COUNTERFACTUAL UPDATES FOR CONFORMAL RETRIEVAL WITH SEMI-BANDIT FEEDBACK

Anonymous authors
Paper under double-blind review

ABSTRACT

Conformal retrieval constructs document sets at a prescribed coverage rate. Under semi-bandit feedback, a successful retrieval reveals the correct document, while a miss reveals only its absence. Existing conservative threshold methods provide guarantees for smooth coverage objectives, leaving the discrete number of returned documents unresolved. Coverage protection also changes the feedback used to learn the threshold. We propose counterfactual updates that recover whether the original proposed set would have contained the correct document. Whenever another miss would exceed the cumulative miss budget, the method returns the full candidate pool and uses the revealed document to evaluate the original proposal. Assuming every candidate pool contains the correct document, this preserves the unprotected learner's threshold trajectory and guarantees cumulative coverage after every query. For i.i.d. queries with fixed scores, a unique continuous target quantile and continuity of expected document cost at that quantile, average returned-set size converges to the best feasible fixed-threshold cost. In contrast, updating from protected hits keeps the threshold at or below its initial value and requires full-pool returns on a positive fraction of queries when initialized below the target quantile. The decreasing-step-size learner tracks that quantile directly, while conservative estimation certifies a threshold through a distributional bound. In the initial ten-order SQuAD comparison, our method returns 2.64 documents per query versus 9.44 for conservative threshold estimation. Protection adds about 0.09 documents per query relative to the same unprotected learner with identical step sizes, while maintaining at least 90% cumulative coverage after every query. With unchanged parameters, evaluation on Natural Questions also produces smaller sets than the compared conservative and constant-step methods. These results show that learning from the original proposal supports efficient retrieval with cumulative coverage protection.

1 INTRODUCTION

Question-answering systems use retrieved documents as evidence, making the number of returned documents a direct component of downstream processing cost. Conformal retrieval controls how often a selected set omits the correct document (Fisch et al., 2021; Ge et al., 2025). In sequential retrieval, that set also determines the feedback available for learning. We consider finite candidate pools that contain the correct document: a successful return reveals its identity, while an unsuccessful return reports only a miss. Under this semi-bandit feedback, we seek small returned sets while maintaining a prescribed fraction of successful retrievals among all queries seen so far, after every query.

Existing approaches address different parts of this problem. Split conformal prediction calibrates a threshold on a separate labeled sample (Vovk et al., 2005; Lei et al., 2018). The Semi-bandit Prediction Set (SPS) algorithm instead learns a conservative threshold from semi-bandit feedback (Ge et al., 2025), with coverage and smooth coverage-loss guarantees under independent and identically distributed inputs. The actual number of documents changes discretely as scores cross the threshold, so these guarantees leave a distinct document-cost objective. OCP-Unlock+ uses threshold monotonicity to recover feedback for unchosen sets (Yang et al., 2026), while adaptive conformal inference methods learn from observed success and use boundary actions for stability (Gollapudi et al., 2026).

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.

Protecting cumulative coverage creates a further interaction between selection and learning. A system can expand an original proposal to ensure a successful return when its remaining miss budget is small. The expanded set may then succeed even though the original proposal would have failed. Updating from this actual success changes the threshold learner's feedback precisely when protection intervenes. Efficient protection therefore requires separating the outcome used for learning from the success of the returned set.

We introduce *counterfactual updates* that recover the original proposal's hit or miss. Before returning a set, a cumulative miss-budget rule checks whether another miss could exceed the allowed total and, when necessary, returns the full candidate pool. Because that pool contains the correct document, protection reveals its identity. We can then determine whether it belonged to the original proposal. When no expansion occurs, the observed hit or miss already supplies the required outcome. Using this feedback with the decreasing-step-size learner of Angelopoulos et al. (2024a) reproduces its unprotected threshold trajectory, while the returned sets satisfy target cumulative coverage after every query. Figure 1 compares the resulting cost with updates based on actual returned-set success.

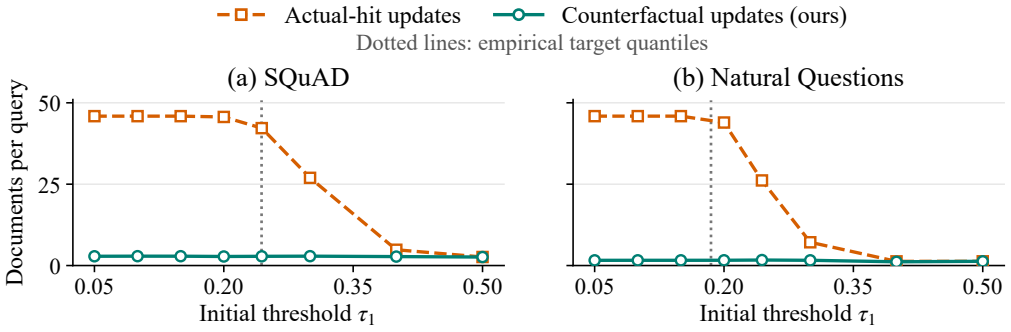

Figure 1: **Feedback choice determines the cost of coverage protection.** Both variants hold 90% cumulative coverage after every query. Counterfactual updates use the original proposal's outcome and actual-hit updates use the returned set's success. Points are means with $\pm$ SE over 20 query permutations (Section 6.3).

The distinction also determines long-run cost. For i.i.d. queries with fixed scores, a unique target score quantile, and continuity of the score distribution and expected document cost at that quantile, counterfactual updates make protection occupy a vanishing fraction of queries (Section 5). The average returned-document count then converges to the best feasible fixed-threshold cost. By comparison, under the same protection rule and nonincreasing step sizes, actual-hit updates keep the threshold at or below its initial value. In the stochastic setting analyzed, a positive initialization below the target quantile therefore requires full-pool returns on a positive fraction of queries.

Experiments quantify both document cost and the additional cost of protection. In the initial ten-order SQuAD comparison (Rajpurkar et al., 2016), the method returns 2.64 documents per query versus 9.44 for SPS with Dense Passage Retrieval scores (Karpukhin et al., 2020) at the same coverage target (Table 7). Protection adds about 0.09 documents per query relative to the underlying learner with identical initialization and step sizes (Table 8). The method maintains at least 90% cumulative coverage after every query, and evaluation on Natural Questions (Kwiatkowski et al., 2019) with the same parameter choices also produces smaller sets than SPS and scalar Primal-ACI (Section 6).

Our contributions are as follows.

- We combine counterfactual feedback recovery with a miss-budget rule to protect every query prefix while preserving the underlying threshold trajectory (Proposition 1 and Observation 1).
- We establish convergence to the best feasible fixed-threshold document cost under the stated stochastic conditions and characterize the persistent protection cost caused by actual-hit feedback at low initial thresholds (Theorem 1 and Proposition 2).
- We evaluate coverage, document count and feedback choice on SQuAD and Natural Questions, using paired comparisons to measure the cost added by protection.

2 Related Work

Split and online conformal prediction. Split conformal prediction provides marginal coverage under exchangeability (Vovk et al., 2005; Lei et al., 2018; Angelopoulos & Bates, 2023; Tailor et al., 2025; Angelopoulos et al., 2024b). Online methods adapt calibration (Gibbs & Candès, 2021), combine step sizes (Zaffran et al., 2022; Gibbs & Candès, 2024; Lin et al., 2022; Areces et al., 2025), select experts (Bhatnagar et al., 2023), control multivalid coverage (Bastani et al., 2022; Ramalingam et al., 2025) or forecast scores (Angelopoulos et al., 2023). We retain the decreasing-step-size quantile tracker of Angelopoulos et al. (2024a) and recover its proposal feedback during full-pool returns.

Conformal prediction with partial feedback. SPS estimates coverage conservatively from censored scores (Ge et al., 2025). OCP-Unlock+ infers every grid threshold's outcome from a revealed label (Yang et al., 2026); we use that inference at one threshold to preserve a quantile tracker under coverage protection. Other approaches use abstention (Lee et al., 2025), unbiased class-specific estimates (Wang & Qiao, 2024) or separate label queries (Skalse et al., 2026; Cesa-Bianchi et al., 2005). Partial monitoring studies incomplete observation signals (Tsuchiya et al., 2023), while logged-feedback methods estimate counterfactual outcomes from recorded actions (Dudik et al., 2011; Swaminathan & Joachims, 2015; Taufiq et al., 2022).

Efficient selection and semi-bandit learning. Primal-ACI learns low-cost sets from returned-action success (Gollapudi et al., 2026); full-feedback methods optimize interval length (Srinivas, 2026). Combinatorial semi-bandits use selected-item feedback (Chen et al., 2013; Audibert et al., 2014; Kveton et al., 2015; Zimmert et al., 2019), while multiple-play and submodular analyses compare fixed subsets (Komiyama et al., 2015; Streeter & Golovin, 2008; Tajdini et al., 2024). Conservative bandits protect cumulative reward with a default arm (Wu et al., 2016). For an always-successful default and unit hit rewards, their rule becomes Equation (3). Our question is which feedback updates the threshold after that return (Proposition 2).

Conformal retrieval and generation. Conformal methods calibrate retrieved evidence (Fisch et al., 2021; Li et al., 2024; Rouzrokh et al., 2024; Xu et al., 2025; Chakraborty et al., 2026; Zhu et al., 2026) for retrieval-augmented systems (Lewis et al., 2020), or filter generated answers and subclaims (Mohri & Hashimoto, 2024; Quach et al., 2024; Feng et al., 2025; Kang et al., 2024). Complementing ranked-relevance evaluation (Järvelin & Kekäläinen, 2002) and cross-task retrieval (Thakur et al., 2021), we study online set selection on SQuAD (Rajpurkar et al., 2016) and Natural Questions (Kwiatkowski et al., 2019) with fixed DPR scores (Karpukhin et al., 2020).

3 Setup

We distinguish proposed and returned sets because coverage protection changes both feedback and document cost.

3.1 Threshold sets and semi-bandit feedback

At query t, Z_t comprises the query and K candidates with precomputed scores. Their local indices $\mathcal{Y} = \{1, \dots, K\}$ need not denote the same documents across queries. The pool contains a designated correct document with hidden index $Y_t \in \mathcal{Y}$. After returning a set, feedback arrives before the next update: a hit reveals Y_t, while a miss reveals only its absence (Ge et al., 2025).

A fixed function s assigns nonconformity scores in $[0, 1]$, with smaller scores indicating better matches. Define the nested family

$$C(Z_t, \tau) = \begin{cases} \varnothing, & \tau \le 0, \\ \{j \in \mathcal{Y} : s(Z_t, j) \le \tau\}, & 0 < \tau < 1, \\ \mathcal{Y}, & \tau \ge 1. \end{cases} \tag{1}$$

Interior thresholds include all ties. At zero the empty-set boundary action overrides score inclusion, even for a zero-scored document. Write $C_t(\tau) = C(Z_t, \tau)$.

Coverage protection may enlarge the proposed set $P_t = C_t(\tau_t)$ to produce the returned set R_t, where τ_t is the current threshold. We distinguish the returned-set and proposal hit indicators:

$A_t = \mathbf{1}\{Y_t \in R_t\}, \qquad B_t = \mathbf{1}\{Y_t \in P_t\}$. The returned hit A_t updates the cumulative miss count. The proposal hit B_t updates the threshold and is recovered as in Section 4.3.

3.2 Cumulative coverage and document cost

Let $\alpha \in (0,1)$ be the permitted miss fraction and $\phi = 1 - \alpha$ the target coverage. The returned sets determine the cumulative miss count and realized coverage: $M_t = \sum_{i=1}^{t}(1 - A_i), \qquad \widehat{\mathrm{Cov}}_t = 1 - M_t/t, \qquad M_0 = 0$. We require $\widehat{\mathrm{Cov}}_t \geq \phi$ after every query, which returning $\mathcal{Y}$ always satisfies. We minimize returned documents subject to this prefix constraint. Over T queries, the actual average cost is $\widehat{J}_T = T^{-1}\sum_{t=1}^{T}|R_t|$, charging each full-pool return at K documents.

The stochastic benchmark is a fixed threshold satisfying population coverage under an unknown distribution of (Z, Y). Define $g(\tau) = \mathbb{E}[|C(Z,\tau)|], \qquad J^\star = \inf_{\tau:\,\mathbb{P}\{Y\in C(Z,\tau)\}\geq\phi} g(\tau)$, where g is the expected document cost and $J^\star$ is the best feasible fixed-threshold cost. Nestedness makes g nondecreasing, so the smallest feasible threshold attains $J^\star$ whenever that threshold exists.

The cost theorem connects this population benchmark with realized prefix protection through vanishing average additional cost.

4 Method

Each query forms a proposal, checks the miss budget, returns a set, recovers the proposal outcome, then updates the miss count and threshold.

4.1 Learning a threshold with decreasing step sizes

We use the decreasing-step-size rule of Angelopoulos et al. (2024a) with the proposal feedback recovered in Section 4.3. Starting from $\tau_1 \in \mathbb{R}$, we update

$$\tau_{t+1} = \tau_t + \eta_t(\phi - B_t), \qquad \eta_t = c\,t^{-\beta}, \qquad c > 0, \quad \tfrac{1}{2} < \beta \leq 1, \tag{2}$$

where η_t is the step size, c is its initial value and β controls its decay. This exponent range supports stochastic convergence (Appendix A.6); prefix protection applies to any proposal sequence. A proposal miss ($B_t = 0$) increases the threshold by $\eta_t\phi$, and a proposal hit ($B_t = 1$) decreases it by $\eta_t\alpha$. The threshold is unprojected; Equation (1) specifies returns outside $[0,1]$. When every proposal is returned, $B_t = A_t$ and the rule is the unprotected decreasing-step-size learner.

4.2 Choosing the returned set from the miss budget

The requirement $\widehat{\mathrm{Cov}}_t \geq \phi$ is equivalent to $M_t \leq \alpha t$. The integer allowance is $L_t = \lfloor \alpha t \rfloor$. We protect the return when its remaining capacity $L_t - M_{t-1}$ cannot accommodate another miss:

$$L_t = \lfloor \alpha t \rfloor, \qquad H_t = \mathbf{1}\{M_{t-1} + 1 > L_t\}, \tag{3}$$

where $H_t = 1$ marks a budget-triggered full-pool return. If $H_t = 0$, even a proposal miss keeps the cumulative count within L_t, so we return $R_t = P_t$. If $H_t = 1$, we return $R_t = \mathcal{Y}$ to ensure a hit. For example, $L_t = 0$ for $t < 1/\alpha$, so this rule makes nine initial full-pool returns at $\alpha = 0.1$.

4.3 Recovering feedback for counterfactual updates

When $H_t = 0$, the proposal is returned and $B_t = A_t$. When $H_t = 1$, the full-pool return gives $A_t = 1$ and reveals Y_t, allowing us to check its membership in P_t. Thus, the proposal hit indicator is computed exactly as

$$B_t = \begin{cases} A_t, & H_t = 0, \\ \mathbf{1}\{Y_t \in P_t\}, & H_t = 1 \end{cases}. \tag{4}$$

This exact membership recovery defines a *counterfactual update*: it evaluates the original proposal even when the full pool is returned. Without protection, a miss supplies $B_t = 0$ directly. For example, a revealed document outside P_t gives $A_t = 1$ and $B_t = 0$. The miss count stays at $M_t = M_{t-1}$, while the threshold increases by $\eta_t\phi$.

Per-query procedure. Fix the score and set interface above, target $\phi = 1 - \alpha$, finite τ_1, $c > 0$ and $\beta \in (1/2, 1]$. Initialize $M_0 = 0$. For each query:

1. Observe Z_t and form $P_t = C_t(\tau_t)$.
2. Compute L_t and H_t, then return $R_t = \mathcal{Y}$ if $H_t = 1$ and $R_t = P_t$ otherwise.
3. Observe A_t and, upon a hit, Y_t. Recover B_t using Equation (4).
4. Set $M_t = M_{t-1} + 1 - A_t$ and update τ_{t+1} using Equation (2).

Computation. Given candidate scores, selection and feedback recovery take $O(K)$ time per query. The threshold, miss count and query index use $O(1)$ persistent state. These costs exclude scoring-model computation and the current K-candidate data and output storage. Membership can reuse the current proposal or test the revealed score under Equation (1).

Feedback and step-size controls. The actual-hit variant keeps the same budget rule but updates from A_t, as analyzed in Section 5. A different control removes protection and uses a constant step size. With the same initialization, scores, ties and boundary actions, that control is the scalar Primal-ACI update (Gollapudi et al., 2026). Its constant step size is outside the decaying-step cost theorem.

5 THEORY

We establish coverage after every query and equality of threshold trajectories, then bound the protection frequency to obtain the average document-cost limit.

5.1 COVERAGE AND THRESHOLD LEARNING

First, we give pathwise statements under the candidate-pool and feedback interface.

Proposition 1 (Cumulative coverage)**.** *For any input sequence satisfying the candidate-pool condition in Section 3.1, apply Equation* (3) *to any sequence of proposed sets. Starting from $M_0 = 0$, the returns $R_t = \mathcal{Y}$ when $H_t = 1$ and $R_t = P_t$ otherwise satisfy $M_t \le \lfloor \alpha t \rfloor$, $\quad \widehat{\mathrm{Cov}}_t \ge \phi$ for every $t \ge 1$.*

The proof is in Appendix A.1. Counterfactual feedback also preserves the threshold trajectory.

Observation 1 (Equality of threshold trajectories)**.** *Let $\bar{\tau}_t$ denote the threshold of the unprotected learner, which always returns $C_t(\bar{\tau}_t)$ and updates from that set's hit indicator using Equation* (2)*. Couple the learners on the same query–label sequence with identical initialization, step sizes, scores, ties and boundary actions. Then the recovered proposal hits give $\tau_t = \bar{\tau}_t$ for every $t \ge 1$. This equality concerns thresholds and proposals, while actual returned sets can differ.*

Appendix A.2 proves this by coupling the recovered feedback.

Vanishing protection frequency. A budget trigger adds $H_t(K - |P_t|)$ documents on query t. It adds none when the proposal already equals the full pool. Let $N_T = \sum_{t=1}^{T} H_t$ count these triggers, including those for which $P_t = \mathcal{Y}$ already. For analysis, let $\mathcal{F}_{t-1}$ contain the complete query variables, labels and learner states through query $t - 1$. This analysis history includes hidden past labels. Define $X_t = 1 - B_t$ and $p_t = \mathbb{P}(X_t = 1 \mid \mathcal{F}_{t-1})$; neither is additional algorithm input. The cost theorem establishes the lemma's convergence premise.

Lemma 1 (Vanishing protection frequency)**.** *If $p_t \to \alpha \in (0, 1)$ almost surely, the decisions in Equation* (3) *satisfy $\frac{N_T}{T} \longrightarrow 0$ almost surely.*

The proof is in Appendix A.3.

5.2 DOCUMENT COST AND FEEDBACK CHOICE

We next establish that the proposal miss probability converges to α and use the lemma to transfer proposal cost convergence to the returned sets. **Stochastic conditions.** We retain the finite candidate pool and fixed scoring function from Section 3.1. **(A1)** The pairs (Z_t, Y_t) are independent and identically distributed, with each pair independent of $\mathcal{F}_{t-1}$. **(A2)** The correct-document score $S_t = s(Z_t, Y_t)$ has cumulative distribution function $F(u) = \mathbb{P}(S_t \le u)$ and a unique target quantile

$\tau^\star$ at which F is continuous. Specifically, $F(\tau^\star) = \phi$, with $F(u) < \phi$ for $u < \tau^\star$ and $F(u) > \phi$ for $u > \tau^\star$. **(A3)** The expected document cost $g(\tau) = \mathbb{E}[|C(Z, \tau)|]$ is continuous at $\tau^\star$. This condition concerns the expected number of documents: individual returned-set sizes can still change in discrete steps as a threshold passes a candidate score.

Under (A1)–(A2), the decreasing-step-size analysis of Angelopoulos et al. (2024a) gives unprotected threshold convergence, transferred by trajectory equality. The continuous target quantile lies in $(0, 1)$, so eventually $p_t = 1 - F(\tau_t) \to \alpha$. Lemma 1 controls protection frequency, and (A3) transfers threshold convergence to expected document cost.

Theorem 1 (Asymptotic document cost)**.** *Under conditions (A1)–(A3), run the method in Section 4 with $c > 0$ and $\beta \in (1/2, 1]$. For any finite initial threshold τ_1, the threshold, protection frequency and actual average document cost satisfy*

$$\tau_t \longrightarrow \tau^\star, \qquad \frac{N_T}{T} \longrightarrow 0, \qquad \widehat{J}_T \longrightarrow g(\tau^\star) = J^\star \quad \textit{almost surely}, \tag{5}$$

where $J^\star$ is the best feasible fixed-threshold cost in Section 3.2.

Relative to the proposals on the same input sequence, the extra average cost satisfies

$$0 \le \widehat{J}_T - \frac{1}{T}\sum_{t=1}^{T} |P_t| = \frac{1}{T}\sum_{t=1}^{T} H_t(K - |P_t|) \le K\frac{N_T}{T}.$$

The trigger-count bound makes this exact additional cost vanish on average, transferring proposal-cost convergence to actual returns. Appendix A.4 gives the full proof.

Feedback on budget-triggered returns. For the actual-hit variant, write N_T^{act} and $\widehat{J}_T^{\mathrm{act}}$ for trigger count and document cost. We first bound its threshold pathwise, then its stochastic protection frequency.

Proposition 2 (Actual-hit thresholds stay at or below their initial value)**.** *Replace Equation* (2) *by $\tilde{\tau}_{t+1} = \tilde{\tau}_t + \eta_t(\phi - A_t)$ with positive nonincreasing step sizes, and apply Equation* (3) *to the proposals $C_t(\tilde{\tau}_t)$. Then $\tilde{\tau}_t \le \tilde{\tau}_1$ for every $t \ge 1$ on every input sequence satisfying the candidate-pool condition. If conditions (A1) and (A2) also hold with $0 < \tilde{\tau}_1 < \tau^\star$, the initial proposal miss probability $p_{\mathrm{init}} = 1 - F(\tilde{\tau}_1)$ exceeds α and*

$$\liminf_{T\to\infty} \frac{N_T^{\mathrm{act}}}{T} \ge 1 - \frac{\alpha}{p_{\mathrm{init}}}, \qquad \liminf_{T\to\infty} \widehat{J}_T^{\mathrm{act}} \ge K\Big(1 - \frac{\alpha}{p_{\mathrm{init}}}\Big) \quad \textit{almost surely}, \tag{6}$$

where N_T^{act} counts the triggers and $\widehat{J}_T^{\mathrm{act}}$ measures the document cost of this variant.

Started below the target quantile, the actual-hit variant keeps proposing sets whose miss probability exceeds α, requiring frequent budget-triggered returns. Each resulting hit decreases its threshold by $\eta_t\alpha$. Counterfactual feedback instead increases the threshold by $\eta_t\phi$ whenever the revealed document lies outside the proposal. The proof is in Appendix A.5.

6 Experiments

We test document cost under prefix coverage, separate the cost of protection from proposal learning, and examine how feedback choice changes the threshold trajectory. Sensitivity studies then vary initialization, step size and retrieval conditions.

6.1 Setup

Data and scores. We use the first 10,000 positive-containing queries from DPR-adapted SQuAD training data (Rajpurkar et al., 2016; Karpukhin et al., 2020) and all 6,515 Natural Questions (NQ) development queries (Kwiatkowski et al., 2019). Each fixed pool contains its first listed positive and 50 supplied ordinary negatives ($K = 51$). The official NQ-trained DPR BERT bi-encoder supplies scores $(1 - \cos)/2$ on both collections. These experiments measure returned-set selection within positive-containing pools.

Comparisons. We compare conservative SPS (Ge et al., 2025), constant-step Primal-ACI (Gollapudi et al., 2026), grid-based OCP-Unlock+ (Yang et al., 2026) and unprotected decreasing-step-size learning (DLR) (Angelopoulos et al., 2024a). Their interfaces and configurations appear in Appendix B.1. Our method and matched unprotected DLR use $.2t^{-0.6}$ and $\tau_1 = .5$. We select $c = .2$ on a separate permutation of the same SQuAD collection and transfer it unchanged to NQ. Table 13 separately tests selection on disjoint development queries. Methods share scores and orders within each comparison; Table 5 distinguishes their updates and evaluation roles.

Measurements and batches. At target $\phi = .9$, a successful run has at least $\lceil .9t \rceil$ returned hits at every query t. We measure documents/query, with excess relative to the empirically feasible fixed threshold chosen using the complete collection. Protection cost counts documents added to proposals; N_T counts budget triggers, including zero-increment returns. Means and SE describe order variation, and cost contrasts use paired 95% t intervals. The main, feedback and initialization studies use separate twenty-order batches. Table 4 indexes these and the initial ten-order results retained in the abstract.

6.2 Smaller Sets With Cumulative Coverage

Table 1: **Document cost and coverage within positive-containing fixed pools.** Costs are mean $\pm$ SE over twenty shared permutations per dataset. The runs column counts permutations meeting 90% cumulative coverage after every query. The separated full-information reference is chosen after observing the complete collection and need not satisfy early prefixes.

	SQuAD		Natural Questions	
Method	Documents/query	Runs	Documents/query	Runs
SPS	9.48 ± 0.06	20/20	7.79 ± 0.01	20/20
OCP-Unlock+	39.59 ± 0.05	0/20	35.63 ± 0.07	0/20
Primal-ACI	3.81 ± 0.01	20/20	2.151 ± 0.001	20/20
DLR ($c = 1$)	2.86 ± 0.01	0/20	0.993 ± 0.004	0/20
DLR ($c = 0.2$)	2.550 ± 0.007	8/20	1.296 ± 0.001	20/20
Counterfactual (ours)	2.65 ± 0.02	20/20	1.296 ± 0.001	20/20
Fixed-threshold reference	2.189	0/20	0.905	0/20

On SQuAD, counterfactual updates return 2.65 documents/query, about 72% fewer than SPS, with every-prefix coverage in all twenty orders (Table 1). SPS has the lowest cumulative weighted coverage loss (CDF loss) and never falls below the empirical feasible threshold (Table 6). The comparison thus measures document cost alongside realized prefix coverage. For our method the miss budget guarantees that coverage on each sequence; the table records its implementation and the other methods' observed performance.

On NQ, the encoder ranks the positive first on 99.2% of queries. Matched unprotected DLR already meets the prefix target in all orders, compared with eight SQuAD orders. Protection therefore repairs observed shortfalls on SQuAD and retains the guarantee with almost no extra cost on NQ. Empty proposals are allowed, explaining the NQ full-information reference below one document/query. That reference meets the whole-collection constraint and can violate earlier prefixes.

Cost of protection. Table 2(a) decomposes excess over the full-information reference. On SQuAD, approximately $.10$ documents/query are added to matched DLR proposals, out of $.46$ total excess; the rest comes from proposal learning. On NQ, all orders have nine startup triggers, but nineteen add no documents and one adds a single document across the entire collection (Table 10). A trigger can return an already-full proposal. Symmetric parameter selection on disjoint queries also retains smaller sets than Primal-ACI (Table 13).

6.3 Feedback Choice During Coverage Protection

Under the same miss budget, we compare recovered proposal outcomes, returned-set hits and skipped protection-round updates. At $\tau_1 = .5$, counterfactual updates save $.21$ documents/query relative to

Table 2: **Protection cost and feedback choice.** Panel (a) decomposes excess documents relative to the fixed-threshold reference. Panel (b) uses a separate batch with $\tau_1 = 0.5$ and identical step sizes. Costs are mean $\pm$ SE over twenty orders per panel. The NQ protection mean and SE are both below 10^{-5}. The runs column counts every-prefix coverage successes. Intervals are paired t intervals for counterfactual cost minus the listed variant.

(a) Protection cost, main-comparison orders

Dataset	Proposal excess	Protection	Total excess
SQuAD	0.362 ± 0.007	0.10 ± 0.02	0.46 ± 0.02
Natural Questions	0.391 ± 0.001	$< 10^{-5}$	0.391 ± 0.001

(b) Feedback variants, separate SQuAD orders

Variant	Documents/query	Runs	Paired 95% CI
Unprotected DLR	2.550 ± 0.007	11/20	$[0.019, 0.111]$
Actual-hit updates	2.62 ± 0.02	20/20	$[-0.0056, 0.0026]$
Skipped updates	2.825 ± 0.007	20/20	$[-0.258, -0.162]$
Counterfactual (ours)	2.61 ± 0.02	20/20	–

skipping, while actual-hit costs are close (Table 2(b)). This high-coverage start often gives both rules identical executed sets (Table 19). We therefore vary initialization to test Proposition 2 below the target quantile.

Figure 1(a,b) shows the sharp rise in actual-hit cost at low starts. At $\tau_1 = .2$ on SQuAD, actual-hit updates return about 46 documents/query versus 2.8 for counterfactual updates. The former cannot raise its threshold above the start: persistent proposal misses require approximately nine out of ten returns to use all 51 documents. Table 17 also records separation at some above-quantile starts. Across the grid, skipped updates can be cheaper or dearer than counterfactual updates (Table 20), which preserve the unprotected trajectory at every start.

Across 2 datasets $\times$ 8 starts $\times$ 20 orders, counterfactual updates reproduce the unprotected threshold sequence exactly and preserve prefix coverage. Table 21 reports the two checks separately; Table 23 details violations and triggers. Figure 3 illustrates the update process for one low-start order.

The common miss budget keeps all three protected variants feasible. Their cost differences arise through future proposals: updating after a protected hit can shrink the next proposal, whereas recovery records whether the original proposal actually missed. The initialization grid makes this feedback effect visible when low thresholds would otherwise require repeated full-pool returns. Skipping removes the protected-round update altogether, so its cost ordering changes with the start. These comparisons identify trajectory preservation as the mechanism supporting the convergence analysis. The finite-sample cost ranking additionally depends on initialization and step size.

6.4 Step Size, Sequence Length and Retrieval Conditions

Step size and horizon. Small steps retain large proposals despite infrequent protection, while larger steps increase added-document cost. The default coefficient has the lowest total cost in this five-order grid (Figure 2a,b). In the resampled streams, early triggers add no documents; narrowing proposals then brings additional protection by $T = 10{,}000$. Both excess and trigger frequency decrease across the later measured horizons, reaching about .13 excess documents/query at $T = 100{,}000$ (panels c,d). These finite-horizon patterns illustrate the cost mechanism. Sampling a finite empirical distribution does not establish the continuity assumptions of Theorem 1.

Distribution changes. SQuAD–NQ switches test whether prefix protection survives a changing query distribution. Carrying state preserves coverage over the complete sequence, while second-collection cost depends on direction (Table 25). Its early protection is frequent toward SQuAD and absent in the reverse recorded windows (Table 24). The boundary-informed restart resets the threshold, misses and step index and uses a fresh segment budget. The post-switch window in Table 24 diagnoses adaptation: its hit rate can fall below 90% while accumulated hits still satisfy the global prefix budget. Restarting therefore changes both the adaptation state and the origin of the coverage constraint.

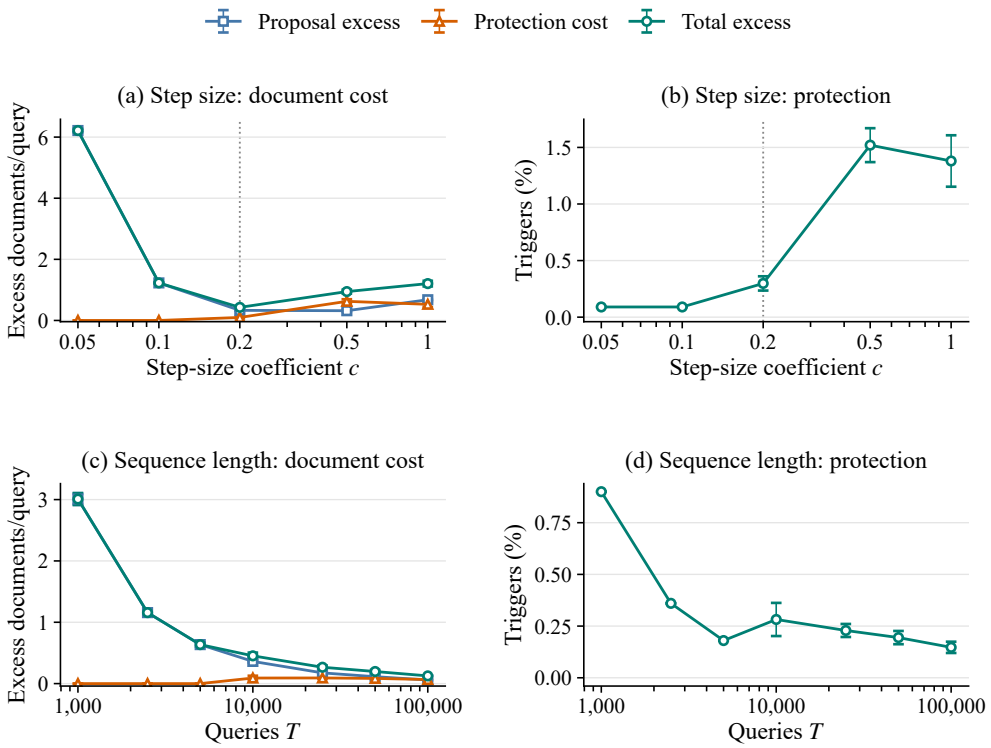

Figure 2: **Proposal learning and protection contribute different costs.** (a,b) Five paired SQuAD orders per coefficient, separate from main parameter selection. (c,d) Five streams resampled from the fixed empirical SQuAD distribution, with nested prefixes at seven horizons. Markers are means with SE; segments join measured settings. Trigger frequency is N_T/T, including the first nine returns that add no documents.

Larger pools and feasible sets. Pool-size controls hold correct-document scores fixed, hence preserving threshold and trigger sequences at each target. Added negatives increase document cost (Figure 4). Quora instead uses a fixed rank-based score map and requires larger feasible sets: the method returns about 38 documents/query against a fixed-threshold reference of 33, with coverage in every order (Table 12). This setting measures learning and protection cost when the reference sets are already large.

7 Conclusion

We introduce counterfactual updates for conformal retrieval with semi-bandit feedback. Recovering the original proposal's outcome preserves the underlying learner's threshold trajectory, while the miss budget guarantees target cumulative coverage after every query when each candidate pool contains the correct document. Under the stated i.i.d. and regularity conditions, budget-triggered returns occupy a vanishing fraction of queries and average document cost converges to the best feasible fixed-threshold cost. Our analysis also shows that actual-hit updates initialized below the target quantile require persistent full-pool protection. On SQuAD and Natural Questions, our method returns smaller sets than SPS and Primal-ACI while maintaining the coverage target at every prefix. Matched comparisons with the underlying learner quantify the small additional document cost of protection. Counterfactual feedback thus enables cumulative coverage protection while preserving the learning process that produces efficient proposals.

AI USE STATEMENT

In this work, we used generative AI tools for tasks with required disclosure. Generative AI tools helped develop the theoretical models and conceptual frameworks, helped formulate the mathematical claims, provided critical ingredients for proving those claims, and assisted in the writing of the proofs. The mathematical claims, the proofs, and every experimental result were checked and verified by the authors. We additionally used generative AI tools for tasks with recommended disclosure, namely to aid and polish the writing and to edit the paper for readability. We have reviewed all AI-assisted work, including the mathematical derivations and the text. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Anastasios N. Angelopoulos and Stephen Bates. Conformal Prediction: A Gentle Introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023. doi: 10.1561/2200000101.

Anastasios N. Angelopoulos, Emmanuel J. Candès, and Ryan J. Tibshirani. Conformal PID Control for Time Series Prediction. In *Advances in Neural Information Processing Systems*, volume 36, pp. 23047–23074, 2023. doi: 10.52202/075280-1000.

Anastasios N. Angelopoulos, Rina Foygel Barber, and Stephen Bates. Online conformal prediction with decaying step sizes. In *International Conference on Machine Learning*, pp. 1616–1630, 2024a. URL https://proceedings.mlr.press/v235/angelopoulos24a.html.

Anastasios N. Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal Risk Control. In *International Conference on Learning Representations*, 2024b. URL https://arxiv.org/abs/2208.02814.

Felipe Areces, Christopher Mohri, Tatsunori Hashimoto, and John Duchi. Online Conformal Prediction via Online Optimization. In *International Conference on Machine Learning*, 2025.

Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in Online Combinatorial Optimization. *Mathematics of Operations Research*, 39(1):31–45, 2014. doi: 10.1287/moor.2013.0598.

Osbert Bastani, Varun Gupta, Christopher Jung, Georgy Noarov, Ramya Ramalingam, and Aaron Roth. Practical Adversarial Multivalid Conformal Prediction. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://arxiv.org/abs/2206.01067.

Aadyot Bhatnagar, Huan Wang, Caiming Xiong, and Yu Bai. Improved Online Conformal Prediction via Strongly Adaptive Online Learning. In *International Conference on Machine Learning*, pp. 2337–2363, 2023. URL https://proceedings.mlr.press/v202/bhatnagar23a.html.

Nicolò Cesa-Bianchi, Gábor Lugosi, and Gilles Stoltz. Minimizing Regret With Label Efficient Prediction. *IEEE Transactions on Information Theory*, 51(6):2152–2162, 2005. doi: 10.1109/TIT.2005.847729.

Debashish Chakraborty, Eugene Yang, Daniel Khashabi, Dawn Lawrie, and Kevin Duh. Principled Context Engineering for RAG: Statistical Guarantees via Conformal Prediction. In *Advances in Information Retrieval*, pp. 537–546. 2026. doi: 10.1007/978-3-032-21300-6_45. URL https://doi.org/10.1007/978-3-032-21300-6_45.

Wei Chen, Yajun Wang, and Yang Yuan. Combinatorial Multi-Armed Bandit: General Framework and Applications. In *International Conference on Machine Learning*, pp. 151–159, 2013. URL https://proceedings.mlr.press/v28/chen13a.html.

Miroslav Dudik, John Langford, and Lihong Li. Doubly Robust Policy Evaluation and Learning. In *International Conference on Machine Learning*, 2011.

Bradley Efron and Robert Tibshirani. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, 1986.

Naihe Feng, Yi Sui, Shiyi Hou, Jesse C. Cresswell, and Ga Wu. Response Quality Assessment for Retrieval-Augmented Generation via Conditional Conformal Factuality. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2832–2836, 2025. doi: 10.1145/3726302.3730244.

Adam Fisch, Tal Schuster, Tommi Jaakkola, and Regina Barzilay. Efficient Conformal Prediction via Cascaded Inference with Expanded Admission. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2007.03114.

Haosen Ge, Hamsa Bastani, and Osbert Bastani. Stochastic Online Conformal Prediction with Semi-Bandit Feedback. In *International Conference on Machine Learning*, pp. 18954–18965, 2025. URL https://proceedings.mlr.press/v267/ge25a.html.

Isaac Gibbs and Emmanuel J. Candès. Adaptive Conformal Inference Under Distribution Shift. In *Advances in Neural Information Processing Systems*, volume 34, pp. 1660–1672, 2021. URL https://arxiv.org/abs/2106.00170.

Isaac Gibbs and Emmanuel J. Candès. Conformal Inference for Online Prediction with Arbitrary Distribution Shifts. *Journal of Machine Learning Research*, 25(162):1–36, 2024. URL https://jmlr.org/papers/v25/22-1218.html.

Sreenivas Gollapudi, Kostas Kollias, Kamesh Munagala, and Ali Sinop. Efficient Online Conformal Selection with Limited Feedback. *arXiv preprint arXiv:2605.14953*, 2026. URL https://arxiv.org/abs/2605.14953v4.

Kalervo Järvelin and Jaana Kekäläinen. Cumulated Gain-Based Evaluation of IR Techniques. *ACM Transactions on Information Systems*, 2002.

Mintong Kang, Nezihe Merve Gürel, Ning Yu, Dawn Song, and Bo Li. C-RAG: Certified Generation Risks for Retrieval-Augmented Language Models. In *International Conference on Machine Learning*, pp. 22963–23000, 2024. URL https://proceedings.mlr.press/v235/kang24a.html.

Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen tau Yih. Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6769–6781, 2020. doi: 10.18653/v1/2020.emnlp-main.550.

Junpei Komiyama, Junya Honda, and Hiroshi Nakagawa. Optimal Regret Analysis of Thompson Sampling in Stochastic Multi-Armed Bandit Problem with Multiple Plays. In *International Conference on Machine Learning*, 2015.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvari. Tight Regret Bounds for Stochastic Combinatorial Semi-Bandits. In *Artificial Intelligence and Statistics*, pp. 535–543, 2015. URL https://proceedings.mlr.press/v38/kveton15.html.

Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, Kristina Toutanova, Llion Jones, Matthew Kelcey, Ming-Wei Chang, Andrew M. Dai, Jakob Uszkoreit, Quoc Le, and Slav Petrov. Natural Questions: A Benchmark for Question Answering Research. *Transactions of the Association for Computational Linguistics*, 7:452–466, 2019. doi: 10.1162/tacl_a_00276.

Minjae Lee, Yoonjae Jung, and Sangdon Park. Online Conformal Abstention for Factuality Control Under Adversarial Bandit Feedback. *arXiv preprint arXiv:2506.14067*, 2025. URL https://arxiv.org/abs/2506.14067.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*, 113 (523):1094–1111, 2018. doi: 10.1080/01621459.2017.1307116.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

Shuo Li, Sangdon Park, Insup Lee, and Osbert Bastani. TRAQ: Trustworthy Retrieval Augmented Question Answering via Conformal Prediction. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 3799–3821, 2024. doi: 10.18653/v1/2024.naacl-long.210.

Zhen Lin, Shubhendu Trivedi, and Jimeng Sun. Conformal Prediction with Temporal Quantile Adjustments. In *Advances in Neural Information Processing Systems*, 2022.

Christopher Mohri and Tatsunori Hashimoto. Language Models with Conformal Factuality Guarantees. In *International Conference on Machine Learning*, pp. 36029–36047, 2024. URL https://proceedings.mlr.press/v235/mohri24a.html.

Victor Quach, Adam Fisch, Tal Schuster, Adam Yala, Jae Ho Sohn, Tommi S. Jaakkola, and Regina Barzilay. Conformal Language Modeling. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2306.10193.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. SQuAD: 100,000+ Questions for Machine Comprehension of Text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2383–2392, 2016. doi: 10.18653/v1/D16-1264.

Ramya Ramalingam, Shayan Kiyani, and Aaron Roth. The Relationship Between No-Regret Learning and Online Conformal Prediction. In *International Conference on Machine Learning*, 2025.

Pouria Rouzrokh, Shahriar Faghani, Cooper U. Gamble, Moein Shariatnia, and Bradley J. Erickson. CONFLARE: CONFormal LArge language model REtrieval. *arXiv preprint arXiv:2404.04287*, 2024. URL https://arxiv.org/abs/2404.04287.

Joar Skalse, Edoardo Pona, Osvaldo Simeone, and Nicola Paoletti. Online Conformal Prediction Beyond Feedback. *arXiv preprint arXiv:2608.07139*, 2026. URL https://arxiv.org/abs/2608.07139.

Vaidehi Srinivas. Online Conformal Prediction with Efficiency Guarantees. In *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 6696–6726, 2026. doi: 10.1137/1.9781611978971.240.

Matthew Streeter and Daniel Golovin. An Online Algorithm for Maximizing Submodular Functions. In *Advances in Neural Information Processing Systems*, 2008.

Adith Swaminathan and Thorsten Joachims. Counterfactual Risk Minimization: Learning from Logged Bandit Feedback. In *International Conference on Machine Learning*, 2015.

Dharmesh Tailor, Alvaro H. C. Correia, Eric Nalisnick, and Christos Louizos. Approximating Full Conformal Prediction for Neural Network Regression with Gauss-Newton Influence. In *International Conference on Learning Representations*, 2025.

Artin Tajdini, Lalit Jain, and Kevin Jamieson. Nearly Minimax Optimal Submodular Maximization with Bandit Feedback. In *Advances in Neural Information Processing Systems*, 2024.

Muhammad Faaiz Taufiq, Jean-Francois Ton, Rob Cornish, Yee Whye Teh, and Arnaud Doucet. Conformal Off-Policy Prediction in Contextual Bandits. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://arxiv.org/abs/2206.04405.

Nandan Thakur, Nils Reimers, Andreas Rücklé, Abhishek Srivastava, and Iryna Gurevych. BEIR: A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

Taira Tsuchiya, Shinji Ito, and Junya Honda. Best-of-Both-Worlds Algorithms for Partial Monitoring. In *Proceedings of The 34th International Conference on Algorithmic Learning Theory*, 2023.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005. doi: 10.1007/b106715.

Zhou Wang and Xingye Qiao. Efficient Online Set-valued Classification with Bandit Feedback. In *International Conference on Machine Learning*, pp. 51328–51347, 2024. URL https://proceedings.mlr.press/v235/wang24bg.html.

Yifan Wu, Roshan Shariff, Tor Lattimore, and Csaba Szepesvari. Conservative Bandits. In *International Conference on Machine Learning*, pp. 1254–1262, 2016. URL https://proceedings.mlr.press/v48/wu16.html.

Yunpeng Xu, Mufang Ying, Wenge Guo, and Zhi Wei. Two-stage Risk Control with Application to Ranked Retrieval. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 9104–9111, 2025. doi: 10.24963/ijcai.2025/1012. URL https://doi.org/10.24963/ijcai.2025/1012.

Junyoung Yang, Kyungmin Kim, and Sangdon Park. Online Conformal Prediction with Adversarial Semi-bandit Feedback via Regret Minimization. In *International Conference on Learning Representations*, volume 2026, pp. 47335–47377, 2026. URL https://arxiv.org/abs/2604.17984.

Margaux Zaffran, Olivier Féron, Yannig Goude, Julie Josse, and Aymeric Dieuleveut. Adaptive Conformal Predictions for Time Series. In *International Conference on Machine Learning*, pp. 25834–25866, 2022. URL https://proceedings.mlr.press/v162/zaffran22a.html.

Zixi Zhu, Jiayuan Su, Jian Zhang, Yu Lin, and Hongwei Wang. CAS: Conformalized Agentic Search via Adaptive Retrieval and Policy Weighting. *arXiv preprint arXiv:2608.20771*, 2026. URL https://arxiv.org/abs/2608.20771.

Julian Zimmert, Haipeng Luo, and Chen-Yu Wei. Beating Stochastic and Adversarial Semi-bandits Optimally and Simultaneously. In *International Conference on Machine Learning*, pp. 7683–7692, 2019. URL https://proceedings.mlr.press/v97/zimmert19a.html.

A PROOFS OF COVERAGE AND DOCUMENT-COST GUARANTEES

We begin with cumulative coverage and threshold-trajectory equality because they separate the returned-set guarantee from the learning dynamics on each input sequence. Neither property alone controls the cost of protection. We next study the frequency of budget-triggered full-pool returns because even an unchanged threshold trajectory can incur a persistent cost if coverage protection repeatedly returns every candidate document while the learner approaches the target coverage. Finally, the asymptotic document-cost result combines that frequency bound with stochastic threshold convergence so that the fixed-threshold benchmark applies to actual returns as well as the learner's proposals.

A.1 PROOF OF PROPOSITION 1

Induction preserves the miss budget because a full-pool return guarantees a hit and a returned proposal can add at most one miss.

Proposition 1 makes cumulative coverage a property of the returned-set decision for every proposed threshold trajectory.

Proof. If $H_t = 1$, returning $\mathcal{Y}$ gives a hit and leaves the miss count unchanged. If $H_t = 0$, the decision rule ensures $M_{t-1} + 1 \leq \lfloor \alpha t \rfloor$ before feedback arrives. Hence either possible outcome preserves the bound by induction. The identity $\widehat{\mathrm{Cov}}_t = 1 - M_t/t$ converts the miss bound into the coverage statement. □

A.2 PROOF OF OBSERVATION 1

The argument couples the learners on the same input sequence and recovers the same proposal hit in both return branches.

Observation 1 transfers properties of the unprotected learner to the protected method's threshold sequence.

Proof. When the thresholds agree, the proposed sets agree. When $H_t = 0$, both learners use the observed proposal hit indicator. When $H_t = 1$, the revealed Y_t determines the hit indicator of the unprotected learner's proposal. The same hit and step size give the same next threshold, completing the induction. □

A.3 PROOF OF LEMMA 1

For this proof only, represent the conditional Bernoulli miss as $X_t = \mathbf{1}\{U_t \leq p_t\}$, where p_t is measurable with respect to the enlarged history $\mathcal{F}_{t-1}$. The variables U_t are independent uniform draws on $(0, 1)$, with U_t independent of $\mathcal{F}_{t-1}$.

The uniform variables express the stochastic assumption and are used only for analysis. The argument uses monotone comparison with independent misses of probability slightly above the target, followed by concentration and budget balance. The conditional Bernoulli representation can be realized with auxiliary uniforms as detailed in the stochastic convergence proof.

To derive the frequency bound, let $Q_t = L_t - M_t$ be the remaining miss budget after query t. The budget increases by $a_t = L_t - L_{t-1} \in \{0, 1\}$ before query t, with $L_0 = Q_0 = 0$. The miss-budget and returned-set rules yield the recursion

$$H_t = \mathbf{1}\{Q_{t-1} + a_t = 0\}, \qquad Q_t = \max\{0, Q_{t-1} + a_t - X_t\}, \tag{7}$$

where a_t supplies new miss allowance and X_t consumes it from the previous balance Q_{t-1}. The maximum at zero expresses that coverage protection prevents a miss when no allowance remains. Thus, H_t records a decision made before the current proposal outcome.

The trigger count N_T records budget-triggered full-pool returns, including triggers when the proposal already equals the full pool.

Proof. Fix $\epsilon \in (0, 1-\alpha)$ and put $q = \alpha + \epsilon$. We first analyze a comparison process started at a deterministic time n with $Q_n^+ = 0$. For $t > n$, define

$$X_t^+ = \mathbf{1}\{U_t \le q\}, \qquad Q_t^+ = \max\{0, Q_{t-1}^+ + a_t - X_t^+\}.$$

Let $H_t^+ = \mathbf{1}\{Q_{t-1}^+ + a_t = 0\}$ and let $R_t^+ = H_t^+ X_t^+$ count the prevented comparison misses. Here R_t^+ is an auxiliary scalar, distinct from the returned document set R_t. The comparison misses are independent Bernoulli variables of mean q, and we show that $Q_T^+/T \to 0$ almost surely by expressing the reflected recursion as a maximum over the suffixes of these independent misses. The reflected recursion has the suffix representation

$$Q_T^+ = \max_{n \le r \le T} \sum_{t=r+1}^{T} (a_t - X_t^+),$$

where the empty suffix at $r = T$ has sum zero. For any suffix of length m, the deterministic arrival count obeys

$$\sum_{t=T-m+1}^{T} a_t = \lfloor \alpha T \rfloor - \lfloor \alpha (T-m) \rfloor \le \alpha m + 1.$$

If $Q_T^+ \ge \delta T$ for a fixed $\delta > 0$, some suffix consequently satisfies

$$\sum_{t=T-m+1}^{T} X_t^+ \le \alpha m + 1 - \delta T.$$

Its mean is qm, so the deviation below its mean is at least $\delta T - 1$. For all sufficiently large T, Hoeffding's inequality and a union bound over at most T suffixes yield

$$\mathbb{P}(Q_T^+ \ge \delta T) \le T \exp\left(-\frac{2(\delta T - 1)^2}{T}\right).$$

The right side is summable in T, so Borel–Cantelli gives $Q_T^+/T \to 0$ almost surely after applying the bound to a countable sequence of positive δ decreasing to zero.

The budget balance is

$$Q_T^+ = \sum_{t=n+1}^{T} a_t - \sum_{t=n+1}^{T} X_t^+ + \sum_{t=n+1}^{T} R_t^+.$$

Divide by T. The arrival average converges to α and the demand average converges to q by the strong law. The sublinear remaining budget therefore implies

$$\frac{1}{T} \sum_{t=n+1}^{T} R_t^+ \longrightarrow q - \alpha = \epsilon.$$

Each H_t^+ depends only on previous uniforms and the deterministic budget increment, which makes $H_t^+(X_t^+ - q)$ a bounded martingale difference whose average converges to zero almost surely. Since $R_t^+ = H_t^+ X_t^+$, we obtain

$$\frac{1}{T} \sum_{t=n+1}^{T} H_t^+ \longrightarrow \frac{\epsilon}{\alpha + \epsilon} \quad \text{almost surely.}$$

This conclusion holds simultaneously for every deterministic integer start n on an event of probability one, so on each path where $p_t \to \alpha$ we may choose a finite integer n after which $p_t \le q$. The comparison conclusion already holds for that integer. This avoids treating the eventual convergence time as a stopping time. For every $t > n$ on this path, the shared uniforms give $X_t \le X_t^+$. The recursion is increasing in its initial budget and decreasing in its miss input, so induction from $Q_n^+ = 0 \le Q_n$ gives $Q_t^+ \le Q_t$ throughout the sequence that starts at the selected integer n. If $H_t = 1$, nonnegativity then forces $Q_{t-1}^+ + a_t = 0$. Hence $H_t \le H_t^+$ for $t > n$ and

$$\limsup_{T \to \infty} \frac{N_T}{T} \le \frac{\epsilon}{\alpha + \epsilon}.$$

Apply this argument to a countable sequence of positive ϵ decreasing to zero. We conclude $N_T/T \to 0$ almost surely. □

Expected trigger count at constant miss probability. Under the stronger condition $p_t = \alpha$ at every query, the expected trigger count satisfies $\mathbb{E}[N_T] = O(\sqrt{T})$. To obtain this rate, write $S_t = \sum_{i=1}^{t}(X_i - \alpha)$. The suffix representation of the remaining budget gives $Q_T \le 1 + \max_{0 \le r \le T} S_r - S_T$. The centered sums form a martingale with $\mathbb{E}[S_T^2] = \alpha(1-\alpha)T$, so the martingale maximal inequality bounds the expected maximum absolute sum by $O(\sqrt{T})$ and hence gives $\mathbb{E}[Q_T] = O(\sqrt{T})$. Finally, budget balance and the fact that each decision precedes the current miss imply $\alpha\mathbb{E}[N_T] = \mathbb{E}[Q_T] - L_T + \alpha T$. Since $0 \le \alpha T - L_T < 1$, the stated expected trigger bound follows.

A.4 Proof of Theorem 1

The argument combines trajectory equality with stochastic quantile convergence and the martingale strong law.

Condition (A1) lets past feedback determine a threshold for a fresh query and condition (A2) makes that threshold's coverage identify a unique target, while condition (A3) connects threshold convergence to average document cost even though individual set sizes change in discrete steps.

Proof. Observation 1 identifies the protected threshold sequence with the unprotected decreasing-step-size learner. We use the fixed-score quantile convergence result of Angelopoulos et al. (2024a) under the stated independent-query and unique-quantile conditions. The step sizes obey $\sum_t \eta_t = \infty$ and $\sum_t \eta_t^2 < \infty$.

To check the endpoint convention used here, let $h(\tau) = \mathbb{P}\{Y \in C(Z, \tau)\}$ be the expected hit at threshold τ. The boundary rule gives $h(\tau) = 0$ for $\tau \le 0$ and $h(\tau) = 1$ for $\tau \ge 1$. For $0 < \tau < 1$, it gives $h(\tau) = F(\tau)$. Continuity at the target quantile and $0 < \phi < 1$ imply $0 < \tau^\star < 1$. Therefore the expected update $\eta_t(\phi - h(\tau_t))$ points toward the same unique $\tau^\star$ on both sides, including at either boundary. The update noise is bounded. These are the drift and step-size properties in the underlying stochastic-approximation argument. In particular, its squared-distance calculation is

$$\mathbb{E}[(\tau_{t+1} - \tau^\star)^2 \mid \mathcal{F}_{t-1}] \le (\tau_t - \tau^\star)^2 - 2\eta_t(\tau_t - \tau^\star)(h(\tau_t) - \phi) + \eta_t^2.$$

The drift term is nonnegative and is bounded away from zero when the threshold stays a fixed distance from $\tau^\star$. The thresholds are bounded because the boundary actions point inward and $\eta_t \le c$, so the standard nonnegative-supermartingale argument gives convergence of the squared distance and summability of the weighted drift. The divergent sum of step sizes then forces the distance to converge to zero. Thus the same quantile-convergence conclusion holds with the explicit boundary convention in the manuscript:

$$\tau_t \longrightarrow \tau^\star \quad \text{almost surely.}$$

The threshold eventually lies in $(0, 1)$ on each convergent path. From then on, independence of the next query gives

$$p_t = \mathbb{P}(X_t = 1 \mid \mathcal{F}_{t-1}) = 1 - F(\tau_t) \longrightarrow 1 - \phi = \alpha.$$

Auxiliary representation for the proof. The algorithm samples no additional random variables. To apply Lemma 1, we represent conditional Bernoulli misses on an enlarged probability space without changing the learner's distribution. For completeness, take independent auxiliary uniforms V_t and set

$$U_t = \begin{cases} p_t V_t, & X_t = 1, \\ p_t + (1 - p_t)V_t, & X_t = 0. \end{cases}$$

Conditionally on the past, this variable is uniform on $(0, 1)$ and $X_t = \mathbf{1}\{U_t \le p_t\}$ almost surely, which makes successive U_t independent uniforms with respect to the enlarged past filtration. These variables only represent the analysis and are not sampled by the algorithm. Lemma 1 therefore gives $N_T/T \to 0$ almost surely.

Proposal cost and its conditional expectation. Define $D_t = |P_t| - g(\tau_t)$. The threshold is determined by the past. Under (A1), the fresh query is independent of that past, so $\mathbb{E}[|P_t| \mid \mathcal{F}_{t-1}] = g(\tau_t)$ and hence $\mathbb{E}[D_t \mid \mathcal{F}_{t-1}] = 0$. Because both costs lie in $[0, K]$, $|D_t| \le K$. The martingale strong law yields

$$\frac{1}{T}\sum_{t=1}^{T} D_t \longrightarrow 0 \quad \text{almost surely.}$$

Limit of the proposal average. Condition (A3) gives $g(\tau_t) \to g(\tau^\star)$. Taking Cesaro averages, we obtain

$$\frac{1}{T}\sum_{t=1}^{T}|P_t| = \frac{1}{T}\sum_{t=1}^{T} g(\tau_t) + \frac{1}{T}\sum_{t=1}^{T} D_t \longrightarrow g(\tau^\star).$$

Additional cost of protection. Since $P_t \subseteq R_t$, the per-query increase is exactly $H_t(K - |P_t|)$. The trigger-frequency lemma gives

$$0 \le \frac{1}{T}\sum_{t=1}^{T}(|R_t| - |P_t|) \le K\frac{N_T}{T} \longrightarrow 0.$$

This proves $\widehat{J}_T \to g(\tau^\star)$ almost surely.

Optimality within the fixed-threshold family. Finally, the unique-quantile condition makes every interior $\tau < \tau^\star$ have population coverage below ϕ. Thresholds at or below zero have zero coverage. The threshold $\tau^\star$ itself has coverage exactly ϕ. Nestedness makes g nondecreasing, so every feasible threshold has cost at least $g(\tau^\star)$. It follows that $J^\star = g(\tau^\star)$, completing the proof. □

Cost decomposition. The additional average document cost obeys

$$0 \le \widehat{J}_T - \frac{1}{T}\sum_{t=1}^{T}|P_t| = \frac{1}{T}\sum_{t=1}^{T} H_t(K - |P_t|) \le K\frac{N_T}{T} \longrightarrow 0. \tag{8}$$

The upper bound charges at most K extra documents for each budget trigger.

Equation (8) separates the cost learned through τ_t from the extra documents controlled by N_T/T. The scoring function and query distribution determine g and $\tau^\star$. The step sizes control threshold adjustment, while counterfactual updates preserve this learning during budget-triggered returns. The resulting guarantees give cumulative coverage and actual document cost as the criteria for the comparisons in Section 6.

A.5 Proof of Proposition 2

This subsection supports the feedback comparison at the end of Section 5. The first claim holds on every input sequence because Proposition 1 keeps the cumulative actual hits at or above the target after every query, and summation by parts turns that record into a bound on the threshold. The second claim holds because a threshold capped at $\tilde{\tau}_1$ leaves every proposal with miss probability at least p_{init}, while the budget admits at most αT misses in total. Within this subsection, A_t, H_t, M_t are the actual-hit variant's states, with trigger count N_T^{act} and cost $\widehat{J}_T^{\text{act}}$. Write $\tilde{P}_t = C_t(\tilde{\tau}_t)$ for the proposals of this variant and $\tilde{X}_t = \mathbf{1}\{Y_t \notin \tilde{P}_t\}$ for their miss indicators.

Proof. Put $D_t = \sum_{s=1}^{t}(A_s - \phi)$ with $D_0 = 0$. Proposition 1 applies to any sequence of proposed sets, so $M_t \le \lfloor \alpha t \rfloor \le \alpha t$ and therefore $\sum_{s \le t} A_s = t - M_t \ge \phi t$, which gives $D_t \ge 0$ for every $t \ge 1$. Using $A_t - \phi = D_t - D_{t-1}$ and summing by parts,

$$\sum_{t=1}^{T}\eta_t(A_t - \phi) = \eta_T D_T + \sum_{t=1}^{T-1}(\eta_t - \eta_{t+1})D_t.$$

Every term on the right is nonnegative because the step sizes are nonincreasing. The update gives $\tilde{\tau}_{T+1} - \tilde{\tau}_1 = -\sum_{t=1}^{T}\eta_t(A_t - \phi)$, so $\tilde{\tau}_{T+1} \le \tilde{\tau}_1$ for every $T \ge 1$, which proves the first claim.

Assume now (A1) and (A2) with $0 < \tilde{\tau}_1 < \tau^\star$, and let $h(\tau) = \mathbb{P}\{Y \in C(Z, \tau)\}$ be the nondecreasing hit function used above, which equals $F(\tau)$ for $0 < \tau < 1$. The threshold $\tilde{\tau}_t$ is determined by $\mathcal{F}_{t-1}$ while (Z_t, Y_t) is independent of it, so $\mathbb{P}(\tilde{X}_t = 1 \mid \mathcal{F}_{t-1}) = 1 - h(\tilde{\tau}_t)$. The first claim and monotonicity of h give $1 - h(\tilde{\tau}_t) \ge 1 - h(\tilde{\tau}_1) = 1 - F(\tilde{\tau}_1) = p_{\text{init}}$, and condition (A2) with $\tilde{\tau}_1 < \tau^\star$ gives $F(\tilde{\tau}_1) < \phi$ and hence $p_{\text{init}} > \alpha$. A query with $H_t = 0$ returns the proposal and contributes the miss $\tilde{X}_t$, whereas a query with $H_t = 1$ returns the full pool and contributes no miss,

so $M_T = \sum_{t=1}^{T}(1 - H_t)\tilde{X}_t$. The indicator H_t is computed from M_{t-1} and L_t and is therefore measurable with respect to $\mathcal{F}_{t-1}$, which makes

$$\xi_t = (1 - H_t)\big(\tilde{X}_t - \mathbb{P}(\tilde{X}_t = 1 \mid \mathcal{F}_{t-1})\big)$$

a martingale difference sequence bounded by one. The martingale strong law gives $T^{-1}\sum_{t\le T}\xi_t \to 0$ almost surely, and the lower bound on the conditional miss probability gives

$$M_T \ge p_{\text{init}}\left(T - N_T^{\text{act}}\right) + \sum_{t=1}^{T}\xi_t.$$

Proposition 1 bounds the left side by αT. Dividing by T and passing to the limit, we obtain $p_{\text{init}}(1 - \liminf_T N_T^{\text{act}}/T) \le \alpha$, which rearranges to the first bound in Equation (6). Each protected query returns K documents, so $\widehat{J}_T^{\text{act}} \ge K N_T^{\text{act}}/T$ and the second bound follows. □

The initial threshold therefore separates two regimes for the actual-hit variant, since $p_{\text{init}} > \alpha$ holds exactly when $\tilde{\tau}_1$ lies below the target quantile.

A population example. Let the correct document carry a nonconformity score uniform on $[0.7, 0.9]$ and let the remaining 50 candidates carry the score 0.95. Then $F(u) = \min\{\max\{(u - 0.7)/0.2, 0\}, 1\}$ and $g(\tau) = F(\tau) + 50 \cdot \mathbf{1}\{\tau \ge 0.95\}$, so conditions (A1)–(A3) hold with $\tau^\star = 0.88$ and $J^\star = 0.9$ at $\phi = 0.9$. Starting the actual-hit variant at $\tilde{\tau}_1 = 0.5$ gives $p_{\text{init}} = 1$, and Equation (6) then forces a cost of at least $K(1 - \alpha) = 45.9$ documents. Table 3 reports twenty independent runs at three horizons. The actual-hit variant attains that bound with equality at every horizon, whereas counterfactual updates fall toward $J^\star$ as the horizon grows.

Table 3: **Returned documents in the population example.** Mean ± SE over $n = 20$ seeds with $\tilde{\tau}_1 = 0.5$, $\alpha = 0.1$ and $K = 51$. The actual-hit cost is 45.9 documents per query in every run.

T	Counterfactual	Actual-hit
1,000	4.4 ± 0.3	45.9
10,000	1.65 ± 0.02	45.9
100,000	1.044 ± 0.006	45.9

A.6 Step-size conditions and threshold-update relations

The per-query procedure is given in Section 4.3. The exponent range in Equation (2) gives the step-size conditions used for stochastic convergence:

$$\sum_{t=1}^{\infty}\eta_t = \infty, \qquad \sum_{t=1}^{\infty}\eta_t^2 < \infty, \tag{9}$$

where the first sum accumulates step sizes and the second accumulates their squares. These conditions allow continued adjustment with decreasing update noise. Under the assumptions of Theorem 1, the threshold approaches the value attaining the target population coverage.

The scalar Primal-ACI rule of Gollapudi et al. (2026) likewise adjusts an unprojected threshold from success feedback with explicit boundary actions. When every proposed set is returned, $B_t = A_t$ and our update is the underlying decreasing-step-size rule. Using a constant step size in that setting gives the scalar Primal-ACI update with the same initialization and boundary convention.

B Experimental details

We organize the additional results by the comparison they support. Tables report means ± standard errors (SE) across query orders unless their captions specify otherwise. Coverage and trigger frequencies use percentages. Means are rounded to the same decimal place as their SE, and coverage success counts use the unrounded values. For repeated-run means, omitted error terms indicate zero observed SE at the reported precision. Fixed-threshold reference costs are deterministic for a fixed query collection.

B.1 Experimental batches and comparison roles

Table 4 separates the query-order batches. In particular, the abstract's initial cost, the main comparison and the feedback ablation are estimates over different permutations of the same fixed queries. Their means and success counts need not coincide. Table 5 specifies the update information used by each compared construction. We evaluate every method on realized prefix coverage and returned-document cost, while retaining SPS's own threshold and CDF diagnostics below.

Table 4: **Experimental batches and their roles.** Main, expanded feedback and initialization orders are distinct. Within a study, methods share orders; initialization also shares orders across starting thresholds. Long streams resample the fixed SQuAD collection. Repeats quantify order or resampling variation, rather than new query collections.

Study	Queries / repeats	Result
Initial comparison	SQuAD, 10 orders	Table 7
Main comparison	SQuAD/NQ, 20 per dataset	Table 1
Initial feedback	SQuAD, 5 orders	Table 15
Expanded feedback	SQuAD, 20 new orders	Table 2(b)
Initialization	2 datasets, 8 starts, 20 orders	Table 17
Coefficient grid	SQuAD, 5 paired orders	Table 9
Disjoint tuning	2,000 dev. / 8,000 eval., 20 each	Table 13
Distribution switch	Both directions, 5 paired orders	Table 25
Longer sequences	5 streams, 7 nested horizons	Table 27
Pool/target grid	3 targets, 4 pools, 20 shared orders	Table 26
Quora	10,000 queries, 20 orders	Table 12

Table 5: **Comparison interfaces.** Online methods use the same scores and semi-bandit interface. Prefix success is our common evaluation metric, distinct from each baseline's original guarantee or regret criterion. Only the last row uses full-information threshold selection.

Method	Update information and state	Role in this comparison
SPS	Censored correct-score history; conservative threshold	Threshold/CDF and document cost
OCP-Unlock+	Revealed outcome; weights on 200 threshold arms	Grid-based partial-feedback control
Primal-ACI	Returned hit; constant-step scalar threshold	Constant-step selection control
DLR	Proposal hit; decreasing-step threshold	Matched unprotected learner
Counterfactual	Recovered proposal hit; threshold and miss count	Prefix-protected proposal learning
Actual-hit control	Returned hit; same threshold schedule and budget	Effect of protection on feedback
Skipped control	No threshold update on protected rounds	Cost of omitting recovered feedback
Fixed-threshold reference	All correct scores in the complete collection	Post-hoc efficiency reference

SPS uses $\delta = 2/T^2$, Primal-ACI uses step size $T^{-1/2}$, and OCP-Unlock+ has 200 threshold choices. Unprotected DLR uses $ct^{-0.6}$ with $c = 1$ or 0.2; its matched feedback control uses $c = 0.2$.

B.2 SPS settings and evaluation criteria

We use the SPS confidence parameter $\delta = 2/T^2$ from its original retrieval setting. Threshold shortfalls count rounds whose executed threshold is more restrictive than the empirical feasible fixed threshold. This criterion measures threshold feasibility against the reference distribution and complements realized cumulative coverage. The cumulative CDF loss sums the piecewise coverage loss of Ge et al. (2025), with weights 0.1 for overcoverage and 10 for undercoverage. We use the empirical right-continuous CDF and measure loss relative to ideal zero, including the reference threshold's deviation at a discrete score boundary.

B.3 Complete SQuAD comparison and protection cost

Table 6 supplements the main cost comparison with final coverage, minimum prefix coverage and SPS's threshold criteria. SPS has zero threshold shortfalls and the lowest CDF loss among the online methods, consistent with its conservative construction. Counterfactual protection maintains realized prefix coverage while allowing the threshold to track the target, producing smaller returned sets.

Table 6: **Complete comparison on SQuAD.** Mean $\pm$ SE over twenty permutations. Final and minimum prefix coverage are percentages. Threshold shortfalls count rounds with a threshold below the empirical feasible reference. CDF loss sums the piecewise coverage loss over the run.

Method	Documents/query	Final (%)	Min. prefix (%)
SPS	9.48 ± 0.06	95.06 ± 0.05	95.04 ± 0.05
OCP-Unlock+	39.59 ± 0.05	91.26 ± 0.07	56 ± 7
Primal-ACI	3.81 ± 0.01	90.254 ± 0.002	90.233 ± 0.001
DLR ($c = 1$)	2.86 ± 0.01	90.009 ± 0.004	84.4 ± 0.5
DLR ($c = 0.2$)	2.550 ± 0.007	90.088 ± 0.008	89.995 ± 0.006
Counterfactual (ours)	2.65 ± 0.02	90.098 ± 0.009	90.010 ± 0.003
Fixed-threshold reference	2.189	90	82 ± 2
Method	**Excess documents**	**Threshold shortfalls**	**CDF loss**
SPS	7.29 ± 0.06	0	51.0 ± 0.3
OCP-Unlock+	37.40 ± 0.05	1266 ± 8	7240 ± 50
Primal-ACI	1.62 ± 0.01	4190 ± 40	1740 ± 30
DLR ($c = 1$)	0.67 ± 0.01	4500 ± 50	1620 ± 30
DLR ($c = 0.2$)	0.362 ± 0.007	4600 ± 90	710 ± 20
Counterfactual (ours)	0.46 ± 0.02	4600 ± 90	710 ± 20
Fixed-threshold reference	0	0	10

Table 7 retains the initial ten-order comparison reported in the abstract. Its paired protection cost is about 0.09 documents per query (Table 8). The paired bootstrap 95% interval (Efron & Tibshirani, 1986) for the reduction relative to SPS is $[71.0\%, 72.8\%]$. Changing the DLR coefficient from 1 to 0.2 reduces the cost of the proposals before protection. The budget triggers on 0.276% of queries on average in this batch. Its document-cost bound KN_T/T exceeds measured protection cost because a full-pool return adds only documents absent from the proposal.

Table 7: **Initial SQuAD comparison reported in the abstract.** Mean $\pm$ SE over the original ten orders, distinct from the main twenty-order comparison.

Method	Documents/query
DLR c=0.2	2.55 ± 0.01
Counterfactual (ours)	2.64 ± 0.03
DLR ($c = 1$)	2.84 ± 0.02
SPS	9.44 ± 0.07

Table 8: **Paired document-cost differences.** The first two rows compare protected counterfactual updates with their matched unprotected learner. The other rows use the twenty-order feedback batch. Entries are mean $\pm$ SE and two-sided 95% paired t intervals.

Comparison / batch	Paired difference	95% CI
SQuAD, initial ten orders	0.088 ± 0.026	$[0.029, 0.146]$
SQuAD, main twenty orders	0.097 ± 0.019	$[0.057, 0.137]$
Feedback: ours $-$ Unprotected DLR	0.065 ± 0.022	$[0.019, 0.111]$
Feedback: ours $-$ Actual-hit updates	-0.0015 ± 0.0019	$[-0.0056, 0.0026]$
Feedback: ours $-$ Skipped updates	-0.210 ± 0.023	$[-0.258, -0.162]$

B.4 Complete step-size comparison

We vary $c \in \{0.05, 0.1, 0.2, 0.5, 1\}$ over five shared SQuAD permutations, keeping $\beta = 0.6$ and $\tau_1 = 0.5$. Table 9 separates proposal excess, protection cost and trigger frequency for every coefficient.

The two smallest coefficients trigger only the nine initial full-pool returns, which add no documents to their proposals. Their different total costs therefore come from proposal learning. At larger coefficients, protection contributes more to the cost. Figure 2(a,b) displays both effects across the full grid.

Table 9: **Complete coefficient comparison on SQuAD.** Mean $\pm$ SE across five paired query orders, with $\tau_1 = 0.5$ and $\beta = 0.6$. All configurations meet the target at every prefix. Costs are documents per query.

c	Proposal excess	Protection	Total excess	Triggers (%)
0.05	6.22 ± 0.06	0	6.22 ± 0.06	0.09
0.1	1.23 ± 0.02	0	1.23 ± 0.02	0.09
0.2	0.330 ± 0.007	0.10 ± 0.03	0.43 ± 0.03	0.30 ± 0.06
0.5	0.32 ± 0.02	0.63 ± 0.07	0.94 ± 0.07	1.5 ± 0.1
1	0.68 ± 0.01	0.5 ± 0.1	1.21 ± 0.09	1.4 ± 0.2

B.5 Complete Natural Questions comparison

Table 11 uses the same encoder, candidate construction and method configurations as the SQuAD comparison. The encoder ranks the designated positive document first on 99.2% of these queries. Threshold selection can return empty sets, so the feasible fixed-threshold reference averages fewer than one document per query. Counterfactual protection and matched unprotected DLR have nearly identical costs in this comparison. Nineteen orders require no added documents, and the remaining order adds one document across its $6{,}515$ queries. All twenty protected runs meet the target after every query. Table 10 distinguishes these startup triggers from actual added documents.

Table 10: **Budget triggers and added documents, main batches.** Trigger ranges include the nine startup returns. Added-document totals sum $H_t(K - |P_t|)$ over all twenty orders, so a trigger with an already-full proposal adds zero. The totals cover $20 \times 10{,}000$ SQuAD or $20 \times 6{,}515$ NQ queries. Per-query means and SE appear in Table 2(a).

Dataset	Triggers/order	Zero-added orders	Added documents, all orders
SQuAD	9–65	5/20	19337
Natural Questions	9–9	19/20	1

Table 11: **Complete comparison on Natural Questions.** Mean $\pm$ SE over twenty permutations. Final and minimum prefix coverage are percentages. Threshold shortfalls count rounds with a threshold below the empirical feasible reference. CDF loss sums the piecewise coverage loss over the run.

Method	Documents/query	Final (%)	Min. prefix (%)
SPS	7.79 ± 0.01	95.97 ± 0.05	95.96 ± 0.06
OCP-Unlock+	35.63 ± 0.07	91.73 ± 0.07	59 ± 7
Primal-ACI	2.151 ± 0.001	90.387 ± 0.003	90.362 ± 0.002
DLR ($c = 1$)	0.993 ± 0.004	90.026 ± 0.004	87.2 ± 0.3
DLR ($c = 0.2$)	1.296 ± 0.001	90.21 ± 0.01	90.134 ± 0.006
Counterfactual (ours)	1.296 ± 0.001	90.21 ± 0.01	90.134 ± 0.006
Fixed-threshold reference	0.905	90.008	75 ± 5
Method	**Excess documents**	**Threshold shortfalls**	**CDF loss**
SPS	6.89 ± 0.01	0	39.0 ± 0.2
OCP-Unlock+	34.73 ± 0.07	693 ± 5	4520 ± 40
Primal-ACI	1.246 ± 0.001	2580 ± 20	1310 ± 20
DLR ($c = 1$)	0.088 ± 0.004	2790 ± 30	1230 ± 20
DLR ($c = 0.2$)	0.391 ± 0.001	3030 ± 50	490 ± 10
Counterfactual (ours)	0.391 ± 0.001	3030 ± 50	490 ± 10
Fixed-threshold reference	0	0	5

B.6 QUORA CANDIDATE POOL

We use the Quora duplicate-question collection from BEIR (Thakur et al., 2021) to examine less separable ranked candidates. For each of its 10,000 test queries, the pool includes one relevant question selected in lexicographic document-ID order. The other candidates are the first 50 non-relevant questions in the saved BM25 ranking. We resolve tied BM25 scores by their original rank position. For zero-based within-pool rank r, nonconformity is $0.6[1 - \sqrt{(K-1-r)/(K-1)}]$, with $K = 51$. This fixed map preserves candidate order and defines the numerical scale used by the threshold updates. Its scale and the unchanged step-size coefficient jointly specify the experiment. Methods use their main-setting configurations over twenty shared query permutations.

Table 12 reports document cost and prefix coverage for every method. The feasible fixed threshold returns 33 documents, reflecting the depth required under these scores. Counterfactual protection saves documents relative to SPS's full-pool returns while meeting the target in every run. Primal-ACI uses smaller mean sets and meets the prefix target in three of the twenty orders.

Table 12: **Rank-based retrieval on Quora.** Mean $\pm$ SE over twenty shared query permutations, with $K = 51$ and target 90%. The runs column counts permutations satisfying the target at every prefix. The fixed-threshold reference costs 33 documents per query.

Method	Documents/query	Min. prefix (%)	Runs
SPS	51	100	20/20
OCP-Unlock+	37.47 ± 0.03	78 ± 6	9/20
Primal-ACI	34.4 ± 0.1	71 ± 7	3/20
DLR ($c = 1$)	32.5 ± 0.2	69 ± 7	0/20
DLR ($c = 0.2$)	37.9 ± 0.6	71 ± 7	2/20
Counterfactual (ours)	38.0 ± 0.6	90.33 ± 0.07	20/20
Fixed-threshold reference	33	70 ± 7	0/20

B.7 PARAMETER SELECTION AND COMPARISON IMPLEMENTATIONS

For the main configuration, we select $c = 0.2$ by mean set size on a separate permutation of the same SQuAD collection. The initial ten-order comparison, main twenty-order comparison, feedback ablation and initialization grid use distinct batches. Methods share query orders within each batch. Natural Questions receives the SQuAD choices unchanged.

Our SPS implementation follows Algorithm 1 of Ge et al. (2025) on fixed candidate scores. OCP-Unlock+ uses 200 threshold arms with its uniform exploration component. This exploration assigns probability to thresholds returning large sets, contributing to its document cost in the evaluated configuration.

For symmetric parameter selection, we allocate the first 2,000 SQuAD queries to development and the remaining 8,000 to evaluation. Each split uses twenty permutations. Each method selects from $c \in \{0.05, 0.1, 0.2, 0.5, 1\}$ by minimizing mean development cost among configurations with mean final coverage of at least 0.9. Ties select the smaller coefficient. The DLR schedules are $ct^{-0.6}$, and Primal-ACI uses $cT^{-1/2}$. Evaluation starts afresh from $\tau_1 = 0.5$. Table 13 reports the selected coefficients, held-out costs and every-prefix coverage counts. The cost advantage over Primal-ACI persists when methods select their coefficients through this common procedure.

Table 13: **Evaluation after symmetric parameter selection.** Each method selects c on 2,000 development queries, then starts afresh on 8,000 held-out queries. Costs are mean $\pm$ SE over twenty evaluation permutations. The runs column counts every-prefix coverage successes.

Method	Selected c	Documents/query	Runs
Counterfactual (ours)	0.2	2.91 ± 0.02	20/20
DLR unprotected	0.5	2.82 ± 0.01	0/20
Primal-ACI	1	4.25 ± 0.02	20/20

B.8 COMPLETE FEEDBACK COMPARISON

These comparisons change only the update made on a protection round. Counterfactual updates use the original proposal's hit indicator, actual-hit updates use the returned set's indicator, and skipped updates leave the threshold unchanged. All protected variants share the miss budget and decreasing-step-size schedule. Unprotected DLR supplies the corresponding proposal-cost reference.

Table 15 retains the initial five-order comparison. The separate twenty-order comparison appears in Table 2(b), with its coverage and trigger statistics in Table 16. Table 8 gives the paired cost contrasts and their intervals. At $\tau_1 = 0.5$, actual-hit and counterfactual costs are close, while using recovered proposal outcomes reduces cost relative to skipping updates. Table 14 reports the signed paired counts and two-sided sign tests, with zero differences excluded from the test sample. The actual-hit comparison uses nine nonzero pairs, and the skipped-update comparison uses all twenty.

Table 14: **Direction of paired feedback-cost differences.** The twenty-order feedback batch compares counterfactual cost minus each listed variant. Counts show positive, negative and zero differences. The two-sided sign test uses only nonzero pairs; the paired differences and intervals appear in Table 8.

Ours minus	$+/-/0$	Nonzero n	Two-sided p
Unprotected DLR	11/0/9	11	0.00098
Actual-hit updates	6/3/11	9	0.51
Skipped updates	2/18/0	20	0.0004

Table 15: **Initial five-order feedback comparison.** Mean $\pm$ SE on SQuAD, with the same initialization and step sizes as the main configuration. The runs column counts every-prefix coverage successes. These orders are separate from Table 2(b).

Variant	Documents/query	Runs
Unprotected DLR	2.519 ± 0.007	3/5
Actual-hit updates	2.7 ± 0.1	5/5
Skipped updates	2.811 ± 0.008	5/5
Counterfactual (ours)	2.62 ± 0.03	5/5

Table 16: **Coverage and protection frequency in the twenty-order feedback comparison.** Mean $\pm$ SE. Startup comprises the first nine full-pool returns. Document costs and paired contrasts appear in Tables 2 and 8.

Variant	Final (%)	Min. prefix (%)	N_T after startup
Unprotected DLR	90.09 ± 0.01	90.010 ± 0.006	0
Actual-hit updates	90.09 ± 0.01	90.017 ± 0.004	15 ± 5
Skipped updates	90.16 ± 0.01	90.096 ± 0.006	0
Counterfactual (ours)	90.09 ± 0.01	90.017 ± 0.004	14 ± 5

B.9 INITIAL THRESHOLD AND FEEDBACK

We evaluate $\tau_1 \in \{0.05, 0.1, 0.15, 0.2, 0.244, 0.3, 0.4, 0.5\}$ with $c = 0.2$ and $\beta = 0.6$. Twenty new permutations per dataset are shared across every initialization and feedback variant, giving 320 matched settings. Table 17 reports all four variants, extending the comparison in Figure 1. The empirical target quantiles are approximately 0.24379 on SQuAD and 0.18490 on Natural Questions. These extra digits distinguish the 0.244 grid point from the SQuAD quantile.

Actual-hit thresholds remain at or below their initial value in all 320 settings, consistent with Proposition 2. For starts below the empirical target quantile, Table 18 compares protection frequencies with $1 - \alpha/p_{\text{init}}$ evaluated using the empirical initial miss rate. The reference is an empirical substitution into the asymptotic expression in Proposition 2. All $(4 + 3) \times 20 = 140$ settings below the corresponding dataset quantile meet or exceed that reference in this finite comparison. Frequent full-pool returns explain their large document costs.

The grid also records separation after some above-quantile starts. Protected hits can reduce the threshold, but the saved aggregate grid does not isolate this transient explanation. The full grid shows

Table 17: **All feedback variants across the initialization grid.** Mean document cost ± SE over twenty shared permutations per dataset and initial threshold. All protected variants satisfy the target at every prefix. Figure 1 displays the two protected update rules central to Proposition 2.

τ_1	Counterfactual	Actual-hit	Skipped	Unprotected DLR
SQuAD				
0.05	2.85 ± 0.03	45.9	3.14 ± 0.02	2.331 ± 0.008
0.1	2.88 ± 0.03	45.9	2.98 ± 0.02	2.325 ± 0.007
0.15	2.87 ± 0.03	45.9	2.84 ± 0.02	2.334 ± 0.008
0.2	2.78 ± 0.03	45.657 ± 0.007	2.76 ± 0.02	2.347 ± 0.008
0.244	2.84 ± 0.04	42.22 ± 0.07	2.70 ± 0.02	2.33 ± 0.01
0.3	2.88 ± 0.03	26.9 ± 0.5	2.66 ± 0.02	2.328 ± 0.007
0.4	2.75 ± 0.03	4.8 ± 0.2	2.60 ± 0.02	2.386 ± 0.009
0.5	2.61 ± 0.02	2.62 ± 0.02	2.837 ± 0.006	2.561 ± 0.008
Natural Questions				
0.05	1.60 ± 0.06	45.9	1.83 ± 0.02	0.9063 ± 0.0006
0.1	1.61 ± 0.06	45.9	1.63 ± 0.02	0.9087 ± 0.0004
0.15	1.61 ± 0.08	45.9	1.51 ± 0.02	0.918 ± 0.004
0.2	1.62 ± 0.07	43.94 ± 0.04	1.41 ± 0.03	0.912 ± 0.005
0.244	1.69 ± 0.07	26.1 ± 0.6	1.33 ± 0.02	0.9065 ± 0.0008
0.3	1.61 ± 0.06	7.1 ± 0.5	1.18 ± 0.02	0.9123 ± 0.0004
0.4	1.15 ± 0.03	1.28 ± 0.05	1.236 ± 0.001	1.0114 ± 0.0008
0.5	1.299 ± 0.002	1.299 ± 0.002	1.727 ± 0.001	1.299 ± 0.002

Table 18: **Actual-hit protection below the empirical target quantile.** The reference substitutes the empirical initial miss rate into the asymptotic expression $1 - \alpha/p_{\text{init}}$ in Proposition 2. Actual-hit trigger percentages are mean ± SE over twenty orders. Each observed run meets or exceeds its reference value.

Dataset	τ_1	p_{init}	Empirical formula reference (%)	Actual-hit (%)
SQuAD	0.05	1.000	90.0	90
SQuAD	0.1	1.000	90.0	90
SQuAD	0.15	0.891	88.8	90
SQuAD	0.2	0.388	74.2	89.51 ± 0.01
Natural Questions	0.05	1.000	90.0	90.008
Natural Questions	0.1	1.000	90.0	90.008
Natural Questions	0.15	0.509	80.4	90.008

that an initial threshold above the empirical quantile can still lead to substantial actual-hit cost. At $\tau_1 = 0.5$, actual-hit and counterfactual variants execute identical sets in fourteen SQuAD orders and all twenty Natural Questions orders (Table 19).

Table 19: **Actual-hit versus counterfactual returned sets.** Counts use the default-start subset of the initialization batch. A matching order has the same executed document set after every query. This comparison differs from counterfactual–unprotected threshold equality in Table 21.

Dataset	Orders	Identical executed sets at $\tau_1 = .5$
SQuAD	20	14/20
Natural Questions	20	20/20

The relative cost of skipped and counterfactual updates also changes across the grid. Table 20 displays counterfactual-minus-skipped cost and its paired SE at each start. The sign changes across the grid: preserving the learner trajectory does not order the document costs of these two protected rules.

Table 20: **Paired cost across initialization.** Counterfactual minus skipped-update documents/query, mean $\pm$ paired SE over twenty shared orders. Negative values favor counterfactual updates. The same orders are reused across starts within each dataset.

τ_1	SQuAD	Natural Questions
0.05	-0.296 ± 0.037	-0.229 ± 0.045
0.1	-0.099 ± 0.033	-0.016 ± 0.058
0.15	0.030 ± 0.031	0.106 ± 0.074
0.2	0.020 ± 0.028	0.207 ± 0.059
0.244	0.133 ± 0.039	0.365 ± 0.064
0.3	0.216 ± 0.034	0.427 ± 0.054
0.4	0.155 ± 0.034	-0.083 ± 0.031
0.5	-0.222 ± 0.017	-0.4279 ± 0.0021

Counterfactual updates preserve the matched unprotected trajectory at every initialization. The stored maximum absolute threshold difference is exactly zero in every run, before display rounding (Table 21).

Table 21: **Trajectory equality and returned-set coverage are separate checks.** Eight starts share twenty query orders per dataset. Exact match means the saved maximum absolute difference across the two threshold sequences equals zero before rounding. Prefix success requires coverage at least 90% after every query.

Dataset	Settings	Exact threshold match	Protected prefix success	Unprotected success
SQuAD	160	160/160	160/160	14/160
Natural Questions	160	160/160	160/160	26/160

Table 22 compares low and default initialization for counterfactual updates. The cumulative cost changes even when the final thresholds are close, showing that initialization affects the transient learning process. Table 23 complements these costs with counts of coverage shortfalls across all orders. Unprotected DLR has a violating prefix in 280 settings, whereas counterfactual protection has none in all 320. On Natural Questions at $\tau_1 = 0.5$, the initial triggers add no documents because the proposals already contain the full pool.

Table 22: **Transient costs under different initializations.** Counterfactual updates at $\tau_1 = 0.05$ minus those at $\tau_1 = 0.5$, using twenty paired orders. Entries are mean $\pm$ SE. Threshold differences are scaled by 10^4 for display.

Dataset	Cost difference	Final-threshold difference (10^{-4})
SQuAD	0.234 ± 0.035	-0.47 ± 0.94
Natural Questions	0.305 ± 0.058	-2.56 ± 0.93

Table 23: **Coverage protection across all initial thresholds.** Failed runs contain at least one prefix below 90% coverage. Failed prefixes and trigger percentages are mean $\pm$ SE over twenty orders per entry. Counterfactual protection has zero failed prefixes in all $2 \times 8 \times 20 = 320$ runs; exact threshold equality is reported separately in Table 21.

τ_1	Unprotected failed runs	Unprotected failed prefixes	Counterfactual triggers (%)	Actual-hit triggers (%)
SQuAD				
0.05	20/20	4900 ± 300	1.13 ± 0.08	90
0.1	20/20	4500 ± 300	1.20 ± 0.07	90
0.15	20/20	3500 ± 200	1.20 ± 0.07	90
0.2	20/20	3000 ± 400	0.98 ± 0.07	89.51 ± 0.01
0.244	20/20	3400 ± 300	1.10 ± 0.09	82.6 ± 0.1
0.3	20/20	3200 ± 300	1.20 ± 0.07	52 ± 1
0.4	20/20	1000 ± 100	0.87 ± 0.07	5.5 ± 0.4
0.5	6/20	15 ± 8	0.20 ± 0.04	0.23 ± 0.05
Natural Questions				
0.05	20/20	2700 ± 200	1.4 ± 0.1	90.008
0.1	20/20	1800 ± 100	1.4 ± 0.1	90.008
0.15	20/20	1700 ± 200	1.4 ± 0.2	90.008
0.2	20/20	1900 ± 200	1.4 ± 0.1	86.08 ± 0.07
0.244	20/20	1800 ± 200	1.6 ± 0.1	50 ± 1
0.3	20/20	990 ± 90	1.4 ± 0.1	12 ± 1
0.4	14/20	40 ± 10	0.41 ± 0.06	0.7 ± 0.1
0.5	0/20	0	0.138	0.138

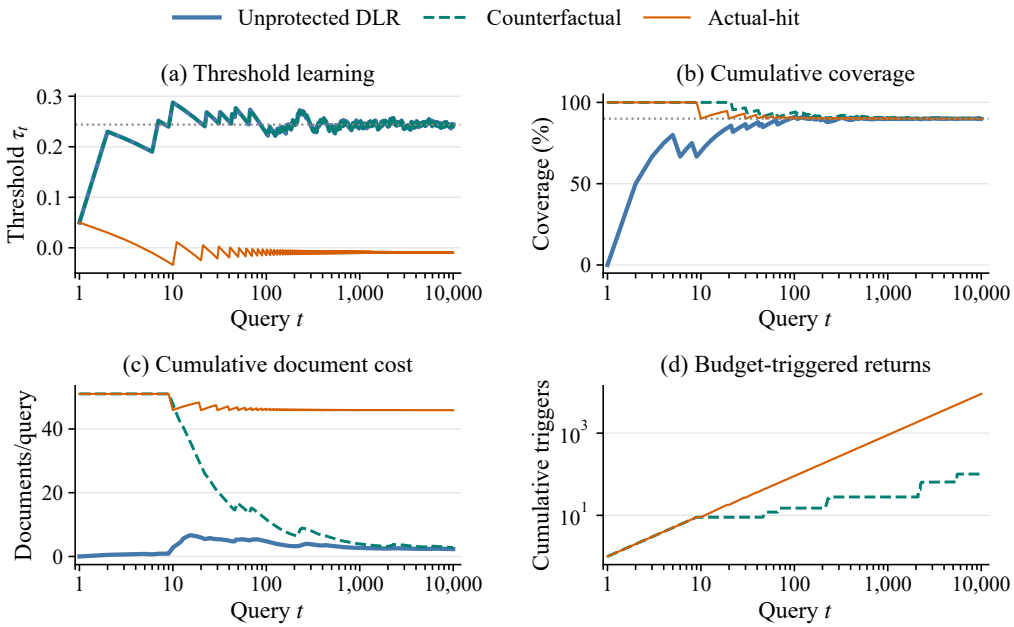

Figure 3: **Learning and protection along one query order.** The first order of the SQuAD initialization study starts at $\tau_1 = 0.05$. Panel (a) overlays counterfactual and unprotected thresholds, with the empirical target quantile dotted. Panel (b) shows cumulative coverage and the 90% target. Panels (c,d) show mean documents returned so far and cumulative budget triggers. The counterfactual and unprotected threshold curves coincide, while their coverage and returned-document costs differ. Every curve uses the individual query outcomes.

Figure 3 illustrates the mechanism behind the aggregate counts. Actual-hit updates repeatedly return the full pool while their threshold remains low. Counterfactual updates follow the unprotected learner's threshold sequence and enlarge returned sets when the cumulative budget requires a hit. The cumulative curves describe this individual order, while Table 23 reports the full initialization study.

B.10 Changing query distributions

We concatenate SQuAD and Natural Questions in both orders, independently permuting each collection with five shared seeds. The primary learner carries its threshold, miss count and step-size index across the boundary. A boundary-informed control restarts on the second collection with $\tau_1 = 0.5$, zero misses and step index one, using the same c and β. Table 25 compares document costs on the second collection.

Carrying state costs more after Natural Questions switches to SQuAD and less in the reverse direction. Both sequences maintain global prefix coverage and the matched unprotected threshold trajectory. Table 24 shows the early response after switching. Carrying state from NQ into SQuAD requires frequent protection in the first $1{,}000$ queries, whereas the reverse direction requires none in those recorded windows. These windows locate the early cost under carried state. The paired carry-minus-restart effect is measured over the full second collection in Table 25. The lower NQ target quantile and the older step-size index offer a plausible explanation for slower adaptation toward SQuAD, but threshold, budget and step size are carried together in this comparison. Global prefix coverage does not require coverage above the target on every prefix starting afresh at the switch.

B.11 Coverage targets and candidate-pool sizes

We cross error rates $\alpha \in \{0.05, 0.1, 0.2\}$ with pool sizes $K \in \{11, 21, 31, 51\}$. Each pool retains the correct document and the first $K - 1$ ordinary negatives in source order. Every setting uses twenty shared SQuAD permutations with the main initialization and step sizes. All runs maintain their target after every query.

Table 24: **First 1,000 post-switch queries under carried state.** Mean ± SE over five orders. Window coverage counts hits only in this window; the guarantee and success counts in Table 25 concern prefixes of the complete concatenation. The window statistics locate early protection cost without separately manipulating the inherited threshold, budget and step index.

Direction	Documents/query	Triggers (%)	Window coverage (%)
NQ → SQuAD	14.5 ± 0.5	26 ± 1	89.4 ± 0.2
SQuAD → NQ	1.16 ± 0.03	0	96.9 ± 0.1

Table 25: **Document cost after a distribution switch.** Mean ± SE across five paired orders. Costs refer to queries after the switch. The difference is carry-state minus a boundary-informed restart of threshold, misses and step index. The runs column counts the carry-state learner's every-prefix coverage successes over the complete concatenated sequence.

Sequence	Carry state	Restart	Paired difference	Runs
NQ → SQuAD	3.8 ± 0.1	2.56 ± 0.02	1.23 ± 0.10	5/5
SQuAD → NQ	0.946 ± 0.005	1.301 ± 0.003	-0.3555 ± 0.0034	5/5

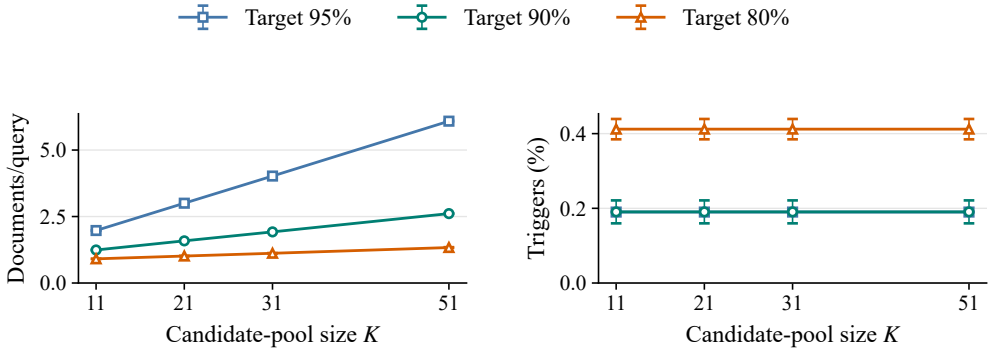

Figure 4: **Document cost across candidate-pool sizes and coverage targets.** Markers are means and error bars are SE over twenty shared SQuAD orders. The correct document remains in every pool. Larger pools increase document cost, while protection frequency is unchanged at a fixed target because the correct-document scores and update sequence are identical.

Figure 4 makes the cost changes visible across the complete grid. For a fixed target, protection decisions are identical across pool sizes. Their cost still increases because more candidates can enter the proposals and protective returns. The initial protection periods contain 19, 9 and 4 queries for $\alpha = 0.05$, 0.1 and 0.2, respectively. Table 26 supplies the numerical comparisons and minimum prefix coverage.

Table 26: **Coverage targets and candidate-pool sizes.** Mean ± SE across twenty SQuAD orders per setting. Each pool contains the correct document. All $3 \times 4 \times 20 = 240$ setting–order combinations meet their own target after every query, sharing twenty orders across configurations. The reference cost is fixed within each query collection.

α	K	Reference cost	Excess documents	Triggers (%)	Min. prefix (%)
0.05	11	1.704	0.270 ± 0.004	0.19	95.128 ± 0.006
0.05	21	2.473	0.529 ± 0.007	0.19	95.128 ± 0.006
0.05	31	3.232	0.79 ± 0.01	0.19	95.128 ± 0.006
0.05	51	4.771	1.32 ± 0.02	0.19	95.128 ± 0.006
0.1	11	1.157	0.085 ± 0.003	0.19 ± 0.03	90.016 ± 0.003
0.1	21	1.417	0.169 ± 0.006	0.19 ± 0.03	90.016 ± 0.003
0.1	31	1.668	0.255 ± 0.009	0.19 ± 0.03	90.016 ± 0.003
0.1	51	2.189	0.42 ± 0.01	0.19 ± 0.03	90.016 ± 0.003
0.2	11	0.849	0.060 ± 0.003	0.41 ± 0.03	80
0.2	21	0.896	0.119 ± 0.005	0.41 ± 0.03	80
0.2	31	0.941	0.177 ± 0.008	0.41 ± 0.03	80
0.2	51	1.040	0.29 ± 0.01	0.41 ± 0.03	80

B.12 Longer sequences

We sample five independent streams with replacement from the fixed SQuAD empirical distribution and evaluate nested prefixes at seven horizons. The experiment describes finite-horizon behavior on this discrete distribution; the population continuity conditions in (A2)–(A3) remain assumptions of the asymptotic theorem. Initialization and the step-size schedule remain unchanged. The reference uses the original empirical fixed threshold and evaluates its cost on each sampled prefix. Table 27 decomposes total excess into proposal excess and documents added by protection.

The first three horizons have nine triggers and zero protection cost because those proposals contain the full pool. Additional protection appears by $T = 10{,}000$, followed by decreasing mean trigger frequencies across the later measured horizons. The full trajectory of these summaries is visible in Figure 2(c,d). At $T = 100{,}000$, proposal excess and protection each contribute approximately half of the remaining total excess.

Table 27: **Protection and document cost across sequence lengths.** Mean ± SE across five independent streams resampled from the fixed empirical SQuAD distribution. Each stream supplies nested prefixes at all seven horizons. Costs are documents per query relative to the original fixed threshold evaluated on the same prefix.

T	Triggers (%)	Protection cost	Proposal excess	Total excess
1 000	0.9	0	3.0 ± 0.1	3.0 ± 0.1
2 500	0.36	0	1.16 ± 0.03	1.16 ± 0.03
5 000	0.18	0	0.64 ± 0.03	0.64 ± 0.03
10 000	0.28 ± 0.08	0.09 ± 0.04	0.36 ± 0.02	0.45 ± 0.05
25 000	0.23 ± 0.03	0.09 ± 0.01	0.17 ± 0.02	0.27 ± 0.03
50 000	0.19 ± 0.03	0.08 ± 0.02	0.11 ± 0.02	0.20 ± 0.03
100 000	0.15 ± 0.03	0.07 ± 0.01	0.06 ± 0.01	0.126 ± 0.002

B.13 Score ties at the target quantile

Write $\widehat{\tau}^\star$ for the target quantile of the stored correct-document scores, distinct from the population $\tau^\star$ in (A2). We examine ties near this empirical quantile. Cosines are stored in `float32` and mapped by $(1 - \cos)/2$ in `float64`, giving a grid spacing of 1.490×10^{-8} near the target quantile. Table 28 reports the distinct fraction, the largest repeated-score mass over each query collection,

and the local tie structure within 10^{-3} of its empirical target quantile. The largest repeated masses are approximately 0.031% on Natural Questions and 0.060% on SQuAD. All 68 and 78 scores in the respective local windows are distinct, so repeated values away from the target account for the observed ties. The empirical quantile comparison therefore uses individually resolved scores in this neighborhood. This finite-sample check describes stored-score ties, rather than continuity of a population distribution.

Table 28: Tie structure of the correct-document nonconformity scores. Distinct fraction and heaviest mass are over the whole query collection, and the last two columns cover the scores within 10^{-3} of $\widehat{\tau}^{\star}$.

Corpus	Scores	Distinct	Heaviest mass (%)	Within 10^{-3}	Distinct there
Natural Questions	6515	99.85%	0.031	68	68
SQuAD	10,000	99.29%	0.060	78	78

Online computation. Table 29 measures the selection and update procedure on precomputed scores. The timed function scans candidates, selects documents and updates the threshold and miss count, corresponding to Section 4.3. Encoding, initial retrieval and downstream generation are outside this measurement. Returned-document count and online selection time measure different resources.

Table 29: **Online selection and state-update time.** Includes candidate scanning, selection, feedback recovery and threshold/miss-count updates on saved scores. Encoder inference, retrieval and generation are outside the timing.

Quantity	Recorded setting or measurement
Input	10,000 queries, 51 precomputed scores/query
Hardware	Intel Core i9-12900K, one CPU thread
Runtime	Python 3.12.4, NumPy 2.3.2
Repetitions	One warm-up, five timed runs
Total time, median [range]	63.1 [62.4, 65.3] ms
Per-query time, median	6.31 microseconds