloor \log_2 K \rfloor\} \geq \frac{1}{2}. \tag{168}$$

Combining Eqs. (166) and (168), there is a universal $c_1 > 0$ such that

$$\mathcal{R}_T(K, 1/2) \geq c_1 \big[1 + \mathfrak{R}_{\geq m_0}(K,T)\big]. \tag{169}$$

Theorem S.3 and Eq. (167) imply

$$1 + R_2(K,T) \leq C[1 + \mathfrak{R}(K,T)] \leq C'[1 + \mathfrak{R}_{\geq m_0}(K,T)].$$

Substitution into Eq. (169) proves the corollary with $\log_2$. Replacing $\log_2$ by the natural logarithm changes only a universal constant. □

S.3 Comparing the critical upper and lower bounds

We next compare the lower bound with the clipped upper bound, both valid at every horizon. For convenience, write the lower rate from Corollary S.1 and the factor relating the two bounds as

$$R(K,T) := \min\left\{T, \left\lceil T^2 \log\left(1 + \frac{K}{T^2}\right)\right\rceil^{1/3}\right\}, \qquad \alpha(K,T) := \left\lceil \frac{\log(2K)}{\log(1 + K/T^2)} \right\rceil^{1/3}. \tag{170}$$

Here and below unadorned logarithms are natural.

Proof of the upper-bound comparison in Corollary S.1. The definition in Eq. (155) uses a base-two logarithm, so

$$R_2(K,T) \asymp R(K,T) \tag{171}$$

with universal constants. Corollary S.1 therefore gives the claimed lower bound.

For the other direction, the critical clipped multiplicative-weights bound in Eq. (104), together with the constant-half horizon cap, gives for all $K \geq 2$ and $T \geq 1$

$$\mathcal{R}_T(K, 1/2) \leq C \min\Big\{T, T^{2/3}(\log K)^{1/3}\Big\}. \tag{172}$$

Indeed, when $T < 64 \log K$, the second entry is a universal constant times at least T, so the constant-half action supplies both bounds.

Put $\Lambda_{\mathrm{e}} = \log(1 + K/T^2)$ and $s_{K,+} = \log(2K)$. Since $T \geq 1$, $1 + K/T^2 \leq 1 + K \leq 2K$, and hence $\alpha = (s_{K,+}/\Lambda_{\mathrm{e}})^{1/3} \geq 1$. If $R(K,T) = T$, the minimum in Eq. (172) is at most $T \leq \alpha R(K,T)$. Otherwise,

$$\alpha(K,T) R(K,T) = T^{2/3} s_{K,+}^{1/3} \geq T^{2/3} (\log K)^{1/3}.$$

This proves the upper bound in Corollary S.1.

To see where the two bounds match, examine the comparison factor. If $T \leq \sqrt{K}$, then $\Lambda_{\mathrm{e}} \geq \log 2$, while $s_{K,+} \leq 2 \log K$ for $K \geq 2$. Thus

$$\alpha(K,T) \leq C (\log K)^{1/3}. \tag{173}$$

More sharply, if $T \leq K^{1/2-\varepsilon}$, then $\Lambda_{\mathrm{e}} \geq 2\varepsilon \log K$, so $\alpha(K,T) = O_\varepsilon(1)$. Thus the bounds agree up to constants under this horizon restriction, while the comparison factor grows as T approaches $\sqrt{K}$. □

S.4 A LOWER BOUND FOR EVERY AUDIT RATE

The same parameter choice can be applied to both audit-rate branches of Proposition O.1. It limits the time spent producing the two predictors to order $\sqrt{K}$, leaving enough targets for the later tests.

Set

$$m_{\mathrm{fd}} := \min\{T, \sqrt{K}\}, \qquad d_{\mathrm{fd}} := \min\left\{m_{\mathrm{fd}}, \log_2\left(1 + \frac{K}{m_{\mathrm{fd}}^2}\right)\right\}, \qquad R_{\mathrm{fd}} := (m_{\mathrm{fd}}^2 d_{\mathrm{fd}})^{1/3}. \tag{174}$$

Because $m_{\mathrm{fd}} \le \sqrt{K}$, the logarithm in Eq. (174) is at least one. Consequently $1 \le d_{\mathrm{fd}} \le m_{\mathrm{fd}}$ and $1 \le R_{\mathrm{fd}} \le m_{\mathrm{fd}}$.

Theorem S.4 (Lower bound for all class sizes, horizons, and audit rates)**.** *There is a universal constant $c > 0$ such that, for all $K \ge 2$, $T \ge 1$, and $\rho \in [0, 1]$, the following statements hold. If $\rho \ge 1/2$ and $\Delta_\rho = 2\rho - 1$, then*

$$\mathcal{R}_T(K, \rho) \ge c \min\left\{R_{\mathrm{fd}}, \frac{d_{\mathrm{fd}}}{\Delta_\rho^2}\right\}, \tag{175}$$

where the second entry is interpreted as $+\infty$ at $\Delta_\rho = 0$. If $\rho < 1/2$ and $\delta_\rho = 1 - 2\rho$, then

$$\mathcal{R}_T(K, \rho) \ge c \min\{m_{\mathrm{fd}}, \delta_\rho m_{\mathrm{fd}} + R_{\mathrm{fd}}\}. \tag{176}$$

Independently, the balanced-tree lower bound $\mathcal{R}_T(K, \rho) \ge \frac{1}{2}\min\{T, \lfloor \log_2 K \rfloor\}$ always applies, so either displayed lower bound may be maximized with it.

Proof. Let $c_{\mathrm{blk}} > 0$ and m_0 be the universal constants in Proposition O.1, and put

$$m_{\mathrm{fd},0} := \max\{2048, 128 m_0\}.$$

Suppose first that $m_{\mathrm{fd}} \ge m_{\mathrm{fd},0}$. Choose the integer parameters

$$m = \left\lfloor \frac{m_{\mathrm{fd}}}{64} \right\rfloor, \qquad d = \max\left\{1, \left\lfloor \frac{d_{\mathrm{fd}}}{2048} \right\rfloor\right\}. \tag{177}$$

Then

$$\frac{m_{\mathrm{fd}}}{128} \le m \le \frac{m_{\mathrm{fd}}}{64}, \qquad \frac{d_{\mathrm{fd}}}{4096} \le d \le 1 + \frac{d_{\mathrm{fd}}}{2048}. \tag{178}$$

Since $d_{\mathrm{fd}} \le m_{\mathrm{fd}}$ and $m_{\mathrm{fd}} \ge 2048$, these estimates give $d \le m_{\mathrm{fd}}/1024 \le m/2$. Moreover $3m \le 3m_{\mathrm{fd}}/64 \le T$ and $m \ge m_0$.

It remains to check the support constraint. Write $\chi_m = K/m_{\mathrm{fd}}^2 \ge 1$. Since $d_{\mathrm{fd}} \le \log_2(1 + \chi_m)$,

$$2^d \le 2\, 2^{d_{\mathrm{fd}}/2048} \le 2(1 + \chi_m)^{1/2048} \le 4\chi_m.$$

Also $m + 1 \le m_{\mathrm{fd}}/32$, and therefore

$$(m + 1)^2 2^d \le \frac{m_{\mathrm{fd}}^2}{1024}\, 4\chi_m \le K.$$

Thus Eq. (177) is feasible for Proposition O.1. Its block scale satisfies

$$(dm^2)^{1/3} \ge 2^{-12} R_{\mathrm{fd}}, \qquad m \ge 2^{-12} m_{\mathrm{fd}}, \qquad d \ge 2^{-12} d_{\mathrm{fd}}. \tag{179}$$

The supercritical line of Proposition O.1 and Eq. (179) give Eq. (175), after changing the universal constant. The subcritical line gives Eq. (176) in the same way.

Finally suppose that $m_{\mathrm{fd}} < m_{\mathrm{fd},0}$. Lemma C.6 gives $\mathcal{R}_T(K, \rho) \ge 1/2$. The minimum in Eq. (175) is at most $R_{\mathrm{fd}} \le m_{\mathrm{fd}}$, and the minimum in Eq. (176) is at most m_{fd}. Reducing the universal constant to at most $1/(2m_{\mathrm{fd},0})$ completes both cases. □

Remark S.2 (Agreement with the lower bound at audit rate one half). At $\rho = 1/2$, Theorem S.4 reduces to $\mathcal{R}_T(K, 1/2) \ge cR_{\mathrm{fd}}$. This is the same scale as the current $R(K, T)$, uniformly over all horizons. Indeed, if $T \le \sqrt{K}$, then $m_{\mathrm{fd}} = T$ and

$$R_{\mathrm{fd}} = \left[T^2 \min\left\{T, \log_2\left(1 + \frac{K}{T^2}\right)\right\}\right]^{1/3} = R_2(K, T) \asymp R(K, T). \tag{180}$$

If $T \geq \sqrt{K}$, then $m_{\mathrm{fd}} = \sqrt{K}$, $d_{\mathrm{fd}} = 1$, and $R_{\mathrm{fd}} = K^{1/3}$. In this range $x = K/T^2 \in (0, 1]$ and $\log(1 + x) \asymp x$, so

$$R(K, T) = \min\{T, [T^2 \log(1 + x)]^{1/3}\} \asymp K^{1/3} = R_{\mathrm{fd}}.$$

Thus the bound for all audit rates agrees with the previous critical lower bound for every horizon. In particular, when $T \geq \sqrt{K}$, it again gives $\mathcal{R}_T(K, 1/2) = \Omega(K^{1/3})$.

T A CUMULATIVE LOSS BOUND FOR EVERY HORIZON

For a fixed finite threshold class, positive audit probability is enough to bound cumulative loss independently of the horizon. We obtain the following bound by adapting the conservative threshold strategy of Dmitriev et al. (2026) to independent hidden audits and loss charged on every round.

Proposition T.1 (A horizon-independent bound for finite threshold classes)**.** *For every $K \geq 2$, $T \geq 1$, and $\rho > 0$,*

$$\mathcal{R}_T(K, \rho) \leq \frac{K - 1}{2\rho}. \tag{181}$$

The bound holds for all K, T and can be combined with the clipped upper bound and the loss from always predicting with probability one half. Every fixed finite threshold class has bounded expected cumulative deployment loss at every positive audit rate.

Proof. First start from the constant-zero threshold, with $b_1 = K$, and deploy $p_t = h_{b_t}$. If $p_t(x_t) = 0$ and $Y_t = 1$, set $b_{t+1} = x_t$. Otherwise, keep $b_{t+1} = b_t$. Inductively $b_t \geq a^*$ and (b_t) is nonincreasing. Whenever the current score is zero, every earlier score at the same query is also zero. Thus, whenever the learner predicts zero, a feedback label of one must come from an audit, or from the forced true first observation. Such a label certifies $x_t \geq a^*$, preserves the invariant, and strictly decreases b_t. All mistakes are false negatives. If U_T is the number of updates, then pathwise $U_T \leq K - a^*$ and

$$\mathbb{E}U_T = \mathbb{E}\ell_1 + \rho \sum_{t=2}^{T} \mathbb{E}\ell_t = \rho \mathbb{E}L_T + (1 - \rho)\mathbb{E}\ell_1.$$

It follows that $\mathbb{E}L_T \leq (K - a^*)/\rho$.

The dual strategy starts at $b_1 = 1$ and increases the cutoff to $x_t + 1$ exactly when its current score is one and $Y_t = 0$. It preserves $b_t \leq a^*$. All historical scores at a current one are also one. A feedback label of zero then certifies the true label, and every update strictly increases b_t. Hence $U_T \leq a^* - 1$, and its loss is at most $(a^* - 1)/\rho$.

Choose between these two strategies with a fair private coin at time zero. The adversary can infer this choice from the first predictor, but the target was fixed before the independent coin was drawn. Each bound holds against any adaptive adversary for that fixed target. Averaging therefore gives

$$\mathbb{E}L_T \leq \frac{(K - a^*) + (a^* - 1)}{2\rho} = \frac{K - 1}{2\rho},$$

uniformly in the fixed target. Taking the minimax operators proves the claim. The private choice between orientations improves the constant by a factor of two. □

Combining Proposition T.1 with the clipped upper bound in Corollary M.4 gives, for a universal constant C and every $K \geq 2$, $T \geq 1$, and $0 < \rho \leq 1$,

$$\mathcal{R}_T(K, \rho) \leq \min\{C r_\rho(T, K), (K - 1)/(2\rho)\}. \tag{182}$$

Known parameters determine which complete learner to use before round one. Appendix S.1 compares this bound with the critical lower bound, including the unresolved long-horizon loss and the time needed to reach it.

U REPLAY IN CLASSES CLOSED UNDER INTERSECTION

A finite class can be easier than thresholds of the same size. We show this by adapting the closure algorithm of Dmitriev et al. (2026, Algorithm 1 and Appendix B.4, Lemma 6 and Theorem 3) to independent audits and cumulative deployment loss, then giving an elementary matching example. Their Appendix B.1, Lemma 2 also establishes the associated connection between depth and threshold dimension.

Proposition U.1 (An upper bound for intersection-closed classes)**.** *Let $\mathcal{H}$ be a nonempty finite binary class closed under pointwise intersection, and let $b = \bigwedge_{h\in\mathcal{H}} h$. Define*

$$d_{\mathcal{H}}(h^*) = \max\{r : b = g_0 < g_1 < \cdots < g_r \le h^*, \quad g_i \in \mathcal{H}\},$$

where inequalities are pointwise and $<$ is strict. For every fixed target, every legal adversary that knows the target, and $\rho > 0$, a proper learner satisfies

$$\rho\mathbb{E}L_T + (1-\rho)\mathbb{E}\ell_1 \le d_{\mathcal{H}}(h^*).$$

The learner does not require knowledge of ρ.

Proof. Maintain a set $\mathcal{P}_t$ of inputs whose true labels have been certified as one, initially empty. Deploy the intersection p_t of all hypotheses that predict one on $\mathcal{P}_t$. Add x_t only when $p_t(x_t) = 0$ and $Y_t = 1$. Inductively $p_t \le h^*$ and every earlier predictor satisfies $p_s \le p_t$. If the current prediction is zero, all earlier predictions at the same input are also zero, so replay cannot return one. Every added point has true label one, preserving the induction and properness. If U_t is the update indicator, then pathwise $U_1 = \ell_1$ and $U_t = A_t\ell_t$ for $t \ge 2$. Each update strictly increases p_t, so $\sum_t U_t \le d_{\mathcal{H}}(h^*)$. Taking expectations and using independence of the fresh audit coin gives the claim. □

Proposition U.2 (Sharp minimax rate for the complete binary class)**.** *Let $\mathcal{H}_d = \{0,1\}^{[d]}$. Uniformly over $d \ge 2$, $T \ge 1$, and $\rho \in (0,1]$, the present game without audit flags or source information satisfies*

$$\mathcal{R}_T(\mathcal{H}_d, \rho) = \Theta(\min\{T, d/\rho\})\,.$$

The upper bound is at most $\frac{1}{2}\min\{T, d/\rho\}$. The lower bound holds even when audit flags and source indices are revealed. At $\rho = 0$, the value is exactly $T/2$. For $d = 1$, the value is instead $1/2$ for every T and ρ.

Proof. For a target with k positive coordinates, the closure learner incurs expected loss at most k/ρ. Reversing zero and one gives a learner that starts from all ones and certifies negative labels, with loss at most $(d-k)/\rho$. A fair private choice between these learners before round one gives $d/(2\rho)$. These bounds hold against an adversary that observes the chosen action, since the target was fixed before the private choice. Comparison with the constant-half policy gives the stated upper bound.

For the lower bound put $m = d - 1$. Sample all d target bits independently and fairly at time zero, query coordinate d on round one, and on every later round query an independent uniform coordinate from $[m]$. Always replay the actual source $S_t = 1$. Grant the learner all audit flags. Conditional on its private seed, the initial score p_1 is independent of the target. Nonaudit feedback at coordinate j has law $\mathrm{Bern}(p_1(j))$, and therefore contains no information about its target bit. Audits reveal only the queried bit. Thus every coordinate not previously audited remains conditionally fair, even given the full history and the current query. Its expected deployment loss is $1/2$. The first-round loss is also $1/2$ in expectation. Consequently every randomized learner has Bayes loss at least

$$\frac{1}{2} + \frac{m}{2\rho}\left[1 - (1-\rho/m)^{T-1}\right].$$

This is a lower bound against a fixed-target prior and a legal policy using only one actual source. Some fixed target therefore attains at least this expected loss. The inequality $1-e^{-z} \ge (1-e^{-1})\min\{z,1\}$ and $m \ge d/2$ show that the display is at least $c\min\{T, d/\rho\}$ for a universal $c > 0$. Its limit at $\rho = 0$ is $T/2$. When $d = 1$, the forced first truth identifies the entire target, giving the separate value $1/2$. □

At criticality this class has loss $\Theta(\min\{T, \log|\mathcal{H}_d|\})$. Thus an upper bound expressed only through class size can be attained by thresholds while other classes of the same size are easier. For K thresholds, the closure argument itself gives the upper bound $(K-1)/\rho$. The preceding comparison does not supply a matching threshold lower bound.

For thresholds, Theorem 3.2 and Appendix W complement this established upper bound with a matching lower bound at every horizon when $0 < \rho \le 1/3 - \delta$ for fixed $\delta > 0$.

V LOWER BOUND FOR ARBITRARY PROBABILITY-VALUED PREDICTORS

This appendix proves Proposition 3.4 and then optimizes its parameters to obtain the improper lower bound in Theorem 3.1. The learner may submit any complete prediction vector in $[0,1]^{K-1}$, but must do so before the query is chosen. Write $\mathcal{R}_T^{\mathrm{imp}}(K,\rho)$ for this game's value. The proper clipped learner of Proposition 3.3 is still available, so its optimized bound and the constant-half strategy give the upper bound.

The lower bound must replace the monotonicity argument used for proper predictions. We draw independent query points from the distribution of possible later test points. Because the learner commits its entire function first, its loss at these earlier queries controls its average prediction probability at the later points. The construction below states exactly what is conditioned on when making this comparison.

We keep the protocol of Section 2. History starts empty, the first label is truthful and arrives after prediction, and every later source is an actual earlier predictor selected before the independent audit. The learner observes only the query and feedback bit, together with the extra messages specified below. The adversary's choices never use the learner's unrevealed seed. This proof requires actions to be submitted before queries. Announcing queries earlier or revealing source or audit information requires separate arguments.

To apply Lemmas C.4 and C.5, we first fix a complete starting history common to both possible values of a fair target bit. The query is then fixed, and no extra message reveals that bit before the last prediction of the test. The only new observations during testing are therefore the Bernoulli feedback labels covered by those lemmas. The message sent at the end of the block is excluded from its pre-action histories. When identical feedback distributions are possible, we realize them by mixing the two actual source indices as in Proposition C.3.

V.1 THE TARGET PRIOR AND INDEPENDENT QUERY DISTRIBUTIONS

Fix integers $m \ge 4$ and $1 \le d \le m/2$ with

$$3m \le T, \qquad (m+1)^2 2^d \le K. \tag{183}$$

Draw independent truncated geometric clocks $I, J \in \{1, \dots, m, \infty\}$ with hazard $\beta = 1/m$. Thus $\mathbb{P}(I = s) = \beta(1-\beta)^{s-1}$ and $\mathbb{P}(I = \infty) = (1-\beta)^m$, and similarly for J.

Order the first-clock blocks by $1, \dots, m, \infty$ from left to right. Within each, order the second-clock subblocks by $\infty, m, \dots, 1$ from left to right. Give each (I, J) subblock 2^d consecutive cutoff leaves. Conditional on the clocks, choose a uniform leaf, equivalently d independent fair descendant bits. These clocks and bits specify a target cutoff at time zero, independently of the learner's private seed and the primitive random seeds used for the independent query probes, source sampling, and feedback. All independent random seeds may be drawn at time zero. Write $x_k(I, J, B_{<k})$ for the balanced-tree pivot at descendant level k, and take B_k to be the target's label at that pivot. The prior is the same for every realization of the learner seed.

Explicit coordinate convention and boundary legality. Put $B_0 = 2^d$, and index the first-clock blocks by $i = 1, \dots, m+1$. Write $\iota(i) = i$ for $i \le m$ and $\iota(m+1) = \infty$ for the represented clock value. Set $q(\infty) = 1$, $q(j) = m + 2 - j$ for finite j. The (i, j) cutoff block is

$$L_{i,j} = (i-1)(m+1)B_0 + (q(j)-1)B_0 + 1, \qquad U_{i,j} = L_{i,j} + B_0 - 1.$$

An internal pivot of $[L, U]$ is $(L + U - 1)/2$ when its number of leaves is even. When $I = s$, query $s(m+1)B_0$, strictly below the last cutoff because the $I = \infty$ block is retained. When $J = s$, query

$L_{i,s} - 1$, at least one because the $J = \infty$ block lies to its left. All queries belong to $[K-1]$, including when the other clock fails. The failure blocks belong to the time-zero target support. They do not supply external predictors.

At a surviving first-clock stage s, let

$$\mu_s = \frac{1}{d}\sum_{k=1}^{d} \mathcal{L}\big(x_k \mid I = s\big). \tag{184}$$

This fixed distribution averages pivots throughout the block $I = s$ over the *hypothetical* second clock and descendant bits. At a second-clock stage s inside a known first block i, put

$$\nu_{i,s} = \frac{1}{d}\sum_{k=1}^{d} \mathcal{L}\big(x_k \mid I = \iota(i), J = s\big). \tag{185}$$

The decoy queries are independent draws from these distributions. Their randomness is independent of the target bits, so drawing a possible later query never requires reading the actual future target bits.

We give the learner the current clock node before prediction and reveal after loss whether the clock has stopped. Thus the first clock's completed value is known when the second begins. The actual query is sampled only after the full action, independently of the real future target bits. These extra messages strengthen the learner, so it suffices to prove the lower bound for every randomized learner receiving them.

First clock. If $I > s$, query a fresh draw from μ_s. Every remaining target lies to its right, so its label is zero. If $I = s$, query the right boundary of that block. Its label is one and reveals no second-clock or descendant bit. Select the predictor submitted on this round as p^{lo}.

Second clock. Run the second clock even if the first failed, in its $I = \infty$ block. If $J > s$, query a fresh draw from $\nu_{i,s}$. The target lies to its left, so the label is one. If $J = s$, query the left boundary of that subblock. Its label is zero and reveals no descendant bit. Select this round's predictor as p^{hi}.

The first phase runs these two clocks to select the predictors. Every round in this phase after round one uses actual source one. The first observation is forced truthful after loss. Each clock runs until its sampled finite value is reached or m stages have passed. If either clock fails, fill the remaining rounds with $x = 1, S = 1$. Otherwise p^{lo} and p^{hi} come from actual rounds I and $I + J$ and are distinct earlier sources when testing starts. Neither predictor must be monotone. After testing, fill remaining rounds with $x = 1, S = 1$. All earlier predictors remain available.

V.2 How earlier loss controls the average score at future queries

Fix a randomized learner in this enlarged experiment. Let $L_{\mathrm{lo}}, L_{\mathrm{hi}}$ be its expected losses during the two clocks, and let $L_{\mathrm{build}} = L_{\mathrm{lo}} + L_{\mathrm{hi}}$. All expectations below, including $\mathbb{E}L_T$, refer to this experiment. Thus L_{build} is the learner's actual expected first-phase loss, not a pathwise quantity or an adversary's budget.

For the hazard calculation, condition on the learner-visible pre-action history, its independent seed, and completed clock outcomes and messages, including the current node. Exclude the unrevealed current outcome and target bits. Every remaining first-clock value gives the same transcript likelihood, so the probability of stopping now remains β. The full action precedes the query, whose distribution differs between $I = s$ and $I > s$. On $I > s$, the expected loss is

$$\bar{r}_s = \mathbb{E}_{x\sim\mu_s} p_s(x).$$

The loss on $I = s$ is nonnegative, so

$$L_{\mathrm{lo}} \geq (1-\beta)\sum_{s=1}^{m} \mathbb{E}[\mathbf{1}\{I \geq s\}\,\bar{r}_s].$$

Conditional on $I = s$, the actual future second clock and descendant bits have exactly the distribution used in (184), independently of this pre-action history. Therefore

$$\mathbb{E}\left[\mathbf{1}\{I \leq m\}\frac{1}{d}\sum_{k=1}^{d} p^{\mathrm{lo}}(x_k)\right] = \beta\sum_{s=1}^{m}\mathbb{E}[\mathbf{1}\{I \geq s\}\,\bar{r}_s] \leq \frac{\beta}{1-\beta}L_{\mathrm{lo}}. \tag{186}$$

The same argument in the reversed second chain, additionally conditioning on the completed first clock, gives

$$\mathbb{E}\left[\mathbf{1}\{J \le m\}\frac{1}{d}\sum_{k=1}^{d}(1-p^{\mathrm{hi}}(x_k))\right] \le \frac{\beta}{1-\beta}L_{\mathrm{hi}}. \tag{187}$$

Possible test points are defined even when a clock fails, although no testing losses are charged there. Each indicator restricts attention to cases where the selected predictor exists. Values assigned elsewhere only make the expectation well defined. Those extensions are never replayed.

Let $E = \{I \le m, J \le m\}$ and $\pi_0 = (1-e^{-1})^2$. On E put

$$\delta_k = 1 - p^{\mathrm{hi}}(x_k) + p^{\mathrm{lo}}(x_k) \ge 0.$$

Restricting the nonnegative terms in (186)–(187) to E yields

$$\mathbb{P}(E) \ge \pi_0, \qquad \mathbb{E}\left[\mathbf{1}_E\frac{1}{d}\sum_{k=1}^{d}\delta_k\right] \le \frac{2L_{\mathrm{build}}}{m}. \tag{188}$$

For $b = 4/\pi_0$, define the levelwise events

$$F_k = E \cap \{\delta_k \le bL_{\mathrm{build}}/m\}.$$

Markov's inequality on the product of the probability space and a uniform level gives

$$\sum_{k=1}^{d}\mathbb{P}(F_k) \ge \frac{\pi_0 d}{2}. \tag{189}$$

If $L_{\mathrm{build}} = 0$, nonnegativity gives the same statement directly.

Equation (189) provides enough probability mass across the testing levels to sum their losses. It is an average guarantee across levels. The proper proof instead obtains one event on which the two selected predictors' scores are suitable at *every* later query.

V.3 Testing unresolved target bits with the same two sources

First-phase queries and their source-one feedback depend on completed clocks and independent query draws, never on a real descendant bit. Conditional on the clocks, learner seed, and complete first-phase transcript, the descendant bits remain independent and fair. At level k, both x_k and δ_k depend only on this transcript and $B_{<k}$. Thus F_k is determined before B_k is exposed, and conditioning on it preserves the bit's initial fairness.

For testing, fix the first-phase transcript, both clocks, the independent learner seed, all completed descendant bits, and earlier levels' observations. The selected predictors' scores at x_k are fixed, while B_k is fair. On a starting history in F_k, the same two predictors are reused throughout the level. Its pivot depends only on the completed prefix, and the genie sends no current-bit information during testing. Lemmas C.4 and C.5 apply to the full observation sequence. Their one-step bounds hold on every supported later history of this conditional experiment, not merely on average over F_k. Feedback may change the posterior of B_k during the test.

Repeat the level-k pivot for $h = \lfloor m/d \rfloor$ rounds. Reveal its bit only after the last loss and feedback in that block. Use p^{lo} under bit one and p^{hi} under bit zero, except for the legal mixture below. Every source choice precedes the independent audit and depends on no deeper bit. Hence each later level starts with a conditionally fair bit. Testing uses $dh \le m$ rounds, with $dh \ge m/2$. Together, the first phase and tests use at most $3m$ rounds.

For $\rho \ge 1/2$, put $\Delta_\rho = 2\rho - 1$. With $z^{\mathrm{lo}} = p^{\mathrm{lo}}(x_k)$ and $z^{\mathrm{hi}} = p^{\mathrm{hi}}(x_k)$, the two feedback means are $\theta_1 = \rho + (1-\rho)z^{\mathrm{lo}}$ and $\theta_0 = (1-\rho)z^{\mathrm{hi}}$. On F_k they have gap

$$\Delta_\rho + (1-\rho)\delta_k \le \Delta_\rho + bL_{\mathrm{build}}/m.$$

If this upper bound is at most $1/4$, the two means straddle $1/2$ and belong to $[1/4, 3/4]$. Otherwise use the first-action branch of Lemma C.4. Summing over levels using (189) gives

$$\mathbb{E}L_T \ge L_{\mathrm{build}} + c\min\left\{m, \frac{d}{(\Delta_\rho + bL_{\mathrm{build}}/m)^2}\right\}. \tag{190}$$

All constants are universal, and division by zero means $+\infty$.

For $\rho < 1/2$, put $\delta_\rho = 1 - 2\rho$. If $L_{\mathrm{build}} \geq m/(4b)$, the first-phase loss is already a constant times m. Otherwise, on each F_k the scores $z^{\mathrm{lo}} = p^{\mathrm{lo}}(x_k)$ and $z^{\mathrm{hi}} = p^{\mathrm{hi}}(x_k)$ satisfy $z^{\mathrm{lo}} \leq 1/4$ and $z^{\mathrm{hi}} \geq 3/4$. If $z^{\mathrm{hi}} - z^{\mathrm{lo}} \geq \rho/(1-\rho)$, use p^{lo} under bit one. Under bit zero, randomize between the actual indices of p^{lo} and p^{hi} to obtain mean score $z^{\mathrm{lo}} + \rho/(1-\rho)$. This mixture lies within $[z^{\mathrm{lo}}, z^{\mathrm{hi}}]$ and is sampled before the audit. The two feedback means coincide:

$$\rho + (1-\rho)z^{\mathrm{lo}} = (1-\rho)\left(z^{\mathrm{lo}} + \frac{\rho}{1-\rho}\right). \tag{191}$$

If the width condition fails, use the low source under one and the high source under zero. For $\rho \in [1/4, 1/2)$, the positive gap is at most $[bL_{\mathrm{build}}/m - \delta_\rho]_+$, while $\theta_1 \in [1/4, 5/8]$ and $\theta_0 \in [3/8, 3/4]$. Applying the interior testing bound when the gap bound is at most $1/4$, and its first-action branch otherwise, gives

$$\mathbb{E}L_T \geq L_{\mathrm{build}} + c \min\left\{m, \frac{d}{[bL_{\mathrm{build}}/m - \delta_\rho]_+^2}\right\}. \tag{192}$$

A level with identical feedback distributions contributes $h/2$ by Lemma C.5. For $\rho < 1/4$, the feedback distributions can be made identical at all good levels because $z^{\mathrm{hi}} - z^{\mathrm{lo}} \geq 1/2 > \rho/(1-\rho)$, and they cost a constant times m in total. This includes $\rho = 0$. Off the good events, use the overlap mixture only when its actual scores permit it and otherwise use the designated endpoint sources. If those scores coincide and $\rho = 0$, the degenerate mixture simply uses p^{lo}. The adversary never tests F_k or uses the unknown expectation L_{build} in its source rule.

Proof of Proposition 3.4. Write $R = R_{m,d} = (dm^2)^{1/3}$. The construction above is valid under (183) and gives the inequalities needed for both audit regimes.

For the first line set $r_0 = \min\{R, d/\Delta_\rho^2\}$. Then $r_0 \leq m$, $r_0^3 \leq dm^2$, and $\Delta_\rho \leq \sqrt{d/r_0}$. If $L_{\mathrm{build}} \geq r_0$, the first-phase loss suffices. Otherwise $(\Delta_\rho + bL_{\mathrm{build}}/m)^2 \leq (1+b)^2 d/r_0$, and (190) gives cr_0 loss.

For the second line put $r_0 = \min\{m, \delta_\rho m + R\}$. A sufficiently large first-phase loss already suffices. If $\delta_\rho m \geq R$ and $L_{\mathrm{build}} < \varepsilon_0 r_0$, then $bL_{\mathrm{build}}/m \leq 2b\varepsilon_0\delta_\rho < \delta_\rho$, so the overlap term contributes m. If $\delta_\rho m < R$, then $r_0 \leq 2R$ and, unless $L_{\mathrm{build}} \geq \varepsilon_0 r_0$, the inverse-square term in (192) is at least a constant times R. Choose a fixed sufficiently small ε_0. The construction with identical feedback distributions already covers $\rho < 1/4$.

These bounds hold for every randomized learner receiving the extra messages, with the target prior independent of its seed. A learner in the original protocol can ignore those messages. Averaging over the fixed target prior gives a target and adversary with at least this expected loss, proving the minimax bound when actions precede queries. Conditioning on the seed was only an analytical device. The source rule uses the target, actual submitted actions, and its own legal randomness. □

V.4 Optimizing the parameters and endpoints

Fix $\varepsilon \in (0, 1/2)$. For $C_\varepsilon \log K \leq T \leq K^{1/2-\varepsilon}$, choose

$$m = \lfloor T/6 \rfloor, \qquad d = \min\left\{\lfloor m/2 \rfloor, \left\lfloor \log_2 \frac{K}{(m+1)^2} \right\rfloor\right\}.$$

For a sufficiently large C_ε, these parameters satisfy (183), $m \asymp T$, and $d \asymp_\varepsilon \log K$. The latter follows from

$$\log_2 \frac{K}{(m+1)^2} \geq \log_2 K - 2\log_2 T \geq 2\varepsilon \log_2 K.$$

Proposition 3.4 gives the improper lower bound in Theorem 3.1 with actions submitted before queries. At shorter horizons, Lemma C.6 gives $\frac{1}{2}\min\{T, \lfloor \log_2 K \rfloor\}$ loss even to an improper learner. Before prediction, the current bit is equally likely to be zero or one, so every score in $[0,1]$ has expected loss $1/2$. Replaying source one reveals no deeper bit. The first truthful observation also reveals only the completed first bit. This gives $\Omega_\varepsilon(T)$ when $T < C_\varepsilon \log K$. The constant-half strategy attains the exact short-horizon upper bound without requiring $T \leq K^{1/2-\varepsilon}$. For $K = 2$, this restriction permits only $T = 1$, already covered above. Together with the proper upper bound, this proves the improper assertions of Theorem 3.1.

W LOW-AUDIT BOUNDS THROUGH EXTENDED THRESHOLD DIMENSION

This appendix proves Theorem 3.2 through a more general finite-class result. Its lower bound also allows arbitrary prediction functions and queries revealed before prediction, as stated in Corollary W.8. The relevant class complexity is the extended threshold dimension of Dmitriev et al. (2026). We first describe finite witnesses for this dimension, then use them to construct targets that remain hard to distinguish after the first truthful label.

Let $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ contain N distinct functions on a finite domain. We use the protocol of Appendix C: the learner starts with no past predictors, chooses its first action freely, and receives a truthful first label after prediction. The target is fixed at time zero independently of the learner's private randomness. Write $\mathcal{R}_T^{\mathrm{imp}}(\mathcal{H}, \rho)$ for arbitrary complete score functions submitted before the query. The superscript qf in $\mathcal{R}_T^{\mathrm{qf,imp}}(\mathcal{H}, \rho)$ means that the query is announced before the learner submits its predictor. Without the imp superscript, scores are restricted to $\mathrm{conv}(\mathcal{H})$. In both cases, the adversary knows the target and, after seeing the full action, selects an actual earlier predictor before the audit. Neither that selection nor audit status is disclosed.

W.1 EXTENDED THRESHOLD DIMENSION AND ITS WITNESSES

Identify each binary function with the set on which it predicts one. The dimension below measures the longest chain obtainable by intersecting such sets, after choosing a coordinatewise label flip. All intersections use nonempty finite collections, and chain height counts strict inclusion *edges*, not vertices.

Definition W.1 (Extended threshold dimension in height form)**.** *For $f \in \{0,1\}^{\mathcal{X}}$, define*

$$\mathrm{Cl}(\mathcal{H} \oplus f) = \left\{ \bigwedge_{h \in S} (h \oplus f) : \varnothing \neq S \subseteq \mathcal{H} \right\}, \qquad d_f(\mathcal{H}) = \mathrm{height}\big(\mathrm{Cl}(\mathcal{H} \oplus f)\big), \qquad d = \min_f d_f(\mathcal{H}) = \mathrm{ExThD}(\mathcal{H}). \tag{193}$$

The equality with the established notation follows from Definition 10 and Appendix B.1, Lemma 1, of Dmitriev et al. (2026).

Lemma W.2 (Witness characterization of extended threshold dimension)**.** *One has $d_f(\mathcal{H}) \geq m$ if and only if there are $h^*, h_1, \ldots, h_m \in \mathcal{H}$ and $x_1, \ldots, x_m \in \mathcal{X}$ such that*

$$h^*(x_i) = 1 - f(x_i), \qquad h_i(x_j) = h^*(x_j) \quad (j < i), \qquad h_i(x_i) = f(x_i).$$

The witness coordinates and the $m+1$ hypotheses are distinct. No condition is imposed on $h_i(x_j)$ for $j > i$.

Proof. Work in the flipped class. Choose a chain $C_0 \subsetneq \cdots \subsetneq C_m$ in its intersection closure, truncating a longer chain if necessary. Pick $x_i \in C_i \setminus C_{i-1}$. Since C_{i-1} is an intersection of original hypotheses, some original h_i contains C_{i-1} but excludes x_i. Any original hypothesis containing C_m can serve as h^*. All x_j for $j < i$ belong to C_{i-1}, giving the witness.

Conversely, constrain $x_1, \ldots, x_i$ to be positive successively. The remaining class is nonempty because it contains h^*. Before adding x_i, its intersection excludes x_i because h_i is still present. After the new constraint it includes x_i. Thus the intersections form a chain of m strict steps. Distinctness follows from the displayed incidence conditions. □

Lemma W.3 (Complexity remaining after the first truthful label)**.** *Write $\mathcal{H}_{xb} = \{h \in \mathcal{H} : h(x) = b\}$. For every nonempty two-sided split and every f,*

$$d_f(\mathcal{H}) \leq d_f(\mathcal{H}_{x0}) + d_f(\mathcal{H}_{x1}) + 1. \tag{194}$$

For $N \geq 3$, define

$$R(\mathcal{H}) = \max_x \min_f \max_{b : \mathcal{H}_{xb} \neq \varnothing} d_f(\mathcal{H}_{xb}).$$

Then

$$1 \leq R(\mathcal{H}) \leq d \leq 2R(\mathcal{H}) + 1 \leq 3R(\mathcal{H}). \tag{195}$$

Proof. Follow a longest witness sequence from Lemma W.2. After flipping labels, restrict its successive positive constraints to the two branches of a fixed split. If both branches remain nonempty, a strict change of the full intersection requires a strict change of at least one branch intersection. At most one step can remove a branch entirely, since the full class never empties. Count the branch changes to prove (194). A one-sided split satisfies the needed bound trivially.

Restriction cannot increase d_f. Choosing a globally minimizing f proves $R \leq d$. For a fixed x, choose f minimizing the maximum branch height and apply (194), or the one-sided inequality. This gives $d \leq 2R+1$. A class of at least three distinct binary functions has a coordinate branch containing at least two distinct functions. The intersection closure of any two distinct flipped functions has height at least one. Hence $R \geq 1$. □

Proposition W.4 (Bounds, products, and examples)**.** *Writing* Ldim *for the Littlestone dimension, for* $N \geq 2$*,*

$$\mathrm{Ldim}(\mathcal{H}) \leq \mathrm{ExThD}(\mathcal{H}) \leq \min\{|\mathcal{X}|, \lfloor N/2 \rfloor\}.$$

For classes on disjoint coordinate sets,

$$\mathrm{ExThD}(\mathcal{H}_1 \times \mathcal{H}_2) = \mathrm{ExThD}(\mathcal{H}_1) + \mathrm{ExThD}(\mathcal{H}_2).$$

The following values use the edge-count convention:

Class	ExThD
K *ordered thresholds*	$\lfloor K/2 \rfloor$
Complete binary cube on r *points*	r
One-hot class on $n \geq 2$ *points*	1
Weight-r *class on* n *points*	$\min(r, n-r)$
r *independent* K*-threshold chains*	$r\lfloor K/2 \rfloor$

Proof. Choose f to be a coordinatewise majority of the class. The first constraint requiring a label opposite to this majority leaves at most $\lfloor N/2 \rfloor$ hypotheses. Each later constraint in a witness sequence removes at least one, giving depth at most $\lfloor N/2 \rfloor$. The coordinate count is immediate. The conservative learner described below predicts intersections and makes at most d errors with noiseless feedback, so the complete binary query-tree lower bound gives $\mathrm{Ldim} \leq d$.

The intersection closure of a Cartesian product is the Cartesian product of the two closures. Every pair of component intersections can be obtained by intersecting the Cartesian product of their representing nonempty subfamilies. The height of a product is the sum of its heights. Minimizing independently over the two coordinate-flip vectors proves additivity.

For thresholds, any original full chain has $K-1$ distinct increment coordinates. For any flip, at least $\lceil (K-1)/2 \rceil = \lfloor K/2 \rfloor$ increments share the same flip direction. On those coordinates the selected original hypotheses form an increasing or decreasing threshold chain, giving that lower bound. Choosing a central threshold as f leaves two disjoint one-sided chains and their common empty intersection, with maximum height $\lfloor K/2 \rfloor$. For the cube, the whole coordinate cube is its own closure and has height r, and $\mathrm{Ldim} = r$. For the weight-r class, an all-zero or all-one default gives height at most $\min(r, n-r)$. A complete query tree of that depth exists by using distinct coordinates: every binary labeling of those coordinates extends to a weight-r vector. This matches the lower bound. One-hot and product examples follow. □

Proposition W.5 (Proper prediction when a label flip gives intersection closure)**.** *Suppose* $\mathcal{H} \oplus f_0$ *is intersection closed for some* $f_0 \in \mathcal{H}$*. Its height* h_0 *satisfies* $h_0 \leq 2\,\mathrm{ExThD}(\mathcal{H})$*. The conservative algorithm based on* f_0 *is proper.*

Proof. Take a chain of h_0 steps in the class after the label flip. Under any other flip, at least half of its increment coordinates have the same flip direction. The restriction to those coordinates is a threshold chain of at least $h_0/2$ steps, hence has closure height at least $h_0/2$. Minimize over flips. Properness follows because every nonempty intersection of remaining flipped hypotheses belongs to the flipped class. If $\mathcal{H}$ is closed under weighted majority, it is closed under the median of three. For any $f_0 \in \mathcal{H}$,

$$\mathrm{median}(h, g, f_0) \oplus f_0 = (h \oplus f_0) \wedge (g \oplus f_0),$$

so flipping by any member of the class gives an intersection-closed class. □

W.2 Sharp minimax rates at low audit probabilities

Theorem W.6 (Low-audit minimax rates in terms of ExThD). *Let $N \geq 3$ and $d = \mathrm{ExThD}(\mathcal{H})$. There is a universal $c > 0$ such that for every $T \geq 1$ and $0 < \rho < 1/3$,*

$$c\frac{1-3\rho}{1-\rho}\min\{T, d/\rho\} \leq \mathcal{R}_T^{\mathrm{qf,imp}}(\mathcal{H}, \rho) \leq \mathcal{R}_T^{\mathrm{imp}}(\mathcal{H}, \rho) \leq \min\{T/2, d/\rho\}. \tag{196}$$

Consequently, for every fixed $\varepsilon \in (0, 1/3)$, uniformly over these finite classes, horizons, and $0 \leq \rho \leq 1/3 - \varepsilon$, the two improper values are $\Theta_\varepsilon(\min\{T, d/\rho\})$. At $\rho = 0$ the quotient is interpreted as infinity. The same order holds for proper actions whenever $\mathcal{H} \oplus f_0$ is intersection closed for some $f_0 \in \mathcal{H}$.

W.2.1 An upper bound using intersections

Choose a flip attaining d. In the flipped problem, predict the intersection of the hypotheses consistent with all certified positive labels. Initially this is the intersection of the whole class. Update only when a prediction of zero is followed by feedback one. Because these intersections increase pointwise, every earlier predictor also gave zero at this coordinate. Such feedback must therefore come from an audit. Each certified label preserves the target and strictly increases the intersection, so at most d updates occur. All prediction errors are false negatives. The truthful first label and independent later audits give

$$\rho\mathbb{E}L_T + (1-\rho)\mathbb{E}\ell_1 \leq d.$$

Flip predictions back to the original labels. An improper learner that always predicts with probability one half gives $T/2$. A learner that sees the query first can ignore this advance information. For the proper classes in Proposition W.5, the intersection strategy gives $h_0/\rho \leq 2d/\rho$. Combining this with the trivial proper cap T gives

$$\mathcal{R}_T(\mathcal{H}, \rho) \leq \min\{T, h_0/\rho\} \leq 2\min\{T/2, d/\rho\}.$$

The constant-half action need not be proper merely because some label flip makes the class intersection closed. The proper lower bound follows from the stronger improper lower bound. The upper construction adapts the established closure method of Dmitriev et al. (2026).

W.2.2 A universal small-class lower bound

We first obtain a lower bound that will cover small dimensions and short horizons. For $N \geq 3$, two distinct hypotheses agree at a coordinate x_0 and disagree at another coordinate x. To see this, any three distinct binary functions include two that agree at a given coordinate, and these two must differ somewhere else. Put a fair time-zero prior on this pair. Query x_0 for the first truthful label, then repeat x while always selecting predictor one. Even if the learner sees every audit flag, the unknown label at x remains fair until its first audit. Thus

$$\mathcal{R}_T^{\mathrm{qf,imp}}(\mathcal{H}, \rho) \geq c\min\{T, 1/\rho\}, \qquad 0 < \rho \leq 1. \tag{197}$$

For $T = 1$, use instead a distinguishing first query and a fair two-target prior, giving loss $1/2$. Together, the two constructions cover the first round as well as longer horizons. At $\rho = 0$, the repeated-query argument gives $(T-1)/2$. Combining it with the first-round bound gives at least $T/4$.

W.2.3 A fixed prior despite the randomized first action

The main construction must accommodate a randomized first predictor without choosing the target after seeing it. Fix a randomized learner that sees the query before submitting its predictor, and represent its private randomization by an independent tape sampled at time zero. We condition on this tape only in the analysis. Neither the target prior nor the adversary's strategy can depend on its realized value. Before the first action, select x_0 attaining R in Lemma W.3. Conditional on x_0, the distribution of the first action is independent of the target. Define the deterministic flip

$$f(x) = \mathbf{1}\{\mathbb{P}(p_1(x) \geq 1/2) \geq 1/2\}.$$

Choose a branch b_0 for which $d_f(\mathcal{V}) \geq R$, where $\mathcal{V} = \mathcal{H}_{x_0 b_0}$. Take a witness in $\mathcal{V}$ of length exactly $s = R$ by truncating a longer one if necessary. Then $d/3 \leq s \leq d$. The witness and the prior below depend on the algorithm's initial *distribution*, not its realized seed.

Let M be an integer with $2M \leq s$. Pair the witness targets as (h_{2j-1}, h_{2j}), with pivot $z_j = x_{2j-1}$ for $1 \leq j \leq M$. In pair j the pivot bit is fair under its uniform prior. Every later pair has the common pivot label

$$b_j = 1 - f(z_j).$$

Sample the pair index I_0 uniformly from $[M]$ and its internal bit fairly at time zero. Every supported target labels x_0 by b_0, so the first forced truth exposes no pair or internal bit.

After seeing p_1, define

$$G = \{j : |p_1(z_j) - b_j| \geq 1/2\}.$$

Each index is included with probability at least $1/2$. Hence $\mathbb{E}|G| \geq M/2$ and $\mathbb{P}(|G| \geq M/4) \geq 1/3$. Grant G and whether the target pair belongs to G after the first observation. On the branch

$$|G| \geq M/4, \qquad I_0 \in G, \tag{198}$$

the target index remains uniform on G, of size $v = |G|$, and its internal bit remains fair, including after conditioning on the complete learner tape. Indeed, G depends only on the first action, whereas the pair index and internal bit were sampled independently of that tape. The branch has probability at least $1/12$, since its probability conditional on the tape is $\mathbf{1}\{|G| \geq M/4\}|G|/M$. The grants only strengthen the learner. They need not be used by an original learner, and nothing is granted before the first action.

W.2.4 Selecting a source from a stage with no audits

On this branch, write $G = \{g_1 < \cdots < g_v\}$ and let $I \in [v]$ be the target pair's rank, so $I_0 = g_I$. Relabel the selected pairs by their ranks, writing z_j, b_j from now on for the pivot and common label of original pair g_j. Each of at most $m = \lfloor v/2 \rfloor$ stages repeats its pivot for $b \geq 1$ rounds, always replaying the first predictor. The query stays the same even if this is the target's stage. After a stage ends, a genie may disclose whether its pair contains the target. All later surviving pairs have the same labels on preceding stages. Once the target's stage is reached, end this sequence of stages and independently select one of that stage's b actual actions uniformly for later replay. Denote its prediction function by p^{sel}. If no target stage is reached, use a legal continuation and discard its nonnegative losses in the lower bound.

The following comparison uses the event E_0 that $I \leq m$ and every audit coin in the target's stage is zero. This event is used only to analyze the strategy. The adversary never observes or tests it. On $I \leq m$, put

$$\delta_I = |p^{\mathrm{sel}}(z_I) - b_I|,$$

and set $\delta_I = 0$ otherwise. Let L_{build} denote the actual expected loss during these stages, conditional on a fixed complete learner tape for which $|G| \geq M/4$ and on $I_0 \in G$. This expectation includes branches outside the event and unsuccessful branches, and averages the remaining target and feedback randomness. The tape fixes G and v, but is never read by the adversary.

Lemma W.7 (Early loss controls the sampled predictor)**.** *For $v \geq 2$,*

$$\mathbb{P}(E_0) \geq \frac{m}{v}(1-\rho)^b, \qquad \mathbb{E}[\mathbf{1}_{E_0}\delta_I] \leq \frac{2L_{\mathrm{build}}}{vb}. \tag{199}$$

Conditional on E_0, the target's stage index, selected source index, complete transcript of these stages, and learner tape, the final pair bit is still fair.

Proof. The probability is immediate. Fix a stage $j \leq m$ and offset $r \leq b$. When stage j is reached, denote its predictor at offset r by $p_{j,r}$. Under $I > j$, the loss at that pivot is its distance from the common label b_j. Its contribution to L_{build} has coefficient $(v-j)/v$. On $I = j$, restrict to the subevent $\mathcal{E}_j$ of no audits in the entire stage. The subprobability law restricted to this no-audit event, and hence of the pre-feedback action at offset r, equals the no-audit subprobability law under $I > j$. Earlier stages have common truths, and every observed label in stage j on this event comes from the same first predictor. Thus the contribution to $\mathbb{E}[\mathbf{1}_{E_0}\delta_I]$ from $I = j$ and selected offset r is

$$\frac{1}{vb}\mathbb{E}[\mathbf{1}_{\mathcal{E}_j}|p_{j,r}(z_j) - b_j| \mid I > j] \leq \frac{1}{vb}\mathbb{E}[|p_{j,r}(z_j) - b_j| \mid I > j].$$

Sum over j, r and use $v - j \geq v/2$. We have compared laws restricted to the no-audit event, including that event's probability. We have not treated conditioning on no audits as leaving the unrestricted action law unchanged. The comparison also stops before the end-of-stage message: whether the current pair contains the target differs between these two worlds. That message arrives after all actions whose scores are compared and is excluded from the identity.

On E_0, earlier labels have a common truth and all feedback during the target's stage comes from replay. Whether this event holds and which offset is selected do not depend on the internal bit. The complete observed transcript and selected action therefore have the same conditional law for the two possible bits, proving fairness. Unlike the comparison according to whether the current stage contains the target, this within-pair comparison may include the terminal membership message: that message is identical for both members of the target pair. □

W.2.5 Making the feedback distributions identical and optimizing

For $0 < \rho < 1/3$, take

$$b = \max\{1, \lfloor 1/(4\rho) \rfloor\}, \qquad a = \rho/(1-\rho), \qquad \delta = 1/2 - a > 0.$$

Then $b \asymp 1/\rho$, $(1-\rho)^b \geq 1/2$, and $m/v \geq 1/3$. Lemma W.7 implies

$$\mathbb{P}(E_0 \cap \{\delta_I \leq \delta\}) \geq \frac{1}{6} - \frac{2L_{\mathrm{build}}}{\delta v b}.$$

If $L_{\mathrm{build}} \geq \delta v b/48$, these stages already cost a constant times $\delta v b$. Otherwise the displayed probability is at least $1/8$. On that event, the first source is at least $1/2$ away from the common label b_I, whereas p^{sel} is at most δ away from it. The two actual sources span at least a. On all subsequent rounds they are both earlier predictors. Repeat the pivot and mix these sources as in Proposition C.3. Explicitly, if their scores span $[\underline{p}, \overline{p}]$, use the source at $\underline{p}$ under true label one and an endpoint mixture of mean $\underline{p} + a$ under true label zero. Both visible means equal $\rho + (1-\rho)\underline{p}$, with source selection before the audit. The observed transcript has the same law under the two internal bits, even after restricting to $\delta_I \leq \delta$. Inductively, its future visible law also remains identical, so each remaining prediction has average loss $1/2$ on this subprobability law with a fair target bit. Select this rule whenever the actual sources have enough range, irrespective of whether E_0 occurred, and otherwise use any legal continuation.

Choose

$$M = \min\left\{\lfloor s/2 \rfloor, \left\lfloor \frac{T-1}{8b} \right\rfloor\right\}.$$

For $M \geq 8$, the initial stages use at most $mb \leq Mb/2 \leq (T-1)/16$ rounds after the first observation. On (198), $vb \geq Mb/4$. Thus either these stages or the subsequent tests contribute at least $c\delta Mb$. This alternative holds for every conditioned tape. It is not a decision that the adversary must make from L_{build}. Integrate it over the learner tape and the branch of probability at least $1/12$. Since $s \geq d/3$, floors change only universal constants and

$$Mb \asymp \min\{d/\rho, T\}.$$

If $M < 8$, either d is universally bounded or $T = O(1/\rho)$, so (197) gives the same result with a smaller constant. Finally $\delta = (1-3\rho)/(2(1-\rho))$, proving Theorem W.6.

Why the learner may see the query first. The current pivot is fixed throughout every stage, even at the target stage. The source-sampling comparison therefore uses the same publicly announced query in both worlds before each action. It does not rely on replacing a pivot by a boundary query after seeing the action. Thus the entire lower-bound construction is legal whether the learner sees the query before or after submitting its predictor.

W.3 The result for thresholds

Corollary W.8 (Sharp threshold rate at low audit probabilities)**.** *Fix $\eta \in (0, 1/3)$. For every $K \geq 3$, $T \geq 1$, and $0 < \rho \leq 1/3 - \eta$, threshold games in either order of action and query, with either proper or unrestricted improper actions, have value*

$$\Theta_\eta(\min\{T, K/\rho\}).$$

At $\rho = 0$ these four values are $\Theta(T)$.

Proof. For the K thresholds, Proposition W.4 gives $d = \lfloor K/2 \rfloor$. Flipping by a central threshold gives the intersection-closed class described there. Hence the proper upper bound and the improper lower bound with queries revealed first in Theorem W.6 apply. The other two values lie between them. The factor $(1-3\rho)/(1-\rho)$ is bounded below by a positive constant depending only on η. Since $\lfloor K/2 \rfloor$ and K differ by universal factors for $K \geq 3$, the stated rate follows. The zero-audit claim follows from the small-class lower bound and the constant-half proper upper bound. □

The condition $K \geq 3$ leaves uncertainty after the first truthful label. For $K = 2$, that label identifies the target and the value is $1/2$ at every horizon and audit rate. The all-horizon matching result is uniform when audit rates stay below $1/3$ by a fixed margin. The displayed lower bound specifies how its constant deteriorates near $1/3$. No matching transition at that point is asserted.

X IMPROPER LEARNING WHEN THE INPUT IS REVEALED FIRST

This appendix studies what happens when two restrictions are relaxed together: the learner sees the query before predicting and may submit an arbitrary complete probability function. It gives an upper bound that need not hold when either properness or the original order of prediction before input selection is retained.

Let $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ contain N distinct targets on a finite domain, with the target fixed before interaction. The learner sees x_t before submitting $p_t \in [0,1]^{\mathcal{X}}$. An adversary that knows the target then selects an actual earlier function, possibly at random, before the new audit and replay randomness. The learner observes neither the source nor audit status. As in Section 2, the first round has no source and gives a truthful label after prediction. Write $\mathcal{R}_T^{\mathrm{qf,imp}}(\mathcal{H}, \rho)$ for the minimax deployment loss. We assume $N \geq 2$, since a one-target class has value zero.

X.1 LEARNING AT ONE COORDINATE WHEN ALL EARLIER SCORES EQUAL ONE HALF

We first learn one coordinate whose earlier prediction probabilities all equal one half. The analysis will also track what happens after its score changes. More generally, if its historical scores lie in $[\underline{p}, \overline{p}]$, Proposition C.3 gives the two possible intervals of feedback means. Their center and signed separation are

$$c_\rho = \frac{\rho + (1-\rho)(\underline{p} + \overline{p})}{2}, \qquad g_\rho = \rho - (1-\rho)(\overline{p} - \underline{p}). \tag{200}$$

Conditional on the target and legal pre-feedback information, the true label B satisfies $(2B-1)\mathbb{E}[Y - c_\rho] \geq g_\rho/2$, for every legal source mixture. Thus centering feedback at c_ρ gives a signal in the direction of the true label whenever $g_\rho > 0$. The next lemma allows the times at which this coordinate is queried to be chosen adaptively.

Lemma X.1 (Concentration up to adaptive completion)**.** *Consider a subsequence whose query times are chosen before the corresponding fresh observations. Suppose a fixed sign $b \in \{-1, 1\}$ and increments Z_i satisfy $\mathbb{E}[bZ_i \mid \text{past}] \geq \gamma/2$, where $\gamma > 0$, and have conditional range of length at most one. For any positive integer m, the probability that m increments occur and their signed sum is nonpositive is at most $\exp(-m\gamma^2/2)$. This is a completion-and-error probability, not a probability conditional on eventual completion.*

Proof. Conditional Hoeffding gives $\mathbb{E}[\exp(-2\gamma bZ_i) \mid \text{past}] \leq \exp(-\gamma^2/2)$. Thus the exponential of minus 2γ times the accumulated signed sum plus $\gamma^2/2$ times the number of observed increments is a nonnegative supermartingale. Stop it at the mth increment or a finite global horizon. On completion with a nonpositive signed sum it is at least $\exp(m\gamma^2/2)$, which proves the bound. Letting the global horizon increase gives the claim even when the subsequence might never reach m. □

Lemma X.2 (Learning one label from an initial guess)**.** *Suppose all earlier scores at a coordinate equal $1/2$, and fix a default label $f \in \{0,1\}$ and $0 < \varepsilon \leq 1/16$. For every $\rho \in [1/2, 1]$, there is a scalar learner keeping every earlier score available for replay such that, for either fixed true label B and every legal adaptive source policy, its cumulative loss L_∞ over all query times at this coordinate satisfies*

$$\mathbb{E}L_\infty \leq C \begin{cases} \varepsilon \log(\mathrm{e}/\varepsilon), & B = f, \\ \log(\mathrm{e}/\varepsilon), & B \neq f. \end{cases} \tag{201}$$

The constant C is universal. The same bound holds on adaptively selected and truncated subsequences of queries, while the coordinate's score is held fixed between local updates.

Proof. First take $\rho = 1/2$. Put $\delta_j = \varepsilon^{3^j}$ for $j \geq 0$. In phase j, choose a direction f_j and deploy score δ_j if $f_j = 0$, or $1 - \delta_j$ if $f_j = 1$. Initially $f_0 = f$. At phase start let $[\underline{p}_j, \overline{p}_j]$ be the smallest interval containing all earlier scores at this coordinate and the new phase score. Including the new score only enlarges the interval used for analysis. It does not make the current action a legal source on its first round. Set

$$g_j = \frac{1 - \overline{p}_j + \underline{p}_j}{2}, \qquad c_j = \frac{1}{4} + \frac{\underline{p}_j + \overline{p}_j}{4}, \qquad m_j = \left\lceil \frac{64 \log(1/\delta_{j+1})}{g_j^2} \right\rceil.$$

Keep this score for m_j local observations, and set $f_{j+1} = 1$ exactly when their sum of $Y - c_j$ is nonnegative. Between local queries the score is unchanged, so intervening global rounds only add repetitions to the history. The selected predictor's score on each round of the phase lies in this fixed interval. Equation (200) gives signed drift at least $g_j/2$. Lemma X.1, conditional on the phase-start history, yields

$$\mathbb{P}(\text{phase } j \text{ completes and } f_{j+1} \neq B \mid \text{phase-start history}) \leq \delta_{j+1}^8. \tag{202}$$

A phase can instead be truncated by the query schedule or horizon, which only lowers its loss. No step conditions on the future event that a phase completes.

Every score through phase j belongs to $[\delta_j, 1 - \delta_j]$, so $g_j \geq \delta_j$ and

$$m_j \leq C\delta_j^{-2} \log(\mathrm{e}/\delta_j). \tag{203}$$

For $j \geq 1$, an incorrectly directed phase can start only after completion and error in the preceding phase. Its expected loss is therefore at most $C\delta_j^8 \delta_j^{-2} \log(\mathrm{e}/\delta_j) = C\delta_j^6 \log(\mathrm{e}/\delta_j)$. The sum of these contributions is at most $C\varepsilon \log(\mathrm{e}/\varepsilon)$.

Returning to the correct direction does not remove a previous wrong score from history. We must therefore also count the cost of correct phases that follow an error. Before the first error, the opposite endpoint is still $1/2$, so $g_j \geq 1/4$ and these correct phases cost at most $C \sum_j \delta_j \log(\mathrm{e}/\delta_j)$. After an incorrectly directed phase r, each correct phase before the next error has gap at least $\delta_r/2$. Their total loss is at most

$$C\delta_r^{-2} \sum_{j>r} \delta_j \log(\mathrm{e}/\delta_j) \leq C\delta_r \log(\mathrm{e}/\delta_r),$$

where $\delta_{r+1} = \delta_r^3$. Charge this run to the error at r. For $r \geq 1$, Eq. (202) supplies the additional probability factor δ_r^8, and the expected charges are summable. For $r = 0$, the charge is at most $C\varepsilon \log(\mathrm{e}/\varepsilon)$.

Finally, $g_0 \geq 1/4$. The initial phase costs at most $C\varepsilon \log(\mathrm{e}/\varepsilon)$ when its direction is correct, and at most $C \log(\mathrm{e}/\varepsilon)$ otherwise. Adding these contributions proves Eq. (201) at $\rho = 1/2$.

For $\rho \geq 1/2$, use the center c_ρ from Eq. (200). Since $g_\rho \geq (1 - \overline{p}_j + \underline{p}_j)/2 = g_j$, the same phase lengths and estimates apply. Expected loss over infinitely many queries is the monotone limit of the finite-horizon losses. If the adaptive schedule queries this coordinate only finitely often, extend it using any legal sources. Discarding the added nonnegative loss gives the claimed bound for the original schedule. □

X.2 A FINITE-CLASS UPPER BOUND

We now combine the coordinatewise learners. Keep every unqueried coordinate at probability one half, so its first feedback has a source-independent distribution. Use this first observation to update weights on the target class. Later queries to that coordinate use its own scalar learner. Seeing the input first lets the learner start the scalar strategy before making its first prediction at that coordinate.

Theorem X.3 (Improper upper bound with the input revealed first)**.** *For every finite binary class $\mathcal{H}$ of size $N \geq 2$, every integer $T \geq 1$, and every $\rho \in [1/2, 1]$,*

$$\mathcal{R}_T^{\mathrm{qf,imp}}(\mathcal{H}, \rho) \leq C \min\{T, \log^2(2N)\}, \tag{204}$$

where C is universal. A single learner achieves this upper bound without knowing the horizon.

Proof. Deploy $1/2$ everywhere on the first round. Its exact label identifies the target label at that queried coordinate, which can thereafter be predicted without loss. Keep every previously unqueried coordinate at $1/2$ in every subsequent full action until its first query. Maintain initially uniform positive weights on the targets consistent with the first truthful label, and update them only at the first ordinary query to each previously unqueried coordinate.

At such a fresh coordinate x, all historical scores equal $1/2$. Consequently, for $B = h^*(x)$ and every legal source choice,

$$Y \sim \mathrm{Bern}\left(\frac{1-\rho}{2} + \rho B\right).$$

Write $\boldsymbol{w} \in \Delta(\mathcal{H})$ for the current normalized weights, with zero mass on targets ruled out by the first truthful label. Let f_x be the weighted-majority label and μ_x the minority weight mass, using a fixed tie convention. If $\mu_x > 0$, start the scalar learner of Lemma X.2 with default f_x and $\varepsilon_x = \mu_x^2/16$. The learner can make this choice before predicting because it already knows the input. If $\mu_x = 0$, all remaining targets agree, so predict their known label. Update the weights at this first observation by

$$w'(h) = \frac{w(h)\exp\{\eta(Y-1/2)h(x)\}}{\sum_{g\in\mathcal{H}} w(g)\exp\{\eta(Y-1/2)g(x)\}}, \qquad \eta = 1/64.$$

On repeated queries, update only that coordinate's scalar learner and leave every other prediction probability unchanged. Although the current score at a newly queried coordinate has changed, its first feedback still has the displayed distribution: every earlier function has score one half there, and the current function becomes available for replay only on later rounds.

For the fixed true target, let $\Phi = \log(1/w(h^*)) \geq 0$ and let e be the wrong-label weight mass at a fresh coordinate. Its exact update satisfies

$$\exp(\Phi' - \Phi) = 1 - e + e\exp\{-\eta(2B-1)(Y-1/2)\}.$$

Conditional on the activation history and the fixed target, $X = (2B-1)(Y-1/2)$ takes values $\pm 1/2$ and has mean $\rho/2 \geq 1/4$. If $a = \Phi' - \Phi$, the inequalities $\log(1+z) \leq z$ and the bounded exponential expansion give $\mathbb{E}a \leq e\,\mathbb{E}(e^{-\eta X} - 1) \leq -\eta e/8$. Also $|a| \leq \eta e$: Jensen's inequality gives the lower side, and $e^{\eta/2} - 1 \leq \eta$ gives the upper side. Hence, for $F(\Phi) = \Phi^2 + 4\Phi$,

$$\mathbb{E}[F(\Phi') - F(\Phi) \mid \text{activation history}, h^*] = (2\Phi + 4)\mathbb{E}a + \mathbb{E}a^2 \leq -\frac{\eta}{8}e(1+\Phi). \tag{205}$$

If the default is correct, $e = \mu_x$, and the conditional total expected loss at this coordinate is at most $C\mu_x^2 \log(\mathrm{e}/\mu_x) \leq Ce$. If the default is wrong, $e \geq 1/2$. The minority includes the true target's weight, so $\mu_x \geq \exp(-\Phi)$, and the scalar cost is at most $C(1 + \log(1/\mu_x)) \leq C(1+\Phi) \leq 2Ce(1+\Phi)$. These bounds hold conditional on the activation history for either fixed true label, uniformly over subsequent adaptive queries and source choices. Coordinates with $\mu_x = 0$ cost zero.

Sum the conditional costs over coordinates first queried before T. Equation (205) bounds their sum by the total decrease of F. Initially $\Phi \leq \log N$, and $F \geq 0$. Including the first-round loss $1/2$ gives $O(\log^2(2N))$. Truncating a scalar learner only reduces its loss. Since each round costs at most one, this horizon-independent learner satisfies $\mathbb{E}L_T \leq \min\{T, C\log^2(2N)\}$, proving Eq. (204). Its updates use only observed labels and its own state, and all past predictors remain available for replay. □

Y Additional Related Work

This appendix expands Section 6, connecting stochastic replay to self-training, recursive estimation, language generation, and learning with restricted feedback.

Y.1 Self-Training and Recursive Synthetic Data

Self-training methods use a model's predictions to supervise subsequent updates, often combining pseudo-labels with augmentation or confidence selection. Noisy Student trains a student on labels supplied by an earlier teacher and then repeats this procedure (Xie et al., 2020). FixMatch combines

confident pseudo-labels with consistency across augmented inputs (Sohn et al., 2020). Theory identifies conditions under which self-training can improve predictions. Wei et al. (2021) analyze self-training under expansion and class-separation assumptions, Oymak & Gulcu (2021) study Gaussian-mixture classification, and Kumar et al. (2020) give guarantees under gradual domain shift. These works explain how distributional structure and update design support successful pseudo-label refinement. Our setting instead asks what is possible uniformly over adversarial queries and target-aware selections from the learner's actual prediction history.

The model-collapse literature studies degradation when generated data enters later training sets. Shumailov et al. (2024) and Alemohammad et al. (2024) document losses in quality or diversity under recursive training. Language-model experiments examine declining diversity in self-consuming training loops (Briesch et al., 2023) and the role of prediction errors and lexical diversity in repeated synthetic translation training (Zhang et al., 2024). Statistical analyses relate synthetic training to changes in scaling laws (Dohmatob et al., 2024), persistent error in regression (Dohmatob et al., 2025), and collapse times for recursive maximum-likelihood estimation (Suresh et al., 2025). Other perspectives describe closed-loop generative dynamics (Marchi et al., 2024) and a transition from generalization to memorization in recursive diffusion training (Shi et al., 2025). These mechanisms depend on the model, estimator, and data-reuse protocol. Our results concern cumulative prediction loss against a fixed target, optimized over all admissible learners.

The choice of data retained between generations can change the outcome. Accumulating real and synthetic data can stabilize recursive learning in settings where replacing the training set leads to deterioration (Gerstgrasser et al., 2024; Kazdan et al., 2025). Stability analyses also identify favorable initialization and real-data mixture conditions (Bertrand et al., 2024; Seddik et al., 2024). Taori & Hashimoto (2023) relate feedback-induced bias amplification to how faithfully model outputs are sampled. Verification provides another intervention. Fu et al. (2025) analyze confidence-based internal filtering under conditions supporting stable recursive training, while Yi et al. (2026) study the benefits and limitations of external verifiers in linear regression. These results motivate examining the information available about synthetic supervision. Our audit and disclosure models make that information explicit and quantify its effect on the minimax loss, including when disclosure is partial or noisy.

Y.2 Replay and Language Generation

Language generation in the limit separates producing new valid examples from identifying the entire target language. Kleinberg & Mullainathan (2024) establish eventual generation for countable language collections. Subsequent work studies uniform and non-uniform generation through combinatorial dimensions (Raman et al., 2025), the roles of breadth and feedback (Charikar & Pabbaraju, 2025), and tradeoffs between validity and output diversity (Kalavasis et al., 2025). Robust variants allow finitely many erroneous examples (Raman & Raman, 2025) or infinite contamination (Mehrotra et al., 2026). Their success criteria concern eventual generation and properties of the output support, whereas our objective charges prediction error throughout a finite interaction.

Racca et al. (2026) introduce replay into language generation by allowing previous outputs to re-enter the example stream. They show that its effect depends on the notion of generation, with robustness for uniform generation and separations for other notions. Both their setting and ours constrain contamination to actual earlier outputs. Our supervised model additionally tracks fresh sampling from stored probability functions and independent truthful audits. The resulting statistical problem is to control cumulative deployment loss, rather than eventual valid novelty.

The closest supervised predecessor is Dmitriev et al. (2026). Their loss counts a prediction error when the received label agrees with the target, including when a replayed label happens to be correct. Our loss counts deployment error on every round. Their closure construction supplies our low-audit upper bound, while the matching lower bound must allow stochastic predictions and the prescribed audit mechanism (Appendix W). Theorem 3.1 treats their proposed stochastic-prediction and audit extensions jointly for finite thresholds within its stated horizon range.

Y.3 Online Learnability, Proper Prediction, and Closure

Classical realizable online learning is governed by Littlestone dimension (Littlestone, 1988). Agnostic online learning instead compares the learner with a best hypothesis in hindsight (Ben-David et al., 2009), while stochastic and constrained adversaries give other ways to restrict the input sequence (Rakhlin et al., 2011). In our protocol, the target is realizable and fixed, but feedback need not reveal its label. Random audits do not make the query sequence stochastic. This distinction explains why ordinary realizable mistake bounds do not by themselves determine the cost of replay.

Proper prediction can introduce additional combinatorial constraints. Helly-type structure controls proper PAC learning guarantees (Bousquet et al., 2020), and Hanneke et al. (2021) characterize online mistake bounds when predictions must come from a restricted family. Our proper actions are distributions over thresholds, whose induced probability functions enter the replay archive. Closure methods have a longer history in learning under malicious noise (Auer & Cesa-Bianchi, 1998) and in optimal guarantees for intersection-closed classes (Auer & Ortner, 2007; Darnstädt, 2015). Refined error bounds connect monotone error regions with sample compression (Hanneke, 2016), and intersection-closed classes admit small unlabelled compression schemes (Rubinstein & Rubinstein, 2022). The replay framework of Dmitriev et al. (2026) uses closure to preserve information that can be trusted despite the learner's past predictions. In our upper bounds, clipping controls the separation between feedback distributions. Appendix U explains how closure can nevertheless improve on the worst-class threshold rate for some finite classes.

Y.4 Selective Feedback and Noisy Supervision

Active learning reduces label requirements by deciding which examples need labels (Cohn et al., 1994; Hanneke & Yang, 2015). Selective classification permits abstention when a prediction cannot be made reliably (El-Yaniv & Wiener, 2010), and its connection to active learning is studied by Gelbhart & El-Yaniv (2019). For streams mixing stochastic and adversarial examples, Goel et al. (2023) analyze a tradeoff between abstentions and mistakes. These models make information acquisition or abstention part of the learner's current decision. Here the audit rate is specified externally, deployment loss is charged on every round, and a current prediction changes the sources available for future feedback. The learner's information-preservation cost therefore persists even without an explicit label-query budget.

Classical label-noise models study corrupted observations of a target label, including independent classification noise (Angluin & Laird, 1988), bounded-noise risk guarantees (Massart & Nédélec, 2006), and asymmetric noise with unknown proportions (Scott et al., 2013). When the corruption law is known, unbiased loss correction can recover useful learning objectives (Natarajan et al., 2013). Our centered-feedback updates have a related statistical purpose. The distinction is that the replay contribution depends on stored actions and may be selected using the target. Knowing the actual selected score permits mean correction, but Section 5 shows that the randomness of a fresh replay draw still affects the attainable rate.

Y.5 Provenance and Recursive Estimation

In recursive distribution estimation, the ability to distinguish real observations from generated ones can change statistical efficiency. Kanabar & Gastpar (2026a) derive minimax bounds for recursive discrete estimation and compare access to sample origins. Their extension studies watermark channels and masking (Kanabar & Gastpar, 2026b). Appendix H relates their channel divergence to our information measure for noisy audit flags. Our model also distinguishes audit status from the identity or score of the replayed predictor. These fields give the same minimax order when the learner designs its history from the outset, but can behave differently when the first predictor is fixed.

Watermarking provides a practical route to information about data origins. Proposed schemes embed detectable signals in language-model outputs (Kirchenbauer et al., 2023), study the reliability of those signals (Kirchenbauer et al., 2024), and demonstrate scalable deployment (Dathathri et al., 2024). Our known noisy-audit channel captures uncertainty about label origins, while disclosure of the selected predictor's identity or score supplies a different kind of information.

Amin et al. (2026) study learning from mixtures of natural and synthetic data with hidden origins. They show that empirical risk minimization can fail in recursive classification even though other

procedures can learn under their model. This supports separating limitations of a chosen training rule from limits of the observation process. Our minimax formulation makes that distinction explicit. It also allows an adversary to choose among actual past predictors, so knowledge of the previous classifier alone need not determine the current replay mean. Appendix F gives guarantees when that conditional mean is known.

Y.6 Action-Dependent Feedback and Deployment Effects

Partial monitoring formalizes situations in which actions determine both loss and the observations available to the learner (Bartók et al., 2014). Apple Tasting is a classification example with action-dependent label feedback (Helmbold et al., 2000; Raman et al., 2024). Our critical-rate upper bound similarly balances immediate prediction cost against future information. Its feedback mechanism also depends on the entire archive of earlier predictors. Establishing the lower bound therefore requires producing the misleading sources within the same interaction, rather than starting from a fixed observation channel.

Performative prediction studies distributions that respond to deployed models (Perdomo et al., 2020), including direct optimization of performative risk (Miller et al., 2021), state-dependent responses (Brown et al., 2022), and regret under performative feedback (Jagadeesan et al., 2022). Policy regret addresses the effect of earlier actions on an adaptive adversary (Arora et al., 2012; Cesa-Bianchi et al., 2013). In our setting, the target and its labeling rule stay fixed, while past deployments influence later supervision. Controlling changes in predictions also has methodological connections to learning with delayed feedback (Joulani et al., 2016). Our bound under a restriction on expected replay age holds a predictor fixed long enough to limit feedback from older sources, then pays for these waiting periods through changes in internal weights. The delay concerns which deployed predictor generates a new label, rather than when an already generated label arrives.