TREE-BASED ADAPTIVE BLOCK ELIMINATION WITH EXEMPLAR SAMPLING FOR FIXED-BUDGET BEST-ARM IDENTIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study fixed-budget best-arm identification over a finite set of candidate arms equipped with a metric. The finite action set may be large, making exhaustive arm-level exploration costly even when nearby arms have similar mean rewards. We propose **TABLE** (Tree-based Adaptive Block Elimination with Exemplar Sampling), which maintains a hierarchical partition of the candidate set, represents each active block by a small set of geometrically selected exemplars, uses top-two posterior sampling to focus exploration on competitive exemplars, and eliminates or refines blocks using confidence evidence. The resulting algorithm trades exhaustive arm-level sampling for a combination of statistical uncertainty and geometric approximation error. We provide a structured certification analysis that makes this trade-off explicit. Under valid adaptive confidence radii, a representability condition, and stated progress and terminal-state conditions, TABLE safely preserves the optimal block and identifies the best arm once statistical and geometric uncertainty fall below the global arm gap. For sub-Gaussian rewards, we give an implementable fixed-parameter confidence calibration and an oracle exponential expression for the error probability intersected with the favorable trajectory event; the unconditional bound retains the probability of its complement. Experiments on Bernoulli metric bandit and a structured dynamic-pricing benchmark show that TABLE is especially effective in low- and moderate-budget regimes.

1 INTRODUCTION

Best-arm identification (BAI) is a canonical pure-exploration problem in stochastic multi-armed bandits (MABs). Unlike cumulative-regret minimization, the learner is evaluated by the recommendation returned after exploration (Audibert & Bubeck, 2010; Bubeck et al., 2011a; Gabillon et al., 2012). We focus on the *fixed-budget* formulation: a sampling budget T is specified in advance and the goal is to minimize the probability of recommending an arm other than the unique best arm (Carpentier & Locatelli, 2016; Tirinzoni & Degenne, 2022). This differs from fixed-confidence BAI, where the target error probability is prescribed and the stopping time is optimized.

We consider a *finite* arm set $\mathcal{A} = \{a_1, \dots, a_K\} \subseteq \mathcal{X}$, where $(\mathcal{X}, d)$ is a metric space and the unknown mean reward $\mu : \mathcal{A} \to \mathbb{R}$ is L-Lipschitz. Thus, our problem is not exact optimization over a continuum: the learner must identify the best element of a finite candidate set. The metric is side information that can be exploited to avoid treating the K candidates as unrelated alternatives. This setting is natural when actions are intrinsically discrete but geometrically organized (LeJeune et al., 2020; Rizk et al., 2021), and when a continuous design domain is implemented through a finite candidate grid (Binois et al., 2020; Gong et al., 2026). Examples include dynamic pricing (Parambath et al., 2025; Gong et al., 2026), hyperparameter search (Jamieson & Talwalkar, 2016; Falkner et al., 2018), clinical trials (Aziz et al., 2021), and sensor-placement problems (Chen et al., 2014).

Although the action space is finite, identifying the best arm can remain challenging when the number of candidates is large, particularly when the available budget is insufficient to obtain reliable estimates for every arm. Classical fixed-budget methods progressively eliminate suboptimal arms but allocate samples at the individual-arm level (Audibert & Bubeck, 2010; Gabillon et al., 2012; Zhuang et al., 2017; Ban & He, 2020; Zhu et al., 2020; Wang et al., 2023). When the candidate arms possess an

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.

underlying geometric structure, however, nearby arms may have similar expected rewards, providing an opportunity to share information across related actions rather than exploring them independently. Exploiting this structure is particularly valuable under limited budgets, as it may allow the learner to identify and discard suboptimal regions without exhaustively sampling their constituent arms. In this work, we assume a unique best arm within the finite candidate set and aim to minimize its misidentification probability by exploiting the metric structure through representative-based exploration and adaptive block elimination.

A natural question is whether one could simply take a continuous-space Lipschitz bandit method and restrict its actions to $\mathcal{A}$. Such a restriction is possible in principle, but it does not automatically produce the mechanism studied here. Many continuous Lipschitz methods target cumulative regret (Magureanu et al., 2014; Podimata & Slivkins, 2021), while infinite-arm pure-exploration methods frequently target ϵ-optimality under fixed confidence (Aziz et al., 2018; Kocák & Garivier, 2021; Jourdan et al., 2023). More importantly, action restriction alone does not imply that the learner will summarize the *realized finite candidate geometry* by a small set of representatives, eliminate a whole block from representative evidence, and refine only the surviving blocks. TABLE is designed specifically around that finite-set, fixed-budget certification problem.

Our approach. We propose **TABLE**, a tree-based adaptive block-elimination method. TABLE begins with a coarse partition of $\mathcal{A}$. Each active node u is represented by at most b exemplar arms chosen to cover the node geometrically. Exploration samples active exemplars. A one-draw top-two posterior sampling (TTPS) heuristic focuses samples on competitive exemplars, while a balancing mode targets under-sampled exemplars; a quantitative time-to-check guarantee is stated separately as a sufficient progress condition. At a check time, TABLE converts exemplar confidence intervals into valid block-level lower and upper scores using Lipschitz continuity. Blocks that can no longer contain the optimum are eliminated; promising but geometrically coarse blocks are recursively split.

Our theoretical analysis separates what is *derived from the algorithm* from what is an *additional sufficient condition*. Under valid adaptive confidence intervals and Lipschitz continuity, block confidence scores are valid and the optimal block cannot be eliminated. To convert this safe-elimination property into a finite-horizon exact-identification bound, the current analysis additionally requires representability, progress, and terminal-state conditions. We make these conditions explicit, classify them by type, and state the probability guarantee accordingly. This clarification is central: the finite-horizon result is a *conditional structured-certification theorem*, not a claim that every execution of TABLE automatically satisfies all refinement and terminal conditions.

Contributions.

- We introduce TABLE for finite-set metric BAI. Our technical mechanism combines multiple-exemplar geometric coverage, Lipschitz block certification, and explicit accounting for balancing and refinement. Recursion and top-two sampling themselves are not claimed as novel.
- We derive block confidence scores that explicitly combine statistical exemplar uncertainty with a geometric coverage term $L\kappa(u)\mathrm{rds}(u)$, allowing blocks to be certified without sampling every contained arm.
- We give a transparent theory that separates safe elimination (derived from confidence validity and Lipschitz continuity) from additional representability, progress, and terminal-state conditions. The theorem is stated as an event-based bound and as an unconditional bound with an explicit unfavorable-trajectory term.
- For conditionally sub-Gaussian rewards, we provide *(i)* an implementable fixed-δ_0 radius that requires neither the unknown arm gap nor the representation time as inputs, and *(ii)* an oracle exponential bound on the intersection with the favorable trajectory event, with the residual term explicit in the unconditional statement.
- Experiments on two structured BAI benchmarks show that TABLE is particularly effective when the budget is tight relative to the cost of accurate arm-level estimation.

Table 1: Main notation used in TABLE.

Symbol	Description
$\mathcal{A}, K, T$	Finite arm set, number of arms, and sampling budget.
d, L	Arm-space metric and Lipschitz constant.
$\mathcal{U}(t), \mathcal{C}(u)$	Active node set at time t and arms contained in node u.
$\mathcal{Z}(u), b$	Exemplar set of node u and maximum number of exemplars per node.
$\mathrm{rds}(u), \kappa(u)$	Node half-diameter and normalized exemplar covering factor.
$N_z(t), \hat{\mu}_z(t)$	Number of pulls and empirical mean of exemplar z before round t.
$\Pi_z(t)$	Sampling distribution of exemplar z at time t.
$\mathrm{rad}_z(t)$	Confidence radius associated with exemplar z at time t.
$\mathrm{LCS}_u(t), \mathrm{UCS}_u(t)$	Lower and upper confidence scores of node u.
$m, n_m, \hat{\tau}_m$	Ladder level, sample threshold, and balancing gate.
$c_0, H, n_{\min}$	Initial number of nodes, maximum tree depth, & minimum split-size parameter.
$\Delta_{\min}, \Delta(u)$	Global arm gap and node gap; definitions in Section 4.
η, β	Split-threshold parameter and top-two sampling probability.

2 Related Work

Classical fixed-budget BAI methods such as Successive Rejects (SR) (Audibert & Bubeck, 2010) and Sequential Halving (SH) (Karnin et al., 2013) progressively eliminate individual arms; UGapE (Gabillon et al., 2012), VBR (Faella et al., 2020), and Continuous Rejects (Wang et al., 2023) make allocation or rejection more adaptive but remain primarily arm-level. Their sharp guarantees are naturally expressed through arm-wise gap complexities. TABLE instead spends budget on a smaller, changing set of representatives, so its certification cost also contains geometric and representation terms.

Structured BAI can share information through global parametric models, e.g., linear or generalized-linear structure (Azizi et al., 2022). TABLE assumes only a metric and Lipschitz continuity. SBE (Shahrampour et al., 2017) studies block elimination using a fixed partition and side observations; TBE (Parambath et al., 2025) provides a closer algorithmic point of comparison for posterior-guided block exploration. Here the analysis focuses on the particular combination of adaptive refinement, exemplar confidence, and Lipschitz geometric transfer, without claiming that block elimination or recursive exploration is new. Continuous/infinite-action Lipschitz methods exploit related geometric ideas (Kleinberg et al., 2008; Bubeck et al., 2011b; Magureanu et al., 2014; Kleinberg et al., 2019; Podimata & Slivkins, 2021), but often target cumulative regret or approximate optimization. Merely restricting such a method to the finite set $\mathcal{A}$ does not automatically produce TABLE's finite-set exemplar representation, block certification, and refinement mechanism. Extended comparisons are in Appendix B.7.

3 TABLE

Let $\mathcal{A} = \{a_1, \ldots, a_K\} \subseteq \mathcal{X}$ be finite, let d be a metric, and assume $|\mu(a) - \mu(a')| \le L d(a, a')$. TABLE maintains active nodes $\mathcal{U}$, where node u contains arms $\mathcal{C}(u)$ and has radius

$$\mathrm{rds}(u) = \tfrac{1}{2} \max_{a,a' \in \mathcal{C}(u)} d(a, a').$$

A deterministic covering rule selects at most b exemplars $\mathcal{Z}(u) \subseteq \mathcal{C}(u)$. Their normalized cover factor is defined as:

$$\kappa(u) = \frac{\max_{a \in \mathcal{C}(u)} \min_{z \in \mathcal{Z}(u)} d(a, z)}{\mathrm{rds}(u)}$$

for $\mathrm{rds}(u) > 0$ and 0 otherwise. Thus $\kappa(u)\mathrm{rds}(u)$ is the node's exemplar covering radius; because exemplars lie in the node, $\kappa(u) \le 2$.

In *explore* mode, TABLE draws posterior samples for the active exemplars, scores each node by the average sampled value of its representatives, selects the highest-scoring node, and applies TTPS to its two most competitive exemplars. In *balance* mode it samples a least-sampled active exemplar. Checks are organized by the ladder $n_m = n_0\, 2^{2^m}$. Once all active exemplars reach the current

Algorithm 1 TABLE (summary)

```
Partition A into c_0 nodes; choose covering exemplars.
for t = 1, ..., T do
    if new exemplars are queued then
        sample the next queued exemplar.
    else if balancing is required then
        sample a least-sampled active exemplar.
    else
        score nodes by posterior samples and apply TTPS in the best node.
    end if
    Update empirical statistics/posterior.
    if all active exemplars have reached n_m then
        Eliminate nodes via (LCS, UCS); split eligible survivors.
        Advance m only if a split occurred; queue new exemplars.
    end if
end for
Final elimination-only check; return empirically best surviving exemplar.
```

threshold, TABLE forms

$$\mathrm{LCS}_u(t) = \max_{z \in \mathcal{Z}(u)} (\hat{\mu}_z(t) - \mathrm{rad}_z(t)) - L\kappa(u)\mathrm{rds}(u), \tag{1}$$

$$\mathrm{UCS}_u(t) = \max_{z \in \mathcal{Z}(u)} (\hat{\mu}_z(t) + \mathrm{rad}_z(t)) + L\kappa(u)\mathrm{rds}(u), \tag{2}$$

and removes u if $\mathrm{UCS}_u(t) < \max_v \mathrm{LCS}_v(t)$.

A surviving node is split when it is sufficiently large, below depth H, budget-feasible, and

$$L\mathrm{rds}(u) \geq \eta \max_{z \in \mathcal{Z}(u)} \mathrm{rad}_z(t).$$

The split rule refines geometry only when the current statistical precision warrants a finer representation. Newly introduced unsampled exemplars are queued for their first pull; their initialization and any accepted split must fit the remaining budget. The ladder advances only when a split occurs, so multiple elimination checks may take place at the same geometric resolution. At $T + 1$ TABLE performs a final elimination-only check and recommends the surviving exemplar with largest empirical mean. If a node contains only one exemplar, explore mode samples it directly; otherwise the highest and second-highest single posterior draws are selected with probabilities β and $1 - \beta$. Algorithm 1 summarizes the control flow; Appendix A gives complete pseudocode and detailed description.

Remark 1 (Relation to canonical TTTS)**.** *The explore-mode rule used here is a one-draw* top-two posterior sampling *heuristic: after drawing one posterior/sampling-distribution value for each active exemplar in the selected node, TABLE mixes between the largest and second-largest sampled values. We therefore use the label TTPS rather than claiming that this step is identical to canonical Top-Two Thompson Sampling, whose challenger construction can use additional posterior resampling. The finite-time certification theorem is not specific to either rule; it relies on valid confidence intervals and the explicit progress/terminal conditions below.*

4 Finite-Horizon Certification Analysis

We derive a finite-horizon misidentification bound for TABLE. Our analysis deliberately separates properties derived from TABLE's confidence construction from additional conditions needed for an explicit finite-horizon exact-identification bound.

For node u, define $\mu^\star(u) = \max_{a \in \mathcal{C}(u)} \mu(a)$ and $\Delta(u) = \mu(a^\star) - \mu^\star(u)$. Define $\Delta_{\min} = \min_{a \in \mathcal{A} \setminus \{a^\star\}} [\mu(a^\star) - \mu(a)] > 0$. Let $\mathcal{Z}_{\mathrm{all}} \subseteq \mathcal{A}$ contain every possible exemplar, and set $M_{\mathrm{conf}} = |\mathcal{Z}_{\mathrm{all}}|$; it is safe to assume $\mathcal{Z}_{\mathrm{all}} = \mathcal{A}$ and $M_{\mathrm{conf}} = K$. Pull counts $N_z(t)$ and sample means $\hat{\mu}_z(t)$ refer to observations available *before* round t, and an arm's observations persist if it ceases to be an exemplar and later reappears.

Assumption 2 (Finite-horizon adaptive confidence)**.** *For any possible exemplar z and any $\alpha \in (0, 1)$, the routine* $\mathrm{rad}_T(n, \alpha)$ *satisfies*

$$\mathbb{P}(\forall t \in \{1, \ldots, T + 1\} : \ |\hat{\mu}_z(t) - \mu(z)| \leq \mathrm{rad}_T(N_z(t), \alpha)) \geq 1 - \alpha.$$

Table 2: Classification and proof roles of the sufficient conditions in the finite-horizon analysis.

Condition	Nature	Role in the analysis
Adaptive confidence validity	Statistical	Ensures valid exemplar confidence intervals.
Representability	Structural	Ensures that $a^\star$ becomes and remains an exemplar along the optimal path.
Balancing progress	Algorithmic progress	Bounds the time required to make ladder levels checkable.
Optimal-path split progress	Instance–trajectory	Ensures sufficient refinement of the optimal branch.
Terminal no-split	Trajectory	Stabilizes the representative sets after T_{rep}.
Terminal sampling balance	Terminal sampling	Controls the final exemplar confidence radii through sample-count lower bounds.
Terminal split-stop	Terminal geometry	Controls the residual geometric approximation error.

The radius is nonincreasing in the count, $\mathrm{rad}_T(0,\alpha) = +\infty$*, and all confidence calculations use the same per-exemplar allocation*

$$\alpha = \delta / M_{\mathrm{conf}}, \qquad \mathrm{rad}_z(t) = \mathrm{rad}_T(N_z(t), \alpha).$$

Definition 3 (Good event)**.** *Let* $\mathcal{T}_{\mathrm{chk}}$ *denote the random set of all elimination check times, including the final forced check at* $T+1$*;* $\mathcal{U}(t)$ *in the following definition is the pre-elimination set at a check. Define*

$$\mathcal{G} := \{\forall t \in \mathcal{T}_{\mathrm{chk}}, \forall u \in \mathcal{U}(t), \forall z \in \mathcal{Z}(u) : |\hat{\mu}_z(t) - \mu(z)| \le \mathrm{rad}_z(t)\}.$$

Because each exemplar event is simultaneous over all counts encountered up to the horizon, a union bound over the M_{conf} possible exemplars gives $\mathbb{P}(\mathcal{G}) \ge 1 - \delta$; no further union bound over checks or ladder levels is needed.

Proposition 4 (Safe block elimination)**.** *On* $\mathcal{G}$*, every active node at every check satisfies*

$$\mathrm{LCS}_u(t) \le \mu^\star(u) \le \mathrm{UCS}_u(t),$$

so the active node containing $a^\star$ *is never eliminated.*

The lower inequality follows because every exemplar lies in the node. For the upper inequality, a node-optimal arm is within $\kappa(u)\mathrm{rds}(u)$ of some exemplar, and Lipschitz continuity transfers that exemplar's confidence bound to the node optimum. The full proof is in Appendix B.

4.1 Additional conditions

The remainder of the theorem uses additional sufficient structural and execution conditions; these are not consequences of Algorithm 1 for every run. In particular, reaching a ladder threshold need not force the optimal branch to split, and a final elimination-only check does not enforce a geometric stopping inequality. Table 2 summarizes the roles of these conditions.

Representability means that a fixed, reward-independent partition tree and deterministic exemplar-selection rule include $a^\star$ after at most $M_{\mathrm{rep}} \le H$ successive optimal-path splits and retains it on subsequent optimal-path descendants. One sufficient construction has depth-H leaves containing at most b arms and selects every arm in a node of size at most b; the size/depth/budget guards must still permit sufficient progress to those leaves. Instance-specific representability of $a^\star$ may nevertheless hold without representing every arm.

The formal milestone-based progress/terminal conditions and their stagewise interpretation are in Appendix B.2. With deterministic stage bounds q_m, $n_{-1} = 0$, and the convention that empty sums vanish when $M_{\mathrm{rep}} = 0$, they give the representation-time envelope

$$T_{\mathrm{rep}} = 1 + 2U_{\max} b\, n_{M_{\mathrm{rep}}-1} + \sum_{m=0}^{M_{\mathrm{rep}}-1} q_m, \qquad U_{\max} = c_0\, 2^H,$$

and an effective post-representation scale

$$s_T = \left\lfloor \frac{(T - T_{\mathrm{rep}})_+}{2U_{\max} b} \right\rfloor.$$

Let $\mathcal{C}_T$ denote the event that the prescribed milestone-based optimal-path progress, terminal no-split, sampling-balance, and split-stop conditions hold; the structural representability requirement is fixed

for the chosen instance and tree. In particular, $\mathcal{C}_T$ is not a shorthand for the confidence event $\mathcal{G}$. Let $\bar{\kappa} \leq 2$ bound all possible cover factors, assume $\eta > 0$, and define

$$r_T(\delta) = \mathrm{rad}_T\left(s_T, \frac{\delta}{M_{\mathrm{conf}}}\right).$$

The absence of a ladder-dependent confidence penalty is deliberate: Assumption 2 covers all times through $T+1$, so the same per-exemplar event covers every check time and every ladder level. Its monotonic-radius requirement is satisfied by the sub-Gaussian routine analyzed below but must be checked separately for any variance-adaptive empirical implementation.

The dependencies for the stronger block-elimination certificate are:

confidence + Lipschitz $\Rightarrow$ optimal node survives, representability + progress $\Rightarrow a^\star$ is represented,
terminal balance $\Rightarrow$ small statistical error, terminal split-stop $\Rightarrow$ small geometric error.

Theorem 5 (Structured fixed-horizon certification)**.** *Under Assumption 2 and persistent structural representability, fix $\delta \in (0,1)$ and deterministic progress bounds defining T_{rep}. If $T \geq T_{\mathrm{rep}}$, $s_T \geq 1$, and $\Delta_{\min} > 2(2+\eta\bar{\kappa})r_T(\delta)$, then*

$$\mathbb{P}(\{\hat{a}_T \neq a^\star\} \cap \mathcal{C}_T) \leq \delta,$$

and therefore

$$\mathbb{P}(\hat{a}_T \neq a^\star) \leq \delta + \mathbb{P}(\mathcal{C}_T^c).$$

If $\mathcal{C}_T$ holds deterministically or almost surely for the considered instance–algorithm class, then $\mathbb{P}(\hat{a}_T \neq a^\star) \leq \delta$.

This is an *event-based* certificate: the proof shows $\hat{a}_T = a^\star$ on $\mathcal{G} \cap \mathcal{C}_T$. It does not imply $\mathbb{P}(\hat{a}_T \neq a^\star \mid \mathcal{C}_T) \leq \delta$ without additional independence or conditional-confidence arguments, because $\mathcal{C}_T$ may depend on the sampled rewards.

4.2 Understanding the certification mechanism

Why representatives can certify blocks. Consider a one-dimensional node with arms $\mathcal{C}(u) = \{0,1,2,3\}$, exemplars $\mathcal{Z}(u) = \{0,3\}$, and metric $d(x,y) = |x-y|$. Its half-diameter is $\mathrm{rds}(u) = 3/2$, its exemplar covering radius is $\max_{a\in\mathcal{C}(u)} \min_{z\in\mathcal{Z}(u)} d(a,z) = 1$, and hence $\kappa(u) = 2/3$. If $L = 1$, every arm in this node has mean at most the mean of its nearest exemplar plus 1. On the confidence event the node's upper score is therefore

$$\mathrm{UCS}_u(t) = \max_{z\in\{0,3\}} \big(\hat{\mu}_z(t) + \mathrm{rad}_z(t)\big) + 1.$$

The corresponding lower score subtracts the same geometric allowance. The scores remain valid even if the node-optimal arm is 1 or 2, neither of which is sampled. Thus the optimal node cannot be discarded merely because its best arm is not yet an exemplar.

From block preservation to an arm recommendation. Safe block elimination is only the first step. If the best arm is 1 in this example, the initial exemplars cannot return it. A refinement that creates $\{0,1\}$ and selects both arms as exemplars makes the optimum representable; continued sampling then controls the final empirical-mean comparison. The structured theorem additionally uses the terminal split-stop condition to show that all suboptimal blocks have been removed. This condition is *sufficient rather than necessary* for selecting the correct exemplar: once $a^\star$ belongs to the surviving exemplar set, a final radius bound r and the simpler condition $\Delta_{\min} > 2r$ already guarantee the empirical-mean recommendation on the good event. Appendix B.4 proves this sharper recommendation-only statement and explains why it does not require the terminal geometric stopping condition. Neither statement establishes that the requisite refinement or sampling trajectory occurs with high probability without additional analysis.

What the theorem does and does not establish. Proposition 4 proves safety of block elimination on the confidence event, and Theorem 5 proves exact identification on $\mathcal{G} \cap \mathcal{C}_T$. It does not prove that the progress/terminal event $\mathcal{C}_T$ has high probability for every metric instance, does not establish instance-optimality, and does not rely on a TTPS-specific allocation theorem. Keeping these claims separate prevents the structural conditions from being mistaken for consequences of the sampling heuristic itself.

Implementable versus oracle calibration. TABLE does not require $\Delta_{\min}$, T_{rep}, or s_T as inputs. For conditionally σ^2-sub-Gaussian rewards, one implementable choice is to fix $\delta_0 \in (0,1)$ and use

$$\mathrm{rad}_T(n,\delta) = \sqrt{\frac{2\sigma^2 \log(2(T+1)/\delta)}{n}}$$

for $1 \le n \le T$, with $\mathrm{rad}_T(0,\delta) = +\infty$ and the observed count $n = N_z(t)$. Theorem 5 applies with δ_0 when its gap and trajectory conditions hold.

For rate interpretation only, fix $\xi \in (0,1)$ and define

$$A_T = 2(T+1)M_{\mathrm{conf}}, \qquad c_\Delta = \frac{(1-\xi)\Delta_{\min}^2}{8\sigma^2(2+\eta\bar{\kappa})^2}.$$

For a specified deterministic representation envelope, when $s_T \ge 1$ and $\delta_T := A_T e^{-c_\Delta s_T} \in (0,1)$, this oracle choice makes the gap condition hold algebraically and yields

$$\mathbb{P}(\{\hat{a}_T \ne a^\star\} \cap \mathcal{C}_T(\delta_T)) \le A_T e^{-c_\Delta s_T}.$$

Unconditionally, the corresponding bound is $\mathbb{P}(\hat{a}_T \ne a^\star) \le A_T e^{-c_\Delta s_T} + \mathbb{P}(\mathcal{C}_T(\delta_T)^c)$. This is an *oracle, event-based rate characterization*, not an implementable tuning prescription: the trajectory event and its representation envelope may depend on the chosen confidence parameter. The full derivation and fixed-δ_0 formulation are given in Appendix B.6.

4.3 Comparison with established fixed-budget bounds

For sorted positive arm gaps $\Delta_{(2)} \le \cdots \le \Delta_{(K)}$, let $H_2 = \max_{i \ge 2} i/\Delta_{(i)}^2$ and $\overline{\log}K = \frac{1}{2} + \sum_{i=2}^{K} i^{-1}$. The original, bounded-reward Successive Rejects (SR) and Sequential Halving (SH) theorems give, under their standard schedules (with K a power of two for the displayed SH constants) (Audibert & Bubeck, 2010; Karnin et al., 2013),

$$P_e^{\mathrm{SR}} \le \tfrac{K(K-1)}{2} e^{-(T-K)/(\overline{\log}K\, H_2)}, \qquad P_e^{\mathrm{SH}} \le 3\log_2 K\, e^{-T/(8H_2 \log_2 K)}.$$

VBR gives an unconditional, variance-sensitive arm-level rate involving H_σ (Faella et al., 2020). TABLE instead certifies $P_e \le \delta + \mathbb{P}(\mathcal{C}_T^c)$: its representation time, number of exemplars, and geometric cover enter the proof, but an unconditional exponential rate needs control of the *additional* trajectory-failure term. We claim neither a tighter bound nor a smaller exponent than SR, SH, or VBR. SBE uses side-observation assumptions, and TBE already studies recursive block elimination. Appendix B.7 compares the exact statements, assumptions, limiting cases, and lack of a matching lower bound.

5 Experiments

We evaluate TABLE on two fixed-budget BAI tasks with finite, metric-structured action sets. The Bernoulli benchmark provides a controlled Lipschitz instance in which the relationship between geometry and reward quality is known. The dynamic-pricing benchmark tests the same algorithmic mechanism on discretized price pairs and a different, unbounded reward distribution. The experiments assess identification error; they do not by themselves establish the trajectory conditions required by the theoretical identification guarantee.

Bernoulli metric bandit: The first task has $K = 1000$ candidate arms in $[1,10]^3$ with Euclidean distance. For an arm x, a Bernoulli success probability is obtained from a sigmoid transform of its distance to a latent optimal location. This produces locally smooth mean rewards, as in Lipschitz and metric bandits (Magureanu et al., 2014; Bubeck et al., 2011b). The objective is to identify the best *candidate arm*, not an optimum over the underlying continuum. The reward construction admits the verified Lipschitz bound $L = 1/4$ for the parameters used here; see Appendix C.2.

Dynamic-pricing bandit: The second task has $K = 1600$ candidate price pairs for normal and express delivery. At each pull, 500 customers choose between the two variants and an outside option according to a multinomial-logit (MNL) model, and the learner observes the corresponding noisy batch revenue. Nearby price pairs have related choice probabilities, making the finite price grid a structured arm set (Javanmard & Nazerzadeh, 2019; Parambath et al., 2025). Multiplicative LogNormal reward noise is unbounded and is *not* sub-Gaussian, so the sub-Gaussian corollary is not asserted here. TABLE uses $L = 1$ as an implementation parameter. Full specifications are in Appendix C.3.

Baselines, protocol, and evaluation. We compare against SH (Karnin et al., 2013), SR (Audibert & Bubeck, 2010), VBR (Faella et al., 2020), SBE (Shahrampour et al., 2017), and TBE (Parambath et al., 2025). TABLE, SBE, and TBE use 16 initial KMeans blocks. For arm-level reward estimation in the low-budget regimes, pull allocations are proportional to the available budget and are charged to T; no separate free initialization data are assumed. In our SBE implementation, each block is sampled through its central arm, following the central-representative approach. All methods use the same definition of budget: one arm pull costs one unit. In pricing, a pull corresponds to one decision on a batch of 500 customers, not 500 separate bandit pulls.

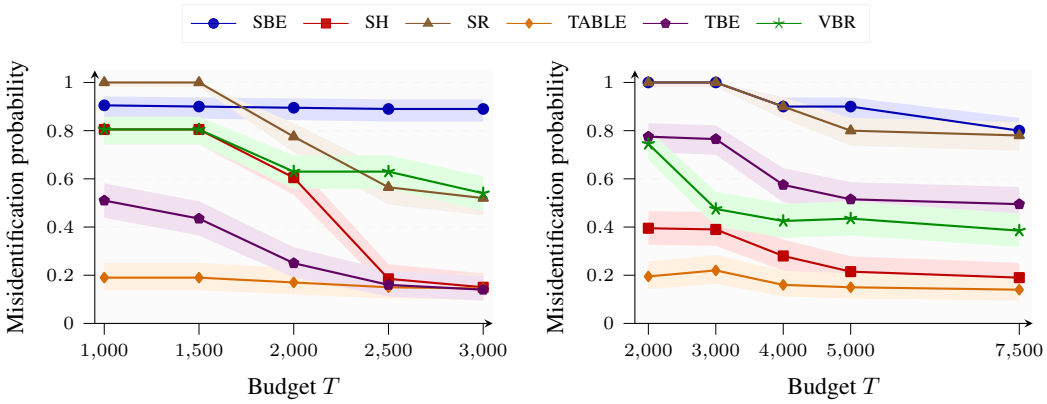
For Bernoulli exploration TABLE uses Beta sampling distributions, while pricing uses LogNormal-based sampling distributions; these govern exploration and are distinct from the confidence radii used in elimination. For confidence-based block elimination, TABLE uses an empirical-Bernstein radius with confidence parameter $\delta_0 = 0.05$ in both benchmarks; the radius uses the observed exemplar counts and empirical reward variability. We estimate the misidentification probability from 200 independent runs at every main-curve budget and hyperparameter-sensitivity configuration. Each plotted point is the fraction of incorrect recommendations. The experiments were run on an ArchLinux CPU server with 64 GB RAM; the original full experimental suite took approximately five hours.

Empirical performance. Figure 1 compares the error rates over five increasing budgets in each benchmark. TABLE obtains the lowest reported Bernoulli point estimate at the first four budgets; at $T = 3000$, TABLE (0.145), TBE (0.140), and SH (0.150) achieve similar observed error rates. TABLE has the lowest reported pricing point estimate at all five budgets. In both benchmarks, the TABLE error rates at the smallest budget are 0.190 and 0.195, respectively. These findings are consistent with the intended advantage of exemplar-based block elimination under restricted budgets, but the experiments do not isolate the contribution of every algorithmic component. Shaded regions show pointwise 95% Clopper–Pearson binomial intervals, interpreted under an independent, identically distributed replication protocol.

Sensitivity to geometric resolution. Figure 2 varies the number of initial blocks c_0 and the number of exemplars per node b. Each displayed configuration was evaluated over 200 independent runs. Increasing c_0 and b generally lowers Bernoulli error across the tested ranges, whereas the pricing errors are non-monotone: in the c_0 sweep the minimum point estimate occurs at $c_0 = 16$ (0.115), and in the b sweep it occurs at $b = 2$ (0.100). These patterns are consistent with a geometry–statistics trade-off: finer representation can reduce covering error but increases the number of exemplars sharing the budget. They are sensitivity results, not ablations that identify the individual effects of balancing, top-two sampling, or refinement. For the b sweep, Bernoulli uses $T = 1000$ and $c_0 = 16$, while pricing uses $T = 2000$ and $c_0 = 16$. For the c_0 sweep, Bernoulli uses $T = 1000$ and $b = 3$, while pricing uses $T = 2000$ and $b = 4$. All four sweeps use 200 independent runs per setting. Runtime complexity is analyzed in Appendix C.5.

5.1 Limitations and future work

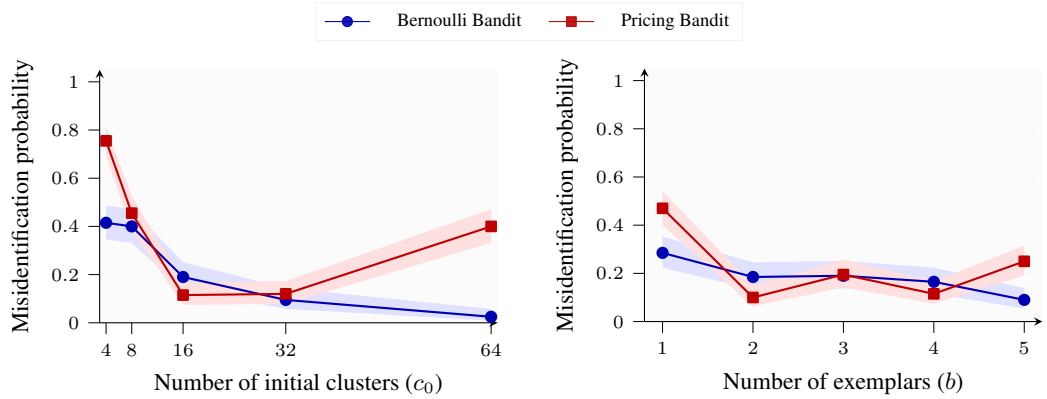
The central theoretical limitation is that we do not bound the probability of the favorable execution event $\mathcal{C}_T$ for arbitrary instances; without such a bound, the oracle exponential term is not a stand-alone unconditional rate. Algorithmically, the number of exemplars b is fixed throughout the run. This is convenient algorithmically, but may be suboptimal in heterogeneous spaces, where some nodes require only a coarse representation while others benefit from a richer local set of exemplars. Future work could therefore consider adaptive exemplar selection, with b varying across nodes and over time according to local complexity or uncertainty. A further limitation is the reliance on a known or chosen Lipschitz constant L, which enters both the confidence scores and the split criterion. Since misspecifying L can make the method either too conservative or too aggressive, a natural next step is to develop data-adaptive or locally estimated smoothness parameters, and more generally algorithms that remain stable under Lipschitz misspecification.



(a) Bernoulli metric bandit.

(b) Dynamic-pricing bandit.

Figure 1: Misidentification probability versus budget T for (a) Bernoulli ($K = 1000$) and (b) pricing ($K = 1600$). Each point uses 200 independent runs; shaded bands are pointwise 95% Clopper–Pearson binomial intervals.



(a) Effect of the number of initial clusters.

(b) Effect of number of exemplars.

Figure 2: Sensitivity of TABLE (200 independent runs per setting). (a) Vary c_0: Bernoulli $T = 1000, b = 3$; pricing $T = 2000, b = 4$. (b) Vary b: Bernoulli $T = 1000, c_0 = 16$; pricing $T = 2000, c_0 = 16$. Shaded regions are pointwise 95% Clopper–Pearson binomial intervals.

6 Conclusion

We proposed TABLE, a fixed-budget BAI algorithm for finite metric-structured arm sets that combines exemplar-based exploration, node-level elimination, and adaptive hierarchical refinement. The method exploits structure without exhaustively sampling all arms, and our analysis makes explicit how geometry, exemplar quality, and refinement progress affect identification. We proved safe block elimination on valid confidence events and an event-based fixed-budget identification guarantee under explicit representation, progress, and terminal conditions; the corresponding oracle exponential term does not alone control the probability of unfavorable trajectories. Experiments on two benchmarks show that TABLE performs particularly well in the low- and moderate-budget regime.

Reproducibility statement. The appendix contains the full pseudocode, formal event definitions and proofs, benchmark models, reported hyperparameters, sensitivity protocols, and complexity analysis. The submission should also include anonymized code and the exact experimental confidence-radius implementation for full reproducibility.

AI use statement. A generative language model was used to assist with language editing, structural reorganization, and presentation. The authors are responsible for independently verifying all technical claims, proofs, citations, experimental results, and final wording.

REFERENCES

Jean-Yves Audibert and Sébastien Bubeck. Best arm identification in multi-armed bandits. In *COLT-23th Conference on Learning Theory-2010*, pp. 13–p, 2010.

Maryam Aziz, Jesse Anderton, Emilie Kaufmann, and Javed Aslam. Pure exploration in infinitely-armed bandit models with fixed-confidence. In *Algorithmic Learning Theory*, pp. 3–24. PMLR, 2018.

Maryam Aziz, Emilie Kaufmann, and Marie-Karelle Riviere. On multi-armed bandit designs for dose-finding trials. *Journal of Machine Learning Research*, 22(14):1–38, 2021.

MohammadJavad Azizi, Branislav Kveton, and Mohammad Ghavamzadeh. Fixed-budget best-arm identification in structured bandits. In Lud De Raedt (ed.), *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pp. 2798–2804. International Joint Conferences on Artificial Intelligence Organization, July 2022.

Yikun Ban and Jingrui He. Generic outlier detection in multi-armed bandit. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 913–923, 2020.

Mickaël Binois, Victor Picheny, Patrick Taillandier, and Abderrahmane Habbal. The kalai-smorodinsky solution for many-objective bayesian optimization. *Journal of Machine Learning Research*, 21(150):1–42, 2020.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19):1832–1852, 2011a.

Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12(5), 2011b.

Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Conference on Learning Theory*, pp. 590–604. PMLR, 2016.

Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. *Advances in neural information processing systems*, 27, 2014.

Marco Faella, Alberto Finzi, and Luigi Sauro. Rapidly finding the best arm using variance. In *ECAI 2020*, pp. 2585–2591. IOS Press, 2020.

Stefan Falkner, Aaron Klein, and Frank Hutter. Bohb: Robust and efficient hyperparameter optimization at scale. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pp. 1437–1446, 2018.

Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. *Advances in Neural Information Processing Systems*, 25, 2012.

Xueping Gong, Wei You, and Jiheng Zhang. Minimax optimality in contextual dynamic pricing with general valuation models. *Operations Research*, 74(2):879–897, 2026.

Kevin Jamieson and Ameet Talwalkar. Non-stochastic best arm identification and hyperparameter optimization. In *Artificial intelligence and statistics*, pp. 240–248. PMLR, 2016.

Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. *Journal of Machine Learning Research*, 20(9):1–49, 2019.

Marc Jourdan, Rémy Degenne, and Emilie Kaufmann. An $varepsilon$-best-arm identification algorithm for fixed-confidence and beyond. *Advances in Neural Information Processing Systems*, 36:16578–16649, 2023.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning*, pp. 1238–1246. PMLR, 2013.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pp. 681–690, 2008.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Bandits and experts in metric spaces. *Journal of the ACM (JACM)*, 66(4):1–77, 2019.

Tomás Kocák and Aurélien Garivier. Epsilon best arm identification in spectral bandits. In *IJCAI*, pp. 2636–2642, 2021.

Daniel LeJeune, Gautam Dasarathy, and Richard Baraniuk. Thresholding graph bandits with grapl. In *International Conference on Artificial Intelligence and Statistics*, pp. 2476–2485. PMLR, 2020.

Stefan Magureanu, Richard Combes, and Alexandre Proutiere. Lipschitz bandits: Regret lower bound and optimal algorithms. In *Conference on Learning Theory*, pp. 975–999. PMLR, 2014.

Shameem Puthiya Parambath, Christos Anagnostopoulos, and Saleh Abdullah M Alfahad. Thompson sampling-based recursive block elimination for dynamic assignment under limited budget in pure-exploration. *Data Mining and Knowledge Discovery*, 39(1):7, 2025.

Chara Podimata and Alex Slivkins. Adaptive discretization for adversarial lipschitz bandits. In *Conference on Learning Theory*, pp. 3788–3805. PMLR, 2021.

Geovani Rizk, Albert Thomas, Igor Colin, Rida Laraki, and Yann Chevaleyre. Best arm identification in graphical bilinear bandits. In *International Conference on Machine Learning*, pp. 9010–9019. PMLR, 2021.

Shahin Shahrampour, Mohammad Noshad, and Vahid Tarokh. On sequential elimination algorithms for best-arm identification in multi-armed bandits. *IEEE Transactions on Signal Processing*, 65 (16):4281–4292, 2017.

Andrea Tirinzoni and Rémy Degenne. On elimination strategies for bandit fixed-confidence identification. In *Advances in Neural Information Processing Systems*, volume 35, pp. 18586–18598. Curran Associates, Inc., 2022.

Po-An Wang, Ruo-Chun Tzeng, and Alexandre Proutiere. Best arm identification with fixed budget: A large deviation perspective. *Advances in Neural Information Processing Systems*, 36:16804–16815, 2023.

Yinglun Zhu, Sumeet Katariya, and Robert Nowak. Robust outlier arm identification. In *International Conference on Machine Learning*, pp. 11566–11575. PMLR, 2020.

Honglei Zhuang, Chi Wang, and Yifan Wang. Identifying outlier arms in multi-armed bandit. *Advances in Neural Information Processing Systems*, 30, 2017.

A Complete TABLE Pseudocode and Operational Conventions

The pseudocode below makes the sampling priority, ladder gate, time indexing, and split-budget guard explicit. All pull counts and empirical means at time t are based on pulls completed before round t; a check performed immediately after round t is indexed by $t+1$. If an arm ceases to be an exemplar and later becomes one again, its historical observations and sampling distribution are *retained*, not reset. Exemplar selection is reward-independent, with any geometric random seed fixed in advance. We assume the budget can pay for the initial exemplars, $T \geq |\cup_{u\in\mathcal{U}(1)} \mathcal{Z}(u)|$, so the final empirical recommendation is defined.

Algorithm 2 TABLE: full fixed-budget control flow

Require: $\mathcal{A}, d, T, L, c_0, b, H, n_{\min}, \eta, n_0, \beta, \delta$; confidence routine rad_T; geometric split and exemplar rules.
Build c_0 nonempty initial nodes; choose $\mathcal{Z}(u)$ with $1 \leq |\mathcal{Z}(u)| \leq b$; compute $\mathrm{rds}(u), \kappa(u)$.
$\mathcal{U} \leftarrow$ initial nodes; $U_{\max} \leftarrow c_0 2^H$; $m \leftarrow 0$; $n_{-1} \leftarrow 0$; $\hat{\tau}_0 \leftarrow 1$.
Initialize counts, empirical means, and sampling distributions for possible exemplars; enqueue each initially unobserved exemplar once in FIFO queue $\mathcal{Q}$.
for $t = 1, \ldots, T$ **do**
  if $\mathcal{Q} \neq \varnothing$ **then**
    $a_t \leftarrow \mathrm{pop}(\mathcal{Q})$.
  else if $t \geq \hat{\tau}_m$ and $\min_{u\in\mathcal{U}, z\in\mathcal{Z}(u)} N_z(t) < n_m$ **then**
    $a_t \leftarrow$ an active exemplar with minimum count (fixed tie-breaking).
  else
    Draw once $\theta_z \sim \Pi_z(t)$ for every active exemplar z.
    Select $u_t \in \arg\max_{u\in\mathcal{U}} \frac{1}{|\mathcal{Z}(u)|} \sum_{z\in\mathcal{Z}(u)} \theta_z$.
    if $|\mathcal{Z}(u_t)| = 1$ **then**
      $a_t \leftarrow$ the unique exemplar of u_t.
    else
      Let $z^{(1)}, z^{(2)}$ have the largest and second-largest θ_z in $\mathcal{Z}(u_t)$.
      $a_t \leftarrow z^{(1)}$ with probability β, else $a_t \leftarrow z^{(2)}$.
    end if
  end if
  Pull a_t once; update its stored count, empirical mean, and sampling distribution.
  if $\min_{u\in\mathcal{U}, z\in\mathcal{Z}(u)} N_z(t+1) \geq n_m$ **then**
    CheckAndRefine$(t+1, \mathrm{true})$.
  end if
end for
CheckAndRefine$(T+1, \mathrm{false})$.
return $\hat{a}_T \in \arg\max_{z\in\cup_{u\in\mathcal{U}}\mathcal{Z}(u)} \hat{\mu}_z(T+1)$, with fixed tie-breaking.

The guard $|\mathcal{Q}| + |\mathcal{J}_u| \leq T - (t-1)$ is evaluated at check time t *after* round $t-1$, so its right-hand side counts the remaining pulls. At a final check $t = T+1$ splitting is disabled. A child whose exemplar has never been sampled has radius $+\infty$, preventing an immediate confidence-based split of that child until it has received observations. The repeat loop may produce several geometric splits in a single check, while m increases only once. Consequently, the number of optimal-path splits need not equal the ladder index; the progress condition below uses *optimal-path milestones* rather than assuming a one-to-one correspondence.

Statistical and geometric bookkeeping. A deterministic farthest-point rule may choose up to b representatives to reduce the largest within-node distance to its nearest exemplar. Its exact optimality is not needed: as long as $\mathcal{Z}(u) \subseteq \mathcal{C}(u)$ is nonempty, $\kappa(u) \leq 2$. The known metric defines $\mathrm{rds}(u)$ and $\kappa(u)$ without accessing rewards. The sampling distributions $\Pi_z(t)$ control only exploration; eliminating a node is justified by the *separate* confidence procedure. The TTPS step above uses one draw per exemplar and must not be identified with canonical TTTS that can resample for a challenger.

Why balancing is a conditional progress mechanism. After its gate opens, balance mode pulls an active minimum-count exemplar, but adaptive splitting can create new exemplars and change n_m; competing sampling modes also consume budget. The update rule therefore motivates, but does not itself establish, a uniform deterministic upper bound on the time required to reach every

Algorithm 3 CHECKANDREFINE$(t, \text{allowSplit})$

Set $\alpha \leftarrow \delta/M_{\text{conf}}$ and $\text{rad}_z(t) \leftarrow \text{rad}_T(N_z(t), \alpha)$ for active exemplars.
for each $u \in \mathcal{U}$ **do**
  $\text{LCS}_u(t) \leftarrow \max_{z \in \mathcal{Z}(u)}(\hat{\mu}_z(t) - \text{rad}_z(t)) - L\kappa(u)\text{rds}(u)$.
  $\text{UCS}_u(t) \leftarrow \max_{z \in \mathcal{Z}(u)}(\hat{\mu}_z(t) + \text{rad}_z(t)) + L\kappa(u)\text{rds}(u)$.
end for
$B(t) \leftarrow \max_{v \in \mathcal{U}} \text{LCS}_v(t)$; simultaneously remove all u with $\text{UCS}_u(t) < B(t)$.
if allowSplit $=$ false **then**
  return.
end if
split $\leftarrow$ false.
repeat
  localSplit $\leftarrow$ false; snapshot the active nodes $\mathcal{U}_{\text{snap}} \leftarrow \mathcal{U}$.
  for each $u \in \mathcal{U}_{\text{snap}}$ still active **do**
    if $|\mathcal{C}(u)| \geq 2n_{\min}$ and $h(u) < H$ and $L\text{rds}(u) \geq \eta \max_{z \in \mathcal{Z}(u)} \text{rad}_z(t)$ **then**
      Construct two nonempty geometric child nodes and their exemplar sets, radii, and cover factors.
      $\mathcal{J}_u \leftarrow \{z \in \cup_{v \in \text{ch}(u)} \mathcal{Z}(v) : N_z(t) = 0,\ z \notin \mathcal{Q}\}$.
      if $|\mathcal{Q}| + |\mathcal{J}_u| \leq T - (t-1)$ **then**
        Replace u by its children; enqueue each $z \in \mathcal{J}_u$ exactly once.
        split $\leftarrow$ true; localSplit $\leftarrow$ true.
      end if
    end if
  end for
until localSplit $=$ false
if split $=$ true **then**
  $m \leftarrow m + 1$; $\hat{\tau}_m \leftarrow (t-1) + U_{\max} b(n_m - n_{m-1})$.
end if

ladder threshold. That quantitative upper bound is explicitly part of the favorable trajectory event in Appendix B.2.

B FORMAL CONDITIONS, FINITE-HORIZON PROOFS, AND INTERPRETATION

This appendix proves the *event-based* theorem in Section 4, with exactly the same confidence allocation $\alpha = \delta/M_{\text{conf}}$ and terminal radius $r_T(\delta) = \text{rad}_T(s_T, \delta/M_{\text{conf}})$. In particular, no ladder-dependent factor 2^{m+1} is introduced. Structural representability is distinguished from the random progress event, and the unconditional error probability retains the term $\mathbb{P}(\mathcal{C}_T^c)$ unless the favorable trajectory conditions hold almost surely.

B.1 MODEL, CONFIDENCE EVENTS, AND BLOCK SCORES

Let $\mathcal{A}$ be a finite metric arm set with a unique best arm $a^\star$, and assume $|\mu(a) - \mu(a')| \leq Ld(a, a')$ with a *valid* known upper bound L. Initial nodes and geometric splits form a disjoint partition of the currently retained arms; elimination can discard nodes but does not change the contents of a surviving node. Let $\mathcal{U}^-(t)$ denote active nodes *immediately before* the elimination at check time t, and $\mathcal{U}^+(t)$ the survivors *immediately after* that elimination, before optional splitting. In the main paper $\mathcal{U}(t)$ refers to the appropriate checked state, and in terminal conditions it means the final surviving state $\mathcal{U}^+(T+1)$.

A fixed reward-independent tree and exemplar rule are sufficient to define a deterministic reservoir $\mathcal{Z}_{\text{all}}$ of arms that could become exemplars. Choosing $\mathcal{Z}_{\text{all}} = \mathcal{A}$ and $M_{\text{conf}} = K$ requires no knowledge of the potential tree. The use of a *single* confidence allocation per arm is legitimate precisely because the underlying radius is uniform over all adaptive sample counts through time $T+1$.

Assumption 6 (Adaptive exemplar confidence)**.** *For every possible exemplar $z \in \mathcal{Z}_{\text{all}}$ and every $\alpha \in (0,1)$, a nonnegative routine $\text{rad}_T(n, \alpha)$ satisfies*

$$\mathbb{P}(\forall t \in \{1, \ldots, T+1\} :\ |\hat{\mu}_z(t) - \mu(z)| \leq \text{rad}_T(N_z(t), \alpha)) \geq 1 - \alpha.$$

We set $\mathrm{rad}_T(0,\alpha) = +\infty$ *and assume* $\mathrm{rad}_T(n,\alpha)$ *is nonincreasing in* n *and nonincreasing in* α*. The statement refers to the actual nonanticipating sampling process; ordinary fixed-sample intervals substituted at optional stopping times need not satisfy it.*

For a fixed analysis confidence parameter $\delta \in (0,1)$, set

$$\alpha = \frac{\delta}{M_{\mathrm{conf}}}, \qquad \mathrm{rad}_z(t) = \mathrm{rad}_T(N_z(t),\alpha).$$

There is no resetting of observations on re-selection of an arm. Let $\mathcal{T}_{\mathrm{chk}}$ be the (possibly random) set of all checks, including $T+1$, and define

$$\mathcal{G} = \left\{\forall t \in \mathcal{T}_{\mathrm{chk}},\ \forall u \in \mathcal{U}^-(t),\ \forall z \in \mathcal{Z}(u) : |\hat{\mu}_z(t) - \mu(z)| \le \mathrm{rad}_z(t)\right\}.$$

For any node define the optimum $\mu^\star(u) = \max_{a\in\mathcal{C}(u)} \mu(a)$ and gap $\Delta(u) = \mu(a^\star) - \mu^\star(u)$. Its covering factor is

$$\kappa(u)\mathrm{rds}(u) = \max_{a\in\mathcal{C}(u)} \min_{z\in\mathcal{Z}(u)} d(a,z), \qquad \kappa(u) \le \bar{\kappa} \le 2,$$

where for a zero-diameter node both sides of the first expression are zero. TABLE computes

$$\mathrm{LCS}_u(t) = \max_{z\in\mathcal{Z}(u)} (\hat{\mu}_z(t) - \mathrm{rad}_z(t)) - L\kappa(u)\mathrm{rds}(u), \tag{3}$$

$$\mathrm{UCS}_u(t) = \max_{z\in\mathcal{Z}(u)} (\hat{\mu}_z(t) + \mathrm{rad}_z(t)) + L\kappa(u)\mathrm{rds}(u), \tag{4}$$

$$B(t) = \max_{u\in\mathcal{U}^-(t)} \mathrm{LCS}_u(t). \tag{5}$$

Lemma 7 (Probability of the good event)**.** *Under Assumption 6,* $\mathbb{P}(\mathcal{G}) \ge 1-\delta$.

Proof. Let E_z be the per-exemplar event that the stated confidence bound holds simultaneously for every $t \le T+1$. By assumption, $\mathbb{P}(E_z^c) \le \delta/M_{\mathrm{conf}}$. Thus

$$\mathbb{P}\left(\bigcap_{z\in\mathcal{Z}_{\mathrm{all}}} E_z\right) \ge 1 - \sum_{z\in\mathcal{Z}_{\mathrm{all}}} \frac{\delta}{M_{\mathrm{conf}}} = 1-\delta.$$

On this intersection every possible exemplar is covered at every time, including all random checks. It is therefore a subset of $\mathcal{G}$. A second union bound over checks or ladder levels would be unnecessary and would weaken the bound. □

Lemma 8 (Validity of the node scores)**.** *On* $\mathcal{G}$*, for every node just before any check,* $\mathrm{LCS}_u(t) \le \mu^\star(u) \le \mathrm{UCS}_u(t)$.

Proof. For every $z \in \mathcal{Z}(u)$, $\hat{\mu}_z - \mathrm{rad}_z \le \mu(z) \le \hat{\mu}_z + \mathrm{rad}_z$ on $\mathcal{G}$. Since $z \in \mathcal{C}(u)$, $\max_z(\hat{\mu}_z - \mathrm{rad}_z) \le \mu^\star(u)$, and subtracting the nonnegative covering term proves the lower bound. Let $a_u \in \arg\max_{a\in\mathcal{C}(u)} \mu(a)$. Some nearest exemplar z_u satisfies $d(a_u, z_u) \le \kappa(u)\mathrm{rds}(u)$. Lipschitz continuity and confidence give

$$\mu^\star(u) \le \mu(z_u) + L\kappa(u)\mathrm{rds}(u) \le \hat{\mu}_{z_u}(t) + \mathrm{rad}_{z_u}(t) + L\kappa(u)\mathrm{rds}(u) \le \mathrm{UCS}_u(t).$$

No assertion about how accurately the representatives approximate the optimal *arm identity* is needed for this node-score validity. □

Lemma 9 (Safe elimination)**.** *On* $\mathcal{G}$*, TABLE never eliminates the currently active node containing* $a^\star$.

Proof. At any check before elimination, let $u^\star$ contain $a^\star$. By the preceding lemma, $B(t) \le \max_{v\in\mathcal{U}^-(t)} \mu^\star(v) = \mu(a^\star)$ while $\mathrm{UCS}_{u^\star}(t) \ge \mu^\star(u^\star) = \mu(a^\star)$. Thus $\mathrm{UCS}_{u^\star}(t) \ge B(t)$ and TABLE's strict-elimination test cannot remove it. Induction over checks applies because splitting replaces a node with a disjoint pair of children, exactly one containing $a^\star$. □

Lemma 10 (Nonempty survivors and a well-defined final recommendation)**.** *Assume each node has at least one exemplar, confidence radii are nonnegative, $n_0 \geq 1$, and the initial queue and split-acceptance guard obey the stated remaining-budget convention. At least one node survives every elimination check. At the final check, every surviving exemplar has at least one observed reward, so the empirical-mean recommendation is defined on every execution, even outside $\mathcal{G}$.*

Proof. Let v maximize LCS_v immediately before a check. Since $\mathrm{UCS}_v \geq \mathrm{LCS}_v = B(t)$, v cannot satisfy the strict elimination condition. Thus the surviving set is nonempty. The queue initially fits the budget, and the split guard preserves the invariant $|\mathcal{Q}| \leq T - (t-1)$ immediately after any check indexed by t. When the queue is nonempty, the next pull removes its first entry, reducing queue length and remaining budget by one. Every newly active exemplar with zero stored observations is queued, and the final check adds none. At time $T+1$ the invariant forces $\mathcal{Q} = \varnothing$, so no surviving exemplar can remain unobserved. Therefore the maximum of the final stored empirical means exists. □

B.2 Structural representation and favorable trajectories

The following event formalizes *sufficient* progress and terminal behavior; it is not claimed to hold for all runs. Indexing progress by the number of completed optimal-path splits avoids the incorrect assumption that each ladder increment corresponds to exactly one such split. Several geometric splits can occur at one check, and one ladder increment may instead result from a split elsewhere.

Assumption 11 (Reward-independent, persistent representability)**.** *The fixed partition tree and exemplar-selection rule admit a deterministic integer $M_{\mathrm{rep}} \in \{0, \dots, H\}$ such that, after M_{rep} optimal-path splits, the resulting node contains $a^\star$ as an exemplar, and* every subsequent optimal-path descendant also selects *$a^\star$. Thus representability persists if further splits occur. A sufficient instance is a depth-H tree whose leaves have at most b arms, with every arm chosen when a node has at most b arms: in that case one may take $M_{\mathrm{rep}} = H$ (or an earlier persistently representative depth). If the geometric random seed changes with rewards, this assumption must instead be checked conditional on that seed.*

Definition 12 (Deterministic representation envelope)**.** *Fix nonnegative* deterministic *integer bounds $q_0, \dots, q_{M_{\mathrm{rep}}-1}$ and put*

$$U_{\max} = c_0 2^H, \quad n_{-1} = 0, \quad c_j = 2U_{\max} b(n_j - n_{j-1}), \quad T_{\mathrm{rep}} = 1 + \sum_{j=0}^{M_{\mathrm{rep}}-1} (c_j + q_j).$$

By telescoping, $T_{\mathrm{rep}} = 1 + 2U_{\max} b\, n_{M_{\mathrm{rep}}-1} + \sum_{j=0}^{M_{\mathrm{rep}}-1} q_j$. For $M_{\mathrm{rep}} = 0$, define $T_{\mathrm{rep}} = 1$ and interpret both sums as empty. This expression is a conditional envelope *based on the progress bounds q_j, not a budget guarantee derived solely from the ladder rule. Define*

$$s_T = \left\lfloor \frac{(T - T_{\mathrm{rep}})_+}{2U_{\max} b} \right\rfloor, \qquad r_T(\delta) = \mathrm{rad}_T\left(s_T, \frac{\delta}{M_{\mathrm{conf}}}\right).$$

Both quantities are deterministic for a fixed instance, horizon, and chosen progress envelope. We require $s_T \geq 1$ when using a finite final-radius guarantee.

Definition 13 (Optimal-path progress event)**.** *Set $t_0 = 1$ before the first pull. For each stage $j = 0, \dots, M_{\mathrm{rep}} - 1$, require the existence of a regular check time g_j and a subsequent regular check time t_{j+1}, with the following three properties:*

- *(i)* Balancing/check progress: *$t_j \leq g_j \leq t_j + c_j$; at g_j all then-active exemplars have attained the* actual current *threshold $n_{m(g_j)}$.*
- *(ii)* Optimal-path split progress: *$g_j \leq t_{j+1} \leq g_j + q_j$; at t_{j+1} the active optimal node is split (or has already been split between these checkpoints) so that by t_{j+1} the optimal branch has accumulated at least $j+1$ splits. The accepted split must meet the size, depth, confidence, and remaining-budget guards.*
- *(iii) The optimal arm remains in an active node at all these checkpoints. When several optimal-path splits occur at one check, the corresponding t_j may coincide; their sample-round cost is zero.*

Let $\mathcal{P}_T$ denote the conjunction of these requirements and $t_{M_{\mathrm{rep}}} \le T+1$ (vacuous when $M_{\mathrm{rep}} = 0$). The constants c_j and q_j are upper bounds assumed on this event*; the min-count balancing rule does not prove them in every trajectory. In particular, we do not assume $m(g_j) = j$: unrelated splits may advance the ladder.*

Assumption 14 (Terminal no-split, balance, and geometry: trajectory conditions)**.** *Consider a run after the milestone progress described above. The terminal conditions are:*

(i) No-split: *no regular geometric split occurs after the deterministic time T_{rep}; in particular, the representative sets are fixed during the post-envelope portion of the run. The final check performs no splits by construction.*

(ii) Sampling balance: *every exemplar in the final surviving nodes has $N_z(T+1) \ge s_T$.*

(iii) Terminal split-stop: *every node surviving final elimination satisfies*

$$L\mathrm{rds}(u) < \eta \max_{z \in \mathcal{Z}(u)} \mathrm{rad}_z(T+1).$$

The last condition is not an invariant imposed by the algorithm at $T+1$, since the check there forbids splitting. Nodes blocked from splitting by depth or size can have positive radius and fail the inequality. The condition does hold for singleton nodes when $\eta > 0$ and the statistical radius is positive. These are additional sufficient terminal-state conditions; none should be silently deduced from the update rules.

Definition 15 (Favorable event)**.** *For a fixed δ and the selected deterministic representation envelope, define*

$$\mathcal{C}_T(\delta) = \mathcal{P}_T \cap \mathcal{N}_T \cap \mathcal{B}_T \cap \mathcal{S}_T,$$

where $\mathcal{N}_T, \mathcal{B}_T, \mathcal{S}_T$ denote, respectively, the no-split, sampling-balance, and terminal split-stop conditions in Assumption 14. When the confidence level is fixed and unambiguous we abbreviate this event as $\mathcal{C}_T$. Structural representability is a property of the fixed instance/tree, not a reward event. Crucially, $\mathcal{C}_T$ does not *include $\mathcal{G}$, and no independence between these events is assumed.*

Lemma 16 (Representation by the envelope)**.** *Under Assumption 11, on $\mathcal{G} \cap \mathcal{P}_T$ with $T \ge T_{\mathrm{rep}}$, the final optimal node contains $a^\star$ among its exemplars.*

Proof. By safe elimination the optimal branch is present at every regular check on $\mathcal{G}$. The stage definitions provide check times t_j for which the optimal branch has been split at least j times. Inductively,

$$t_{j+1} \le g_j + q_j \le t_j + c_j + q_j \quad \Longrightarrow \quad t_{M_{\mathrm{rep}}} \le 1 + \sum_{j=0}^{M_{\mathrm{rep}}-1} (c_j + q_j) = T_{\mathrm{rep}}.$$

Consequently the optimal branch has experienced at least M_{rep} splits by that time. Persistent representability ensures that $a^\star$ is an exemplar at this depth and after any additional optimal-path splits. The final no-split regime is consistent with this persistence but is not used to assume that the branch splits exactly once per ladder level. For $M_{\mathrm{rep}} = 0$, representability holds already in the initial partition and persists by assumption. □

Lemma 17 (Final statistical and geometric bounds)**.** *On $\mathcal{C}_T$ and for $s_T \ge 1$, every final surviving exemplar satisfies* $\mathrm{rad}_z(T+1) \le r_T(\delta)$*, and every final surviving node satisfies $L\kappa(u)\mathrm{rds}(u) \le \eta\bar{\kappa} r_T(\delta)$.*

Proof. Terminal sampling balance gives $N_z(T+1) \ge s_T$. Because the same $\alpha = \delta/M_{\mathrm{conf}}$ is used at every check, monotonicity in n immediately yields $\mathrm{rad}_z(T+1) \le \mathrm{rad}_T(s_T, \alpha) = r_T(\delta)$. The terminal split-stop inequality and $\kappa(u) \le \bar{\kappa}$ yield

$$L\kappa(u)\mathrm{rds}(u) < \eta\bar{\kappa} \max_{z \in \mathcal{Z}(u)} \mathrm{rad}_z(T+1) \le \eta\bar{\kappa} r_T(\delta).$$

No ladder union bound or factor $2^{\bar{m}+1}$ appears. □

B.3 STRUCTURED BLOCK CERTIFICATE AND MAIN THEOREM

Write $r = r_T(\delta)$ and $G = \eta\bar{\kappa}r$. For a node surviving the final elimination check, compare its score to the surviving optimal node. The following lemma is the key reason the main theorem retains a geometric factor even though the final recommendation is an exemplar.

Lemma 18 (Gap of a surviving suboptimal block)**.** *Suppose the optimal node survives the final check, $a^\star$ is one of its exemplars, every final exemplar has confidence radius at most r, and every final surviving node has $L\kappa(u)\mathrm{rds}(u) \le G$. On $\mathcal{G}$, if a distinct node u survives, then $\Delta(u) \le 4r + 2G$.*

Proof. Because u survives, $\mathrm{UCS}_u(T+1) \ge B(T+1) \ge \mathrm{LCS}_{u^\star}(T+1)$. On $\mathcal{G}$, each exemplar $z \in \mathcal{Z}(u)$ satisfies $\hat{\mu}_z + \mathrm{rad}_z \le \mu(z) + 2\mathrm{rad}_z \le \mu^\star(u) + 2r$. Hence

$$\mathrm{UCS}_u(T+1) \le \mu^\star(u) + 2r + G.$$

Since $a^\star$ is an exemplar of the optimal node, $\hat{\mu}_{a^\star} - \mathrm{rad}_{a^\star} \ge \mu(a^\star) - 2r$, and therefore

$$\mathrm{LCS}_{u^\star}(T+1) \ge \mu(a^\star) - 2r - G.$$

Combining the two inequalities yields $\mu(a^\star) - \mu^\star(u) \le 4r + 2G$. □

Lemma 19 (Gap of a wrong empirical recommendation)**.** *Suppose $a^\star$ is a final surviving exemplar and every final surviving exemplar has radius at most r. On $\mathcal{G}$, if the empirical-mean recommendation differs from $a^\star$, then $\mu(a^\star) - \mu(\hat{a}_T) \le 2r$.*

Proof. The final maximization is over *all* surviving exemplars, including $a^\star$. Thus $\hat{\mu}_{\hat{a}_T} \ge \hat{\mu}_{a^\star}$. Confidence implies

$$\mu(\hat{a}_T) + r \ge \hat{\mu}_{\hat{a}_T} \ge \hat{\mu}_{a^\star} \ge \mu(a^\star) - r,$$

which gives the claimed inequality. This argument does not require elimination of all suboptimal blocks. □

Theorem 20 (Structured fixed-horizon certification: event-based form)**.** *Fix a horizon T, $\delta \in (0,1)$, a deterministic representation envelope, and the structural condition in Assumption 11. Assume valid adaptive confidence and $T \ge T_{\mathrm{rep}}$, $s_T \ge 1$. If*

$$\Delta_{\min} > 2(2 + \eta\bar{\kappa})r_T(\delta),$$

then

$$\mathbb{P}\big(\{\hat{a}_T \ne a^\star\} \cap \mathcal{C}_T\big) \le \delta, \qquad \mathbb{P}(\hat{a}_T \ne a^\star) \le \delta + \mathbb{P}(\mathcal{C}_T^c).$$

In particular $\mathbb{P}(\hat{a}_T \ne a^\star) \le \delta$ if $\mathcal{C}_T$ holds almost surely. The theorem is identical to Theorem 5 in the main text.

Proof. Work on $\mathcal{G} \cap \mathcal{C}_T$. Safe elimination preserves the optimal node at every check. By Lemma 16, $a^\star$ is a final surviving exemplar. By Lemma 17, the confidence and geometric radii of all final survivors are at most $r = r_T(\delta)$ and $G = \eta\bar{\kappa}r$, respectively.

Suppose a distinct, suboptimal node survives the final check. Lemma 18 yields

$$\Delta(u) \le 4r + 2G = 2(2 + \eta\bar{\kappa})r.$$

Because the node cannot contain $a^\star$, every arm in it is suboptimal and $\Delta(u) \ge \Delta_{\min}$, contradicting the strict gap condition. Thus only the node containing $a^\star$ survives. If the returned exemplar nevertheless differs from $a^\star$, Lemma 19 implies $\Delta_{\min} \le 2r$, again contradicting the stated condition. Therefore $\hat{a}_T = a^\star$ on $\mathcal{G} \cap \mathcal{C}_T$.

The event inclusion $\{\hat{a}_T \ne a^\star\} \cap \mathcal{C}_T \subseteq \mathcal{G}^c$ and Lemma 7 give the first probability bound. Splitting the error event according to $\mathcal{C}_T$ yields the unconditional bound. We do not divide by $\mathbb{P}(\mathcal{C}_T)$ or treat the trajectory event as independent of the confidence event. □

B.4 SHARPER RECOMMENDATION-ONLY OBSERVATION

Proposition 21 (Direct final-recommendation certificate)**.** *Let $\mathcal{C}_T^{\mathrm{rec}}$ be the event that (i) $a^\star$ is a final surviving exemplar and (ii) every final surviving exemplar is sampled at least $s_T \geq 1$ times. Under valid adaptive confidence, if*

$$\Delta_{\min} > 2r_T(\delta),$$

then

$$\mathbb{P}\big(\{\hat{a}_T \neq a^\star\} \cap \mathcal{C}_T^{\mathrm{rec}}\big) \leq \delta, \qquad \mathbb{P}(\hat{a}_T \neq a^\star) \leq \delta + \mathbb{P}((\mathcal{C}_T^{\mathrm{rec}})^c).$$

The terminal split-stop and no-suboptimal-block conditions are not required for this statement. On $\mathcal{G}$, structural representability together with $\mathcal{P}_T$ implies the first part of $\mathcal{C}_T^{\mathrm{rec}}$, while the full event $\mathcal{C}_T$ implies both parts.

Proof. On $\mathcal{G} \cap \mathcal{C}_T^{\mathrm{rec}}$, monotonicity of the radius gives $\mathrm{rad}_z(T+1) \leq r_T(\delta)$ for every surviving exemplar. If $\hat{a}_T \neq a^\star$, Lemma 19 yields $\Delta_{\min} \leq 2r_T(\delta)$, a contradiction. The probability statements follow by the same event inclusion and union decomposition as above. □

This result explains the distinction between a *block-elimination certificate* and exact final recommendation. The main theorem's larger threshold $2(2+\eta\bar{\kappa})r_T$ certifies that suboptimal nodes have disappeared under terminal split-stop. The direct recommendation guarantee has a smaller threshold because once the best arm is actually represented, selecting the largest empirical mean is simply a comparison among sampled arms. Neither result establishes a high-probability bound for its own favorable trajectory event. In particular, $\mathcal{C}_T^{\mathrm{rec}}$ cannot be assumed likely merely because it is weaker than $\mathcal{C}_T$.

Why terminal split-stop can fail. A depth-limited leaf may contain two distinct arms but be prohibited from splitting at depth H. Its half-diameter remains positive; as samples accumulate, $\max_z \mathrm{rad}_z(t)$ tends to zero for well-sampled exemplars. Therefore the inequality $L\mathrm{rds}(u) < \eta \max_z \mathrm{rad}_z(T+1)$ eventually fails whenever $L\mathrm{rds}(u) > 0$. A fixed depth cap or a size threshold $|\mathcal{C}(u)| < 2n_{\min}$ does not imply terminal split-stop. This explains why the main theorem explicitly conditions on that geometric event and why the recommendation-only observation can be useful for non-singleton terminal leaves whose exemplar sets already contain $a^\star$.

B.5 WORKED CASES AND CHECKS ON THE FAVORABLE CONDITIONS

A capacity sanity check for representability. A complete binary forest with c_0 roots and maximum additional depth H has at most $c_0 2^H$ terminal leaves. If each leaf is required to contain at most b candidate arms, its total capacity is at most $c_0 2^H b$. This is a sufficient *global* covering construction, not the definition of instance-specific representability of $a^\star$. In the Bernoulli bandit settings, under the reported default values $c_0 = 16$, $H = 3$, and $b = 3$, that capacity is 384, below both $K = 1000$ and $K = 1600$. Hence that particular sufficient construction cannot be asserted for either entire experimental candidate set. An optimal branch could still become representative after fewer than H splits if the geometric selection happens to include $a^\star$; checking this requires the actual realized tree, exemplar choices, and best-arm identity. Eliminating other nodes does not turn the global capacity example into a guarantee for the originally fixed tree.

A non-vacuous class satisfying the terminal geometric condition. Consider a finite instance initialized with one singleton node per arm ($c_0 = K$), choose $H = 0$ and $b = 1$, and use a positive, finite radius and $\eta > 0$. Every arm, including $a^\star$, is an exemplar from the beginning ($M_{\mathrm{rep}} = 0$). No split is permitted and every node has $\mathrm{rds}(u) = 0$, so the no-split and strict terminal split-stop conditions hold automatically. If the initialization and least-sampled balancing phase provide at least $s_T = \lfloor (T-1)/(2K) \rfloor$ pulls to each final exemplar, then the terminal balance condition also holds. For example, this sample-count condition is enforced by an initial least-count allocation that reaches s_T before unrestricted exploration, provided its cost fits T. This is an example of a sufficient configuration, not a statement that the reported $c_0 = 16$ experiments satisfy it or that TABLE automatically enforces the sample-count condition for arbitrary parameter choices.

Why a fixed small depth may make split-stop unlikely at large budgets. If a final node contains two distinct arms, has $L > 0$, and is prevented from splitting by the depth or size guard, its half-diameter is positive. As sufficiently sampled representatives acquire shrinking confidence radii, $L\mathrm{rds}(u) < \eta \max_z \mathrm{rad}_z(T+1)$ fails eventually. Therefore, a proof of a general unconditional asymptotic rate would need either a refinement mechanism permitting vanishing-diameter terminal cells, a different geometric stopping certificate, or control of the probability of the appropriate favorable event. Merely increasing T at fixed H does not establish the required terminal condition.

B.6 Finite-horizon sub-Gaussian calibration and oracle rate

Suppose that observations from each arm are conditionally σ^2-sub-Gaussian and the algorithm's arm choices are nonanticipating. Let $Y_{z,n}$ be the nth observed reward of arm z, and let $\hat{\mu}_{z,n}$ average its first n observations. For a fixed arm and fixed n,

$$\mathbb{P}\left(|\hat{\mu}_{z,n} - \mu(z)| > \varepsilon\right) \le 2\exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right).$$

As the number of pulls of an arm is chosen adaptively, we first union-bound over every possible count $n \le T$ and only then evaluate at the observed count. Define

$$\mathrm{rad}_T(n,\alpha) = \begin{cases} \sqrt{\dfrac{2\sigma^2 \log(2(T+1)/\alpha)}{n}}, & 1 \le n \le T, \\ +\infty, & n = 0. \end{cases}$$

For every $1 \le n \le T$, the fixed-count failure probability at this radius is $\alpha/(T+1)$; hence

$$\mathbb{P}\left(\exists n \le T : \ |\hat{\mu}_{z,n} - \mu(z)| > \mathrm{rad}_T(n,\alpha)\right) \le \frac{T\alpha}{T+1} < \alpha.$$

It follows that the radius satisfies Assumption 6 for all realized adaptive counts $N_z(t) \le T$. This proof establishes coverage for the displayed sub-Gaussian radius, not for an unspecified empirical-Bernstein implementation. Choosing $\alpha = \delta/M_{\mathrm{conf}}$ yields

$$r_T(\delta) = \sqrt{\frac{2\sigma^2 \log(A_T/\delta)}{s_T}}, \qquad A_T = 2(T+1)M_{\mathrm{conf}}, \quad s_T \ge 1.$$

The exponential confidence factor is A_T; there is *no* factor depending on the number of ladder levels because the individual-arm bound is already uniform over time.

Corollary 22 (Fixed-parameter confidence calibration)**.** *Choose any fixed $\delta_0 \in (0,1)$ before execution and run TABLE with the preceding radius and allocation $\delta_0/M_{\mathrm{conf}}$. For $T \ge T_{\mathrm{rep}}$, $s_T \ge 1$, and*

$$s_T > \frac{8\sigma^2(2+\eta\bar{\kappa})^2}{\Delta_{\min}^2} \log\left(\frac{A_T}{\delta_0}\right),$$

we have $\mathbb{P}(\{\hat{a}_T \ne a^\star\} \cap \mathcal{C}_T(\delta_0)) \le \delta_0$ and $\mathbb{P}(\hat{a}_T \ne a^\star) \le \delta_0 + \mathbb{P}(\mathcal{C}_T(\delta_0)^c)$. The run does not require $\Delta_{\min}$ or T_{rep} as an input; these quantities appear only in the post hoc sufficient condition.

Proof. The displayed inequality is equivalent to $\Delta_{\min} > 2(2+\eta\bar{\kappa})r_T(\delta_0)$ after substituting the sub-Gaussian radius and squaring. Theorem 20 gives the conclusion. □

Corollary 23 (Oracle event-based exponential characterization)**.** *Fix $\xi \in (0,1)$ and the deterministic representation/progress envelope for the configured algorithm, and define*

$$c_\Delta = \frac{(1-\xi)\Delta_{\min}^2}{8\sigma^2(2+\eta\bar{\kappa})^2}, \qquad \delta_T = A_T e^{-c_\Delta s_T}.$$

If $T \ge T_{\mathrm{rep}}$, $s_T \ge 1$, and $\delta_T \in (0,1)$, then for the TABLE run configured with δ_T,

$$\mathbb{P}\big(\{\hat{a}_T \ne a^\star\} \cap \mathcal{C}_T(\delta_T)\big) \le A_T e^{-c_\Delta s_T},$$

and

$$\mathbb{P}(\hat{a}_T \ne a^\star) \le A_T e^{-c_\Delta s_T} + \mathbb{P}(\mathcal{C}_T(\delta_T)^c).$$

If the favorable event holds almost surely for the considered instance–algorithm class, the second term vanishes. Otherwise no unconditional exponential rate follows from this corollary alone. The dependence of the progress envelope on the configured confidence level must be accounted for when interpreting s_T; δ_T is an oracle algebraic characterization, not an implementable tuning rule.

Proof. For the configured oracle radius,

$$r_T(\delta_T) = \sqrt{\frac{2\sigma^2 \log(A_T/\delta_T)}{s_T}} = \sqrt{2\sigma^2 c_\Delta} = \frac{\sqrt{1-\xi}\,\Delta_{\min}}{2(2+\eta\bar{\kappa})}.$$

Thus $2(2+\eta\bar{\kappa})r_T(\delta_T) = \sqrt{1-\xi}\,\Delta_{\min} < \Delta_{\min}$, which verifies the strict gap condition. The event-based and unconditional conclusions follow from Theorem 20. The use of a predetermined progress envelope is essential to this algebra; if T_{rep} changes with the proposed tuning, the definition of δ_T can be implicit and should not be presented as a practical selection procedure. □

What "exponential" means here. Since $s_T = \lfloor (T - T_{\mathrm{rep}})_+ / (2U_{\max} b) \rfloor$, the *intersection probability* decays at the displayed exponential rate in s_T, up to the polynomial prefactor A_T. An asymptotic rate in T follows only along classes of horizons for which the deterministic envelope is controlled (for example, a fixed T_{rep}, $U_{\max}$ and b), and it does not transfer to the unconditional error unless $\mathbb{P}(\mathcal{C}_T^c)$ is separately controlled. No instance-optimal or matching-lower-bound claim is made.

B.7 Theoretical comparison with existing fixed-budget results

Scope and conventions. The goal of this comparison is to identify what each published theorem controls, rather than to rank algorithms by constants from different statistical models. Write $P_e(T) = \mathbb{P}(\hat{a}_T \neq a^\star)$, order the means as $\mu_{(1)} > \mu_{(2)} \geq \cdots \geq \mu_{(K)}$, and put $\Delta_{(i)} = \mu_{(1)} - \mu_{(i)}$ for $i \geq 2$. The standard unstructured hardness quantities are

$$H_1 = \sum_{i=2}^{K} \frac{1}{\Delta_{(i)}^2}, \qquad H_2 = \max_{2 \leq i \leq K} \frac{i}{\Delta_{(i)}^2}, \qquad \overline{\log} K = \frac{1}{2} + \sum_{i=2}^{K} \frac{1}{i}.$$

They use *all* suboptimal arm gaps, unlike the deliberately conservative global $\Delta_{\min} = \Delta_{(2)}$ in our sufficient certificate. The symbol $\overline{\log} K$ is the harmonic factor in SR, not an unspecified logarithmic-order term. We reproduce the following bounds as statements from the cited literature, under their original reward assumptions and sampling schedules; they are not new results of this paper.

Successive Rejects (SR): unconditional arm-wise elimination. For independent $[0,1]$ rewards, the standard SR analysis (Audibert & Bubeck, 2010) gives, for its specified schedule with $T > K$,

$$P_e^{\mathrm{SR}}(T) \leq \frac{K(K-1)}{2} \exp\left[-\frac{T-K}{\overline{\log} K \, H_2} \right]. \tag{6}$$

SR samples all surviving arms phase by phase and rejects one arm at a time. Its probability statement applies to the entire run; it does not condition on a separate representability or refinement event. The factor $T - K$ reflects the standard schedule's initial arm-level allocation. In particular, substituting $T = K$ gives only a vacuous numerical bound, rather than a meaningful assurance for an arbitrary reduced-initialization adaptation. TABLE avoids the requirement to explicitly estimate all K arms in order to *construct its block scores*, but pays for this choice with geometric-transfer error and an unproved favorable-trajectory probability. This comparison does not imply an improved worst-case sample complexity for TABLE.

Sequential Halving (SH): an arm-gap rate with a different schedule. SH distributes pulls uniformly among remaining arms over halving rounds. For bounded rewards and K a power of two, Theorem 4.1 of Karnin et al. (2013) states

$$P_e^{\mathrm{SH}}(T) \leq 3 \log_2 K \exp\left[-\frac{T}{8 H_2 \log_2 K} \right]. \tag{7}$$

The original paper discusses extensions to other arm counts. Equation equation 7 is a fixed-budget, unconditional claim for the original halving schedule. A low-budget implementation with modified per-arm allocations, rounding, or unobserved arms need not inherit these exact constants. Unlike TABLE, SH does not use the metric to certify arms that have never been directly sampled. SH's logarithmic number of phases and SR's $K - 1$ phases also explain why their displayed prefactors differ; the prefactors alone do not determine actual error rates.

Table 3: Scope of the compared theoretical statements. Bounds refer to the cited papers' original procedures and assumptions, not to our low-budget baseline adaptations.

Method	Information and allocation	Guarantee / controlling quantity	Crucial distinction
SR	Direct observations; reject one arm per phase	Unconditional P_e via H_2 and harmonic phase factor	No metric transfer or trajectory event
SH	Direct observations; uniform pulls in halving rounds	Unconditional P_e via $H_2 \log K$	Original schedule needed for quoted bound
VBR	Direct observations; variance-informed arm confidence	Unconditional P_e via H_σ	Variance adaptation is not unique to TABLE
TABLE	Exemplars; metric transfer; adaptive nodes	$P_e \le \delta + \mathbb{P}(\mathcal{C}_T^c)$; oracle rate on favorable intersection	Requires representability and progress/terminal conditions

Variance-Based Rejects (VBR): variance adaptivity is established elsewhere. The VBR result of Faella et al. (2020) allows arm-dependent sub-Gaussian reward scales σ_i^2 and uses confidence scores informed by empirical variance. Its stated complexity is

$$H_\sigma = \max_{2 \le i \le K} \frac{\sigma_{(1)}^2 + \sigma_{(i)}^2}{\Delta_{(i)}^2},$$

and its Theorem 1 gives the unconditional expression

$$P_e^{\mathrm{VBR}}(T) \le \psi(K) \exp\left[-\frac{\phi(T,K)}{2H_\sigma}\right], \qquad \psi(K) = \frac{(K-1)(K+2)}{2}, \tag{8}$$

with $\phi(T,K) = (T-K)/[(K+1)\overline{\log} K]$ as defined in that work. The variances/scales in H_σ belong to the original arm-level model; the VBR theorem does not supply TABLE with a calibrated block-level empirical-Bernstein radius. Conversely, our generic safe-elimination argument can accept a *valid* confidence routine, but the explicit oracle rate proved here uses a common sub-Gaussian proxy σ^2 and is *not* a variance-adaptive improvement on equation 8. Both the sub-Gaussian assumption and correct confidence implementation must be checked before importing either theorem to a new reward model.

TABLE: precisely which probability is exponentially controlled? For a *configured* finite-horizon routine satisfying the confidence assumption, Theorem 20 gives

$$P_e^{\mathrm{TABLE}}(T) \le \delta + \mathbb{P}(\mathcal{C}_T(\delta)^c), \quad \Delta_{\min} > 2(2 + \eta\bar{\kappa}) r_T(\delta). \tag{9}$$

The sub-Gaussian oracle inversion yields $\delta_T = A_T \exp(-c_\Delta s_T)$, with $A_T = 2(T+1)M_{\mathrm{conf}}$ and $s_T = \lfloor (T - T_{\mathrm{rep}})_+ / (2U_{\max} b) \rfloor$, but only

$$\mathbb{P}\big(\{\hat{a}_T \ne a^\star\} \cap \mathcal{C}_T(\delta_T)\big) \le \delta_T.$$

Without an independently established decay rate for $\mathbb{P}(\mathcal{C}_T(\delta_T)^c)$, *this is not an unconditional exponential error bound*. In particular, T_{rep} is a deterministic bound arising from stipulated progress milestones, not a sample-complexity theorem derived from the TTPS heuristic. The event may itself change with the confidence tuning. By comparison, equation 6–equation 8 directly bound the total error under their assumptions and standard algorithms. Comparing just their exponential coefficients to c_Δ would discard the unresolved remainder term and would therefore be invalid.

Why the complexity measures differ. An arm-wise elimination proof monitors direct discrimination between $a^\star$ and individual competitors, producing quantities such as H_1, H_2, or H_σ. TABLE instead first bounds a node optimum using up to b observed exemplars:

$$\mu^\star(u) \le \max_{z \in \mathcal{Z}(u)} \big(\hat{\mu}_z + \mathrm{rad}_z\big) + L\kappa(u)\mathrm{rds}(u).$$

This inequality applies even when the best arm within u has not yet been sampled. A separate representability argument is then necessary because TABLE recommends an *exemplar*, not an arbitrary unsampled arm. The post-representation sample scale s_T, cover factor $\bar{\kappa}$, and trajectory event $\mathcal{C}_T$ account for those different proof stages. The coarse factor $U_{\max} b$ bounds the number of simultaneous representatives; it is not an asserted effective dimension or an established replacement for H_2. Nor does the current exponent prove a favorable dependence on K for all metric instances.

Worked limiting case: recovering an unstructured arm-wise situation. Take $c_0 = K$ singleton nodes, $H = 0$, and $b = 1$. Every arm is immediately represented, $\kappa(u)\mathrm{rds}(u) = 0$, and $U_{\max} b = K$. The representation cost disappears apart from the time-index convention, and $s_T = \lfloor (T-1)/(2K) \rfloor$. Conditional on the stated final sample-count requirement, the sharper direct-recommendation proposition yields the sufficient gap $\Delta_{\min} > 2r_T(\delta)$; its sub-Gaussian substitution has the schematic rate

$$\mathbb{P}\big(\{\hat{a}_T \neq a^\star\} \cap \mathcal{C}_T^{\mathrm{rec}}\big) \lesssim TK \exp\left(-c\,\frac{\Delta_{\min}^2 T}{\sigma^2 K}\right)$$

for a positive numerical constant c when the corresponding fixed deterministic balance envelope holds. This shows that the certificate does not automatically beat classical H_2-dependent bounds even in the absence of geometric approximation. It is an algebraic limiting illustration, not a new guarantee for the actual sampling rule without that balance event. A different metric or representative rule may substantially improve the observed trajectory, but the present proof does not quantify such improvement unconditionally.

Other structured comparisons: side observations and recursion. The side-observation part of Shahrampour et al. (2017) establishes improvements when pulling an arm reveals rewards of related arms. That is an additional feedback model, not a consequence of having a metric: TABLE observes only the chosen exemplar's reward and transfers *bounds*, not extra observations. Our experiment uses a central-arm SBE adaptation and should not be interpreted as verifying the side-observation theorem. The more closely related TBE procedure of Parambath et al. (2025) already explores posterior-based, recursive block elimination on discretized action spaces. Its article explicitly places a theoretical analysis of that algorithm among its future directions. Our distinct analytical element is a finite-set Lipschitz exemplar-cover certificate with clearly stated conditional progress requirements; neither recursive blocking nor posterior-guided exploration is claimed as a first in itself. Restricting a continuous-space method to $\mathcal{A}$ is also possible, but a bound for continuum regret or approximate identification does not become an exact fixed-budget BAI theorem without reconciling the objective, sampling feedback, discretization error, and output rule.

Lower bounds, interpretation, and remaining limitations. The classical literature relates H_1 and H_2 to arm-wise identification hardness under its particular stochastic models (Audibert & Bubeck, 2010; Karnin et al., 2013). These lower bounds cannot be re-labelled as sharp lower bounds for TABLE's metric-structured class, nor does our theorem contradict them: the structured assumptions, information available to the learner, and performance statements differ. We give no matching lower bound and no order-optimality or dominance claim. The principal theoretical open question is to bound the chance of failing the optimal-path representation and terminal milestones from checkable geometric and stochastic assumptions. A sharper proof might bypass terminal split-stop using the direct recommendation certificate, but it would still need a representability and sampling-progress argument.

C Experimental Specifications and Reproducibility

C.1 What the plotted intervals and elimination radii measure

The misidentification indicator in replication i is $E_i = \mathbf{1}\{\hat{a}_T^{(i)} \neq a_i^\star\}$ and the displayed error estimate is $\hat{p} = \sum_{i=1}^{200} E_i/200$. Under independent, identically distributed replications with a common error probability p, the count $S = \sum_i E_i$ is binomial. For an observed count $S = k$, the two-sided pointwise Clopper–Pearson interval is

$$I_{0.95}(k) = \begin{cases} [0, \mathrm{Beta}^{-1}(0.975; 1, 200)], & k = 0, \\ [\mathrm{Beta}^{-1}(0.025; 200, 1), 1], & k = 200, \\ [\mathrm{Beta}^{-1}(0.025; k, 201-k),\ \mathrm{Beta}^{-1}(0.975; k+1, 200-k)], & 0 < k < 200. \end{cases}$$

Here $\mathrm{Beta}^{-1}(q; a, b)$ denotes the q-quantile of the Beta distribution with shape parameters a, b. These intervals quantify Monte Carlo uncertainty in the *reported error rate*. They are not the exemplar-level radii $\mathrm{rad}_z(t)$ used by TABLE to eliminate nodes and are not simultaneous confidence intervals across methods and budgets. If simulations instead use a prespecified heterogeneous, non-identically distributed group of instances, a pooled binomial interval need not have the claimed exact coverage.

The experimental elimination routine used empirical Bernstein with input $\delta_0 = 0.05$. That input does not determine the bound: the variance estimator, leading constants, treatment of small sample sizes, and adjustment across exemplars and adaptive check times all matter. A conventional empirical-Bernstein formula requires bounded observations, whereas pricing multiplies batch revenue by a very large LogNormal factor. Moreover, a radius expressed through the *current* empirical variance need not be pathwise nonincreasing as the sample count grows, unlike the deterministic sub-Gaussian radius used in the proofs. Accordingly, neither the experimental routine's simultaneous coverage nor the final-radius monotonicity assumption can be inferred from our experiments alone. The oracle and implementable theoretical bounds use a separately specified valid radius; they are not fitted to the plotted binomial intervals.

C.2 Bernoulli Metric Bandit Benchmark

Our first benchmark is a synthetic Bernoulli bandit on a metric arm space. The purpose of this environment is to provide a controlled setting in which reward quality varies smoothly over the geometry of the arm set. This type of construction is standard when evaluating methods designed to exploit local regularity in large structured action spaces.

Action space. We work on the compact Euclidean domain

$$\mathcal{X} = [1, 10]^3 \subset \mathbb{R}^3,$$

equipped with the usual Euclidean metric

$$d(x, y) = \|x - y\|_2.$$

The learner itself does not operate on the full continuum, but on a finite candidate set

$$\mathcal{A} = \{a_1, \dots, a_K\} \subset \mathcal{X}, \qquad K = 1000,$$

where each arm is identified with its three-dimensional coordinate.

Mean-reward model. The Bernoulli success probability is generated from the distance to a latent optimum $x^\star \in \mathcal{X}$. Before clipping, define

$$\widetilde{\mu}(x) = \sigma\big(\alpha - \beta\|x - x^\star\|_2\big), \qquad \sigma(z) = \frac{1}{1 + e^{-z}}.$$

Here, $\alpha \in \mathbb{R}$ controls the overall reward level and $\beta > 0$ determines how quickly rewards decay away from the optimum. In all experiments, unless otherwise stated, we use

$$\alpha = 3, \qquad \beta = 1.$$

If $x^\star$ is not manually fixed, it is drawn uniformly from $\mathcal{X}$ using the experimental random seed. The original description does not say whether the candidate grid and latent optimum were fixed across the 200 repetitions; this should be specified from the released code, along with the seeds and grid-generation procedure.

Observed rewards. To prevent degenerate Bernoulli parameters, we use the clipped mean

$$\mu(x) = \min\{1 - 10^{-3},\, \max\{10^{-3}, \widetilde{\mu}(x)\}\}.$$

Then, whenever arm $a \in \mathcal{A}$ is selected, the observed reward is drawn as

$$Y_t \sim \mathrm{Bernoulli}(\mu(a)).$$

The corresponding best arm is therefore defined relative to the finite candidate set $\mathcal{A}$, namely

$$a^\star \in \arg\max_{a \in \mathcal{A}} \mu(a).$$

Why this benchmark is structured. This environment induces a smooth spatial reward geometry: arms that are close in Euclidean distance have similar expected rewards, while the mean decreases gradually as one moves away from the optimal region. The logistic derivative is at most $1/4$, the distance-to-$x^\star$ function is 1-Lipschitz, and clipping is nonexpansive. Thus μ is $\beta/4$-Lipschitz, establishing the reported choice $L = 1/4$ when $\beta = 1$. This makes the benchmark particularly appropriate for evaluating whether TABLE can benefit from local representative coverage, hierarchical partitioning, and adaptive refinement.

C.3 Dynamic Pricing Benchmark

Our second benchmark is a two-product dynamic pricing problem under a multinomial-logit (MNL) demand model. Compared with the synthetic Bernoulli setting, this environment is more application-driven and provides a structured discrete action space with a clear economic interpretation.

Action space. Each action is a price pair

$$a = (p_n, p_e) \in [10, 50] \times [30, 70],$$

where p_n and p_e denote the prices of the normal-delivery and express-delivery variants, respectively. To fit the fixed-budget best-arm identification setting, we discretise this two-dimensional price region and obtain a finite arm set

$$\mathcal{A} \subset [10, 50] \times [30, 70], \qquad |\mathcal{A}| = 1600.$$

The optimal arm is the price pair in $\mathcal{A}$ that maximises expected revenue.

Choice model. Customer behaviour is generated using an MNL model with an outside option. In every round, a batch of 500 customers independently chooses among three alternatives: no purchase, the normal-delivery option, and the express-delivery option. The corresponding latent utilities are

$$u_0 = 0, \qquad u_n = a_n - b_n p_n + \gamma_n p_e, \qquad u_e = a_e - b_e p_e + \gamma_e p_n.$$

The constants a_n and a_e are baseline preference parameters, b_n and b_e are own-price sensitivities, and γ_n, γ_e are cross-price coefficients capturing substitution between the two variants. In the default configuration we use

$$a_n = 2.0, \qquad b_n = 0.18, \qquad a_e = 2.3, \qquad b_e = 0.06,$$

and set

$$\gamma_n = \gamma_e = 0,$$

so that each utility depends only on its own price. Applying the softmax transformation to (u_0, u_n, u_e) gives the choice probabilities, from which we sample the realised counts

$$(c_0, c_n, c_e)$$

from a multinomial distribution.

Reward model. Given the realised customer counts, the noiseless revenue in round t is

$$R_t = p_n c_n + p_e c_e.$$

To introduce positive multiplicative noise with unit expectation before flooring, we multiply R_t by an independent LogNormal factor:

$$\xi_t \sim \mathrm{LogNormal}\left(-\tfrac{1}{2}\sigma^2, \sigma^2\right), \qquad \sigma = 0.15.$$

The strictly positive observed reward is defined as

$$Y_t = \max\{R_t \xi_t, \varepsilon\}, \qquad \varepsilon = 10^{-12},$$

which ensures strictly positive rewards. Since $\mathbb{E}[\xi_t] = 1$ and ξ_t is independent of the customer choices, the *unfloored* reward satisfies $\mathbb{E}[R_t \xi_t \mid a] = \mathbb{E}[R_t \mid a]$. Flooring can change the mean: the actual arm mean is

$$\mu(a) = \mathbb{E}[Y_t \mid a] = \mathbb{E}[R_t \mid a] + \mathbb{E}[(\varepsilon - R_t \xi_t)_+ \mid a],$$

where $\mathbb{E}[R_t \mid a] = 500[p_n P_n(a) + p_e P_e(a)]$. The floored observation is unbounded above and does not satisfy a standard sub-Gaussian or bounded empirical-Bernstein assumption. A valid finite-set Lipschitz bound on μ exists if the grid consists of distinct points, but the experimental choice $L_=1$ was chosen arbitrarily.

Why this benchmark is structured. This pricing environment induces meaningful local regularity in the action space. Nearby price vectors tend to generate similar utilities, similar purchase probabilities, and hence similar revenues. As a result, the discretised price grid forms a large but structured set of actions in which local geometry is informative. This makes the benchmark well suited to TABLE, whose design relies on representative-based exploration, adaptive elimination of weak regions, and recursive refinement of promising parts of the price space.

Table 4: TABLE implementation settings

Hyperparameter	**Value**
Number of initial clusters c_0	16
Maximum depth H	3
Default exemplar limit b	3 (pricing c_0 sweep: 4)
Minimum split size $n_{\min}$	8
Split aggressiveness η	0.7
Bernoulli Beta sampling parameter α	1
Bernoulli Beta sampling parameter β_{prior}	6
Pricing sampling parameter (reported μ)	0
Pricing sampling parameter (reported κ)	1
Other reported pricing sampling parameter α	2
Other reported pricing sampling parameter β	1
Lipschitz constant L (Bernoulli experiment)	$1/4$
Pricing implementation coefficient L	1 (arbitrary)
TTPS mixing probability β	0.7
Ladder base n_0	10
Confidence parameter δ_0	0.05
Empirical-Bernstein constants	2 and 3

C.4 TABLE Implementation Details

For TABLE, the initial partition of the arm set is constructed using KMeans clustering. In the Bernoulli benchmark we use $c_0 = 16$ initial clusters, while in the pricing benchmark the same clustering procedure is applied to the discretised price grid before the adaptive refinement stage begins. Within each active node u, TABLE selects a representative set $\mathcal{Z}(u) \subseteq \mathcal{C}(u)$ of size at most b using a greedy farthest-point heuristic. Starting from an initial seed arm in $\mathcal{C}(u)$, the procedure repeatedly adds the arm whose distance to the current representative set is largest, thereby encouraging the representatives to spread out across the node. This gives a computationally simple approximation to the k-centre objective and provides a good geometric cover of the node, which is important for both node-level scoring and the Lipschitz transfer term used in elimination. Whenever a node is refined, the split operation is implemented using a 2Means clustering routine, consistent with the canonical pseudo-code.

The remaining experimental choices are fixed across runs unless stated otherwise. For Bernoulli rewards, exploration uses a Beta sampling distribution; pricing uses a LogNormal-based sampling distribution. These exploration laws should not be conflated with elimination confidence bounds. The experimental confidence parameter was fixed at $\delta_0 = 0.05$, and an empirical-Bernstein-style radius was used. The exact numerical constants are given in the above table. The recorded hyperparameters are summarized in Table 4, and the four sensitivity protocols are specified in the main figure caption.

C.5 Time Complexity of TABLE

The computational cost of TABLE is driven by four main components: the initial clustering step, representative selection, per-round posterior-based scoring, and occasional elimination/refinement checks.

The initial partition is obtained by k-Means clustering on the full arm set. If K denotes the number of arms, l the ambient dimension, c_0 the number of initial clusters, and I_0 the number of KMeans iterations, this stage costs

$$\mathcal{O}(Klc_0I_0).$$

In our experiments, l is small and c_0 is fixed, so this cost is incurred only once at the start of the run.

After the initial partition is formed, TABLE selects b representatives in each active node using the farthest-point heuristic. With a standard implementation that maintains the current nearest-representative distance for each arm, selecting b representatives in a node u uses $\mathcal{O}(|\mathcal{C}(u)|\, b)$ distance lookups, or $\mathcal{O}(|\mathcal{C}(u)|\, b\, l)$ operations if each Euclidean distance costs $\mathcal{O}(l)$, so across all active nodes

at a given stage the total cost is linear in the number of arms times the number of representatives. The same observation applies whenever representatives are re-selected after a split.

During each exploration round, TABLE samples once from the posterior of each active representative and computes one score per active node. If $|\mathcal{U}(t)|$ is the number of active nodes at round t, the per-round cost of this step is

$$\mathcal{O}(|\mathcal{U}(t)|\, b).$$

The balancing step is cheaper, since it only requires identifying the least-sampled active representative, which is also bounded by $\mathcal{O}(|\mathcal{U}(t)|\, b)$ under a direct scan.

At elimination checks, TABLE computes representative-level confidence radii and node-level lower and upper confidence scores, again requiring at most

$$\mathcal{O}(|\mathcal{U}(t)|\, b)$$

operations. If a node is split, the additional cost depends on the size of that node: applying 2Means to a node u with $|\mathcal{C}(u)|$ arms costs

$$\mathcal{O}(|\mathcal{C}(u)|\, l\, I_{\mathrm{split}}),$$

where I_{split} is the number of 2Means iterations. Since splitting is performed only on selected active nodes and the active set shrinks through elimination, this cost is incurred adaptively rather than at every round.

Overall, the dominant per-round cost of TABLE is linear in the number of active representatives, namely $\mathcal{O}(|\mathcal{U}(t)|\, b)$, while the more expensive clustering operations occur only at initialisation and when a node is explicitly refined. This separates a per-round upper bound in active exemplars from occasional full-node distance and clustering costs. A comprehensive runtime comparison with the baselines would additionally require measured timings; the preceding operation counts alone do not establish a wall-clock advantage.