PRIOR-FREE ADAPTATION IN IMPROVING BANDITS: FREE WITHOUT NOISE, COSTLY UNDER NOISE

Anonymous authors
Paper under double-blind review

ABSTRACT

Improving bandits model decision problems in which rewards increase with repeated investment in an arm while exhibiting diminishing returns. Existing guarantees rely on prior knowledge of problem-dependent quantities such as reward scale, a lower-envelope parameter, or the horizon. We ask whether such prior information is fundamentally necessary. Our first result shows that unknown scale is essentially free. We remove the logarithmic overhead in previous scale-oblivious guarantees and obtain the optimal all-horizon competitive ratio $\Theta(\sqrt{k} + k/T)$ without knowing the reward scale. More generally, in the noiseless setting, a single anytime algorithm simultaneously adapts to unknown scale, unknown lower-envelope parameter β, and unknown horizon, achieving the optimal ratio $\Theta(k^{\beta/(1+\beta)} + k/T)$ for every $\beta \in (0, 1]$ whose lower-envelope condition holds for an optimal arm. Thus, the price of prior-free adaptation is constant. We then show that this picture changes under bounded multiplicative noise, with observations fixed by an oblivious adversary. Unknown scale alone remains essentially free, but simultaneously adapting to both scale and β incurs a growing cost. For any fixed $\varepsilon \in (0, 1/2]$, we establish a structural asymptotic lower bound of order $\Omega_\varepsilon(\sqrt{\log k / \log\log k})$ on the uniform price of adaptation, and give a matching algorithm achieving $O_\varepsilon(\sqrt{\log k / \log\log k})$. The uniform price is the minimax loss relative to $k^{\beta/(1+\beta)}$ over β, instances, noise tables, and horizons $T \geq 16k$, among algorithms knowing neither scale nor β. Knowing either the scale or β alone restores a constant price. Together, these results reveal a sharp separation between noiseless and noisy improving bandits. Prior-free adaptation is possible at constant cost without noise, while multiplicative noise creates a growing multiscale adaptation cost when scale and envelope information are unknown.

1 INTRODUCTION

Improving bandits formalize sequential allocation when investing effort in an option raises its future rewards, with diminishing increments. In this model (Heidari et al., 2016; Patil et al., 2023), there are k arms; pulling arm i for the t-th time yields $f_i(t)$, where the unknown reward curve f_i is nondecreasing and discretely concave, with $f_i(0) = 0$. Its argument counts that arm's pulls, not calendar time. Pulling one arm throughout is optimal in hindsight (Heidari et al., 2016, Prop. 1), giving the benchmark $\mathrm{OPT}_T = \max_i \sum_{t \leq T} f_i(t)$. For fixed bounded curves, Heidari et al. (2016, Thm. 2) establish instance-dependent sublinear policy regret. A uniform guarantee is different: our delayed-reveal lower bound excludes uniform $o(T)$ expected regret even for randomized algorithms with known T (Corollary C.6: already for $k = 16$, every algorithm faces an instance with expected regret at least $(T + 1)/8$). The stochastic *rising bandit* literature encounters the same barrier in its own model (Song et al., 2026, Thm. 4: $\Omega(LT)$ regret for combinatorial rising bandits). Following Patil et al. (2023) and Blum & Ravichandran (2025), we therefore study competitive ratios $\mathrm{OPT}_T / \mathbb{E}[\mathrm{ALG}_T]$ and their horizon dependence.

Blum & Ravichandran (2025) proved the value of randomization: no algorithm beats $\Omega(\sqrt{k})$; random round robin with an absolute threshold $m\,t/T$ achieves $O(\sqrt{k})$ when given the scale $m = f^*(T)$ of the optimal arm ($T \geq 2k$), whereas without m their guarantee is $O(\sqrt{k}\log k)$ for $T > 4k$. Blum et al. (2026) sharpened this picture through a lower-envelope exponent $\beta \in (0, 1]$:

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.

Noiseless	Unknown scale m + exponent β + horizon T	$\longrightarrow$	**Constant price:** $\Theta(1)$
Noisy	Unknown scale m + exponent β	$\longrightarrow$	**Growing price:** $\Theta_\varepsilon\left(\sqrt{\log k/\log\log k}\right)$

Figure 1: Adaptation price with missing priors. Noiseless: one anytime policy preserves the rate $k^{\beta/(1+\beta)}+k/T$ for every admissible optimal-arm envelope. Noisy: the uniform price over $T \geq 16k$ grows asymptotically in k, for fixed $\varepsilon \in (0, \frac{1}{2}]$ under bounded multiplicative, obliviously fixed noise (Section 2). The lower bound holds at the hard horizons of Theorem 5.3, even with T, ε known.

the condition $f(t) \geq f(T)(t/T)^\beta$ controls how quickly rewards rise relative to the terminal value. This exponent is the *curvature* in our title, not strong concavity. For envelope exponent β_I, Blum et al. (2026, Thm. 3.5) give a known-scale guarantee $O(k^{\alpha/(1+\alpha)})$ for $\alpha \in (\beta_I, 1)$ and $T \geq 2k$, with a matching β-dependent lower bound for sufficiently large horizons. Their Theorem B.15 gives $O(k^{\alpha/(1+\alpha)} \log k)$ with both m and T unknown for $0 < \beta_I < \alpha < 1$ and $T > 4k$, but still requires a suitable exponent parameter. Their data-driven parameter selection uses offline instances; the per-instance worst-case caveat concerns the regret/competitive-ratio hybrid in their Appendix D. We ask which priors m, β and T are necessary, and what their absence costs.

Learning-curve allocation connects this model to automated hyperparameter search: arms represent configurations and pulls allocate further training epochs (Blum et al., 2026). Under the monotone, concave reward assumptions and cumulative-reward objective, our results separate the difficulty of allocating training effort from the cost of missing calibration or curve-shape information. The classification identifies when that information is dispensable and when, under noise, its joint absence imposes an unavoidable loss.

Contributions. Without noise, one anytime algorithm attains the optimal all-horizon order without reading m or β. At any fixed $\varepsilon \in (0, \frac{1}{2}]$, either m or β alone suffices for constant uniform adaptation price, whereas knowing neither costs $\tilde{\Theta}_\varepsilon(\Phi(k))$ as $k \to \infty$. This lower bound holds for $k \geq k_0(\varepsilon)$ at the hard horizons of Theorem 5.3, not at every fixed $T \geq 16k$. The results separate three questions:

1. **Unknown scale is free (Section 3).** *Probe-and-commit* samples $s = \lfloor\sqrt{k}\rfloor$ arms, pulls each $\lfloor T/(2s)\rfloor$ times, and commits to the arm with the largest probe value. It guarantees a competitive ratio $4\sqrt{3}\sqrt{k} < 7\sqrt{k}$ for all $T \geq 2\lfloor\sqrt{k}\rfloor$, without knowing m (Theorem 3.1), removing the $\log k$ factor of Blum & Ravichandran (2025). The optimal ratio for every $k, T \geq 1$ is $\Theta(\sqrt{k} + k/T)$ (Theorem 3.2). One anytime algorithm reading only k attains this order at *every* horizon (Theorem 3.3); the naive doubling schedule of Example C.9 does not.
2. **Without noise, every prior is free (Section 4).** *Random-marginal probing* draws independent weights Z_i, greedily probes by last observed increment divided by Z_i, then commits. Without m or β, it achieves $\mathbb{E}[\mathrm{ALG}_T] \geq \frac{3}{256}\,\mathrm{OPT}_T\, k^{-\beta/(1+\beta)}$ simultaneously for every β whose envelope condition holds for the *optimal arm*, for $T \geq 16k$ (Theorem D.4). Independent sparse inclusion extends the optimal order to all horizons, $\Theta(k^{\beta/(1+\beta)} + k/T)$ (Theorem 4.1); an anytime version also removes T (Theorem 4.2). Thus withholding m, β and T costs only $\Theta(1)$.
3. **Noise makes joint ignorance costly (Section 5).** Under the bounded multiplicative noise of Blum & Ravichandran (2025, App. C) (observations fixed by an oblivious adversary; Section 2), probe-and-commit retains the bounds, constants and horizon conditions of Theorem 5.1 without reading ε: reward bounds acquire a factor a^2/b against R (and a^2/b^2 against $\hat{R}$), $a = 1-\varepsilon$, $b = 1+\varepsilon$. The unknown-scale noisy extension of Blum & Ravichandran (2025) reads ε and pays logarithmic overhead. Random-marginal probing and two increment-based repairs fail with unbounded ratio (exact counterexamples, Section 5; not a claim against every possible repair). More fundamentally, for fixed $\varepsilon \in (0, \frac{1}{2}]$, every algorithm knowing neither m nor β — even if given T and ε — loses $\Omega_\varepsilon(\Phi(k))$, $\Phi(k) = \sqrt{(1+\ln k)/\ln(e+\ln k)}$, relative to $k^{\beta/(1+\beta)}$ on some instance at the construction's hard horizons (structurally and asymptotically in k, for $k \geq k_0(\varepsilon)$; Theorem 5.3). Nested random-permutation probing matches this loss (Theorem 5.2); either m alone (Theorem G.1) or β alone restores constant price. These are the quantifiers of the

uniform price of adaptation $\phi_\varepsilon(k)$ in Section 2: choose the algorithm first, then β, the instance and the noise table.

Table 1 classifies the information each algorithm needs and its cost. Table 2 (Appendix A) records constants and horizon ranges; the appendix contains proofs and supporting arguments for all new theoretical results.

Why the logarithm disappears. The algorithms of Blum & Ravichandran (2025); Blum et al. (2026) eliminate arms below an *absolute* threshold $m\,t/T$ (or $m(t/T)^\alpha$). They therefore require scale information, and the random dyadic guess of m in Blum & Ravichandran (2025) costs $\log k$. Probe-and-commit ranks probe values *relative to each other*: its $\arg\max$ is scale-free, with a $\sqrt{k}$ subsample paying for missing information. Random-marginal probing instead ranks weighted observed increments; its key lemma charges expected greedy work on decoys to their total value, avoiding a multi-scale union bound. Relative value comparisons also make probe-and-commit noise-robust without reading ε.

Table 1: The uniform price of adaptation $\phi_\varepsilon(k)$ of Section 2, i.e. $\inf$ over algorithms (or anytime policies) of the worst case over $T \geq 16k$, over β, over instances of $\mathcal{I}_\beta$ and over admissible noise tables, of $\hat{R}\,q_\beta/\mathbb{E}[\widehat{\mathrm{ALG}}_T]$. The algorithm is chosen before β and the instance. The noisy lower bound is established at the hard horizons $T = 16k'$, $32k'$ of Theorem 5.3 and is asymptotic in k; it is not claimed at every fixed $T \geq 16k$. Each indicated information class has its own infimum; reading T or ε does not change the order. In known-scale cells, the advice refers to the same designated optimal arm as the envelope hypothesis and $\hat{R}$; the arm's index is not supplied. The unread-T guarantees obtained from Theorem G.1 use advice tied to the evaluation horizon and do not assert one fixed-advice anytime policy; the prior-free entries admit single-policy anytime versions reading only k (Theorem 4.2; the anytime clause of Theorem 5.2). Further conventions for the cells: Appendix B.

Unknown parameters	noiseless price (Section 4)	price under fixed $\varepsilon \in (0, \frac{1}{2}]$
none / only m / only β	$\Theta(1)$	$\Theta(1)$ (Thms. 5.1, G.1)
both m and β	$\Theta(1)$ (Thms. D.4, 4.1)	$\Theta_\varepsilon(\Phi(k))$ (Thms. 5.2, 5.3)

2 Model and conventions

We follow the specification of Patil et al. (2023); Blum & Ravichandran (2025). An instance consists of k arms; arm i has a reward function $f_i : \{0, 1, \ldots, T\} \to \mathbb{R}_{\geq 0}$ with $f_i(0) = 0$, nondecreasing, and with *diminishing returns*:

$$f_i(t+1) - f_i(t) \leq f_i(t) - f_i(t-1) \qquad \text{for every integer } 1 \leq t \leq T-1$$

(discrete concavity). Pulling arm i for the t-th time yields $f_i(t)$; the argument is the number of pulls of that arm, not calendar time. The functions are unknown; the algorithm observes only the rewards it collects. The horizon is T. The adversary is *oblivious*: all f_i are fixed after T is fixed and before the algorithm runs.

Write $P_i(n) = \sum_{t=1}^{n} f_i(t)$, $R = \mathrm{OPT}_T = \max_i P_i(T)$, and let $*$ denote a *comparator*: an arm attaining the maximum, chosen as part of the statement of each theorem (by default the smallest index). Set $m = f_*(T)$ and $M = \max_i f_i(T)$. ALG_T denotes the algorithm's cumulative reward; expectations are over the algorithm's internal randomness only. An algorithm has *competitive ratio* ρ if $\mathrm{OPT}_T \leq \rho\,\mathbb{E}[\mathrm{ALG}_T]$ on every instance (Blum & Ravichandran, 2025, Def. 1). Instances with $R = 0$ make every guarantee trivial and are excluded from ratios. The comparator convention in full (when several arms are cumulatively optimal), the computational model and two further conventions are recorded in Appendix B.

Envelope exponent. Following Blum et al. (2026, Def. 3.2), for $\beta \in (0, 1]$ we say that arm i satisfies $\mathrm{LE}(\beta)$ if the inequality $f_i(t) \geq f_i(T)\,(t/T)^\beta$ holds for all $t \leq T$. Every concave nondecreasing

f with $f(0) = 0$ satisfies $\mathrm{LE}(1)$. Our positive results require the envelope condition *only for the optimal arm*; the lower bounds of Blum et al. (2026) use instances in which every arm satisfies it, so both directions apply to the class where all arms do. We write

$$\gamma = \frac{\beta}{1+\beta}, \qquad q_\beta = k^{-\gamma}, \qquad K_\beta = k^{1/(1+\beta)} = kq_\beta, \qquad Q_\beta(k,T) = \min\{q_\beta,\, T/k\}.$$

For $\beta \in (0,1]$ let $\mathcal{I}_\beta(k,T)$ be the class of instances in which at least one cumulatively optimal arm satisfies $\mathrm{LE}(\beta)$ (an upper-envelope class, not the class with envelope exponent exactly β; when several cumulatively optimal arms exist, the comparator $*$ of an instance in $\mathcal{I}_\beta$ is by convention one satisfying $\mathrm{LE}(\beta)$, and m, $\hat{R}$ and the envelope hypothesis all refer to that same arm, as fixed in Appendix B), and define the minimax ratio $\rho_\beta(k,T) = \inf_A \sup_{I\in\mathcal{I}_\beta(k,T),\, R_I>0} R_I/\mathbb{E}[\mathrm{ALG}_T(A,I)]$, the infimum being over the algorithms of the knowledge class under discussion; $\rho(k,T) = \rho_1(k,T)$.

Knowledge. An algorithm is *scale-oblivious* if it does not read m (nor any estimate of it), β-*oblivious* if it does not read β (nor an upper bound), and *anytime* if it does not read T. We always list explicitly what each algorithm reads. Being oblivious to a parameter forbids receiving it as input; it does not forbid estimating quantities from the observations already paid for. Our anytime guarantees are expectation guarantees of a single policy at each fixed horizon, not high-probability statements uniform over data-dependent stopping times.

Two adaptation criteria. For a fixed β, $\rho_\beta(k,T)$ is a *class-wise* minimax value: the infimum may be attained by a different algorithm for each β, so a family of bounds "$\forall\beta\ \exists A_\beta$" does not by itself produce one algorithm that is near-optimal for all β at once. The latter, stronger property — "$\exists A\ \forall\beta$" — is what we call *simultaneous adaptation*; it is what Theorems D.4, 4.1 and 5.2 provide, and what Theorem 5.3 limits under noise. Quantitatively, for a class $\mathfrak{A}$ of algorithms (specified by what they read) define the *uniform price of adaptation*

$$\phi_\varepsilon(k) \;=\; \inf_{\{A_T\}_{T\ge 16k}\in\mathfrak{A}} \;\sup_{T\ge 16k}\; \sup_{\beta\in(0,1]}\; \sup_{\substack{I\in\mathcal{I}_\beta(k,T),\, R_I>0\\ \hat{f}\in\mathcal{N}_\varepsilon(I)}} \frac{\hat{R}_I\, q_\beta}{\mathbb{E}[\widehat{\mathrm{ALG}}_T(A_T,I,\hat{f})]},$$

where $\mathcal{N}_\varepsilon(I)$ is the set of admissible noise tables for I (a single table $\hat{f} = f$ when $\varepsilon = 0$, in which case $\hat{R}_I = R_I$). The order of the quantifiers is essential: the algorithm is chosen before β, the instance and the noise.

Noise model (Section 5). In the multiplicative model of Blum & Ravichandran (2025, App. C), "when we pull an arm, instead of getting the exact value of $f_i(t_i)$, we get a reward value $\hat{f}_i(t_i) \in [(1-\varepsilon)f_i(t_i), (1+\varepsilon)f_i(t_i)]$"; the algorithm observes and earns $\hat{f}$. The noisy values are fixed by the oblivious adversary before the run, for every (i,t), with no further structure (they may be non-monotone and non-concave). We compare, as in Blum & Ravichandran (2025), with the policy that plays the best uncorrupted arm: $\hat{R} = \sum_{t\le T} \hat{f}_*(t)$, which satisfies $aR \le \hat{R} \le bR$ with $a = 1-\varepsilon$, $b = 1+\varepsilon$, and $\hat{R} \ge \frac{a}{b}\max_\pi \hat{V}(\pi)$. Throughout we assume $\varepsilon \in [0, \frac{1}{2}]$; this is a convention of the present paper (the additional restriction $\varepsilon \le \frac{1}{2}$ is not imposed in Blum & Ravichandran (2025)), used only by the algorithm RET of Section 5; the probe-and-commit and NPA bounds hold for all $\varepsilon < 1$.

3 Unknown scale is free

Sample coupling (Lemma C.2) yields the core upper bound (Lemma C.3); delayed reveal (Lemma C.5) supplies the lower bounds. History conditioning (Lemma C.7) gives the anytime guarantees. Appendix C gives the proofs, starting from the shared concavity facts of Lemma C.1: for a curve f with prefix sums P, $f(t) \ge \frac{t}{T}f(T)$ for $1 \le t \le T$, $P(n) \ge f(T)\frac{n(n+1)}{2T}$, $R \le mT$, $P_*(r) \ge R/4$ for $\lceil T/2\rceil \le r \le T$, $M \le 2m$, and $\sum_i P_i(n_i) \le R$ for every allocation $\sum_i n_i = T$ of pulls.

Probe-and-commit. $\mathrm{PC}(s,\tau)$, for integers $1 \le s \le k$ and $\tau \ge 1$ with $s\tau \le \lfloor T/2 \rfloor$, draws a uniformly random set S of s distinct arms, pulls each arm of S exactly τ times, and pulls $c = \arg\max_{i \in S} f_i(\tau)$ (any fixed tie-breaking rule) for the remaining $r = T - s\tau \ge \lceil T/2 \rceil$ steps. It reads only k and T.

Theorem 3.1 (unknown scale)**.** *Let $k \ge 2$, $T \ge 2\lfloor\sqrt{k}\rfloor$, $s = \lfloor\sqrt{k}\rfloor$, $\tau = \lfloor T/(2s) \rfloor$. Then $\mathrm{PC}(s,\tau)$ satisfies*

$$\mathbb{E}[\mathrm{ALG}_T] \;\ge\; \frac{s}{8k} R \;\ge\; \frac{R}{16\sqrt{k}}, \qquad \textit{and in fact} \qquad \mathbb{E}[\mathrm{ALG}_T] \;\ge\; \frac{R}{4\sqrt{3}\,\sqrt{k}}.$$

For $k = 1$ the ratio is 1. The algorithm uses no information about m.

Proof sketch. Let $F = \{i \ne * : f_i(\tau) \ge f_*(\tau)\}$ be the arms that can beat the optimal arm at the probe depth, $A = \{* \in S\}$ and $B = \{S \cap F \ne \emptyset\}$. On B the committed arm satisfies $f_c(\tau) \ge f_*(\tau) \ge m\tau/T$ (Lemma C.1(1)), so the commit phase earns at least $r\,m\tau/T \ge R\,r\tau/T^2$; on $A \cap \neg B$ the committed arm is $*$ and the commit phase earns $\sum_{j \le r} f_*(\tau + j) \ge \sum_{j \le r} f_*(j) \ge R\,r(r+1)/(2T^2)$ (Lemma C.1(2)). Conditioned on $\neg B$, S is a uniform s-subset of a set containing $*$, so $\mathbb{P}(A \cap \neg B) \ge \frac{s}{k}\mathbb{P}(\neg B)$ (Lemma C.2); the two disjoint events give $\mathbb{E}[\mathrm{ALG}_T] \ge R\min\{r\tau/T^2,\ s\,r(r+1)/(2kT^2)\}$ for every admissible (s,τ) (Lemma C.3). With $r \ge T/2$, $\tau \ge T/(4s)$ and $s^2 \le k$ this is at least $\frac{s}{8k}R \ge R/(16\sqrt{k})$. The constant $4\sqrt{3}$ comes from the sharper forms $P(u)/P(T) \ge u(u+1)/(T(T+1))$ and $f(\tau)/P(T) \ge 2\tau/(T(T+1))$ for $\tau \le (T+1)/2$, together with $k \le s^2 + 2s \le 3s^2$. □

Theorem 3.2 (all horizons)**.** *For all integers $k, T \ge 1$, the minimax ratio of scale-oblivious algorithms satisfies*

$$\frac{1}{4}\Big(\sqrt{k} + \frac{k}{T}\Big) \;\le\; \rho(k,T) \;\le\; 24\Big(\sqrt{k} + \frac{k}{T}\Big).$$

The upper bound is Lemma C.3 with the truncated parameters $s = \min\{\lfloor\sqrt{k}\rfloor, \lfloor T/2 \rfloor\}$, $\tau = \lfloor T/(2s) \rfloor$ (for $T = 1$, pull a uniformly random arm). The lower bound comes from a "delayed-reveal" lemma (Lemma C.5, a restatement with integer bookkeeping of the hard-distribution argument of Blum & Ravichandran, 2025): a uniformly random arm has reward g and every other arm has $g(\min\{t,d\})$, so an algorithm can only tell the good arm apart after $d+1$ pulls of it. With $g(t) = t/T$ and $d = \lfloor T/\sqrt{k} \rfloor$ this gives $\sqrt{k}/3$; with $d = 0$ (one linear arm among $k-1$ zero arms) it gives k/T (Corollary C.6).

Unknown horizon. With $s = \lfloor\sqrt{k}\rfloor$, a doubling schedule that reruns PC on a fresh sample in phases of geometrically growing length, never resetting true pull counts, reads only k and keeps the order for $T \ge 2s$ (Theorem C.8); the restriction is real (Example C.9). The *anytime probe-and-commit* algorithm APC (Appendix C) instead starts at length 1: it runs phases of planned lengths $H_j = 2^j$, where a phase of length 1 pulls one uniformly random arm and a phase of length $H \ge 2$ runs PC on a fresh sample with the truncated parameters $s = \min\{\lfloor\sqrt{k}\rfloor, \lfloor H/2 \rfloor\}$; it reads only k, true pull counts never reset.

Theorem 3.3 (unknown horizon, all horizons)**.** *For all integers $k, T \ge 1$,* APC *satisfies*

$$\mathbb{E}[\mathrm{ALG}_T] \;\ge\; \frac{R}{1536\,(\sqrt{k} + k/T)}.$$

Proof sketch. For a last completed phase of length $H \ge 2$, put $x = H/T > 1/4$; its sample is fresh, and because rewards are nondecreasing the shifted curves $t \mapsto f_i(n_i + t)$, where n_i is arm i's pull count before the phase, dominate the original ones, so the one-shot guarantee applies to that phase conditionally on the history (Lemma C.7). That phase alone earns the stated order, and not knowing T costs only a bounded power of x. The case $H = 1$ (i.e. $T \le 2$) is checked directly (Appendix C). □

Known envelope exponent. When β is known and the comparator arm satisfies $\mathrm{LE}(\beta)$, PC with $s = \lceil K_\beta \rceil$ and $\tau = \lfloor T/(2s) \rfloor$ gives $\mathbb{E}[\mathrm{ALG}_T] \geq R/(16k^\gamma)$ for $k \geq 2$ and $T \geq 2\lceil K_\beta \rceil$; the class-wise minimax ratio satisfies $\frac{k^\gamma + k/T}{\beta+3} \leq \rho_\beta(k,T) \leq 24(k^\gamma + k/T)$ for all $T \geq 1$; and a doubling version gives $R/(128k^\gamma)$ for every $T \geq 2s$ with T unknown (Theorem C.10). Being class-wise, this does not by itself produce one algorithm that is simultaneously near-optimal for all β; that stronger property is the subject of the next section.

4 Without noise, every prior is free

The β-tuned algorithms of Theorem C.10 use β to choose the probe depth. We now remove this requirement. Throughout this section the instance is noiseless and we assume only that *the optimal arm* satisfies $\mathrm{LE}(\beta)$; the other arms need only satisfy the model assumptions. The proof separates monotone coupling (Lemma D.1) from the work estimate (Lemma D.2). They support the threshold guarantee of Lemma D.3 and its sparse counterpart, Lemma D.9. The latter directly supports both Theorems 4.1 and 4.2, retaining the original comparator on shifted curves. Proofs are in Appendix D.

Random-marginal probing. $\mathrm{RMP}(H)$ reads k and a length $H \geq 16k$, neither m nor β: it draws independent weights $Z_i \sim \mathrm{Unif}(0,1)$, spends a probe budget greedily on the arm maximizing its last observed increment divided by Z_i, and commits to the arm with the largest observed value (Appendix D). Each arm's key sequence is nonincreasing by discrete concavity, so the probe phase is a k-way merge (Lemma D.1); and after truncating a curve at a level v, the expected number of its items with key at least λ is at most v/λ (Lemma D.2) — the *total* increment v, not "number of scales times v", which is what removes the logarithm. Hence, for $k \geq 2$ and $\beta \in (0,1]$, $\mathbb{E}[\mathrm{ALG}_T] \geq \frac{3}{256} R\, q_\beta$ for $T \geq 16k$ and an optimal arm satisfying $\mathrm{LE}(\beta)$ (Theorem D.4); phases with fresh weights give the same order without reading T (Theorem D.6).

All horizons: independent sparse inclusion. RMP initializes every arm; the all-horizon construction instead uses independent inclusion. $\mathrm{SRMP}(k,H)$ reads only k and H. If $k = 1$ it pulls the arm, and if $H < 32$ it pulls a uniformly random arm for all H steps. Otherwise it sets $N = \lfloor H/2 \rfloor$, $r = H - N$, $\theta = \min\{1, H/(128k)\}$, draws independent $X_i \sim \mathrm{Bernoulli}(\theta)$ and $Z_i \sim \mathrm{Unif}(0,1)$, and runs the greedy probe-and-commit rule of RMP on the included set $S = \{i : X_i = 1\}$ (initializing each $i \in S$ with one pull), falling back to a uniformly random arm for all H steps if $S = \emptyset$ or $|S| > N$.

Theorem 4.1 (prior-free, all horizons)**.** *For all integers $k, T \geq 1$, all $\beta \in (0,1]$ and all instances whose optimal arm satisfies $\mathrm{LE}(\beta)$, $\mathrm{SRMP}(k,T)$ satisfies*

$$\mathbb{E}[\mathrm{ALG}_T] \;\geq\; \frac{R}{c_{\mathrm{SR}}} \min\Big\{k^{-\beta/(1+\beta)},\; \frac{T}{k}\Big\}, \qquad c_{\mathrm{SR}} := 2^{13} = 8192.$$

Consequently, the optimal competitive ratio of algorithms reading neither m nor β is $\Theta(k^{\beta/(1+\beta)} + k/T)$ for every horizon: $\frac{k^\gamma + k/T}{\beta+3} \leq \rho_\beta(k,T) \leq \sup_{I \in \mathcal{I}_\beta,\, R_I > 0} \frac{R_I}{\mathbb{E}[\mathrm{ALG}_T(\mathrm{SRMP}, I)]} \leq c_{\mathrm{SR}}(k^\gamma + k/T)$ — one and the same SRMP *attains all of these upper bounds simultaneously — the lower bound being the known-β baseline of Theorem C.10(ii).*

Proof sketch. For $T \geq 32$ fix an integer depth $1 \leq d \leq \lfloor T/4 \rfloor$; the case $T < 32$ is treated at the end. Throughout, $*$ is the comparator arm and (X_*, Z_*) its inclusion indicator and weight from the previous paragraph. Two Markov inequalities, carried out in the proof of Lemma D.9, show that with probability at least $\frac{1}{2}$ the other included arms are few and the work they can command at high keys is a small fraction of the probe budget; call this event C, independent of (X_*, Z_*). On C, if the comparator is included with a small enough weight and no other arm reaches the level $v = f_*(d)/2$, the merge commits to it. After fixing the other arms' randomness and the fallback choice, on C both the high-priority indicator and the indicator that no other arm reaches v during the probe prefix are nondecreasing in the single variable Y ($Y = 0$ if $*$ is excluded, $Y = 1/Z_*$ otherwise), so the conditional positive-correlation inequality applies (Lemma D.9). The depth d is chosen in the analysis only, never by the algorithm; with $v = f_*(d)/2$ and d chosen in Appendix D to balance the two terms (the choice depends on K_β; a fixed d does not suffice), this gives the bound, while for $T < 32$ the uniform single arm earns $\geq R/k \geq RQ_\beta/31$ (Appendix D). □

Theorem 4.2 (prior-free anytime, all horizons)**.** *Run* SRMP *in phases of lengths* $1, 2, 4, \ldots$*, each with fresh randomness and with the first observation of the phase serving as the first increment. The algorithm reads only* k*, and for every* $T \geq 1$*,* $\mathbb{E}[\mathrm{ALG}_T] \geq 2^{-19}\, R\, Q_\beta(k, T)$.

The phases have lengths $1, 2, 4, \ldots$ (Theorem 4.2), so if H is the length of the last completed phase then the completed prefix has length $2H - 1$ and the next phase is unfinished, whence $T \leq 4H - 2$ and $x = H/T > 1/4$. For $H \geq 32$, the proof applies Lemma D.9 to that phase, on the reset curves $g_i(t) = f_i(n_i + t)$, which are again nondecreasing and concave with $g_*(d) \geq f_*(d)$: that phase earns $\geq \frac{Rx^2}{8192} Q_\beta(k, H) \geq \frac{Rx^3}{8192} Q_\beta(k, T)$, and $x^3 > 1/64$ yields the constant 2^{-19}. If $H < 32$ then $T \leq 62$ — since H is a power of two, $H \leq 16$ — and the uniform-arm phase earns $\geq R/(16k) \geq 2^{-19} R Q_\beta$ (Appendix D).

5 Multiplicative noise: the price of priors jumps

We work in the noise model of Section 2: $\hat{f}_i(t) \in [a f_i(t), b f_i(t)]$, $a = 1 - \varepsilon$, $b = 1 + \varepsilon$, fixed by the oblivious adversary; the algorithm observes and earns $\hat{f}$; the benchmark is $\hat{R} = \sum_{t \leq T} \hat{f}_*(t) \in [aR, bR]$. We never treat $\hat{f}$ as monotone or concave; all uses of Lemma C.1 are for the true curves. For an observed prefix let $F_i(n) = \max_{t \leq n} \hat{f}_i(t)$ (running maximum); then $a f_i(n) \leq F_i(n) \leq b f_i(n)$ and, once arm c has been pulled n_c times, every later pull earns $\hat{f}_c(n_c + j) \geq a f_c(n_c + j) \geq a f_c(n_c) \geq \frac{a}{b} F_c(n_c)$. Let $\Phi(k) = \sqrt{(1 + \ln k)/\ln(e + \ln k)}$. Permutation coupling and depth-budget feasibility (Lemmas F.2 and F.3) give the prior-free upper bound of Theorem 5.2; Lemma F.4 handles the anytime case. The common-history lower bound and the known-scale result have independent proofs (Appendices F and G).

Probe-and-commit is robust and needs no ε.

Theorem 5.1 (noisy probe-and-commit)**.** *Let* PC *commit to the arm with the largest observed probe value. For every instance and admissible* (s, τ)*,*

$$\mathbb{E}[\widehat{\mathrm{ALG}}_T] \;\geq\; \frac{a^2}{b}\, R \,\min\Big\{\frac{r\tau}{T^2}, \frac{s\, r(r+1)}{2kT^2}\Big\}.$$

The same estimate holds phase-wise, conditionally on the history, exactly as in Lemma C.7. Consequently each algorithm of Section 3 keeps its guarantee up to the factor a^2/b*: with* $\mathbb{E}[\widehat{\mathrm{ALG}}_T] \geq \frac{a^2}{b} \frac{R}{CG} \geq \frac{a^2}{b^2} \frac{\hat{R}}{CG}$ *one has*

algorithm	CG	*horizon*
PC*, T known (Thm. 3.1)*	$16\sqrt{k}$	$T \geq 2\lfloor\sqrt{k}\rfloor$
PC*, truncated parameters (Thm. 3.2)*	$24(\sqrt{k} + k/T)$	*all* $T \geq 1$
APC*, T unknown (Thm. 3.3)*	$1536(\sqrt{k} + k/T)$	*all* $T \geq 1$
doubling PC*, T unknown (Thm. C.8)*	$256\sqrt{k}$	$T \geq 2\lfloor\sqrt{k}\rfloor$
PC*, β known (Thm. C.10)*	$16/q_\beta$	$T \geq 2\lceil K_\beta\rceil$
doubling PC*, β known, T unknown (Thm. C.10(iii))*	$256/q_\beta$	$T \geq 2\lceil K_\beta\rceil$

The algorithm reads neither m *nor* ε*; the lower bounds of Section 3 persist (zero noise is admissible), so the noisy minimax ratio of scale-oblivious algorithms is* $\Theta(\sqrt{k} + k/T)$ *at every horizon, and* $\Theta(\sqrt{k})$ *when* $T = \Omega(\sqrt{k})$.

Let S be the probe sample, let $\hat{F} = \{i \neq * : \hat{f}_i(\tau) \geq \hat{f}_*(\tau)\}$, and set $A = \{* \in S\}$ and $B = \{S \cap \hat{F} \neq \emptyset\}$. On B, the committed arm c satisfies $\hat{f}_c(\tau) \geq a f_*(\tau)$, and each of its subsequent observed rewards is at least $(a/b)\hat{f}_c(\tau)$; on $A \cap B^c$, the committed arm is $*$. Since $a^2/b \leq a$, the two branches of the argument of Lemma C.3 admit the common factor a^2/b. Only relative comparisons of observed values are used, which is why no knowledge of ε is needed; the algorithm of Blum & Ravichandran (2025, App. C) "must have knowledge of the value of ε".

Random-marginal probing breaks. Under noise the observed increments can be negative and non-monotone, so the merge structure of Lemma D.1 is lost: a hidden growth prefix defeats RMP and two increment-based repairs, by an explicit family with unbounded ratio whose observed table is in band and nondecreasing in each arm's pull count for $1 \le t \le T$ and whose comparator satisfies $\mathrm{LE}(1/2)$ (Example F.1, Appendix F). It refutes the running-maximum *increment* key and the half-window slope key, both fixed precisely there, not every conceivable repair (two variants it does not refute are named there); Theorem 5.3, not this example, is what rules out a constant uniform price of adaptation for algorithms reading neither m nor β.

A matching pair. The upper bound is a noise-robust probe-and-commit with unequal depths (Lemma F.2): fix nonincreasing integer depths $\ell_1 \ge \cdots \ge \ell_k \ge 1$ with $\sum_j \ell_j \le H/2$, permute the arms uniformly at random, pull the j-th arm ℓ_j times, and commit to the arm with the largest running maximum. *Nested permutation probing* $\mathrm{NPA}(k, H)$ (Algorithm 5; $k \ge 2$, $H \ge 16k$) takes all depths from one explicit envelope D with $h = 8\Phi(k)$, setting $\ell_j = \max\{1, \lceil H\, D((j-1)/k)\rceil\}$, reading neither m, β nor ε; for every β its first $s = \lceil k q_\beta/h\rceil$ positions then have depth at least $d_H = \lceil H(q_\beta/h)^{1/\beta}\rceil$ (analysis witnesses, not inputs to NPA; for the known-horizon theorem set $H = T$ and $d = d_T$).

Why $\Phi(k)$. Both directions reflect the cost of covering many possible reveal depths in a single run. For $k \ge 2$ and $H \ge 16k$ the algorithm uses one envelope for all β: with $h = 8\Phi(k)$ the continuous budget obeys $k\int_0^1 D \le \frac{1}{32}$, and with the discretization terms of Lemma F.3 the integer depths satisfy $\sum_j \ell_j \le \frac{7H}{32}$. The lower bound charges the same scarcity: its types form a set $\mathcal{J}$ whose successive indices differ by two, type j having hidden-prefix depth d_j, so consecutive depths obey $d_{j+2} \ge 4d_j$. Together with the construction's scale choices and type weights this yields Theorem 5.3.

Theorem 5.2 (noisy prior-free upper bound)**.** *For every $\beta \in (0, 1]$, every instance whose optimal arm satisfies $\mathrm{LE}(\beta)$, and $T \ge 16k$,*

$$\mathbb{E}[\widehat{\mathrm{ALG}}_T] \;\ge\; \frac{a^2}{64b}\,\frac{R\,q_\beta}{\Phi(k)} \;\ge\; \frac{a^2}{64b^2}\,\frac{\hat{R}\,q_\beta}{\Phi(k)}.$$

With T unknown (phases $16k \cdot 2^j$, conditioning on history as in Lemma D.5) the constant becomes 1024.

Proof sketch. With s and d as above, $s/k \ge q_\beta/h$ and $f_*(d)/m \ge (d/T)^\beta \ge q_\beta/h$; Lemma F.2 with $h = 8\Phi(k)$ gives $\frac{a^2}{8b} \cdot \frac{q_\beta}{8\Phi(k)}$. The anytime clause is proved in Appendix F.4 (Lemma F.4). □

Theorem 5.3 (noisy prior-free lower bound; asymptotic)**.** *Fix $\varepsilon \in (0, \frac{1}{2}]$ and let*

$$r_\varepsilon = \max\{2, \lceil \log_2(1/\varepsilon)\rceil\}, \qquad k_0(\varepsilon) = 2^{\max\{256,\, 3r_\varepsilon\}}.$$

For every $k \ge k_0(\varepsilon)$ and every algorithm that does not read m or β (it may read T and ε), there exist $\beta \in (0, 1]$, an instance whose optimal arm satisfies $\mathrm{LE}(\beta)$, and a fixed in-band noise table such that

$$\mathbb{E}[\widehat{\mathrm{ALG}}_T] \;\le\; \frac{8192}{\varepsilon\,\Phi(k)}\,\hat{R}\,q_\beta.$$

The bound is proved at the horizon $T = 16k'$ (for $k = k'$ a power of two) or $T = 32k'$ (general k, $k' = 2^{\lfloor \log_2 k\rfloor}$, remaining arms identically zero); it is an asymptotic statement in k. Together with Theorem 5.2, the uniform price of not knowing m and β under fixed positive noise is $\phi_\varepsilon(k) = \Theta_\varepsilon(\Phi(k)) = \Theta_\varepsilon\big(\sqrt{\log k/\log\log k}\big)$, with ϕ_ε exactly as defined in Section 2: the supremum runs over all $T \ge 16k$ and the algorithm is chosen before β. The lower bound is established only at the stated hard horizons, which suffices for that supremum; the theorem does not assert the same lower bound at every fixed $T \ge 16k$; the quantifier is "for every algorithm there exist β and an instance", and the $\beta = 1$ guarantee of Theorem 5.1 is not contradicted. Since $k_0(\varepsilon) \ge 2^{256}$, the statement is structural: it separates a bounded price (Theorem 4.1, noiseless) from one that is unbounded in k; at the threshold the explicit constant is tiny (for $\varepsilon = \frac{1}{2}$, $\varepsilon\Phi(2^{256})/8192 \approx 3.6 \cdot 10^{-4}$), so no numerically significant loss is claimed at practical values of k.

Proof sketch. Roadmap: unknown $\beta \to$ many possible reveal depths $\to$ a common pre-reveal history $\to$ one probing budget that must cover all depths $\to$ the factor $\Phi(k)$. In the hard construction the types are those of $\mathcal{J}$ above, type j having dyadic growth length D_j; multiplicative noise flattens the star arm's observations on a hidden prefix $t \le d_j = \eta D_j$ with $\eta \in [\varepsilon/2, \varepsilon]$, so until that prefix ends a growing arm is indistinguishable from a flat one and all types of the hard distribution share one observed history (the hard observation tables are themselves nondecreasing in every arm's pull count, so the separation is not driven by oscillatory corruption). For dyadic k, run with a fixed seed against the all-ones response and let $N(d)$ be the number of arms pulled at least $d+1$ times. Since $d_{j+2} \ge 4d_j$, one run has $\sum_{j\in\mathcal{J}} d_j N(d_j) \le \frac{4}{3}T$ for all types at once, so it cannot be deep enough for every type; weighting the types by Yao's principle then gives the factor $\Phi(k)$ (Appendix F). □

Knowing either prior alone restores a constant price. Knowing β alone, PC tuned to it keeps its guarantee up to the noise factor (Theorem 5.1). Knowing m alone, the *random-exponent elimination* algorithm $\mathrm{RET}(H, \mu)$ (Appendix G; reads k, H and a scale μ, not β or ε) tests the arms in a uniformly random order and eliminates the current arm at the first n where its observed value falls below $\frac{\mu}{2}(n/H)^\alpha$ for an exponent $\alpha \in (0, 1]$ randomized per arm, order and exponent coupled through the same weight U_i; for $k \ge 2$, $\varepsilon \in [0, \frac{1}{2}]$, $H \ge 16k$, $\beta \in (0, 1]$ and a comparator whose curve in this phase satisfies $g_*(t) \ge \mu(t/H)^\beta$ on $1 \le t \le H$, this costs only a constant (Theorem G.1); its unread-T, known-scale form uses advice tied to the evaluation horizon, as in Table 1. Finally, one algorithm that reads none of m, β, T, ε — select uniformly at random one of APC, the phased NPA and the anytime SRMP, and run it — inherits each branch's guarantee divided by 3 (Corollary G.4). Further comparisons and the full discussion of related work are in Appendix I.

6 Discussion

Simultaneous adaptation uses one policy across unknown exponents (Section 2). In the noiseless work bound, depth is chosen only in the analysis (Lemma D.9); shifting retains the original comparator (Theorem 4.2). Under noise, relative-value comparisons survive but increment ordering can fail; the adaptation lower bound concerns the entire information class (Theorem 5.3), while known scale restores constant price (Theorem G.1).

Learning-curve interpretations require our reward, observation and cumulative-objective assumptions. Probe counts differ from running time (Appendix B); Appendix D records a finite-bit variant.

7 Conclusion, Limitations and Open Problems

Conclusion. Without noise, one anytime policy achieves the optimal ratio $\Theta(k^{\beta/(1+\beta)} + k/T)$ simultaneously at every horizon and for every β whose envelope condition holds for an optimal arm (Theorems 4.1, 4.2). Without any envelope assumption, the optimal anytime ratio is $\Theta(\sqrt{k} + k/T)$ (Theorems 3.1–3.3). Under fixed $\varepsilon \in (0, \frac{1}{2}]$, knowing either m or β yields constant price (Theorems 5.1, G.1), while knowing neither costs $\Theta_\varepsilon(\Phi(k))$ in the asymptotic regime below (Theorems 5.2, 5.3).

Limitations. Noise is bounded, multiplicative and obliviously fixed (Blum & Ravichandran, 2025, App. C); only RET uses $\varepsilon \le \frac{1}{2}$. With independent fixed-variance additive Gaussian observations, for every algorithm and fixed k, T, a single constant-mean δ arm in some location forces ratio at least $k - o(1)$ as $\delta \downarrow 0$ against the uncorrupted benchmark; a uniformly selected arm achieves at most k (Appendix H). Known-scale unread-T results use advice tied to the evaluation horizon, not one fixed-advice anytime policy (Appendix B). Theorem 5.3 holds at its hard horizons $T = 16k', 32k'$ for $k \ge k_0(\varepsilon) \ge 2^{256}$; it establishes neither practical-size loss nor an each-horizon lower bound. Constants are unoptimized; Theorem 5.2 requires $T \ge 16k$, with no matching lower bound established for $32 \le T < 16k$.

Open problems. Can the lower bound's linear ε factor be replaced by a constant, closing the gap between $\varepsilon\Phi(k)$ and $64(1-\varepsilon)^{-2}(1+\varepsilon)^2\Phi(k)$? Which signal or benchmark conditions permit prior-free competitive guarantees under stochastic noise?

AI USE STATEMENT

In this work, we used generative AI tools to help develop theoretical models or conceptual frameworks, assist in the writing of proofs, design and provide feedback on research methodology, implement methods, assist with translation, and interpret the theoretical results and their implications. We have not used generative AI tools to formulate mathematical claims, provide critical ingredients for proving mathematical claims, or propose or refine hypotheses. Generating synthetic datasets, cleaning and reformatting datasets, and supporting qualitative and thematic data analysis are not applicable to this theoretical work, which does not involve research datasets or empirical data analysis.

Additionally, we used generative AI tools to create or modify scientific figures, create or edit software code, draft parts of the paper, summarize or analyse existing literature, discover research topics or identify gaps, brainstorm, source and search for information, edit the paper to improve readability, identify relevant literature, format references, suggest the structure of the paper, and propose a title or keywords.

We have reviewed all AI-assisted work. In particular, all mathematical derivations and any AI-assisted code were manually checked and verified for correctness by three human reviewers. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

This is a theoretical paper. Proofs and supporting arguments for our new theoretical results are collected in the appendix: Lemma C.1 and Theorems 3.1, 3.2, C.8, 3.3 and C.10 in Appendix C; Theorems D.4, D.6, 4.1 and 4.2 in Appendix D; Theorems 5.1, 5.2 and 5.3 and Example F.1 in Appendix F; Theorem G.1, Proposition G.2 and Corollary G.4 in Appendix G; the deterministic bounds and the additive-noise argument in Appendix H. All assumptions are stated in Section 2 and Appendix B, and the sources of every quoted prior statement are given in Appendix E. The scripts that check numerical constants and finite-parameter inequalities (including the rounding chain of Theorem 5.3 over the ranges stated in `VERIFICATION_SCOPE.md`) are provided as supplementary material; that file lists, per script, what is checked, over which parameter ranges, and what is not covered by any script.

ETHICS STATEMENT

This work is purely theoretical: it involves no data, no human subjects and no experiments, and we are not aware of any direct ethical concern raised by its content.

REFERENCES

Avrim Blum and Kavya Ravichandran. Nearly-tight approximation guarantees for the improving multi-armed bandits problem. In *Proceedings of The 36th International Conference on Algorithmic Learning Theory*, volume 272 of *Proceedings of Machine Learning Research*, pp. 228–245. PMLR, 2025. arXiv:2404.01198.

Avrim Blum, Marten Garicano, Kavya Ravichandran, and Dravyansh Sharma. Algorithm design and stronger guarantees for the improving multi-armed bandits problem. In *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 585–619. PMLR, 2026. arXiv:2511.10619v2.

Marco Fiandri, Alberto Maria Metelli, and Francesco Trovò. Rising rested bandits: Lower bounds and efficient algorithms. arXiv:2411.14446, 2024.

Hoda Heidari, Michael Kearns, and Aaron Roth. Tight policy regret bounds for improving and decaying bandits. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 1562–1570. AAAI Press / International Joint Conferences on Artificial Intelligence, 2016.

Alberto Maria Metelli, Francesco Trovò, Matteo Pirola, and Marcello Restelli. Stochastic rising bandits. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 15421–15457. PMLR, 2022.

Marco Mussi, Alessandro Montenegro, Francesco Trovò, Marcello Restelli, and Alberto Maria Metelli. Best arm identification for stochastic rising bandits. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 36953–36989. PMLR, 2024.

Vishakha Patil, Vineet Nair, Ganesh Ghalme, and Arindam Khan. Mitigating disparity while maximizing reward: Tight anytime guarantee for improving bandits. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pp. 4100–4108. International Joint Conferences on Artificial Intelligence Organization, 2023.

Seockbean Song, Youngsik Yoon, Siwei Wang, Wei Chen, and Jungseul Ok. Combinatorial rising bandits. In *International Conference on Learning Representations (ICLR)*, 2026. arXiv:2412.00798.

Andrew Chi-Chih Yao. Probabilistic computations: Toward a unified measure of complexity. In *Proceedings of the 18th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 222–227, 1977.

A Summary of the guarantees

Table 2 summarizes selected competitive-ratio guarantees proved in this paper, with their constants, the horizon range on which each is stated, and the theorem it is taken from.

B Conventions in full

Comparator convention. Several arms may be cumulatively optimal with different terminal values $f_i(T)$, so the choice matters and is never left implicit: in the prior-free theorems the comparator is *any* cumulatively optimal arm satisfying the envelope condition of Section 2 (the algorithm does not read this choice, which only enters the analysis), whereas in Theorem G.1, where a scale is supplied to the algorithm, the comparator is the arm that scale refers to. All quantities m, $\hat{R}$ (Section 5) and the envelope hypotheses refer to the same comparator; since $P_*(T) = R$ for every admissible choice, the noisy benchmark changes by at most a factor b/a between comparators. Pulling $*$ throughout is optimal in hindsight (Heidari et al., 2016, Prop. 1); Lemma C.1(6) gives a self-contained proof using only nondecreasing rewards.

Computational model. We idealise the evaluation of the displayed real-valued keys and count exact comparisons. The algorithms with proved positive guarantees use $O(k)$ working storage; standard heap implementations of RMP and SRMP use $O(k)$ setup and $O(\log k)$ time per probe, and sorting the weights for RET takes $O(k \log k)$. The randomness used is $O(k)$ independent draws per phase — not $O(k)$ random bits — and Appendix D records a finite-bit variant. We stress that a small *number of probes* does not mean a small running time: in the sparse regime SRMP may probe far fewer than k arms, but it includes every arm once $H \geq 128k$, and generating its k inclusion variables coordinatewise already costs $\Theta(k)$.

Uniform price of adaptation: further conventions. For known horizons $\mathfrak{A}$ consists of families $\{A_T\}$, one algorithm per horizon; for unknown horizons it consists of single anytime policies, evaluated at every T. Whether the algorithms may read ε is stated in each result; our upper bounds on ϕ_ε do not read it, and the lower bound holds even for algorithms that do. Table 1 is stated in terms of ϕ_ε. Ratios are set to $+\infty$ when the numerator is positive and the expected reward is zero. Whether T or ε may be read does not change the order in any cell: the upper bounds do not read them, the lower bound allows them. Each cell is a separate infimum: ϕ_ε is defined relative to a class $\mathfrak{A}$ fixed by what the algorithms may read, so the entries of different rows are values of the same functional taken over different $\mathfrak{A}$. In the known-scale cells the scale advice refers to the same designated optimal arm as the envelope hypothesis and the noisy benchmark; the arm's index itself is not supplied

Table 2: Selected competitive-ratio guarantees. Except for entries explicitly stated as bounds on ρ or as impossibility results, payoff entries are lower bounds on expected reward against $R = \mathrm{OPT}_T$. For these positive noisy bounds, replacing R by $\hat{R}/b$ gives the corresponding guarantee against $\hat{R}$. The entries citing Theorem 5.3 are existential upper bounds on algorithmic reward and are already stated against $\hat{R}$. Throughout, $\varepsilon \in [0, \frac{1}{2}]$ as in Section 2. In the right-hand column, T is not an input. The known-scale entry uses advice $m/4$ tied to the evaluation horizon, as specified in Appendix B; it does not assert that one fixed-advice policy satisfies this reward bound at every fixed evaluation horizon. Notation as in Section 2: $\gamma = \beta/(1+\beta)$, $q_\beta = k^{-\gamma}$, $K_\beta = kq_\beta$, $Q_\beta = \min\{q_\beta, T/k\}$; $\mathrm{LE}(\beta)$ is assumed for an optimal arm only. Matching lower bounds, for all $k, T \geq 1$: $\rho(k,T) \geq \frac{1}{4}(\sqrt{k} + k/T)$ for scale-oblivious algorithms (Thm. 3.2) and $\rho_\beta(k,T) \geq (k^\gamma + k/T)/(\beta+3)$ (Thm. C.10(ii)). †Theorem 5.3 is asymptotic in k: it holds for $k \geq k_0(\varepsilon) = 2^{\max\{256,\, 3r_\varepsilon\}}$, at the horizons $T = 16k'$ ($k = k'$ a power of two) or $T = 32k'$ ($k' = 2^{\lfloor \log_2 k \rfloor}$), for every algorithm that does not read m or β (it may read T and ε), with β, the instance and the noise table chosen against the algorithm.

inputs when T is known	ε	T known	T not an input
k, T; not m	0	$R/(4\sqrt{3}\sqrt{k})$, $T \geq 2\lfloor\sqrt{k}\rfloor$ (Thm. 3.1); $\rho(k,T) \leq 24(\sqrt{k} + k/T)$, all T (Thm. 3.2)	$R/(1536(\sqrt{k} + k/T))$, all T (Thm. 3.3)
	> 0	PC: $\frac{a^2}{b}\frac{R}{16\sqrt{k}}$ for $T \geq 2\lfloor\sqrt{k}\rfloor$; truncated PC: $\frac{a^2}{b}\frac{R}{24(\sqrt{k}+k/T)}$ for all $T \geq 1$; APC: $\frac{a^2}{b}\frac{R}{1536(\sqrt{k}+k/T)}$ for all $T \geq 1$. None of these algorithms reads ε (Thm. 5.1)	
k, T, β; not m	0	$Rq_\beta/16$, $T \geq 2\lceil K_\beta \rceil$; $\rho_\beta(k,T) \leq 24(k^\gamma + k/T)$, all T (Thm. C.10)	$Rq_\beta/128$, $T \geq 2\lceil K_\beta \rceil$ (Thm. C.10(iii))
	> 0	$\frac{a^2}{b}\frac{Rq_\beta}{16}$ and, anytime, $\frac{a^2}{b}\frac{Rq_\beta}{256}$, both for $T \geq 2\lceil K_\beta \rceil$ (Thm. 5.1)	
k, T; not m, β	0	$\frac{3}{256}Rq_\beta$, $T \geq 16k$ (Thm. D.4); $RQ_\beta/8192$, all T (Thm. 4.1)	$\frac{3}{4096}Rq_\beta$, $T \geq 16k$ (Thm. D.6); $2^{-19}RQ_\beta$, all T (Thm. 4.2)
	> 0	$\frac{a^2}{64b}\frac{Rq_\beta}{\Phi(k)}$, $T \geq 16k$ (Thm. 5.2); no algorithm beats $\frac{8192}{\varepsilon\Phi(k)}\hat{R}q_\beta$ on some instance† (Thm. 5.3)	$\frac{a^2}{1024b}\frac{Rq_\beta}{\Phi(k)}$, $T \geq 16k$ (Thm. 5.2); the same lower bound†
k, T, m; not β	≥ 0	$\frac{a}{2048b}Rq_\beta$, $T \geq 16k$ (Thm. G.1)	$\frac{a}{32768b}Rq_\beta$, $T \geq 16k$, with advice $m/4$ tied to the evaluation horizon (Thm. G.1; see the caption)

to the algorithm. If instead the advice is allowed to come from an arbitrary cumulatively optimal arm, with terminal value $m_0 \leq 2m^\circ$ by Lemma C.1(5), running $\mathrm{RET}(T, m_0/2)$ for known T and phases with advice $m_0/8$ for unknown T gives the constants 4096 and 65536 respectively, so the $\Theta(1)$ entries are unchanged. The known-scale, unread-T entries use advice tied to the evaluation horizon (paragraph "Scale advice and unknown horizons"), as in Table 2; they are not single-policy anytime guarantees.

Scale advice and unknown horizons. The scale $m = f_*(T)$ depends on the horizon. A result of the form "m known, T unknown" therefore means: the algorithm does not read T, but is given one number μ — an advice tied to the horizon at which it will be evaluated — and the guarantee is asserted at that horizon. It is not the stronger statement that a single policy with a single advice does well at every stopping time simultaneously, and we never claim the latter.

Versions. All quotations from Blum & Ravichandran (2025) follow the ALT 2025 proceedings version (PMLR 272), whose appendix on noisy rewards is not contained in the 12-page arXiv v1 of April 2024; quotations from Blum et al. (2026) follow arXiv v2 (May 2026).

C Proofs for Section 3 (unknown scale)

Proof architecture. Lemma C.1 supplies the deterministic inequalities used throughout. The sampling inequality of Lemma C.2 yields the two-event guarantee of Lemma C.3, from which Theorem 3.1 and the upper half of Theorem 3.2 follow. The independent delayed-reveal argument, Lemma C.5 and Corollary C.6, supplies the lower half. Lemma C.7 preserves the sampling argument conditional on the past and supports both horizon-free schedules. Theorem C.10 then applies these upper- and lower-bound mechanisms to the envelope classes.

C.1 Consequences of discrete concavity

Lemma C.1 (basic facts)**.** *Let f be nondecreasing, discretely concave, $f(0) = 0$, and $P(n) = \sum_{t\le n} f(t)$. Then:*

(1) $f(t) \ge \frac{t}{T} f(T)$ for $1 \le t \le T$;

(2) $P(n) \ge f(T)\frac{n(n+1)}{2T}$ for $0 \le n \le T$; more precisely $P(u) \ge \frac{u(u+1)}{T(T+1)}P(T) \ge (u/T)^2 P(T)$ for $1 \le u \le T$ (a ratio form is available when $P(T) > 0$);

(3) $R \le mT$;

(4) if $\lceil T/2 \rceil \le r \le T$ then $P_*(r) \ge R/4$;

(5) $M \le \frac{2T}{T+1} m \le 2m$ and $R \ge \frac{T+1}{2} M$;

(6) for any allocation $\sum_i n_i = T$ of pulls, $\sum_i P_i(n_i) \le R$.

Proof. Let $\Delta_t = f(t) - f(t-1)$; then $\Delta_1 \ge \Delta_2 \ge \cdots \ge \Delta_T \ge 0$ and $f(t)/t$ is the average of $\Delta_1, \dots, \Delta_t$, hence nonincreasing in t; this gives (1) and, summing, the first form of (2). For the second form write $P(n) = \sum_{t\le n} t \cdot (f(t)/t)$; the weighted average $P(n)/\sum_{t\le n} t$ of the nonincreasing sequence $f(t)/t$ with weights t is nonincreasing in n (appending a smaller term cannot raise the average), so $P(u)/P(T) \ge \frac{u(u+1)/2}{T(T+1)/2}$. (3) is monotonicity. (4) follows from (2) with $u = r \ge T/2$: $\frac{r(r+1)}{T(T+1)} \ge \frac{(T/2)(T/2+1)}{T(T+1)} \ge \frac{1}{4}$. For (5), (2) with $n = T$ gives $P_i(T) \ge f_i(T)\frac{T+1}{2}$, so $M\frac{T+1}{2} \le R \le mT$. For (6), the prefix averages $P_i(n)/n$ of the nondecreasing sequence f_i are nondecreasing, so $P_i(n_i) \le \frac{n_i}{T} P_i(T) \le \frac{n_i}{T} R$ and summing over i gives R. □

C.2 Probe-and-commit

Algorithm 1 $\mathrm{PC}(s, \tau)$ (probe-and-commit). Inputs: k, T, integers $1 \le s \le k$, $\tau \ge 1$ with $s\tau \le \lfloor T/2 \rfloor$.

1: Draw a uniformly random set S of s distinct arms; pull each arm of S exactly τ times.
2: Let $c = \arg\max_{i\in S} f_i(\tau)$ (any fixed tie-breaking rule).
3: Pull c for the remaining $r = T - s\tau \ge \lceil T/2 \rceil$ steps.

The algorithm reads only k and T. The analysis rests on a negative-correlation inequality for the random sample.

Lemma C.2 (sample negative correlation)**.** *Fix $F \subseteq [k] \setminus \{*\}$ with $h = |F|$, and let $A = \{* \in S\}$, $B = \{S \cap F \ne \emptyset\}$. Then $\mathbb{P}(A \cap \neg B) \ge \frac{s}{k}\mathbb{P}(\neg B)$.*

Proof. If $s > k - h$ then $\mathbb{P}(\neg B) = 0$. Otherwise, conditioned on $\neg B$, S is a uniform s-subset of the $k - h$ arms outside F, a set containing $*$, so $\mathbb{P}(A \mid \neg B) = \frac{s}{k-h} \ge \frac{s}{k}$. □

Lemma C.3 (core guarantee)**.** *For every instance and every admissible (s, τ), with $r = T - s\tau$,*

$$\mathbb{E}[\mathrm{ALG}_T] \;\ge\; R \cdot \min\Big\{\frac{r\tau}{T^2},\; \frac{s\,r(r+1)}{2kT^2}\Big\} \;\ge\; R \cdot \min\Big\{\frac{\tau}{2T},\; \frac{s}{8k}\Big\}.$$

Proof. Let $F = \{i \neq * : f_i(\tau) \geq f_*(\tau)\}$, a set determined by the instance, and let A, B be as in Lemma C.2. On B (some "fooling" arm was sampled) the committed arm satisfies $f_c(\tau) \geq f_*(\tau) \geq m\tau/T$ by Lemma C.1(1), and since f_c is nondecreasing the commit phase earns at least $r\, m\tau/T \geq R\, r\tau/T^2$ by Lemma C.1(3). On $A \cap \neg B$ the optimal arm is the unique maximizer, so $c = *$ and the commit phase earns $\sum_{j=1}^{r} f_*(\tau+j) \geq \sum_{j=1}^{r} f_*(j) \geq R\frac{r(r+1)}{2T^2}$ by Lemma C.1(2),(3). The two events are disjoint; with $p = \mathbb{P}(B)$ and Lemma C.2, $\mathbb{E}[\mathrm{ALG}_T]/R \geq \frac{r\tau}{T^2}p + \frac{s\, r(r+1)}{2kT^2}(1-p)$, which is at least the minimum of the two coefficients. The second form uses $r \geq T/2$. Probe rewards are nonnegative and were dropped. □

Proof of Theorem 3.1. Since $\lfloor x \rfloor \geq x/2$ for $x \geq 1$, $\tau \geq T/(4s)$ and $s\tau \leq T/2$; since $s^2 \leq k$, $\frac{1}{8s} \geq \frac{s}{8k}$; and $\lfloor\sqrt{k}\rfloor \geq \sqrt{k}/2$. Lemma C.3 gives the first bound. For the sharper constant, use the two branches of Lemma C.3 with the refined Lemma C.1(2) and the following inequality: for $1 \leq \tau \leq (T+1)/2$, $f(\tau)/P(T) \geq \frac{2\tau}{T(T+1)}$. (Write $f = \sum_{d=1}^{T} c_d \min\{t, d\}$ with $c_d = \Delta_d - \Delta_{d+1} \geq 0$, $\Delta_{T+1} := 0$; for each basis function $\min\{t, d\}$ the ratio of its prefix sum $dT - d(d-1)/2$ to its value at τ is at most $T(T+1)/(2\tau)$ when $d \geq \tau$ (maximized at $d = T$) and equals $T - (d-1)/2 \leq T \leq T(T+1)/(2\tau)$ when $d < \tau$; a nonnegative combination preserves the bound.) Hence $\mathbb{E}[\mathrm{ALG}_T]/R \geq \min\{\frac{2r\tau}{T(T+1)}, \frac{s}{k}\frac{r(r+1)}{T(T+1)}\}$. With $r \geq T/2$ and $\tau \geq T/(4s)$ the first term is at least $\frac{T}{4s(T+1)} \geq \frac{1}{6\sqrt{k}}$ (as $T \geq 2$), the second at least $\frac{s}{4k} \geq \frac{1}{4\sqrt{3}\sqrt{k}}$ because $k \leq s^2 + 2s \leq 3s^2$. The smaller is the second. □

Remark C.4. Theorem 3.1 should be compared with (Blum & Ravichandran, 2025, Thm. 11), which requires $T > 4k$ and loses $\log k$. The gap to the lower bound $\sqrt{k}/3$ of Blum & Ravichandran (2025) is a constant factor $4\sqrt{3} \cdot 3 < 21$. Extrapolation-based thresholds cannot be made scale-free by simply halving a guessed threshold: for the instance $T = k^2$, $f_*(t) = 3t/(4T)$ and $k-1$ decoys $\min(t, k)/T$, a threshold $\mu = 1$ rejects $*$ after one pull and each decoy absorbs $k+1$ pulls, exhausting the horizon with ratio $\Theta(k)$; guessing the level at random is exactly the $\log k$ of Blum & Ravichandran (2025).

C.3 Lower bounds and the full horizon dependence

The following "delayed-reveal" lemma is a restatement, with integer bookkeeping, of the hard-distribution argument of Blum & Ravichandran (2025) (see also (Blum et al., 2026, Lemma B.9), of which it is a variant); we include it because it yields the constant 3 without divisibility conditions and the short-horizon term.

Lemma C.5 (delayed reveal)**.** *Let g be nonnegative, nondecreasing, discretely concave with $g(0) = 0$, and $0 \leq d \leq T$ an integer. Consider the distribution $\mathcal{D}_{g,d}$ over instances in which a uniformly random arm G has reward g and every other arm has $h_d(t) = g(\min\{t, d\})$; let $R_g = \sum_{t \leq T} g(t)$. For every randomized algorithm A,*

$$\mathbb{E}_{I \sim \mathcal{D}_{g,d}} \mathbb{E}[\mathrm{reward}(A, I)] \;\leq\; T\, g(d) + \frac{\lfloor T/(d+1) \rfloor}{k}\, R_g,$$

the inner expectation being over the internal randomness of A. Consequently, for every A there is an instance in the support of $\mathcal{D}_{g,d}$ on which $\mathbb{E}[\mathrm{reward}]$ is at most this quantity, while $\mathrm{OPT}_T = R_g$ on every instance of the support.

Proof. Fix the algorithm's random seed. Run it against the virtual instance in which every arm is h_d; let D be the set of arms pulled at least $d+1$ times there, so $|D| \leq \lfloor T/(d+1) \rfloor$. Let E be the event that, in the real run, the good arm is pulled for the $(d+1)$-st time within the horizon (this is not required to be observationally recognizable). Before that pull the real and virtual runs coincide, hence E holds iff $G \in D$, and $\mathbb{P}_G(E) \leq |D|/k$. On E^c every pull yields at most $g(d)$; on E the total reward is at most R_g by Lemma C.1(6) applied to the real instance. Averaging over the seed gives the displayed bound, and the last sentence is the usual averaging (Yao's principle Yao (1977)) step: a random variable is somewhere below its mean. □

Corollary C.6. *For every randomized algorithm there is an instance on which its competitive ratio is at least (a) $\sqrt{k}/3$, for all $k, T \geq 1$ (take $g(t) = t/T$, $d = \lfloor T/\sqrt{k} \rfloor$); and (b) k/T (take $g(t) = t/T$, $d = 0$: one linear arm and $k-1$ arms identically zero).*

Proof. (a) $R_g = \frac{T+1}{2}$; $2d/(T+1) < 2/\sqrt{k}$ and $T/(k(d+1)) < 1/\sqrt{k}$ since $d+1 > T/\sqrt{k}$; if $T < \sqrt{k}$ then $d = 0$ and the construction remains legal. (b) is the case $d = 0$ of Lemma C.5, where $g(0) = 0$ kills the first term. Since Blum & Ravichandran (2025) phrase reward functions as "monotone increasing", note that the zero decoys may be replaced by arms of slope $\eta \to 0$ without changing the order of magnitude. Both families consist of linear-plus-truncation curves; they belong to $\mathcal{I}_1$, and for $T \geq 2$ the linear good arm violates $\mathrm{LE}(\beta)$ whenever $\beta < 1$, so these instances may not be used for the corresponding ρ_β (see Theorem C.10(ii) for the version that stays inside $\mathcal{I}_\beta$). At $T = 1$ every admissible arm satisfies every $\mathrm{LE}(\beta)$, so the restriction is vacuous there. □

Proof of Theorem 3.2. Upper bound: for $T = 1$ pull a uniformly random arm (ratio $\leq k$). For $T \geq 2$ run PC with $s = \min\{\lfloor\sqrt{k}\rfloor, \lfloor T/2\rfloor\}$ and $\tau = \lfloor T/(2s)\rfloor$ (so $s \geq 1$, $\tau \geq 1$ and $s\tau \leq \lfloor T/2\rfloor$); the computation of Theorem 3.1 applied to Lemma C.3 gives $\mathbb{E}[\mathrm{ALG}_T] \geq \frac{s}{8k}R$, and $\lfloor T/2\rfloor \geq T/3$ for $T \geq 2$, so $R/\mathbb{E}[\mathrm{ALG}_T] \leq 8k/s \leq \max\{16\sqrt{k}, 24k/T\}$. Note that this is a corollary of Lemma C.3 with truncated parameters; the hypothesis $T \geq 2\lfloor\sqrt{k}\rfloor$ of Theorem 3.1 is not needed. Lower bound: Corollary C.6 gives $\max\{\sqrt{k}/3, k/T\} \geq \frac{1}{4}(\sqrt{k} + k/T)$. □

C.4 Unknown horizon

Lemma C.7 (conditioning on history)**.** *Suppose that at the start of a phase $t_0 = \sum_i n_i$ pulls have been made (arm i has been pulled n_i times), and the phase runs $\mathrm{PC}(s, \tau)$ with a fresh independent sample, $1 \leq s \leq k$, and commit length r, where $t_0 + s\tau + r \leq T$ (so every value $f_i(n_i + \tau)$ used below is within the domain). Then $\mathbb{E}[$phase reward $\mid$ history$] \geq R \min\{\frac{r\tau}{T^2}, \frac{s\,r(r+1)}{2kT^2}\}$.*

Proof. Define $F_H = \{i \neq * : f_i(n_i + \tau) \geq f_*(n_* + \tau)\}$, a fixed set given the history; Lemma C.2 applies to the fresh sample. On B the committed arm's probe value is at least $f_*(n_* + \tau) \geq f_*(\tau)$; on $A \cap \neg B$ the committed arm is $*$ and earns $\sum_{j \leq r} f_*(n_* + \tau + j) \geq \sum_{j \leq r} f_*(j)$. Conclude as in Lemma C.3. □

The obvious doubling schedule, with phases $2s, 4s, 8s, \ldots$, works as soon as the horizon exceeds the first phase.

The natural doubling schedule (phases of lengths $2s \cdot 2^j$, $s = \lfloor\sqrt{k}\rfloor$, each running PC with a fresh sample, pull counts never reset) reads only k and gives $\mathbb{E}[\mathrm{ALG}_T] \geq R/(256\sqrt{k})$ for every $T \geq 2s$ (Theorem C.8); the restriction $T \geq 2s$ is real: for $k = n^2$ and $T = n$ its ratio is $n^2(n+1)/2$, while $\Theta(n)$ is attainable (Example C.9). The *anytime probe-and-commit* algorithm APC (Appendix C) starts at length 1 instead: phases of planned lengths $H_j = 2^j$, where a phase of length 1 pulls one uniformly random arm and a phase of length $H \geq 2$ runs PC on a fresh sample with the truncated parameters $s = \min\{\lfloor\sqrt{k}\rfloor, \lfloor H/2\rfloor\}$, $\tau = \lfloor H/(2s)\rfloor$, $r = H - s\tau$, committing to the arm of S whose τ-th observation *in this phase* is largest; true pull counts are never reset.

Theorem C.8 (unknown horizon, $T \geq 2\lfloor\sqrt{k}\rfloor$)**.** *Let $s = \lfloor\sqrt{k}\rfloor$ and run phases of lengths $L_j = 2s \cdot 2^j$, $j = 0, 1, \ldots$, each running $\mathrm{PC}(s, L_j/(2s))$ with a fresh sample and without resetting pull counts. The algorithm reads only k, and for every $T \geq 2s$, $\mathbb{E}[\mathrm{ALG}_T] \geq R/(256\sqrt{k})$.*

Proof. Let L be the length of the last completed phase; $2L - 2s \leq T < 4L - 2s$, so $L > T/4$. Lemma C.7 with $\tau = L/(2s)$, $r = L/2$ gives $\mathbb{E} \geq R(L/T)^2 \min\{\frac{1}{4s}, \frac{s}{8k}\} = R(L/T)^2 \frac{s}{8k} \geq R/(256\sqrt{k})$. Compare (Blum & Ravichandran, 2025, Lemma 12): $\mathrm{OPT}_T/(8192\sqrt{k}\log(128k))$ for $T > 4k$. □

This schedule guarantees nothing before its first phase ends, and that restriction is real, not an artefact of the analysis.

Example C.9 (the schedule, not the horizon, is the obstruction). Let $n \geq 2$, $k = n^2$, $T = n$, $f_*(t) = t$ and all other arms identically zero. The algorithm of Theorem C.8 has $s = n$ and first phase $2n > T$: it spends the whole horizon on the n initial probes of the sampled arms and never commits, so $\mathbb{E}[\mathrm{ALG}_T] = \mathbb{P}(* \in S) \cdot f_*(1) = 1/n$ while $R = n(n+1)/2$; its ratio is $n^2(n+1)/2$. Yet the optimal ratio at these parameters is $\Theta(\sqrt{k} + k/T) = \Theta(n)$ by Theorem 3.2, attained by PC

with truncated parameters when T is known. (On a parameter range, the all-horizon minimax value is $O(\sqrt{k})$ if and only if $T = \Omega(\sqrt{k})$ uniformly on that range; equivalently, an $O(\sqrt{k})$ guarantee is ruled out exactly when $\sqrt{k}/T$ is unbounded on the range. In particular $T = o(\sqrt{k})$ rules it out, while the range $T = \Omega(\sqrt{k})$, which still contains parameters with $T < 2s$, does not.) The next algorithm removes the restriction by starting at length 1.

Algorithm 2 APC (anytime probe-and-commit). Reads only k; it may be stopped at any time, and its true pull counts are never reset.

1: If $k = 1$: always pull the unique arm. Otherwise run phases $j = 0, 1, 2, \ldots$ of planned lengths $H_j = 2^j$.
2: **Phase of length H:** if $H = 1$, pull one uniformly random arm once.
3: Otherwise let $s = \min\{\lfloor\sqrt{k}\rfloor, \lfloor H/2\rfloor\}$, $\tau = \lfloor H/(2s)\rfloor$ and $r = H - s\tau$.
4: Draw a fresh uniformly random set S of s distinct arms and pull each of them τ times within this phase.
5: Commit the remaining r pulls of the phase to the arm of S whose τ-th observation *in this phase* is largest (ties broken by index).

Let $H = 2^j$ be the length of the last phase completed within the horizon; then $x := H/T > 1/4$. The core guarantee holds conditionally on the history (Lemma C.7: the sample is fresh, and since rewards are nondecreasing the shifted curves $t \mapsto f_i(n_i + t)$ dominate the f_i), so that phase alone earns $\mathbb{E} \geq \frac{Rx^2}{8} \min\{1/s,\, s/k\}$. This is at least $Rx^2/(16\sqrt{k})$ when $s = \lfloor\sqrt{k}\rfloor$, and at least $Rx^3T/(24k)$ when $s = \lfloor H/2\rfloor < \lfloor\sqrt{k}\rfloor$; both are at least $Rx^3/(24(\sqrt{k} + k/T))$, and $x^3 > 1/64$ gives $1536 = 24 \cdot 64$. The case $H = 1$ (i.e. $T \leq 2$) is checked directly.

Proof of Theorem 3.3. For $k = 1$ the algorithm earns R. Let $k \geq 2$. Each phase of length $H \geq 2$ is admissible: $s \geq 1$, $H/(2s) \geq 1$ so $\tau \geq 1$, and $s\tau \leq H/2$, whence $r \geq \lceil H/2\rceil$. Let $H = 2^j$ be the length of the last phase completed within the horizon. The first $j + 1$ phases occupy $2H - 1$ steps, so $2H - 1 \leq T$; if the next phase also completed we would have $4H - 1 \leq T$, so $T < 4H - 1$. Hence $x := H/T > 1/4$.

Case $H \geq 2$. The phase satisfies the hypotheses of Lemma C.7 (fresh sample, and $t_0 + s\tau + r = t_0 + H \leq T$ because the phase is completed), so, using $r \geq H/2$ and $\tau \geq H/(4s)$ (valid since $H/(2s) \geq 1$ and $\lfloor y\rfloor \geq y/2$ for $y \geq 1$),

$$\mathbb{E}[\mathrm{ALG}_T] \;\geq\; \mathbb{E}[\text{this phase}] \;\geq\; R \min\left\{\frac{r\tau}{T^2}, \frac{s\,r(r+1)}{2kT^2}\right\} \;\geq\; \frac{Rx^2}{8} \min\left\{\frac{1}{s}, \frac{s}{k}\right\},$$

all other phases contributing nonnegative rewards. If $s = \lfloor\sqrt{k}\rfloor$ then $1/s \geq 1/\sqrt{k}$ and $s/k \geq 1/(2\sqrt{k})$, so the last factor is at least $1/(2\sqrt{k})$ and $\mathbb{E}[\mathrm{ALG}_T] \geq Rx^2/(16\sqrt{k})$. If instead $s = \lfloor H/2\rfloor < \lfloor\sqrt{k}\rfloor$ then $s < \sqrt{k}$, so $\min\{1/s, s/k\} = s/k \geq H/(3k)$ (as $\lfloor H/2\rfloor \geq H/3$ for $H \geq 2$), and $\mathbb{E}[\mathrm{ALG}_T] \geq Rx^2H/(24k) = Rx^3T/(24k)$. Since $x \leq 1$, both cases give

$$\mathbb{E}[\mathrm{ALG}_T] \;\geq\; Rx^3 \min\left\{\frac{1}{16\sqrt{k}}, \frac{T}{24k}\right\} \;\geq\; \frac{Rx^3}{24\,(\sqrt{k} + k/T)} \;\geq\; \frac{R}{1536\,(\sqrt{k} + k/T)},$$

where the middle inequality is checked separately for the two branches ($24(\sqrt{k} + k/T) \geq 16\sqrt{k}$; and $(\sqrt{k} + k/T)\,T/k \geq 1$), and the last uses $x^3 > 1/64$.

Case $H = 1$. Then $1 \leq T < 3$. The completed phase of length 1 pulls a uniformly random arm, so $\mathbb{E}[\mathrm{ALG}_T] \geq \frac{1}{k} f_*(1) \geq \frac{2R}{kT(T+1)}$ by Lemma C.1(2), which is R/k for $T = 1$ and $R/(3k)$ for $T = 2$; both exceed $R/(1536(\sqrt{k} + k/T)) \leq R/(768k)$. □

C.5 Known envelope exponent

Theorem C.10 (known β, unknown scale)**.** *Let $\beta \in (0, 1]$ be known and suppose the comparator arm satisfies* $\mathrm{LE}(\beta)$. *For $k \geq 2$, $T \geq 2\lceil K_\beta\rceil$, $s = \lceil K_\beta\rceil$, $\tau = \lfloor T/(2s)\rfloor$, $\mathrm{PC}(s, \tau)$ satisfies $\mathbb{E}[\mathrm{ALG}_T] \geq R/(16k^\gamma)$. Moreover:*

(i) for every $k \geq 2$ and $T \geq 2$ there is an explicit distribution over instances, all of whose arms satisfy $\mathrm{LE}(\beta)$ *and whose envelope exponent equals β, on which* every *(possibly randomized) algorithm has expected reward at most $(\beta+2)k^{-\gamma}\mathrm{OPT}_T$; hence for every algorithm some instance of that family forces ratio at least $k^\gamma/(\beta+2)$, and $\rho_\beta(k,T) \geq k^\gamma/(\beta+2)$;*

(ii) for all $T \geq 1$, $\frac{k^\gamma + k/T}{\beta+3} \leq \rho_\beta(k,T) \leq 24(k^\gamma + k/T)$;

(iii) with T unknown: for $k \geq 2$ and $s = \lceil K_\beta \rceil$, independent PC *phases of lengths $2s \cdot 2^j$ ($j = 0, 1, \dots$) without resetting pull counts give $\mathbb{E}[\mathrm{ALG}_T] \geq R/(128k^\gamma)$ for every $T \geq 2s$ (for $k = 1$ pull the only arm).*

Proof. In Lemma C.3 replace the bound $f_*(\tau) \geq m\tau/T$ by $f_*(\tau) \geq m(\tau/T)^\beta$, which gives

$$\mathbb{E}[\mathrm{ALG}_T]/R \;\geq\; \min\Big\{\frac{1}{2(4s)^\beta},\; \frac{s}{8k}\Big\};$$

since $K_\beta \leq s \leq 2K_\beta$, $\frac{1}{2(4s)^\beta} \geq \frac{1}{2\cdot 8^\beta k^\gamma} \geq \frac{1}{16k^\gamma}$ and $\frac{s}{8k} \geq \frac{1}{8k^\gamma}$.

(i) Apply Lemma C.5 with $g(t) = (t/T)^\beta$ (discretely concave: its increments are integrals of the decreasing function $\beta x^{\beta-1}$) and $d = \lfloor T/K_\beta \rfloor$. Every arm of the support satisfies $\mathrm{LE}(\beta)$: the good arm with equality, and $h_d(t) = g(\min\{t,d\})$ because $h_d(t) = g(t) \geq h_d(T)(t/T)^\beta$ for $t \leq d$ and $h_d(t) = h_d(T) \geq h_d(T)(t/T)^\beta$ for $t > d$. The good arm violates $\mathrm{LE}(\beta')$ for every $\beta' < \beta$ at $t = 1$ (here $T \geq 2$), so the envelope exponent of each instance is exactly β. With $R_g = \sum_{t\leq T}(t/T)^\beta \geq T/(\beta+1)$, Lemma C.5 bounds the expected reward, averaged over the hard distribution, by $(\beta+1)(d/T)^\beta R_g + \frac{T}{k(d+1)}R_g \leq (\beta+2)k^{-\gamma}R_g$; the last sentence of Lemma C.5 converts this into a fixed instance for each algorithm. This reuses the power-curve/truncation construction of (Blum et al., 2026, Thm. 3.6), with Lemma C.5 supplying a self-contained integer version valid for all $T \geq 2$; (Blum et al., 2026, Thm. B.14) states a different explicit threshold, and neither threshold dominates the other for all β, k.

(ii) For $T = 1$ pull a uniformly random arm; the one-positive-arm distribution of Lemma C.5 with $d = 0$ gives the matching lower bound, so $\rho_\beta(k,1) = k$ and both displayed bounds hold. Assume henceforth $T \geq 2$. Upper bound: run PC with $s = \min\{\lceil K_\beta \rceil, \lfloor T/2 \rfloor\}$ and $\tau = \lfloor T/(2s) \rfloor$, so that $s \geq 1$ and $\tau \geq 1$. Lemma C.3 with $f_*(\tau) \geq m(\tau/T)^\beta$ gives $\mathbb{E}[\mathrm{ALG}_T]/R \geq \min\{\frac{1}{2(4s)^\beta}, \frac{s}{8k}\}$; using $s \leq 2K_\beta$ in the first term and $s \geq \min\{K_\beta, T/3\}$ in the second (valid since $\lfloor T/2 \rfloor \geq T/3$ for $T \geq 2$) this is at least $\min\{\frac{q_\beta}{16}, \frac{T}{24k}\} \geq \frac{1}{24(k^\gamma + k/T)}$. Lower bound: besides (i) we need a k/T bound *inside* $\mathcal{I}_\beta$, so we cannot invoke Corollary C.6(b), whose good arm is linear and violates $\mathrm{LE}(\beta)$ for $\beta < 1$. Instead apply Lemma C.5 with $g(t) = (t/T)^\beta$ and $d = 0$: the good arm satisfies $\mathrm{LE}(\beta)$ with equality, the other arms are identically zero and satisfy every $\mathrm{LE}(\beta')$, and $g(0) = 0$ leaves $\mathbb{E}[\text{reward}] \leq \frac{T}{k}R_g$, i.e. ratio at least k/T. Finally $\max\{k^\gamma/(\beta+2),\, k/T\} \geq (k^\gamma + k/T)/(\beta+3)$.

(iii) As in Theorem C.8: in the last completed phase $L > T/4$ the probe depth is exactly $\tau = L/(2s)$ and $r = L/2$ (no rounding); using $f_*(\tau) \geq m(\tau/T)^\beta$ and $\sum_{j\leq r} f_*(j) \geq m\, r^{1+\beta}/((1+\beta)T^\beta)$ in Lemma C.7, the phase earns at least $R(L/T)^{1+\beta}\min\{\frac{1}{2(2s)^\beta}, \frac{s}{2^{1+\beta}(1+\beta)k}\} \geq R(L/T)^{1+\beta}/(8k^\gamma)$, and $(L/T)^{1+\beta} \geq 1/16$ gives 128. □

Remark C.11. Part (ii) is a *class-wise* statement: for each β separately, ρ_β compares algorithms that may depend on β. It does not by itself produce one algorithm that is simultaneously near-optimal for all β; that stronger property is the subject of Section 4.

Remark C.12 (maximum single pull). For the objective of maximizing the largest single reward ((Blum & Ravichandran, 2025, App. B)), the last pull of PC is its maximum observed reward and $\mathbb{E}[\max] \geq \mathbb{E}[\mathrm{ALG}_T]/T \geq M/(32\sqrt{k})$ by Lemma C.1(5); the transfer argument is that of (Blum & Ravichandran, 2025, App. B); the resulting guarantee (no scale knowledge, no logarithm, $T \geq 2\lfloor\sqrt{k}\rfloor$) is a direct consequence of Theorem 3.1 and is not identical to a statement in Blum & Ravichandran (2025).

D PROOFS FOR SECTION 4 (PRIOR-FREE, NO NOISE)

Proof architecture. Lemmas D.1 and D.2 separate the ordering argument from the work estimate. Lemma D.3 combines them into a guarantee for a designated arm; the envelope hypothesis gives Theorem D.4, and the history-conditioned form in Lemma D.5 gives Theorem D.6. Independent inclusion leads to Lemma D.9, the common parent of the all-horizon results, Theorems 4.1 and 4.2. Its designated-arm formulation is essential for the latter: the original comparator need not remain optimal after shifting the curves.

D.1 RANDOM-MARGINAL PROBING

Algorithm 3 $\mathrm{RMP}(H)$ (random-marginal probe-and-commit). Input: k and a length $H \geq 16k$. Reads neither m nor β.

1: Draw independent $Z_1, \ldots, Z_k \sim \mathrm{Unif}(0,1)$. Let $N = \lfloor H/2 \rfloor$ (probe budget) and $r = H - N$ (commit budget).
2: Pull every arm once (these k pulls count towards N); set $n_i = 1$ and $\delta_i(1) = g_i(1)$.
3: For the remaining $N - k$ probes: pull the arm maximizing the lexicographic key $\big(\delta_i(n_i)/Z_i,\ 1/Z_i\big)$, where $\delta_i(n_i) = g_i(n_i) - g_i(n_i - 1)$ is the *last observed* increment; update n_i and δ_i.
4: Commit the r remaining pulls to the arm maximizing $\big(g_i(n_i), 1/Z_i\big)$.

Here g_i is the curve observed in the current run ($g_i = f_i$ in a single run; for a phase starting after a_i earlier pulls, $g_i(t) = f_i(a_i + t)$ with $g_i(0) = 0$). Residual ties (probability zero) are broken by arm index. A finite-randomness version with $Z_i \sim \mathrm{Unif}\{1/M, \ldots, 1\}$, $M = 2^{\lceil \log_2(32k) \rceil}$, and index tie-breaking changes the constants below to $9/1024$ and $9/16384$. For grid-valued weights and $x \geq 0$ one has $\mathbb{P}(Z \leq x) = \min\{1, \lfloor Mx \rfloor / M\} \leq \min\{1, x\}$, so the upper bound in Lemma D.2 is unchanged, and the monotone-coupling argument holds verbatim on the discrete ordered support. For the lower direction, $d \leq H/4$ forces the threshold parameter $p = \min\{1, H/(32kd)\}$ to satisfy $p \geq \frac{1}{8k}$, so with $M \geq 32k$ we get $\mathbb{P}(Z \leq p) \geq p - \frac{1}{M} \geq \frac{3}{4}p$. The discretisation therefore costs at most a factor $\frac{3}{4}$ in the positive guarantees, which is what the two constants record.

Lemma D.1 (merge structure and monotone coupling)**.** *For each arm consider the finite sequence of items $Q_{i,t} = (\delta_i(t)/Z_i,\ 1/Z_i,\ -i,\ -t)$, $1 \leq t \leq H$, which is nonincreasing in t by discrete concavity. The probe phase of* RMP *consumes exactly the $N - k$ largest items of the union in this total order, and $n_i = 1 + \#\{$items of arm i consumed$\}$. Consequently, for fixed curves and fixed Z_{-*}, decreasing Z_* does not increase any n_i, $i \neq *$; in particular, for every $v > 0$ the event $B = \{\exists i \neq * : g_i(n_i) \geq v\}$ is nondecreasing in Z_* (pointwise, for all positive weights, including ties).*

Proof. At every step the head of each arm's sequence is its current key, and the greedy rule takes the globally largest head; since each sequence is nonincreasing, this is a k-way merge and the consumed set is the top $N - k$ items of the union. Decreasing Z_* raises every item of arm $*$ and changes no other item, so no other arm's item can move up in the order. □

Lemma D.2 (truncated marginal work)**.** *Let $\bar{g}_i = \min\{g_i, v\}$ with increments $\bar{\delta}_i$, let $\lambda > 0$, and $L_i = \#\{1 \leq t \leq H : \bar{\delta}_i(t)/Z_i \geq \lambda\}$. Then $\mathbb{E}[L_i] = \sum_t \min\{1, \bar{\delta}_i(t)/\lambda\} \leq \bar{g}_i(H)/\lambda \leq v/\lambda$.*

Proof. $\mathbb{P}(Z_i \leq \bar{\delta}_i(t)/\lambda) = \min\{1, \bar{\delta}_i(t)/\lambda\}$ and the increments telescope. □

The right-hand side is the *total* increment v, not "number of scales times v"; this is what removes the logarithm.

Lemma D.3 (generic threshold guarantee)**.** *Let $k \geq 2$, $H \geq 16k$, curves nonnegative, nondecreasing, discretely concave with $g_i(0) = 0$, and a designated arm $*$. Suppose an integer $1 \leq d \leq H/4$ and $v > 0$ satisfy $g_*(d) \geq 2v$. Let $r = H - \lfloor H/2 \rfloor$, $V_* = \sum_{t \leq r} g_*(t)$ and $p = \min\{1, H/(32kd)\}$. Then* $\mathrm{RMP}(H)$ *satisfies $\mathbb{E}[\mathrm{ALG}_H] \geq \frac{3}{4}\min\{rv,\ pV_*\}$.*

Proof. Let $\lambda = 32kv/H$ and $C = \{\sum_{i\neq *} L_i \le H/8\}$. By Lemma D.2 and Markov, $\mathbb{P}(C) \ge 3/4$; C depends only on Z_{-*}. Let $A = \{Z_* \le p\}$ and B be as in Lemma D.1.

Step 1: on $A \cap C \cap \neg B$ the committed arm is $*$. Suppose $*$ has not reached value v by the end of probing; then $n_* < d$ and, for all $n \le n_*$, concavity gives $\delta_*(n) \ge \frac{g_*(d)-g_*(n)}{d-n} \ge v/d$, so on A the key of $*$ is at least $(v/d)/p \ge \lambda$ throughout. Every extra probe given to another arm i therefore used an item with key $\ge \lambda$, and on $\neg B$ (all observed values of the other arms below v) these items have $\bar{\delta}_i = \delta_i$; hence the extra probes on other arms number at most $\sum_{i\neq *} L_i \le H/8$. The total number of probes is then at most $k + d + H/8 \le \frac{H}{16} + \frac{H}{4} + \frac{H}{8} = \frac{7H}{16} < \lfloor H/2 \rfloor = N$ for $H \ge 32$, a contradiction. So $*$ reaches v while every other arm stays below v, and $*$ is committed.

Step 2: correlation. Fix Z_{-*} (which fixes C). By Lemma D.1, $\mathbf{1}_{\neg B}$ is nonincreasing in Z_*, so $\int_0^p \mathbf{1}_{\neg B}(z)\,dz \ge p\int_0^1 \mathbf{1}_{\neg B}(z)\,dz$; integrating over Z_{-*} on C, $\mathbb{P}(A \cap C \cap \neg B) \ge p\mathbb{P}(C \cap \neg B) \ge p\,(3/4 - \mathbb{P}(B))_+$.

Step 3: rewards. On B the committed arm has current value $\ge v$, so the commit phase earns $\ge rv$; on $A \cap C \cap \neg B$ it earns $\sum_{j\le r} g_*(n_*+j) \ge V_*$. With $b = \mathbb{P}(B)$, $\mathbb{E}[\mathrm{ALG}_H] \ge rv\,b + pV_*(3/4-b)_+ \ge \frac{3}{4}\min\{rv, pV_*\}$ (the expression is linear in b on $[0, 3/4]$). □

$\mathrm{RMP}(H)$ reads k and a length $H \ge 16k$, neither m nor β: it draws independent $Z_1, \dots, Z_k \sim \mathrm{Unif}(0,1)$, pulls every arm once, spends the rest of a probe budget $N = \lfloor H/2 \rfloor$ greedily on the arm maximizing the lexicographic key $\big(\delta_i(n_i)/Z_i,\ 1/Z_i\big)$, where $\delta_i(n_i) = g_i(n_i) - g_i(n_i - 1)$ is the *last observed* increment of the curve g_i observed in the current run, and commits the remaining $r = H - N$ pulls to the arm maximizing $\big(g_i(n_i), 1/Z_i\big)$. Each arm's key sequence is nonincreasing by discrete concavity, so the probe phase is a k-way merge consuming the $N - k$ largest items of the union, and decreasing Z_* does not increase any other arm's *probe count* (Lemma D.1); and after truncating a curve at a level v, the expected number of its items with key at least λ is at most v/λ (Lemma D.2) — the *total* increment v, not "number of scales times v", which is what removes the logarithm. Hence (Theorem D.4) for $k \ge 2$, $\beta \in (0, 1]$, $T \ge 16k$ and an optimal arm satisfying $\mathrm{LE}(\beta)$, $\mathbb{E}[\mathrm{ALG}_T] \ge \frac{3}{256}\,R\,k^{-\beta/(1+\beta)}$; phases $16k \cdot 2^j$ with fresh weights give $\mathbb{E}[\mathrm{ALG}_T] \ge \frac{3}{4096} R q_\beta$ for every $T \ge 16k$ without reading T (Theorem D.6). The restriction $H \ge 16k$ is the range used for the RMP guarantee; the all-horizon guarantee is provided by SRMP.

Theorem D.4 (prior-free, known horizon)**.** *Let $k \ge 2$, $\beta \in (0, 1]$, $T \ge 16k$, and suppose the optimal arm satisfies $\mathrm{LE}(\beta)$. Then $\mathrm{RMP}(T)$, which reads neither m nor β, satisfies*

$$\mathbb{E}[\mathrm{ALG}_T] \ \ge\ \frac{3}{256}\,R\,k^{-\beta/(1+\beta)}.$$

Moreover, for every integer $1 \le d \le T/4$ simultaneously,

$$\frac{\mathbb{E}[\mathrm{ALG}_T]}{R} \ \ge\ \frac{3}{1024}\min\Big\{\frac{f_*(d)}{m},\ \frac{T}{kd}\Big\} \qquad \textit{(depth-oracle guarantee).}$$

Proof. If several arms are cumulatively optimal, fix one that satisfies $\mathrm{LE}(\beta)$ as the comparison arm i° and let $m^\circ = f_{i^\circ}(T)$ play the role of m in the analysis (the algorithm reads neither; the other optimal arms fall into the event B); we keep writing $*$ and m. Let $q = k^{-\gamma}$, $K = kq$, $d = \lceil T/(8K) \rceil \le T/(4K) \le T/4$ (as $T \ge 16k \ge 8K$) and $v = mq/16$. By $\mathrm{LE}(\beta)$, $f_*(d) \ge m(d/T)^\beta \ge m(8K)^{-\beta} = mq8^{-\beta} \ge mq/8 = 2v$. Also $p \ge K/(8k) = q/8$, $rv \ge (T/2)(mq/16) \ge Rq/32$ and $V_* \ge R/8$ by Lemma C.1(2). Lemma D.3 gives $\frac{3}{4}\min\{Rq/32, Rq/64\} = 3Rq/256$. For the depth-oracle bound take $v = f_*(d)/2$: $rv \ge Rf_*(d)/(4m)$ and $pV_* \ge \frac{R}{8}\min\{1, T/(32kd)\}$, and $f_*(d)/m \le 1$. □

Lemma D.5 (conditioning on history)**.** *Suppose a phase of length $H \ge 16k$ starts after t_0 pulls with $t_0 + H \le T$ and arm i pulled a_i times, and runs $\mathrm{RMP}(H)$ with fresh weights on the local curves $g_i(t) = f_i(a_i + t)$, $g_i(0) = 0$. These curves are nondecreasing and discretely concave (the first increment $f_i(a_i + 2) - f_i(a_i + 1) \le f_i(1) \le g_i(1)$; later ones inherit), $g_*(t) \ge f_*(t) \ge m(t/T)^\beta$, and*

$$\mathbb{E}[\textit{phase reward} \mid \textit{history}] \ \ge\ \frac{3}{256}\,R\,q_\beta\,(H/T)^{1+\beta}.$$

Proof. Apply Lemma D.3 with $d = \lceil H/(8K) \rceil$ and $v = \frac{mq}{16}(H/T)^\beta$; now $rv \geq \frac{Rq}{32}(H/T)^{1+\beta}$ and $V_* \geq \frac{m}{T^\beta} \sum_{t \leq r} t^\beta \geq \frac{m\, r^{1+\beta}}{(1+\beta)T^\beta} \geq \frac{R}{8}(H/T)^{1+\beta}$ using $2^{1+\beta}(1+\beta) \leq 8$. □

Theorem D.6 (prior-free anytime, $T \geq 16k$)**.** *Run* $\mathrm{RMP}(H_j)$ *in phases* $H_j = 16k \cdot 2^j$ *with fresh weights, without resetting the true pull counts. The algorithm reads only* k*, and for every* $\beta \in (0, 1]$*, every instance whose optimal arm satisfies* $\mathrm{LE}(\beta)$*, and every* $T \geq 16k$*,* $\mathbb{E}[\mathrm{ALG}_T] \geq \frac{3}{4096} R q_\beta$*.*

Proof. The last completed phase has length $H_j > T/4$ since $\sum_{\ell \leq j+1} H_\ell = 4H_j - 16k > T$; apply Lemma D.5 and $4^{-(1+\beta)} \geq 1/16$. □

Corollary D.7 (price of priors, noiseless, $T \geq 16k$)**.** *Define the price of priors* φ *as the smallest factor by which an algorithm that reads neither* m *nor* β *falls short of* $k^{\beta/(1+\beta)}$*, uniformly over* β*. Then* $\frac{1}{3} \leq \varphi \leq 128$ *with* T *known and* $\frac{1}{3} \leq \varphi \leq 2048$ *with* T *unknown (the lower bound is the* $\beta = 1$ *class of Corollary C.6). Every combination of unknown* m*,* β*,* T *has a* $\Theta(1)$ *price.*

Remark D.8 (comparisons). (i) A one-level randomization of PC over $O(\log k)$ sample sizes ($s_j = \min\{k, \lfloor\sqrt{k}\rfloor 2^j\}$ for $j = 0, \ldots, J$ with $J = \min\{j : s_j = k\}$ and $h = J + 1$, chosen uniformly) gives $\mathbb{E} \geq R/(16h\, k^\gamma)$ with $h \leq \lceil \frac{1}{2} \log_2 k \rceil + 2$ for $T \geq 2k$; this is the same "guess the level" technique as (Blum & Ravichandran, 2025, Thm. 11) and, on the common horizon range $T \geq 16k$, is dominated in order by Theorem D.4; we record it only as a comparison. (ii) On the pure truncation families of Lemma C.5, a platform-detection rule (visit the arms in a uniformly random order, pulling each until a zero increment is observed; if the order is exhausted, spend the remaining probes on the current best candidate; always keep the arm with the largest observed value and commit to it for the second half) is a natural candidate; we do not analyse it here and claim no bound for it. Keeping the best arm matters: the variant that drops the good arm when it plateaus earns only $O(n^2)$ against $R_g = \Theta(n^3)$ on $k = T = n^2$, $g = \min(t, n)$, decoys $\min(t, 1)$.

D.2 All horizons: independent sparse inclusion

RMP initializes every arm; its guarantee is stated for $H \geq 16k$. Restricting it to a random subset of size $\lfloor T/16 \rfloor$ conditioned on containing $*$ yields only $\mathbb{E} \geq \frac{3}{256} R\, s^{1/(1+\beta)}/k$, which is not optimal for $\sqrt{k} < T < k$. The fix is to include arms *independently* and to account for the "good decoy" branch without multiplying by the inclusion probability of $*$.

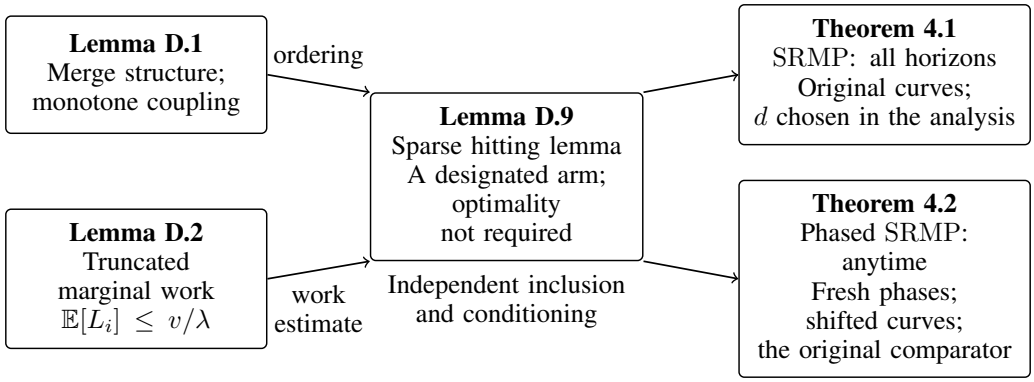


Figure 2: Selected proof inputs for the noiseless all-horizon results in Section 4. Both proofs use the designated-arm form of Lemma D.9; Theorem 4.2 does not require the original comparator to remain optimal after shifting. Envelope and concavity inputs, Lemma C.1, depth balancing and short-horizon cases remain in the proofs. Matching lower bounds come separately from Theorem C.10(ii). This is a local reading guide, not an exhaustive dependency graph.

For the analysis, in all execution paths (including the fallback) the first N pulls are the "probe prefix", and B denotes the event that by the end of this prefix some arm other than the designated one has an observed value $\geq v$; on the fallback path, if B occurs the chosen arm is not the designated one and its true value at pull N is at least v, so the last r pulls earn at least rv as well.

Algorithm 4 SRMP(k, H) (sparse random-marginal probing). Reads only k and H.

1: If $k = 1$: pull the arm. If $H < 32$: pull a uniformly random arm for all H steps.
2: Otherwise let $N = \lfloor H/2 \rfloor$, $r = H - N$, $\theta = \min\{1, H/(128k)\}$; draw independent $X_i \sim \text{Bernoulli}(\theta)$, $Z_i \sim \text{Unif}(0, 1)$; $S = \{i : X_i = 1\}$.
3: If $S = \emptyset$ or $|S| > N$: pull a uniformly random arm for all H steps.
4: Otherwise initialize each $i \in S$ with one pull (counted in N); spend the remaining probes greedily on the key $(\delta_i(n_i)/Z_i, 1/Z_i)$ over $i \in S$; commit the r remaining pulls to $\arg\max_{i \in S}(g_i(n_i), 1/Z_i)$.

Lemma D.9 (sparse hitting lemma)**.** *Let $H \geq 32$, let $*$ be a designated arm (not necessarily optimal) with $g_*(d) \geq 2v$ for some $v > 0$ and integer $1 \leq d \leq \lfloor H/4 \rfloor$. Then* SRMP *satisfies*

$$\mathbb{E}[\text{ALG}_H] \;\geq\; \tfrac{1}{2} \min\{rv,\; \alpha P_*(r)\}, \qquad \alpha = \min\Big\{\theta,\; \frac{H}{32kd}\Big\}.$$

Proof. Counting. Let $\lambda = 32\theta kv/H$, L_i as in Lemma D.2, and $C = \{\sum_{i \neq *} X_i \leq H/32,\; \sum_{i \neq *} X_i L_i \leq H/8\}$. Since $\mathbb{E}\sum_{i \neq *} X_i \leq \theta k \leq H/128$ and $\mathbb{E}\sum_{i \neq *} X_i L_i \leq \theta(k-1)v/\lambda \leq H/32$ (X_i and Z_i are independent), two Markov inequalities give $\mathbb{P}(C) \geq 1/2$; C is independent of (X_*, Z_*). On C, $|S| \leq H/32 + 1 \leq H/16 \leq N$ even if $*$ is included, so the fallback does not occur on $C \cap \{X_* = 1\}$. Let $p = \min\{1, H/(32\theta kd)\}$ and $A = \{X_* = 1,\; Z_* \leq p\}$; $\mathbb{P}(A) = \theta p = \alpha$. On $A \cap C \cap \neg B$, if $*$ had not reached v then $n_* < d$, its key would be at least $(v/d)/p \geq \lambda$ throughout, every extra probe on another arm would use an item with key $\geq \lambda$ and (on $\neg B$) untruncated increment, so the probe count would be at most $|S| + d + \sum_{i \neq *} X_i L_i \leq \frac{H}{16} + \frac{H}{4} + \frac{H}{8} < N$, a contradiction. Hence $*$ reaches v and is committed.

Correlation. Fix all randomness of the other arms (and the uniform choice used on the fallback path) and condition on C; this fixes C and leaves only the pair (X_*, Z_*) random. Parametrize it by the single variable

$$Y = \begin{cases} 0, & X_* = 0, \\ 1/Z_*, & X_* = 1, \end{cases}$$

so that larger Y means "$*$ has higher priority": $Y = 0$ is the lowest atom and, among included states, a smaller Z_* gives a larger Y. Increasing Y does not increase any other arm's pull count. Indeed, on C the run never takes the fallback branch once $*$ is included, because $|S| \leq H/32 + 1 \leq H/16 \leq N$ and $S \neq \emptyset$; for two included states this is Lemma D.1; and passing from "not included" to "included" adds one stream to the merge and removes one merge slot, which can only shorten the prefix consumed by the other streams (if the others' subset was empty, the run before inclusion is the fallback on a uniformly chosen arm, and after inclusion it probes only $*$ and commits to it, so no other arm is pulled at all). Since B is the event that some *other* arm reaches the level v, and other arms' pull counts do not increase, $\mathbf{1}_{\neg B}$ is *nondecreasing* in Y; $\mathbf{1}_A = \mathbf{1}\{Y \geq 1/p\}$ is nondecreasing in Y as well. Two bounded nondecreasing functions of one real random variable are nonnegatively correlated, so

$$\mathbb{P}(A \cap \neg B \mid C, Z_{-*}, X_{-*}) \;\geq\; \mathbb{P}(A)\,\mathbb{P}(\neg B \mid C, Z_{-*}, X_{-*}) \;=\; \alpha\,\mathbb{P}(\neg B \mid \cdots),$$

and averaging over the other arms' randomness on C gives $\mathbb{P}(A \cap C \cap \neg B) \geq \alpha\mathbb{P}(C \cap \neg B) \geq \alpha(1/2 - \mathbb{P}(B))_+$. (The same one-dimensional argument, with $Y = 1/Z_*$, is what is used in Lemma D.3.)

Branches. On B the commit phase earns $\geq rv$; on $A \cap C \cap \neg B$ it earns $\geq P_*(r)$. Hence $\mathbb{E} \geq rv\, b_0 + \alpha P_*(r)(1/2 - b_0)_+ \geq \frac{1}{2}\min\{rv, \alpha P_*(r)\}$. □

The sparse hitting lemma (Lemma D.9) states: for $H \geq 32$ and any designated arm $*$ with $g_*(d) \geq 2v$ for some $v > 0$ and integer $1 \leq d \leq \lfloor H/4 \rfloor$, $\mathbb{E}[\text{ALG}_H] \geq \frac{1}{2}\min\{rv,\; \alpha P_*(r)\}$ with $\alpha = \min\{\theta, H/(32kd)\}$. Two Markov inequalities show that with probability at least $\frac{1}{2}$ the other included arms number at most $H/32$ and their high-key work is at most $H/8$ (an event C independent of (X_*, Z_*)); on C, if $*$ is included with $Z_* \leq p = \min\{1, H/(32\theta kd)\}$ (probability $\theta p = \alpha$) and no other arm reaches level v, a probe count shows that $*$ reaches v and is committed. After fixing the other arms' randomness and the fallback choice, on C both the high-priority indicator and

the indicator that no other arm reaches v during the probe prefix are nondecreasing in the single variable Y ($Y = 0$ if $*$ is excluded, $Y = 1/Z_*$ otherwise), so the conditional positive-correlation inequality applies (Lemma D.9). For $T \geq 32$ take $v = f_*(d)/2$: $rv \geq \frac{R}{4}\frac{f_*(d)}{m}$, $\alpha P_*(r) \geq \alpha R/4$ and $\alpha \geq \frac{1}{128}\min\{1, T/(kd)\}$, so $\mathbb{E}/R \geq \frac{1}{1024}\min\{f_*(d)/m,\, T/(kd)\}$ for every $1 \leq d \leq T/4$. Choosing d in the analysis only ($d = 1$ if $T \leq K_\beta$; $d = \lfloor T/K_\beta \rfloor$ if $T > K_\beta \geq 4$; $d = \lfloor T/4 \rfloor$ if $K_\beta < 4$) gives $\min\{(d/T)^\beta, T/(kd)\} \geq Q_\beta(k,T)/8$ in all cases, and $\mathrm{LE}(\beta)$ gives $f_*(d)/m \geq (d/T)^\beta$, whence $c_{\mathrm{SR}} = 1024 \cdot 8$; for $T < 32$ the uniform single arm earns $\geq R/k \geq RQ_\beta/31$.

Proof of Theorem 4.1. For $T \geq 32$ take $v = f_*(d)/2$ in Lemma D.9: $rv \geq \frac{R}{4}\frac{f_*(d)}{m}$, $\alpha P_*(r) \geq \alpha R/4$ (Lemma C.1(4)) and $\alpha \geq \frac{1}{128}\min\{1, T/(kd)\}$, so $\mathbb{E}/R \geq \frac{1}{1024}\min\{f_*(d)/m,\, T/(kd)\}$ for every $1 \leq d \leq T/4$. Choose d (in the analysis only): if $T \leq K_\beta$ take $d = 1$, then $T^{-\beta} \geq T/k = Q_\beta$; if $T > K_\beta \geq 4$ take $d = \lfloor T/K_\beta \rfloor \in [T/(2K_\beta), T/4]$, then $(d/T)^\beta \geq q_\beta/2$ and $T/(kd) \geq q_\beta$; if $K_\beta < 4$ take $d = \lfloor T/4 \rfloor \geq T/8$, then $(d/T)^\beta \geq 1/8 \geq q_\beta/8$ and $T/(kd) \geq 4/k \geq q_\beta$. In all cases $\min\{(d/T)^\beta, T/(kd)\} \geq Q_\beta(k,T)/8$, and $\mathrm{LE}(\beta)$ gives $f_*(d)/m \geq (d/T)^\beta$. For $T < 32$ the uniform single arm earns $\geq R/k \geq RQ_\beta/31$. The cases $k = 1$ and $R = 0$ are trivial. If several arms are optimal, fix one satisfying $\mathrm{LE}(\beta)$ as $*$; the others fall into the good-decoy branch. □

The proof applies Lemma D.9 to the last completed phase, of length H with $x = H/T > 1/4$, on the reset curves $g_i(t) = f_i(n_i + t)$, which are nondecreasing and concave with $g_*(d) \geq f_*(d)$: with $v = f_*(d)/2$ the phase earns at least $\frac{Rx^2}{8192}Q_\beta(k,H) \geq \frac{Rx^3}{8192}Q_\beta(k,T)$, and $x^3 > 1/64$ gives 2^{-19}; if $H < 32$ then $T \leq 62$ and the uniform-arm phase earns $\geq R/(16k) \geq 2^{-19}RQ_\beta$.

Proof of Theorem 4.2. The reset curves $g_i(t) = f_i(n_i + t)$ are nondecreasing and concave, and $g_*(d) \geq f_*(d)$; Lemma D.9 applies to the designated arm $*$ without requiring it to be optimal for the shifted curves. Let H be the last completed phase, $x = H/T > 1/4$. With $v = f_*(d)/2$: $rv/R \geq \frac{x^{1+\beta}}{4}(d/H)^\beta \geq \frac{x^2}{4}(d/H)^\beta$ and $P_*(r)/R \geq x^2/4$ by Lemma C.1(2), so the phase earns at least $\frac{Rx^2}{8192}Q_\beta(k,H) \geq \frac{Rx^3}{8192}Q_\beta(k,T)$, and $x^3 > 1/64$. If $H < 32$ then $T \leq 62$ and the uniform-arm phase earns $\geq \frac{1}{k}P_*(H) \geq R/(16k) \geq 2^{-19}RQ_\beta$. □

Remark D.10. Two natural alternatives fail by a factor $\Theta(\sqrt{k})$ on the instance $T = k/2$ with $k \geq 16$ a multiple of 8, $f_*(t) = t$, decoys $\equiv 2$ for $t \geq 1$ (so $R = k(k+2)/8$): screening $k/8$ arms drawn uniformly without replacement by one pull each and then pulling the arm with the largest first observation keeps a decoy, for an expected reward $k - 1/8$ (ratio $\frac{k(k+2)}{8k-1}$), and giving uninitialized arms infinite priority spends the whole probe budget $\lfloor T/2 \rfloor = k/4$ on initializing distinct arms in a uniformly random order and again commits to a decoy, for an expected reward $k - 1/4$ (ratio $\frac{k(k+2)}{8k-2}$). Independent sparse inclusion avoids both. Theorem 4.1 also rules out any Yao-type lower bound showing an unbounded noiseless price of priors: the per-instance guarantee averages over any mixture of β-classes (using $B_0 = 1 + k/T$, $Q_0 = \min\{1, T/k\}$ for instances with envelope exponent 0, e.g. all arms constant).

E ATTRIBUTIONS, COMPARISONS AND THE LITERATURE IN FULL

Throughout, *curvature* in the title refers to the lower-envelope exponent β of Blum et al. (2026)—how fast a curve is required to rise relative to its terminal value—and not to strong concavity in any second-derivative sense; all our curves are merely discretely concave. Patil et al. (2023) study deterministic horizon-unaware improving bandits and state an $O(k)$ competitive guarantee. We use their deterministic lower-bound construction as background — the lower bound we restate in Appendix H uses only determinism, so it covers horizon-aware deterministic algorithms as well — but we establish our range-qualified deterministic bounds independently there, and we do not invoke the literal all-horizon upper bound of the algorithm displayed in arXiv:2208.09254v1: its two-pulls-per-arm initialisation already incurs ratio $k(2k+1)/3$ at $T = 2k$ on a single linear arm with zero decoys. We use this characterisation only on the sufficient range $T \geq 2k$, and assert no sharp boundary: Appendix H shows that for $T \leq k$ the deterministic ratio is $k(k+1)/2$ or unbounded, while at $T = 2k - 1$ it is still $\Theta(k)$. The horizon qualification is ours to state and is not

attributed to Patil et al. (2023). Blum & Ravichandran (2025) proved that randomization helps: no algorithm beats $\Omega(\sqrt{k})$; a random round-robin with an absolute threshold $m\,t/T$ achieves $O(\sqrt{k})$ *if it is told* the value $m = f^*(T)$ of the optimal arm in advance; and, in their words, one can "remove this assumption at the cost of an extra $O(\log k)$ approximation factor, achieving an overall $O(\sqrt{k}\log k)$ approximation relative to optimal" (for $T > 4k$). Blum et al. (2026) refined the picture with a concavity-envelope exponent $\beta \in (0, 1]$: if every arm satisfies $f_i(t) \geq f_i(T)(t/T)^\beta$, then a thresholded round robin achieves the optimal $\Theta(k^{\beta/(1+\beta)})$, again assuming m is known, and they handle unknown β by learning a parameter from offline instances; in the context of their data-driven hybrid they write: "We note that as a result of this approach, we do not get per-instance worst-case guarantees. This is an interesting direction for future work."[1]

For the general concave class, Blum & Ravichandran (2025) gives a scale-oblivious $O(\sqrt{k}\log k)$ guarantee for $T > 4k$. The envelope-dependent competitive guarantees of Blum et al. (2026) require a suitable exponent parameter, and their data-driven parameter selection uses offline instances. Our noiseless bounds in Theorems 4.1 and 4.2 instead adapt simultaneously to every admissible β on each instance, without reading m or β and without using offline instances.

Why was the logarithm there? The algorithms of Blum & Ravichandran (2025); Blum et al. (2026) abandon an arm when its current reward falls below an *absolute* threshold $m\,t/T$ (or $m(t/T)^\alpha$); the scale m must be known, and the treatment of unknown m in Blum & Ravichandran (2025)—a random dyadic guess over $O(\log k)$ dyadic levels—costs $\log k$. (We describe the technical difference only and make no claim about why this route was taken.) Our *scale-oblivious* algorithms do not compare rewards against a supplied absolute reward scale: probe-and-commit compares probe values *relative to each other* (an $\arg\max$ is scale-free) and pays for the missing information with a $\sqrt{k}$ subsample, whose hit probability $\lfloor\sqrt{k}\rfloor/k = \Theta(k^{-1/2})$ is of the order the lower bound charges anyway. The random-marginal algorithm goes further and turns "is this arm still worth pulling?" into a comparison of observed increments weighted by fixed random priorities; its key lemma bounds the expected greedy work spent on decoys by their total value, with no multi-scale union bound.

Prior results used. We use the following statements from the literature only as *comparisons* or as *lower bounds*; all upper bounds in this paper are self-contained.

- (Blum & Ravichandran, 2025, Thm. 3, Cor. 4): for any randomized algorithm there is an instance with approximation factor at least $\sqrt{k}/3$ (the hard distribution has $f^*(T) = 1$, so the bound holds even when the scale is known).
- (Blum & Ravichandran, 2025, Thm. 8): let $\tau = T - k$ and let the supplied scale parameter μ satisfy $\mu \in [f^*(\tau)/c_2, f^*(\tau)]$ for a constant $c_2 > 1$; for $T \geq 2k$, their random round robin run with internal horizon parameter τ and scale parameter μ earns at least $\mathrm{OPT}_T/(8c_2\sqrt{k})$. Here μ is the cited algorithm's parameter and is not our $m = f^*(T)$. (Blum & Ravichandran, 2025, Thm. 11, Lemma 12): with m unknown and $T > 4k$, $\Omega(\mathrm{OPT}/(\sqrt{k}\log k))$; with T unknown as well, $\mathrm{OPT}_T/(8192\sqrt{k}\log(128k))$.
- (Blum et al., 2026, Thm. 3.5): with m known and $T \geq 2k$, PTRR_α for $\alpha \in (\beta_I, 1)$ has ratio $O(k^{\alpha/(\alpha+1)})$. (Blum et al., 2026, Thm. 3.6): for every algorithm and T sufficiently large there is an instance with envelope exponent β and $\mathbb{E}[\mathrm{ALG}_T]/\mathrm{OPT}_T \leq C_\beta k^{-\beta/(\beta+1)}$ (Theorem B.14 there gives the explicit condition $T \geq 2[\beta(\beta+1)]^{1/(\beta+1)}k^{1/(\beta+1)}$; as printed it admits, e.g., $k = 2$, $\beta = 1/8$, $T = 1$, where an instance with envelope exponent exactly β cannot exist, so its domain needs a mild restriction; our Theorem C.10(i) is self-contained and unaffected). Section 3.2 of Blum et al. (2026) also removes scale knowledge at an $O(\log k)$ cost, and with both the scale and the horizon unknown (Blum et al., 2026, Thm. B.15) states an $O(k^{\alpha/(\alpha+1)}\log k)$ guarantee for $\beta_I \in (0, 1)$, $\alpha \in (\beta_I, 1)$ and $T > 4k$; that guarantee still requires choosing a suitable exponent

[1] Blum et al. (2026), introduction of Appendix D (discussion of the data-driven regret/competitive-ratio hybrid and of when to switch); see also Appendix D.2 there, in the same context: "Thus, that algorithm may not provide the worst-case fallback guarantee on a fixed instance. As mentioned earlier, this is an interesting consideration for future work." Both remarks concern their hybrid algorithm; they do not name the unknown-scale question. Our per-instance prior-free guarantees address one instance of the direction they describe.

parameter α, and what we remove is both the logarithmic overhead and the need to select such a parameter from prior information. That unknown-horizon result uses a doubling bookkeeping step in its proof whose displayed inequality relating the final time to the last completed round appears to need a small local adjustment; the order of its conclusion is not affected and we make no claim about it.

- Patil et al. (2023) studies deterministic horizon-unaware improving bandits and states an $O(k)$ competitive guarantee. We use its deterministic lower-bound construction as background, not the displayed algorithm's literal all-horizon upper bound. The lower bound restated in Appendix H uses only determinism and therefore also covers horizon-aware algorithms. The horizon-qualified deterministic bounds used here are proved independently in that section. Some horizon qualification is essential — the deterministic minimax ratio is unbounded for $T < k$ and equals $k(k+1)/2$ at $T = k$ (Appendix H) — but we do not claim that $2k$ is a sharp boundary, and the qualification is ours rather than attributed to Patil et al. (2023). The following observation concerns Appendix C.1 of arXiv:2208.09254v1 (19 August 2022). Its deterministic lower-bound argument selects the instance after fixing the deterministic algorithm, and the remark that follows does not justify extending that argument to the randomized competitive ratio $R/\mathbb{E}[\mathrm{ALG}]$. At $T = k$ every instance of the family displayed there is nondecreasing, has diminishing returns and may be taken to vanish at 0, so Theorem 3.3 applies to each member and gives ratio at most $1536(\sqrt{k}+1)$, which is below the asserted $k/2$ once $k \geq 2^{24}$. The asserted randomized extension of that construction is therefore incompatible with our guarantee. The comparison is restricted to that family and does not rely on any containment of the full instance class of Patil et al. (2023).

Why relative comparisons remove the scale dependence. The random round-robin of Blum & Ravichandran (2025) and the PTRR family of Blum et al. (2026) decide whether to keep an arm by an *absolute* test $f_i(t) \geq m(t/T)^\alpha$; the test needs m, and the route taken in Blum & Ravichandran (2025)—estimating m from $T/(2k)$-pull explorations, which localizes it within a factor $4k$, then guessing a dyadic level—costs the $\log k$. We describe the technical difference only. Probe-and-commit replaces the absolute test by a relative one ($\arg\max$ over probe values), which is invariant to rescaling, and buys the missing information with a $\sqrt{k}$ subsample; the $\lfloor\sqrt{k}\rfloor/k = \Theta(k^{-1/2})$ probability of sampling the optimal arm is of the order the $\Omega(\sqrt{k})$ lower bound already charges. The random-marginal and sparse variants make the "keep or leave" decision relative as well (observed increments weighted by fixed random priorities), and Lemma D.2 shows that the work wasted on decoys is paid by their total value, with no union bound over scales. The same relative comparisons are what make Theorem 5.1 robust to noise *without* knowing ε, in contrast to (Blum & Ravichandran, 2025, App. C), where the threshold is deflated by $(1-\varepsilon)$ and "the algorithm therefore must have knowledge of the value of ε".

Positioning. Blum et al. (2026) obtain instance-adaptive guarantees through a parameter learned from offline instances and through a data-driven regret/competitive-ratio hybrid; in the introduction of their Appendix D they write about the hybrid "We note that as a result of this approach, we do not get per-instance worst-case guarantees. This is an interesting direction for future work," and in Appendix D.2, in the same context, "that algorithm may not provide the worst-case fallback guarantee on a fixed instance." These remarks concern their hybrid algorithm rather than the unknown-scale question. Theorems D.4, 4.1 and G.1 provide per-instance worst-case guarantees that adapt to β without reading it, addressing one instance of the direction they describe, for the noiseless model and for the noisy model with known scale; Theorem 5.3 shows that in the noisy model the per-instance adaptation to an unknown exponent β is impossible without a growing price once the scale is also unknown. The stochastic rising-bandit literature (e.g. Metelli et al., 2022) studies instance-dependent regret with sub-Gaussian noise; our objective and adversary are different, and we do not compare constants.

F Proofs for Section 5: probe-and-commit, the counterexample, and the matching pair

Proof architecture. Theorem 5.1 reuses the two-event sampling argument with multiplicative reward comparisons. Example F.1 diagnoses the failure of specific increment keys. The matching upper and lower bounds have separate proofs: Lemma F.2 converts a feasible depth schedule into

reward, while Lemma F.3 certifies that the proposed schedule fits the budget; together they support Theorem 5.2. Theorem 5.3 instead uses a common pre-reveal history, a budget shared across reveal depths, and a weighted hard distribution. Its impossibility claim does not follow from the counterexample. The anytime extension of the upper bound is proved separately in Appendix F.4.

F.1 PROBE-AND-COMMIT IS ROBUST AND NEEDS NO ε

Proof of Theorem 5.1. Let $\hat{F} = \{i \neq * : \hat{f}_i(\tau) \geq \hat{f}_*(\tau)\}$ (fixed once the noise is fixed) and A, B as in Lemma C.2. On B, $\hat{f}_c(\tau) \geq \hat{f}_*(\tau) \geq a f_*(\tau)$ and the commit phase earns $\geq \frac{a}{b} r \hat{f}_c(\tau) \geq \frac{a^2}{b} r f_*(\tau) \geq \frac{a^2}{b} R \frac{r\tau}{T^2}$. On $A \cap \neg B$ the committed arm is $*$ and earns $\geq a \sum_{j \leq r} f_*(\tau + j) \geq aR\frac{r(r+1)}{2T^2}$. Conclude as in Lemma C.3 using $a^2/b \leq a$. Compare (Blum & Ravichandran, 2025, App. C), whose algorithm "must have knowledge of the value of ε" and, without m, is stated there to have approximation factor $O\big((\frac{1-\varepsilon}{1+\varepsilon})^2 \sqrt{k} \log(\frac{1+\varepsilon}{1-\varepsilon} k)\big)$ (as printed; the first factor is presumably meant as its reciprocal). □

F.2 RANDOM-MARGINAL PROBING BREAKS

Under noise the observed increments can be negative and non-monotone, so the merge structure of Lemma D.1 is lost: for $k = 2$, $H = 32$, both true curves $\equiv 1$, $\hat{f}_1 = (1, \frac{3}{4}, \frac{3}{4}, \dots)$, $\hat{f}_2 = (\frac{5}{4}, \frac{3}{4}, \dots)$ and $Z_2 = \frac{1}{2}$, decreasing Z_1 from $\frac{3}{4}$ to $\frac{1}{8}$ changes the probe counts from $(14, 2)$ to $(2, 14)$. More seriously, a hidden growth prefix defeats RMP and two of its natural repairs. We fix the variants precisely, since the phrase admits several readings. The *running-maximum increment* key replaces $\delta_i(n)$ by $\max_{t \leq n} \hat{f}_i(t) - \max_{t \leq n-1} \hat{f}_i(t)$, which guards against non-monotone observations; the *half-window slope* key replaces it by $\big(F_i(n) - F_i(\lfloor n/2 \rfloor)\big)/\big(n - \lfloor n/2 \rfloor\big)$, where $F_i(0) = 0$ and $F_i(n) = \max_{1 \leq t \leq n} \hat{f}_i(t)$. Both keys vanish for the comparator at its second pull in the example below, so the example defeats both. A third reading, which keeps the *largest* dyadic window — $\hat{f}_i(n) - \hat{f}_i(n - 2^{\lfloor \log_2 n \rfloor})$, and $\hat{f}_i(n)/n$ when that index vanishes — is a different algorithm: at $n = 2$ the window covers the whole prefix, so the comparator's key there is $\hat{f}_*(2)/2 > 0$ rather than zero, and the zero-key argument above does not apply to this variant; we do not analyse its competitive ratio here and claim nothing about it. A variant keyed on the *level* of the running maximum rather than on its increment is a different algorithm: there the comparator's key remains $1/Z_*$, comparable to the decoy's, so it is probed further with constant probability and the growth is revealed; the example does not defeat that variant and we claim nothing about it.

Under noise the observed increments can be negative and non-monotone, so the merge structure of Lemma D.1 is lost, and a hidden growth prefix defeats RMP: Example F.1 (Appendix F) gives a family indexed by integers $m \geq 3$, with $k = 2$, $\varepsilon = \frac{1}{4}$ and $T = H = 4m^2 \geq 16k$; its fixed observed reward table is in band and nondecreasing in each arm's pull count for $1 \leq t \leq T$, its comparator satisfies $\mathrm{LE}(1/2)$, the decoy is always committed, and $\hat{R}_m / E_m \sim \frac{8}{9} m \to \infty$. The same example refutes the running-maximum *increment* key and the half-window slope key, both fixed precisely there, but not every conceivable repair (two variants it does not refute are named there, and we claim nothing about them); Theorem 5.3, rather than this example, is what rules out a constant uniform price of adaptation for algorithms reading neither m nor β.

Example F.1 (fake growth). Let $m \geq 3$ be an integer, so that $T = H = 4m^2 \geq 16k$ as required by RMP; let $k = 2$, $T = H = 4m^2$, $\varepsilon = \frac{1}{4}$, $f_*(t) = \min\{m, 1 + \frac{t-1}{4}\}$, $f_2 \equiv 1$ $(t \geq 1)$, $\hat{f}_*(1) = \hat{f}_*(2) = 1$, $\hat{f}_*(t) = f_*(t)$ for $t \geq 3$, and $\hat{f}_2(t) = 1 + \frac{t-1}{4T}$. All observations lie in the prescribed multiplicative bands, and each observed reward sequence is nondecreasing; the optimal arm satisfies $\mathrm{LE}(1/2)$. The second observation of $*$ has zero increment while the decoy's increments stay positive, so after its second probe $*$ is never probed again and the decoy is committed (the event that $*$ receives only one probe has probability $1/(8T)$ and also ends with the decoy committed). Exactly, $E_m = T + \frac{(T-2)(T-3)}{8T} + \frac{T-2}{32T^2}$ while $\hat{R}_m = mT - \frac{(m-1)(4(m-1)+1)}{2} - \frac{1}{4}$, so $\hat{R}_m / E_m \sim \frac{8}{9} m \to \infty$.

The monotone-coupling property itself does not require nonincreasing streams. Consider keys of the form $(\kappa_i(n_i)/Z_i,\ 1/Z_i,\ -i)$, where $\kappa_i \geq 0$ depends only on the arm's own count, with ties broken by index; then decreasing Z_* never increases other arms' counts (delete the moves of $*$; the others'

moves form a prefix of one fixed greedy sequence). What kills the increment-based algorithms is a single zero key at the wrong moment. Theorem 5.3, rather than this example, is what rules out a constant uniform price of adaptation for algorithms reading neither m nor β, at a fixed positive noise level and asymptotically in k; it does not assert that every repair is defeated by Example F.1.

F.3 A MATCHING PAIR: NESTED PERMUTATION PROBING AND A HIDDEN-GROWTH LOWER BOUND

Lemma F.2 (permutation coupling)**.** *Fix nonincreasing integer depths $\ell_1 \geq \cdots \geq \ell_k \geq 1$ with $\sum_j \ell_j \leq H/2$. Permute the arms uniformly at random, pull the j-th arm ℓ_j times, and commit to the arm with the largest running maximum. If the first s positions have depth at least d, then (with $H = T$)*

$$\mathbb{E}[\widehat{\mathrm{ALG}}_T] \;\geq\; \frac{a^2}{8b}\, R \,\min\Big\{\frac{f_*(d)}{m},\; \frac{s}{k}\Big\}.$$

Proof. Take $v = a f_*(d)$, $A = \{* \text{ among the first } s \text{ positions}\}$, $B = \{\text{some other arm ends with running maximum} \geq v\}$. Fix the relative order of the other arms; moving $*$ forward pushes the overtaken arms back, which does not increase their depths or running maxima, so $\neg B$ is a prefix in the position of $*$ and $\mathbb{P}(A \cap \neg B) \geq \frac{s}{k}\mathbb{P}(\neg B)$. On B the commit earns $\geq \frac{a}{b}\frac{T}{2}v \geq \frac{a^2}{2b}R\frac{f_*(d)}{m}$; on $A \cap \neg B$ the unique maximizer is $*$ and the commit earns $\geq aR/8$. □

Algorithm 5 $\mathrm{NPA}(k, H)$ (nested permutation probing), $k \geq 2$, $H \geq 16k$; reads neither m, β nor ε.

1: Let $h = 8\Phi(k)$ and define the depth envelope $D(x) = \frac{1}{h\sqrt{k}}$ for $x \leq \frac{1}{h\sqrt{k}}$; $D(x) = x^{1/\alpha(x)}$ with $\alpha(x) = \frac{\ln(1/(hx))}{\ln(khx)}$ for $\frac{1}{h\sqrt{k}} < x < \frac{1}{h}$; $D(x) = 0$ for $x \geq \frac{1}{h}$.
2: Set $\ell_j = \max\{1, \lceil H\, D((j-1)/k)\rceil\}$; run the procedure of Lemma F.2. (For $k = 1$, pull the arm.)

The envelope is nonincreasing and $\alpha(x) \in (0, 1)$ on the middle piece. The probe budget is controlled by the following lemma, which we isolate because it carries the feasibility of the whole schedule.

Lemma F.3 (depth-budget feasibility)**.** *Let $k \geq 2$, $H \geq 16k$ and $h \geq 8$, and let D be the depth envelope of Algorithm 5. Then*

$$k\int_0^1 D(x)\,dx \;\leq\; \frac{1}{h^2} + \frac{\ln k}{4h^2 \ln h}, \qquad \sum_{j\leq k}\ell_j \;\leq\; \Big(k\int_0^1 D + D(0) + \frac{k}{H}\Big)H.$$

For $h = 8\Phi(k)$ the first bound is at most $\frac{1}{32}$, and consequently $\sum_j \ell_j \leq \frac{7H}{32}$.

Proof. The plateau. On $[0, \frac{1}{h\sqrt{k}}]$ the envelope is constant equal to $\frac{1}{h\sqrt{k}}$, so this piece contributes $k \cdot \frac{1}{h\sqrt{k}} \cdot \frac{1}{h\sqrt{k}} = \frac{1}{h^2}$ to $k\int_0^1 D$. This term is exactly the scaled contribution of the left plateau; it must not be dropped.

The middle piece. Writing $p_\beta = q_\beta/h$ one has $D(p_\beta) = p_\beta^{1/\beta}$, and parametrising the middle piece by $\beta \in (0, 1]$ turns its contribution into $\ln k \int_0^1 h^{-(1+1/\beta)}(1+\beta)^{-2}\,d\beta$. Substituting $y = 1/\beta$ (so $d\beta = -dy/y^2$ and $(1+\beta)^{-2} = y^2(1+y)^{-2}$, the sign being absorbed by the reversal of the limits) gives

$$\ln k \int_1^\infty \frac{h^{-(1+y)}}{(1+y)^2}\,dy \;\leq\; \frac{\ln k}{4}\int_1^\infty h^{-(1+y)}\,dy \;=\; \frac{\ln k}{4h^2 \ln h},$$

using $(1+y)^2 \geq 4$ for $y \geq 1$. Adding the two pieces gives the first display.

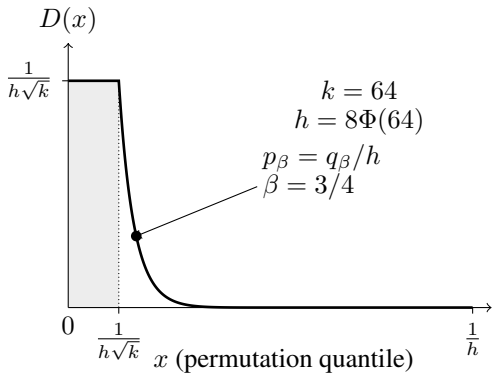
Discretisation. Since D is nonincreasing, the left-endpoint sum obeys $\sum_{j\leq k} D(\frac{j-1}{k}) \leq k\int_0^1 D + D(0)$, and $\ell_j = \max\{1, \lceil H\, D(\frac{j-1}{k})\rceil\} \leq H\, D(\frac{j-1}{k}) + 1$, whence the second display.

The numerical bound. For $h = 8\Phi(k)$ we have $h \geq 8$ (as $\Phi(k) \geq 1$) and $\ln h \geq \frac{1}{4}\ln(e + \ln k)$: with $x = \ln k$, the inequalities $\ln^2(e+x) \leq e+x$ and $1+x \geq (e+x)/e$ give $h^4 = 4096\,(1+x)^2/\ln^2(e+$

$x) \geq \frac{4096}{e^2}(e+x) \geq e+x$. Hence $\frac{1}{h^2}\big(1+\frac{\ln k}{4\ln h}\big) \leq \frac{1}{h^2}\big(1+\frac{\ln k}{\ln(e+\ln k)}\big) \leq \frac{2}{h^2}\cdot\frac{1+\ln k}{\ln(e+\ln k)} = \frac{2\Phi(k)^2}{h^2} = \frac{1}{32}$, where the middle inequality is $\ln(e+\ln k) \leq 2+\ln k$. Finally $D(0) = \frac{1}{h\sqrt{k}} \leq \frac{1}{8}$ and $\frac{k}{H} \leq \frac{1}{16}$, so $\sum_j \ell_j \leq (\frac{1}{32}+\frac{1}{8}+\frac{1}{16})H = \frac{7H}{32}$. □

The upper bound is a noise-robust probe-and-commit with unequal depths (Lemma F.2): fix non-increasing integer depths $\ell_1 \geq \cdots \geq \ell_k \geq 1$ with $\sum_j \ell_j \leq H/2$, permute the arms uniformly at random, pull the j-th arm ℓ_j times, and commit to the arm with the largest running maximum; if the first s positions have depth at least d, then (with $H = T$) $\mathbb{E}[\widehat{\mathrm{ALG}}_T] \geq \frac{a^2}{8b} R \min\{f_*(d)/m,\ s/k\}$. *Nested permutation probing* $\mathrm{NPA}(k, H)$ (Algorithm 5; $k \geq 2$, $H \geq 16k$; reads neither m, β nor ε) takes all depths from one explicit envelope D with $h = 8\Phi(k)$, $\ell_j = \max\{1, \lceil H\, D((j-1)/k)\rceil\}$, built so that, for a run of length H and every β, the first $s = \lceil kq_\beta/h\rceil$ positions have depth at least $d_H = \lceil H(q_\beta/h)^{1/\beta}\rceil$ (analysis witnesses, not inputs to NPA; for the known-horizon theorem set $H = T$ and $d = d_T$), while $k\int_0^1 D \leq \frac{1}{h^2} + \frac{\ln k}{4h^2 \ln h}$ (a plateau term plus a middle piece), which the choice $h = 8\Phi(k)$ makes at most $\frac{1}{32}$, whence $\sum_j \ell_j \leq \frac{7H}{32}$ (Lemma F.3).

Proof of Theorem 5.2. Let $s = \lceil kq_\beta/h\rceil$ and $d = \lceil T(q_\beta/h)^{1/\beta}\rceil$. For $j \leq s$, $(j-1)/k < q_\beta/h$, so $\ell_j \geq \lceil T\, D(q_\beta/h)\rceil = d$; also $s/k \geq q_\beta/h$ and $f_*(d)/m \geq (d/T)^\beta \geq q_\beta/h$. Apply Lemma F.2 with $h = 8\Phi(k)$. □



One envelope; no β input.

The marked point is an analysis witness:

$$D(p_\beta) = p_\beta^{1/\beta}.$$

The integer certificate is

$$s = \lceil kp_\beta\rceil, \quad d = \lceil Hp_\beta^{1/\beta}\rceil,$$

$$j \leq s \implies \frac{j-1}{k} < p_\beta \implies \ell_j \geq d.$$

For $x \geq 1/h$, $D(x) = 0$; nevertheless every integer depth ℓ_j is at least one.

Budget upper bounds (Lemma F.3; $k \geq 2$, $H \geq 16k$, $h = 8\Phi(k)$):

$$\frac{1}{H}\sum_{j\leq k}\ell_j \;\leq\; \underbrace{k\int_0^1 D(x)\,dx}_{\text{continuous area}} + \underbrace{D(0)}_{\text{endpoint correction}} + \underbrace{\frac{k}{H}}_{\substack{\text{rounding and}\\ \text{minimum-one cost}}} \;\leq\; \frac{1}{32}+\frac{1}{8}+\frac{1}{16} = \frac{7}{32}.$$

Figure 3: The depth envelope of Algorithm 5, a reading aid for Theorem 5.2 in Section 5. The curve evaluates the existing piecewise formula on $0 \leq x \leq 1/h$; it does not show rewards or simulation results. The shaded plateau contributes $1/h^2$ to $k\int D$. The marked β supplies an analysis witness, not an algorithm input. The three budget terms are proof upper bounds, not an observed allocation.

Let $k = 2^L$ with $L \geq \max\{256, 3r_\varepsilon\}$, $\eta = 2^{-r_\varepsilon} \in [\varepsilon/2, \varepsilon]$, $T = 16k$, $B = \lceil \frac{1}{2}\log_2 \frac{L}{\log_2 L}\rceil$ and $A = 2^B$. The hard distribution mixes *types* $j \in \mathcal{J} = \{j_{\min}, j_{\min}+2, \cdots \leq \lfloor L/2\rfloor\}$, $j_{\min} = \lceil \frac{L}{2} - \frac{L}{16B}\rceil$, with exponents $\beta_j = \frac{j}{L-j}$ (so $q_j = 2^{-j}$), scales $m_j = A2^j$ and dyadic growth lengths D_j chosen as in the construction in the proof of Theorem 5.3 in Appendix F; that choice supplies the envelope condition and the scale separation used below. All arms but a uniformly random star arm U have $f = \hat{f} \equiv 1$; the star arm of type j has $f_U(t) = \min\{m_j, 1 + \frac{t-1}{D_j}\}$, satisfies $\mathrm{LE}(\beta_j)$, and its observations are flattened to $\hat{f}_U(t) = 1$ on a *hidden prefix* $t \leq d_j = \eta D_j$, in band because $f_U \leq 1+\eta \leq \frac{1}{1-\varepsilon}$ there (for $\varepsilon = 0$ the prefix is illegal). Run the algorithm with a fixed seed against the all-ones response and let $N(d)$ be the number of arms pulled at least $d+1$ times: type j is

revealed with probability exactly $N(d_j)/k$, so $\mathbb{E}[\widehat{\mathrm{ALG}}_j] \le T + \frac{\mathbb{E}N(d_j)}{k}\hat{R}_j$. Since $d_{j+2} \ge 4d_j$, one run has $\sum_j d_j N(d_j) \le \frac{4}{3}T \le 2T$ for all types at once, and weighting type j by $w_j \propto d_j k q_j$ (Yao's principle, Yao, 1977) gives $\sum_j w_j\, \mathbb{E}[\widehat{\mathrm{ALG}}_j]/(\hat{R}_j q_j) \le \frac{4}{A} + \frac{2T}{S}$ with $S = \sum_j d_j k q_j \ge \frac{\eta|\mathcal{J}|T}{4A}$. The finite-parameter checks, in particular $|\mathcal{J}| \ge \frac{L}{32B} - 1 \ge \frac{L}{64B} \ge \frac{L}{32\log_2 L}$ (using $64B \le 32\log_2 L \le L$), give $\frac{4}{A} + \frac{8A}{\eta|\mathcal{J}|} \le \frac{516}{\eta}\sqrt{\log_2 L/L}$; some type and some position of the star arm attain this bound, and $\eta \ge \varepsilon/2$, $\Phi(2^L) \le 2\sqrt{L/\log_2 L}$ give $2064/\varepsilon$ for powers of two. Padding with zero arms and $T = 32k'$ loses at most $\sqrt{2}\cdot\sqrt{2}$, so 8192 suffices.

Proof of Theorem 5.3. Construction. Let $k = 2^L$, $L \ge \max\{256, 3r_\varepsilon\}$, $\eta = 2^{-r_\varepsilon} \in [\varepsilon/2, \varepsilon]$, $T = 16k$, $B = \lceil \frac{1}{2}\log_2 \frac{L}{\log_2 L} \rceil$, $A = 2^B$. Types $j \in \mathcal{J} = \{j_{\min}, j_{\min}+2, \cdots \le \lfloor L/2 \rfloor\}$, $j_{\min} = \lceil \frac{L}{2} - \frac{L}{16B} \rceil$, with $\beta_j = \frac{j}{L-j}$, $q_j = 2^{-j}$, $m_j = A2^j$, $D_j = 2^{\lfloor \log_2 T - (L-j) - B(L-j)/j \rfloor}$, $d_j = \eta D_j$ (positive integers). Set $f_i(0) = 0$ for every arm; the following reward formulas apply to integer pull counts $1 \le t \le T$, and the hidden observation prefix is $1 \le t \le d_j$. The star arm U is uniform in $[k]$; all other arms have $f = \hat{f} \equiv 1$ for $t \ge 1$; the star arm has $f_U(t) = \min\{m_j, 1 + \frac{t-1}{D_j}\}$ and $\hat{f}_U(t) = 1$ for $t \le d_j$ (hidden prefix; in band since $f_U \le 1 + \eta \le \frac{1}{1-\varepsilon}$ there), $\hat{f}_U(t) = f_U(t)$ for $t > d_j$. The star arm is nondecreasing and discretely concave, satisfies $\mathrm{LE}(\beta_j)$ (as $m_j \le (T/D_j)^{\beta_j}$ and $1 + \frac{t-1}{D} \ge \max\{1, t/D\} \ge (t/D)^\beta$), reaches m_j before T, and $\hat{R}_j \ge aR_j \ge m_j T/4$; the constant decoys satisfy every $\mathrm{LE}(\beta)$. For $\varepsilon = 0$ the hidden prefix is illegal: the second pull of the star arm reveals a positive increment. The hard observation tables are nondecreasing in the pull count of every arm: the star arm is observed as the constant 1 on its hidden prefix and as $f_U(t) \ge 1$ afterwards, and every other arm is constant. The lower bound therefore already holds with nondecreasing observation tables; the noise is used only to conceal growth over a type-dependent prefix.

Common history. Fix the algorithm's seed and run it against the all-ones response; let $N(d)$ be the number of arms pulled at least $d+1$ times. The real run coincides with this one until the star arm is pulled for the (d_j+1)-st time, so type j is revealed with probability exactly $N(d_j)/k$. Unrevealed, every pull earns at most 1; revealed, the total is at most $\hat{R}_j$ since $\hat{f}_U \ge 1$ is nondecreasing. Hence $\mathbb{E}[\widehat{\mathrm{ALG}}_j] \le T + \frac{\mathbb{E}N(d_j)}{k}\hat{R}_j$.

Multi-scale budget and Yao mixture Yao (1977). Consecutive types satisfy $d_{j+2} \ge 4d_j$, so $\sum_j d_j N(d_j) \le \frac{4}{3}T \le 2T$. Let $S = \sum_j d_j k q_j$ and weights $w_j = d_j k q_j / S$. Then $\sum_j w_j \frac{\mathbb{E}[\widehat{\mathrm{ALG}}_j]}{\hat{R}_j q_j} \le \frac{4}{A} + \frac{2T}{S}$. The finite-parameter checks behind this display are the following, and we record them in full since they carry the construction. From $L \ge 256$ we get $B \ge 3$, since $L/\log_2 L$ is increasing and already 32 at $L = 256$. For the upper bound, $\lceil x \rceil \le x + 1$ gives $2B \le \log_2 \frac{L}{\log_2 L} + 2 = \log_2 L - \log_2\log_2 L + 2$, so that

$$\log_2 L - 2B \;\ge\; \log_2\log_2 L - 2 \;\ge\; 1 \qquad (L \ge 256),$$

and in particular $B \le \frac{1}{2}\log_2 L$. The analytic lower bound $\log_2\log_2 L - 2$ is increasing for $L \ge 256$, so the displayed bound holds on the full, unbounded parameter range; no finite numerical check is used. Every admissible type index satisfies $j \ge \frac{L}{2} - \frac{L}{16B} \ge \frac{23L}{48}$, the last step using $B \ge 3$. Next $\frac{1}{\beta_j} \le \frac{8B+1}{8B-1} = 1 + \frac{2}{8B-1} \le 1 + \frac{2}{7B}$, valid for $B \ge 1$. With $e_j = \log_2 D_j = \lfloor j + 4 - B/\beta_j \rfloor$ this gives $e_j \ge \frac{23L}{48} + 3 - B - \frac{2}{7} \ge \frac{L}{3}$, and since $r_\varepsilon \le \frac{L}{3}$ the depth $d_j = 2^{e_j - r_\varepsilon}$ is a positive integer. For consecutive admissible types the unrounded exponent gap is $2 + \frac{2BL}{j(j+2)} \ge 2$, and $\lfloor x+d \rfloor - \lfloor x \rfloor \ge \lfloor d \rfloor$ for $d \ge 2$, so $e_{j+2} - e_j \ge 2$, i.e. $d_{j+2} \ge 4d_j$ as used above. Finally $e_j \le j + 4 - B/\beta_j$ gives $B + j + e_j \le 4 + 2j + B(1 - \frac{1}{\beta_j}) \le L + 4 = \log_2 T$, i.e. $m_j D_j \le T$, so the star arm reaches its plateau inside the horizon; the bound on the number of types follows from the length of the admissible index range minus the two rounding losses at its ends. The remaining finite checks are equally short. Since $A = 2^B$ with $B = \lceil \frac{1}{2}\log_2 \frac{L}{\log_2 L} \rceil$ we have $A \in [\sqrt{L/\log_2 L}, 2\sqrt{L/\log_2 L}]$. For the number of types put $u = \lfloor L/2 \rfloor$ and $\ell = \lceil L/2 - L/(16B) \rceil$, so that $|\mathcal{J}| = \lfloor \frac{u-\ell}{2} \rfloor + 1 \ge \frac{L}{32B} - 1$; the two rounding losses at the ends of the range are what the -1 absorbs. Since $L \ge 256$ and $B \le \frac{1}{2}\log_2 L$ give $64B \le 32\log_2 L \le L$, we may drop the -1 at the cost of a factor 2, obtaining $|\mathcal{J}| \ge \frac{L}{64B} \ge$

$\frac{L}{32\log_2 L}$, which is the bound used below. Next $A^{1/\beta_j} = 2^{B/\beta_j} \le 2^{B+2/7} = A \cdot 2^{2/7} \le 2A$ by the bound on $1/\beta_j$ above. Finally $d_j k q_j = \eta 2^{e_j+L-j} \ge \eta 2^{L+3-B-2/7} = \frac{\eta T}{2\cdot 2^{2/7} A} \ge \frac{\eta T}{4A}$. With these, we get $d_j k q_j \ge \frac{\eta T}{4A}$ and $S \ge \frac{\eta |\mathcal{J}| T}{4A}$, and so the weighted average is at most $\frac{4}{A} + \frac{8A}{\eta|\mathcal{J}|}$; here $\frac{4}{A} \le 4\sqrt{\log_2 L/L}$ and $\frac{8A}{\eta|\mathcal{J}|} \le \frac{16\sqrt{L/\log_2 L}\cdot 32\log_2 L}{\eta L} = \frac{512}{\eta}\sqrt{\log_2 L/L}$, so the sum is at most $\frac{516}{\eta}\sqrt{\log_2 L/L}$ using $\eta \le 1$. Some type j (and then some fixed position of the star arm) attains this bound; $\eta \ge \varepsilon/2$ and $\Phi(2^L) \le 2\sqrt{L/\log_2 L}$ (for $L \ge 256$: $1 + L\ln 2 \le 2L\ln 2$ and $\ln(e + L\ln 2) \ge \frac{1}{2}\ln L$, so $\Phi(2^L)^2 \le \frac{2L\ln 2}{\frac{1}{2}\ln L} = \frac{4L}{\log_2 L}$) give the constant $2064/\varepsilon$ for powers of two. For general $k \ge k_0(\varepsilon)$ let $n = k' = 2^{\lfloor \log_2 k \rfloor}$, so $L = \log_2 n \ge \max\{256, 3r_\varepsilon\}$ and $n \le k < 2n$, and take $T = 32n$. Run the construction above with the n active arms and this horizon: B, A, $\mathcal{J}$, β_j, $q_j = 2^{-j}$ and $m_j = A2^j$ are unchanged, while $D_j = 2^{\lfloor \log_2 T - (L-j) - B(L-j)/j \rfloor}$ and $d_j = \eta D_j$ are computed with the new T and are therefore exactly twice their values for $T = 16n$. Since $e_j = \log_2 D_j$ and $\log_2 T$ both increase by one: the star arm is still in band on its hidden prefix, satisfies $\mathrm{LE}(\beta_j)$ (as T/D_j is unchanged), reaches m_j before T ($B + j + e_j \le \log_2 T$), d_j is a positive integer, $d_{j+2} \ge 4d_j$, and $d_j n q_j \ge \eta T/(4A)$. The remaining $k - n$ arms have $f = \hat{f} \equiv 0$; they are nondecreasing and concave, so the instance stays admissible, the star arm (uniform among the n active arms) is still the comparator, and pulls of zero arms earn nothing. Fix the algorithm's seed ξ and run it against the type-independent virtual response table that returns 1 on the n active arms and 0 on the $k - n$ padded arms. Let M_i^ξ be the number of pulls of active arm i in this virtual run, and put $N_\xi(d) = |\{i \le n : M_i^\xi \ge d + 1\}|$. For each type j, the actual and virtual histories agree before the revealing observation on the star arm's $(d_j + 1)$-st pull. Thus

$$\mathbb{P}_U(\text{reveal of type } j \mid \xi) = \frac{N_\xi(d_j)}{n}, \qquad \mathbb{E}_{\xi,U}[\widehat{\mathrm{ALG}}_j] \;\le\; T + \frac{\mathbb{E}_\xi N_\xi(d_j)}{n}\,\hat{R}_j.$$

Moreover, since $d_{j+2} \ge 4d_j$,

$$\sum_j d_j N_\xi(d_j) \;=\; \sum_{i=1}^{n} \sum_{j:\, d_j < M_i^\xi} d_j \;\le\; \frac{4}{3}\sum_{i=1}^{n} M_i^\xi \;\le\; \frac{4}{3}T.$$

The weighted-average bound therefore holds with n in place of k: some type j and some star position give $\mathbb{E}[\widehat{\mathrm{ALG}}_j] \le \frac{2064}{\varepsilon\,\Phi(n)}\hat{R}_j\, n^{-\gamma_j}$ with $\gamma_j = \beta_j/(1+\beta_j)$. Finally $n^{-\gamma_j} = (k/n)^{\gamma_j} k^{-\gamma_j} \le \sqrt{2}\, q_{\beta_j}(k)$ since $\gamma_j \le \frac{1}{2}$, and $\Phi(k)^2/\Phi(n)^2 \le \frac{1+\ln k}{1+\ln n} \le 2$ because $\ln(e + \ln n) \le \ln(e + \ln k)$ and $k < 2n$, so $\Phi(n) \ge \Phi(k)/\sqrt{2}$; the two factors lose at most 2, and $2 \cdot 2064 = 4128 \le 8192$. The horizon $T = 32n > 16k$ lies in the range of the supremum defining ϕ_ε. □

F.4 The anytime part of Theorem 5.2

The proof of Theorem 5.2 above treats a single run with $H = T$. The anytime clause runs $\mathrm{NPA}(k, H_j)$ in phases $H_j = 16k \cdot 2^j$ with a fresh permutation in each phase and without resetting the true pull counts.

Lemma F.4 (NPA phase lemma)**.** *Let $k \ge 2$ and $H \ge 16k$. Suppose a phase of length H starts after t_0 pulls with $t_0 + H \le T$, arm i having been pulled n_i times, and runs $\mathrm{NPA}(k, H)$ with a fresh uniformly random permutation on the local curves $g_i(t) = f_i(n_i + t)$, observing the fixed noise table, $\hat{g}_i(t) = \hat{f}_i(n_i + t) \in [a g_i(t), b g_i(t)]$. If the comparator satisfies $\mathrm{LE}(\beta)$, i.e. $f_*(t) \ge m(t/T)^\beta$ for $1 \le t \le T$, then*

$$\mathbb{E}[\textit{noisy phase reward} \mid \textit{history}] \;\ge\; \frac{a^2}{64b}\,\frac{R\, q_\beta}{\Phi(k)}\Big(\frac{H}{T}\Big)^{1+\beta}.$$

Proof. The local curves are nonnegative and nondecreasing, and $g_*(t) \ge f_*(t) \ge m(t/T)^\beta$. Let $h = 8\Phi(k)$, $p = q_\beta/h$, $s = \lceil kp \rceil$ and $d = \lceil H\, p^{1/\beta} \rceil$. As in the proof of Theorem 5.2, for $j \le s$ one has $(j-1)/k < p$, hence $\ell_j \ge \lceil H\, D(p) \rceil = d$ by the choice of the depth envelope (Lemma F.3: $D(p) = p^{1/\beta}$), and $\sum_j \ell_j \le \frac{7H}{32}$, so the commit phase has length $r \ge H/2$. Put $v = a\, g_*(d)$, let A be the event that $*$ occupies one of the first s positions, and let B be the event that some other

arm ends the probe phase with running maximum at least v. The coupling of Lemma F.2 applies verbatim to the local curves and the fixed noise table: fixing the relative order of the other arms and moving $*$ forward does not increase their depths or running maxima, so $\mathbb{P}(A \cap \neg B) \geq \frac{s}{k}\mathbb{P}(\neg B)$.

On B the committed arm c has running maximum at least v, hence true local value at least v/b at its current count, and each of the r commit pulls earns at least $a \cdot v/b = \frac{a^2}{b}\, g_*(d)$; since $g_*(d) \geq m(d/T)^\beta \geq m\,(H/T)^\beta p$ and $m \geq R/T$, the commit phase earns at least $\frac{a^2}{2b}\, R\,(H/T)^{1+\beta}\, p$. On $A \cap \neg B$ the arm $*$ has running maximum at least $\hat{g}_*(d) \geq v$ while every other arm stays below v, so $*$ is committed, and its commit pulls earn at least $a \sum_{t \leq r} g_*(\ell + t) \geq a \sum_{t \leq r} f_*(t) \geq a\, \frac{m\, r^{1+\beta}}{(1+\beta)T^\beta} \geq \frac{a}{8}\, R\,(H/T)^{1+\beta}$, using $r \geq H/2$, $2^{1+\beta}(1+\beta) \leq 8$ and $m \geq R/T$. Both branch coefficients are at least $\frac{a^2}{8b} p$, because $s/k \geq p$ and $a^2/b \leq a$; the two disjoint events then give $\mathbb{E} \geq \frac{a^2}{8b}\, p\, R\,(H/T)^{1+\beta}$, which is the claim with $p = q_\beta/(8\Phi(k))$. □

Proof of the anytime clause of Theorem 5.2. For $T \geq 16k$ the last completed phase has length $H_j > T/4$, since $\sum_{\ell \leq j+1} H_\ell = 4H_j - 16k > T$. Lemma F.4 applied to that phase, all other phases contributing nonnegative rewards, gives $\mathbb{E}[\widehat{\mathrm{ALG}}_T] \geq \frac{a^2}{64b}\frac{Rq_\beta}{\Phi(k)}\, 4^{-(1+\beta)} \geq \frac{a^2}{1024\, b}\frac{Rq_\beta}{\Phi(k)}$. □

G Proofs for Section 5: known scale, and the blocked confidence key

Proof architecture. Theorem G.1 is a separate positive result: Lemmas G.5 and G.6 protect the comparator, Lemmas G.7, G.8 and G.9 bound competing work, and Lemma G.10 completes the two-event reward argument. Proposition G.2 addresses a different limit through the confidence-key, inflated-work and blocked-hitting lemmas (Lemmas G.11–G.13). Corollary G.4 combines already established guarantees by random selection; it is not an additional proof of the noisy lower bound.

G.1 Knowing either prior alone restores a constant price

Knowing β alone, PC with $s = \lceil K_\beta \rceil$ keeps its guarantee up to a^2/b (Theorem 5.1). Knowing m alone, the *random-exponent elimination* algorithm $\mathrm{RET}(H, \mu)$ (Appendix G; reads k, H and a scale μ, not β or ε) tests the arms in a uniformly random order, eliminates the current arm at the first n where its observed value falls below $\frac{\mu}{2}(n/H)^\alpha$ for an exponent $\alpha \in (0, 1]$ randomized per arm, and commits the rest of the phase to the arm with the largest running maximum; the test order and each arm's exponent are coupled through the same random weight U_i, and Appendix G.2 specifies this coupling. For $k \geq 2$, $\varepsilon \in [0, \frac{1}{2}]$, $H \geq 16k$, $\beta \in (0, 1]$, a comparator with $g_*(t) \geq \mu(t/H)^\beta$ for $1 \leq t \leq H$ and remaining true curves nonnegative and nondecreasing, its expected noisy reward is at least $\frac{a}{2048\, b}\mu H q_\beta$ (Theorem G.1); with $H = T$ and $\mu = m$ this reads $\mathbb{E}[\widehat{\mathrm{ALG}}_T] \geq \frac{a}{2048b} R q_\beta$ simultaneously for every such β, and phases $H_j = 16k \cdot 2^j$ with the advice $\mu = m/4$ give $\frac{a}{32768\, b} R q_\beta$ for every $T \geq 16k$. For fixed k, Proposition G.2 (Appendix G.3) gives a bound uniform in T using an algorithm supplied with a valid noise upper bound $\bar{\varepsilon}$; its loss approaches a constant as $\bar{\varepsilon} \downarrow 0$, a different limit from the fixed-ε, $k \to \infty$ regime of Theorem 5.3. Finally, one algorithm that reads none of m, β, T, ε — select uniformly at random one of APC, the phased NPA and the anytime SRMP, and run it — inherits each branch's guarantee divided by 3: $\mathbb{E}[\widehat{\mathrm{ALG}}_T] \geq \frac{a^2}{b}\frac{R}{4608(\sqrt{k} + k/T)}$ for all $T \geq 1$, $\mathbb{E}[\widehat{\mathrm{ALG}}_T] \geq \frac{a^2}{3072b}\frac{Rq_\beta}{\Phi(k)}$ for $T \geq 16k$, and, when $\varepsilon = 0$, $\mathbb{E}[\mathrm{ALG}_T] \geq RQ_\beta(k, T)/(3 \cdot 2^{19})$ for every $T \geq 1$ (Corollary G.4).

Theorem G.1 (known scale)**.** *Let $k \geq 2$, $\varepsilon \in [0, \frac{1}{2}]$, $H \geq 16k$ and $\beta \in (0, 1]$. Suppose the comparator arm satisfies $g_*(t) \geq \mu(t/H)^\beta$ for all $1 \leq t \leq H$, where $\mu > 0$ is the number supplied to the algorithm, and that the remaining true curves are nonnegative and nondecreasing. Then*

$$\mathbb{E}[\textit{noisy reward of } \mathrm{RET}(H, \mu)] \;\geq\; \frac{a}{2048\, b}\, \mu H q_\beta.$$

In particular, when the scale of a cumulatively optimal arm is known — take $H = T$ and $\mu = m = f_*(T)$, so that $\mathrm{LE}(\beta)$ for that arm supplies the hypothesis — we get $\mathbb{E}[\widehat{\mathrm{ALG}}_T] \geq \frac{a}{2048b} R q_\beta$ simultaneously for every such β, without reading β or ε (using $R \leq mT$). If in addition T is

Algorithm 6 $\mathrm{RET}(H, \mu)$ (random-exponent elimination); reads k, H and a scale μ; not β or ε.

1: Draw $U_i \sim \mathrm{Unif}(0,1)$ and test the arms in decreasing order of U_i. Let $\alpha(u) = 1$ if $u \le k^{-1/2}$ and $\alpha(u) = \frac{\ln(1/u)}{\ln k - \ln(1/u)}$ otherwise.

2: Keep pulling the current arm until the first n with $\hat{g}_i(n) < \frac{\mu}{2}(n/H)^{\alpha(U_i)}$; then eliminate it and move on. Probe budget $\lfloor H/2 \rfloor$; when it is exhausted, commit the rest of the phase to the arm with the largest running maximum. If all arms are eliminated before the probe budget is exhausted, *immediately* commit all remaining pulls of the phase to the arm with the largest running maximum observed so far.

unknown, running $\mathrm{RET}(H_j, \mu)$ *in phases* $H_j = 16k \cdot 2^j$ *with the advice* $\mu = m/4$ *gives* $\mathbb{E}[\widehat{\mathrm{ALG}}_T] \ge \frac{a}{32768\, b} R q_\beta$ *for every* $T \ge 16k$*, in the advice model of Section 2.*

The proof is given in full in Appendix G.2; it is the only place where the convention $\varepsilon \le \frac{1}{2}$ is used. Its shape is: with $q = q_\beta$, $p = q/2$, $c = 1/512$ and $v = c\mu q$, the comparator is never eliminated once $U_* \le q$; the expected work spent on the other arms before one of them reaches the level v is controlled by an integral estimate for the exponent function $\alpha(\cdot)$; and two disjoint events — some other arm reaches v, or the comparator is tested in time and is the only arm at that level — give the two branches.

Proposition G.2 (small noise: a blocked confidence-key algorithm)**.** *For* $\bar{\varepsilon} \in [0, \frac{1}{2}]$ *let* $c_{\bar{\varepsilon}} = \frac{1+\bar{\varepsilon}}{1-\bar{\varepsilon}} - \frac{1-\bar{\varepsilon}}{1+\bar{\varepsilon}}$ *and* $D(k, T, \bar{\varepsilon}) = 1 + c_{\bar{\varepsilon}} \min\{T, 32k\}$*. There is an algorithm* $\mathrm{SRMP}^{\mathrm{cf}}$ *(Appendix G.3) which reads* k*,* T *and any valid upper bound* $\bar{\varepsilon} \ge \varepsilon$ *on the noise level — but neither* m*, nor* β*, nor* ε *itself — and which satisfies, for all integers* $k, T \ge 1$*, every* $\beta \in (0, 1]$ *and every instance whose comparator satisfies* $\mathrm{LE}(\beta)$*,*

$$\mathbb{E}[\widehat{\mathrm{ALG}}_T] \;\ge\; \frac{a^2}{65536\, b\, D(k, T, \bar{\varepsilon})}\, R\, Q_\beta(k, T).$$

Consequently, for every fixed k*,*

$$\limsup_{\varepsilon \downarrow 0} \sup_{T \ge 1} \sup_{\beta, I, \hat{f}} \frac{R\, Q_\beta(k, T)}{\mathbb{E}[\widehat{\mathrm{ALG}}_T]} \;\le\; 65536,$$

uniformly in T*: for a fixed number of arms a vanishing noise level restores a constant price at every horizon. This is a different limit from the fixed-*ε*,* $k \to \infty$ *regime of Theorem 5.3; the two regimes do not interact.*

Bounds in Proposition G.2 are stated against the uncorrupted benchmark R; against $\hat{R}$ multiply the ratio by b.

Remark G.3 (what the counterexample does and does not show). Example F.1 refutes the raw-increment keys, the running-maximum *increment* keys and the half-window slope keys fixed above; as noted there, a key based on the *level* of the running maximum, and the reading that keeps the largest dyadic window, are different algorithms that the example does not refute. For some weight vectors the *unblocked* confidence key ($\mathrm{SRMP}^{\mathrm{cf}}$ with $L = 1$ and $\bar{\varepsilon} = \varepsilon$, i.e. $\mathrm{SRMP}(k, T)$ with the probe key $\mu_i(n_i)$, committing to the largest running maximum) also commits to the decoy there; this is not a competitive-ratio counterexample, since for $T \ge 256$, where both arms are included, its probability of committing to the comparator is exactly $4T/(8T + 5) \to 1/2$. Blocking is what the proof of Proposition G.2 needs in order to replace the factor $1 + c_\varepsilon T$ by $1 + O(\varepsilon k)$, that is, to obtain a bound uniform in T; we do not claim that blocking is necessary.

Corollary G.4 (one algorithm that reads nothing)**.** *Let* A *be the algorithm that, before the run, selects uniformly at random one of the three anytime algorithms* APC *(Theorem 3.3), the phased* NPA *of Theorem 5.2 and the anytime* SRMP *of Theorem 4.2, and then runs the selected one to the end. It reads none of* m, β, T, ε*, and*

$$\mathbb{E}[\widehat{\mathrm{ALG}}_T] \;\ge\; \frac{a^2}{b} \cdot \frac{R}{4608(\sqrt{k} + k/T)} \quad (\text{all } T \ge 1), \qquad \mathbb{E}[\widehat{\mathrm{ALG}}_T] \;\ge\; \frac{a^2}{3072\, b} \cdot \frac{R\, q_\beta}{\Phi(k)} \quad (T \ge 16k),$$

and, when $\varepsilon = 0$*, also* $\mathbb{E}[\mathrm{ALG}_T] \ge R\, Q_\beta(k, T)/(3 \cdot 2^{19})$ *for* every $T \ge 1$*. Replacing the third branch by the phased* RMP *of Theorem D.6 improves the noiseless constant to* $R\, q_\beta/4096$ *but restricts that guarantee to* $T \ge 16k$*; the two mixtures are different algorithms and one must choose.*

Proof. Each branch is selected with probability $1/3$ and all rewards are nonnegative, so each guarantee of a branch survives division by 3: $1536 \cdot 3 = 4608$ for the first line (the noisy APC bound $\frac{a^2}{b}\frac{R}{1536(\sqrt{k}+k/T)}$ of Theorem 5.1), $1024 \cdot 3 = 3072$ for the second, and $2^{-19} \cdot \frac{1}{3}$ for the third. The third line uses Theorem 4.2, which is proved in the noiseless model only; under noise that branch contributes only through the nonnegativity of its rewards. Since Theorem 4.2 holds for every $T \geq 1$ and is stated with $Q_\beta(k,T) = \min\{q_\beta, T/k\}$, the third line inherits both properties; the variant using Theorem D.6 instead would give $\frac{3}{4096} \cdot \frac{1}{3} = \frac{1}{4096}$ but only for $T \geq 16k$ and with q_β in place of Q_β. □

G.2 Proof of Theorem G.1 (known scale, unknown β and ε)

Throughout, $k \geq 2$, $\varepsilon \in [0, \frac{1}{2}]$ (so $a = 1-\varepsilon \geq \frac{1}{2}$ and $b = 1+\varepsilon \leq \frac{3}{2}$), $H \geq 16k$, $\beta \in (0,1]$, and the comparator satisfies $g_*(t) \geq \mu(t/H)^\beta$ for $1 \leq t \leq H$; as for every arm of the model (Section 2), all true curves, the comparator's included, are nonnegative and nondecreasing, which both branches of the proof use for the committed arm. Write $K = \ln k$,

$$q = q_\beta = k^{-\gamma}, \qquad p = \frac{q}{2}, \qquad c = \frac{1}{512}, \qquad v = c\mu q, \qquad w = \frac{2v}{\mu} = 2cq = \frac{q}{256}.$$

Since $k \geq 2$ and $\beta > 0$ we have $q < 1$, hence $0 < w \leq p \leq 1$. Recall $\alpha(u) = 1$ for $u \leq k^{-1/2}$ and $\alpha(u) = \frac{\ln(1/u)}{K - \ln(1/u)}$ for $u > k^{-1/2}$; in the substitution $x = \ln(1/u)$ the second branch is $\alpha = x/(K-x)$, which increases from 0 to 1 as x goes from 0 to $K/2$, so α is continuous at $u = k^{-1/2}$, takes values in $[0,1]$, and is nonincreasing in u. The value $\alpha(u) = 0$ occurs only at $u = 1$; there the elimination threshold is the constant $\mu/2 > v$, so the truncated test of Lemma G.7 ends after a single pull, and we use throughout the convention $w^{1/\alpha} := 0$ for $\alpha = 0$ and $0 < w < 1$, which is consistent with that lemma. The set $\{u = 1\}$ is null and does not affect the integral estimates.

Lemma G.5. *If $u \leq q$ then $\alpha(u) \geq \beta$.*

Proof. $x = \ln(1/u) \geq \ln(1/q) = \gamma K$. If $x \geq K/2$ then $\alpha(u) = 1 \geq \beta$. Otherwise $\alpha(u) = x/(K-x)$ is increasing in x, so $\alpha(u) \geq \gamma K/(K - \gamma K) = \gamma/(1-\gamma) = \beta$. □

Lemma G.6 (the comparator is never eliminated)**.** *If $U_* \leq q$ then the elimination test never fires for the comparator: $\hat{g}_*(n) \geq \frac{\mu}{2}(n/H)^{\alpha(U_*)}$ for every $1 \leq n \leq H$.*

Proof. $\hat{g}_*(n) \geq a\, g_*(n) \geq \frac{1}{2}\mu(n/H)^\beta \geq \frac{\mu}{2}(n/H)^{\alpha(U_*)}$, using $a \geq \frac{1}{2}$, the envelope hypothesis, $n/H \leq 1$ and Lemma G.5. The test eliminates only on a strict inequality. □

For an arm i and a weight $u \in (0,1]$ let $\ell_i(u)$ be the *truncated test length*: the number of pulls that arm i would receive if it were tested with weight u, until it is eliminated or its observed value first reaches v, whichever happens first (and H if neither happens). This depends only on the curve of arm i, on the (fixed) noise table and on u; it does not depend on the run.

Lemma G.7 (truncated test length)**.** $\ell_i(u) \leq 1 + H\, w^{1/\alpha(u)} \leq 1 + Hw$ *for every i and u.*

Proof. For $\alpha(u) > 0$ put $n_0 = \lceil H w^{1/\alpha(u)} \rceil$. At the n_0-th pull the threshold $\frac{\mu}{2}(n_0/H)^{\alpha(u)}$ is at least $\frac{\mu w}{2} = v$: indeed $n_0 \geq H w^{1/\alpha(u)}$ gives $(n_0/H)^{\alpha(u)} \geq w$. Hence at that pull either the observation is at least the threshold, and therefore at least v, so the truncated test has ended; or it is strictly below the threshold and the arm is eliminated, which also ends it. Therefore $\ell_i(u) \leq n_0 \leq 1 + H w^{1/\alpha(u)} \leq 1 + Hw$, the last step because $w \leq 1$ and $1/\alpha(u) \geq 1$. For $\alpha(u) = 0$ (only $u = 1$) the test ends after one pull. If a pull both reaches v and triggers elimination, we record it as reaching v, consistently with the algorithm's use of the running maximum. □

Lemma G.8 (integral estimate)**.** *For $k \geq 2$ and $0 < w \leq p \leq 1$, $\quad k \int_p^1 w^{1/\alpha(u)}\, du \leq 2w/p$.*

Proof. Write $L = \ln(1/p)$ and $W = \ln(1/w)$, so $W \geq L \geq 0$. Substituting $u = e^{-x}$, the integral becomes $\int_0^L k\, w^{1/\alpha(e^{-x})} e^{-x}\, dx$. On the uncapped range $x < K/2$ we have $1/\alpha = K/x - 1$, so the integrand equals $e^{E(x)}$ with

$$E(x) = K + W - \frac{WK}{x} - x, \qquad E'(x) = \frac{WK}{x^2} - 1.$$

Case 1: $L \leq K/2$. The whole range is uncapped. For $0 < x \leq L$, $x^2 \leq L^2 \leq LW \leq \frac{K}{2}W$, so $E'(x) \geq 1$ and

$$\int_0^L e^{E(x)} dx \leq \int_0^L E'(x) e^{E(x)} dx = e^{E(L)} - \lim_{x \downarrow 0} e^{E(x)} = e^{E(L)} = k\, p\, w^{1/\alpha(p)}.$$

It remains to check $k\, p\, w^{1/\alpha(p)} \leq w/p$, i.e. $kp^2 \leq w^{-(1/\alpha(p)-1)}$. Here $1/\alpha(p) - 1 = (K - 2L)/L \geq 0$, so the claim reads $e^{K-2L} \leq e^{W(K-2L)/L}$, which holds because $W/L \geq 1$ (for $L = 0$ the integration range is empty).

Case 2: $L > K/2$, i.e. $p < k^{-1/2}$. Split at $u = k^{-1/2}$. On the uncapped part $x \in (0, K/2]$ we have $x^2 \leq K^2/4 \leq WK/2$ because $W \geq L > K/2$; hence $E' \geq 1$ again and the integral is at most $e^{E(K/2)} = k \cdot k^{-1/2} \cdot w = \sqrt{k}\, w \leq w/p$, using $p < k^{-1/2}$. On the capped part $u \in [p, k^{-1/2}]$ the integrand is kw, contributing at most $k\, w\, k^{-1/2} = \sqrt{k}\, w \leq w/p$. Adding the two parts gives $2w/p$. □

Lemma G.9 (work of the other arms)**.** *Let $Y = \sum_{i \neq *} \mathbf{1}\{U_i \geq p\}\, \ell_i(U_i)$ and $C = \{Y \leq H/4\}$, and let D be the event that for some $i \neq *$ with $U_i \geq p$ the truncated test of arm i ends by reaching v. Then $\mathbb{E}[Y] \leq 5H/64$ and $\mathbb{P}(C) \geq 11/16$; moreover C and D are determined by $(U_i)_{i \neq *}$ and are therefore independent of U_*.*

Proof. By Lemma G.7 and Lemma G.8,

$$\mathbb{E}[Y] = \sum_{i \neq *} \int_p^1 \ell_i(u)\, du \leq k + H \cdot k \int_p^1 w^{1/\alpha(u)} du \leq k + H \frac{2w}{p} = k + \frac{H}{64},$$

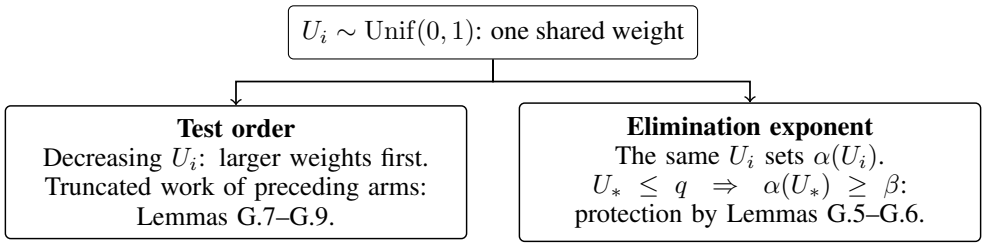
since $2w/p = 2 \cdot \frac{q}{256} \cdot \frac{2}{q} = \frac{1}{64}$. With $k \leq H/16$ this is at most $5H/64$, and Markov's inequality gives $\mathbb{P}(Y > H/4) \leq \frac{5H/64}{H/4} = \frac{5}{16}$. Both Y and D are functions of $(U_i)_{i \neq *}$ and of the fixed curves and noise table. □

Lemma G.10 (the comparator reaches v quickly)**.** *Let $n_1 = \lceil H(v/(a\mu))^{1/\beta} \rceil$. Then $\hat{g}_*(n_1) \geq v$ and $n_1 \leq H/8$.*

Proof. $g_*(n_1) \geq \mu(n_1/H)^\beta \geq v/a$, so $\hat{g}_*(n_1) \geq a g_*(n_1) \geq v$. Also $v/(a\mu) = cq/a \leq 2cq = w \leq 1$ and $1/\beta \geq 1$, so $(v/(a\mu))^{1/\beta} \leq w \leq \frac{1}{256}$ and $n_1 \leq 1 + \frac{H}{256} \leq \frac{H}{8}$ for $H \geq 32$. □

Proof of Theorem G.1. All arms are tested in decreasing order of their weights. We bound the expected reward on two disjoint events.

Branch 1: $C \cap D$. Let j be the arm with $U_j \geq p$, $j \neq *$, whose truncated test ends by reaching v and whose weight is largest among such arms; D says that j exists. Consider the arms tested before j (those with $U_i > U_j$). Every such arm $i \neq *$ has $U_i \geq p$ and, by maximality of U_j, its truncated test ends by elimination, so it consumes exactly $\ell_i(U_i)$ pulls; together they consume at most $Y \leq H/4$ pulls. The comparator, if tested before j, consumes at most $\ell_*(U_*) \leq 1 + Hw \leq H/8$ pulls unless it reaches v; if it does reach v, the committed arm has running maximum at least v and we are done. Note that Y already accounts for j's own truncated test, since $U_j \geq p$ and $j \neq *$. Hence, if the comparator has not reached v, the total work spent up to and including the test of j is at most $Y + \ell_*(U_*) \leq \frac{H}{4} + \frac{H}{8} = \frac{3H}{8} \leq \lfloor H/2 \rfloor$, the last step because $H \geq 32$; the probe budget is therefore not exhausted before j reaches v. In either case the arm committed to has running maximum at least v, hence true value at least v/b at the time of commitment; as the true curves are nondecreasing,



Two cases in the proof

Case 1 $C \cap D$	Some arm reaches v within the probe budget; the committed arm has running maximum at least v.
Case 2 $C \cap D^c \cap \{p \le U_* \le q\}$	The comparator is reached, is never eliminated, and reaches v by Lemma G.10; it uniquely has running maximum at least v and is committed.

Figure 4: The two roles of randomization in the single-phase proof of Theorem G.1. Inputs remain k, H, μ, with $k \ge 2$, $H \ge 16k$ and $\varepsilon \le 1/2$. Here $q = q_\beta$, $p = q/2$, $v = \mu q/512$, and C, D are the events of Lemma G.9. Protection alone does not ensure timely access: the second branch needs its full event and the remaining-budget argument. The elimination test retains its strict inequality.

every one of the $r = H - \lfloor H/2 \rfloor \ge H/2$ committed pulls earns at least $a \cdot v/b$. The branch therefore contributes at least $\frac{a}{b}\frac{H}{2}v$.

Branch 2: $C \cap \neg D \cap \{p \le U_* \le q\}$. All arms tested before the comparator have $U_i > U_* \ge p$ and, by $\neg D$, are eliminated, consuming at most $Y \le H/4$ pulls in total; since $\lfloor H/2 \rfloor - H/4 \ge H/8$ for $H \ge 32$, the comparator is tested with at least $H/8$ of the probe budget left. By Lemma G.6 it is never eliminated, so it is pulled until the probe budget is exhausted, and by Lemma G.10 it reaches observed value v within $n_1 \le H/8$ pulls. Arms with $U_i < U_*$ are never tested, and the other arms with $U_i \ge p$ did not reach v; hence the comparator is the unique arm with running maximum at least v and is committed. Its $r \ge H/2$ committed pulls earn at least

$$a\sum_{j=1}^{r} g_*(n_* + j) \;\ge\; a\sum_{j=1}^{r} g_*(j) \;\ge\; \frac{a\mu}{H^\beta}\cdot\frac{r^{1+\beta}}{1+\beta} \;\ge\; \frac{a\mu H}{2^{1+\beta}(1+\beta)} \;\ge\; \frac{a\mu H}{8}.$$

Combining. Put $x = \mathbb{P}(C \cap D)$ and $y = \mathbb{P}(C \cap \neg D)$; by Lemma G.9 the event $\{p \le U_* \le q\}$, of probability $q - p = q/2$, is independent of C and D. The two branches are disjoint, so

$$\mathbb{E}[\text{reward}] \;\ge\; \frac{a}{b}\frac{H}{2}v\,x + \frac{a\mu H}{8}\cdot\frac{q}{2}\,y = \mu H q\Big[\frac{ac}{2b}x + \frac{a}{16}y\Big] \;\ge\; \mu H q\cdot\frac{ac}{2b}\,(x+y),$$

because $\frac{ac}{2b} \le \frac{a}{16}$. Finally $x + y = \mathbb{P}(C) \ge \frac{11}{16} \ge \frac{1}{2}$, and $\frac{c}{2}\cdot\frac{1}{2} = \frac{1}{2048}$, giving $\mathbb{E}[\text{reward}] \ge \frac{a}{2048b}\mu H q$.

Known scale, known horizon. Take $H = T$, $\mu = m$: the hypothesis is exactly $\mathrm{LE}(\beta)$ for the comparator, and $R \le mT$ gives $\mathbb{E}[\widehat{\mathrm{ALG}}_T] \ge \frac{a}{2048b} R q_\beta$.

Known scale, unknown horizon. Run $\mathrm{RET}(H_j, \mu)$ in phases $H_j = 16k \cdot 2^j$ with $\mu = m/4$, never resetting true pull counts. Let H be the last completed phase; as in Theorem D.6, $\sum_{\ell \le j+1} H_\ell = 4H - 16k > T$, so $H > T/4$. Conditionally on the history, the local curves $g_i(t) = f_i(a_i + t)$ are nonnegative, nondecreasing and discretely concave (the first increment dominates the second because $f_i(a_i+2) - f_i(a_i+1) \le f_i(1) \le f_i(a_i+1)$), and

$$g_*(t) \;\ge\; f_*(t) \;\ge\; m(t/T)^\beta \;\ge\; \frac{m}{4}\Big(\frac{t}{H}\Big)^\beta$$

because $(H/T)^\beta \ge H/T > 1/4$. The theorem applied to this phase gives $\mathbb{E}[\text{phase}] \ge \frac{a}{2048b}\frac{m}{4}Hq_\beta \ge \frac{a}{2048b}\cdot\frac{mT}{16}q_\beta \ge \frac{a}{32768\,b}Rq_\beta$, all other phases contributing nonnegative rewards. □

G.3 A BLOCKED CONFIDENCE-KEY ALGORITHM FOR SMALL NOISE

We now prove Proposition G.2. Let $\bar{\varepsilon} \in [\varepsilon, \frac{1}{2}]$ be the noise bound supplied to the algorithm, $\bar{a} = 1 - \bar{\varepsilon}$, $\bar{b} = 1 + \bar{\varepsilon}$ and $c_{\bar{\varepsilon}} = \frac{\bar{b}}{\bar{a}} - \frac{\bar{a}}{\bar{b}}$.

The algorithm $\mathrm{SRMP}^{\mathrm{cf}}(k, T, \bar{\varepsilon})$**.** If $k = 1$, always pull the unique arm. Otherwise set

$$L = \max\{1, \lfloor T/(16k) \rfloor\}, \qquad H = \lfloor T/L \rfloor,$$

and run $\mathrm{SRMP}(k, H)$ on *macro-steps*, where one macro-step consists of L consecutive pulls of the chosen arm, with the following two changes: (i) the observation attached to the n-th macro-step of arm i is the observation of the last pull of that block, written $\hat{g}_i(n)$; (ii) the probe key is the *confidence key*

$$\mu_i(n) = \min_{t \le n} u_i(t), \qquad u_i(t) = \frac{F_i(t)}{\bar{a}} - \frac{F_i(t-1)}{\bar{b}}, \qquad F_i(t) = \max_{t' \le t} \hat{g}_i(t'),\ F_i(0) = 0,$$

used in the pair $(\mu_i(n_i)/Z_i, 1/Z_i)$; the commitment goes to the arm with the largest $F_i(n_i)$. Any pulls left over after H macro-steps are given to the committed arm. The algorithm reads k, T and $\bar{\varepsilon}$ only.

Macro curves. Let $\tilde{g}_i(n) = f_i(nL)$ for $0 \le n \le H$. These are nonnegative, nondecreasing, discretely concave (an increment of $\tilde{g}_i$ is a sum of L consecutive increments of f_i, and those sums are nonincreasing) and $\tilde{g}_i(0) = 0$; the observation of the n-th macro-step lies in $[a\tilde{g}_i(n), b\tilde{g}_i(n)] \subseteq [\bar{a}\tilde{g}_i(n), \bar{b}\tilde{g}_i(n)]$.

Lemma G.11 (confidence key)**.** *For every arm and every n: (i) $\mu_i(n) \ge 0$; (ii) $\mu_i(n)$ is nonincreasing in n; (iii) $\mu_i(n) \ge \tilde{\delta}_i(n) := \tilde{g}_i(n) - \tilde{g}_i(n-1)$; (iv) if $\tilde{g}_i(n) < v$ then $\mu_i(n) \le \bar{\delta}_i(n) + c_{\bar{\varepsilon}} v$, where $\bar{\delta}_i$ is the increment of* $\min\{\tilde{g}_i, v\}$.

Proof. (i) F_i is nondecreasing and $\bar{a} \le \bar{b}$, so $u_i(t) \ge F_i(t)(\frac{1}{\bar{a}} - \frac{1}{\bar{b}}) \ge 0$. (ii) is the definition of a running minimum. (iii) $F_i(t) \ge \bar{a}\tilde{g}_i(t)$ and $F_i(t-1) \le \bar{b}\tilde{g}_i(t-1)$ give $u_i(t) \ge \tilde{\delta}_i(t)$; the true increments are nonincreasing, so $\mu_i(n) = \min_{t \le n} u_i(t) \ge \min_{t \le n} \tilde{\delta}_i(t) = \tilde{\delta}_i(n)$. (iv) $u_i(n) \le \frac{\bar{b}}{\bar{a}}\tilde{g}_i(n) - \frac{\bar{a}}{\bar{b}}\tilde{g}_i(n-1) \le \tilde{\delta}_i(n) + c_{\bar{\varepsilon}}\tilde{g}_i(n) \le \bar{\delta}_i(n) + c_{\bar{\varepsilon}} v$, using $\tilde{g}_i(n) < v$ (whence $\bar{\delta}_i(n) = \tilde{\delta}_i(n)$), and $\mu_i(n) \le u_i(n)$. □

Lemma G.12 (inflated work)**.** *Let $D_H = 1 + c_{\bar{\varepsilon}} H$, $v > 0$, $\lambda > 0$ and $L_i^D = \#\{n \le H : (\bar{\delta}_i(n) + c_{\bar{\varepsilon}} v)/Z_i \ge \lambda\}$. Then $\mathbb{E}[L_i^D] \le D_H v/\lambda$.*

Proof. $\mathbb{E}[L_i^D] = \sum_{n \le H} \min\{1, (\bar{\delta}_i(n) + c_{\bar{\varepsilon}} v)/\lambda\} \le \frac{1}{\lambda}\big(\sum_n \bar{\delta}_i(n) + H c_{\bar{\varepsilon}} v\big) \le \frac{v + H c_{\bar{\varepsilon}} v}{\lambda}$. □

Lemma G.13 (blocked hitting lemma)**.** *Let $H \ge 32$, let $*$ be a designated arm with $\tilde{g}_*(d) \ge 2v$ for some $v > 0$ and integer $1 \le d \le \lfloor H/4 \rfloor$, and let $r = H - \lfloor H/2 \rfloor$. Then*

$$\mathbb{E}[\widehat{\mathrm{ALG}}] \ \ge\ \tfrac{1}{2} \min\Big\{ \frac{a^2}{b}\, rLv,\ a\, \alpha_D\, P_*(rL) \Big\}, \qquad \alpha_D = \min\Big\{\theta,\ \frac{H}{32\, D_H\, kd}\Big\},$$

where $\theta = \min\{1, H/(128k)\}$ is the inclusion probability of SRMP *and $P_*(n) = \sum_{t \le n} f_*(t)$.*

Proof. This is the proof of Lemma D.9 with four substitutions, which we verify one by one.

(1) *Merge structure.* By Lemma G.11(i)–(ii) the keys $\mu_i(n)$ are nonnegative and nonincreasing in the arm's own macro-count, and (the noise table being fixed in advance) they do not depend on the algorithm's randomness; hence Lemma D.1 applies verbatim to the streams $(\mu_i(n)/Z_i, 1/Z_i, -i, -n)$, and so does the monotone-coupling step of Lemma D.9.

(2) *Counting.* Take $\lambda = 32\theta k D_H v/H$ and let B be the event that some arm other than $*$ ends the probe prefix with running maximum at least av; on $\neg B$, every observed non-comparator arm has true macro-value below v, so Lemma G.11(iv) applies to each such observed prefix. With $C = \{\sum_{i \ne *} X_i \le H/32,\ \sum_{i \ne *} X_i L_i^D \le H/8\}$ we get, by Lemma G.12, $\mathbb{E}\sum_{i \ne *} X_i L_i^D \le$

$\theta k D_H v/\lambda = H/32$, so $\mathbb{P}(C) \geq 1/2$ exactly as before. If the comparator has not reached true macro-value v, then $n_* < d$ and Lemma G.11(iii) with concavity gives $\mu_*(n) \geq \tilde{\delta}_*(n) \geq v/d$, so on $A = \{X_* = 1, Z_* \leq p\}$ with $p = \min\{1, H/(32\theta k D_H d)\}$ its key is at least λ throughout. Under this contradiction hypothesis and on $A \cap C \cap \neg B$, every extra probe of a non-comparator arm uses a key at least λ, and hence is charged to a distinct index counted by L_i^D; therefore the same count $|S| + d + \sum_{i \neq *} X_i L_i^D \leq \frac{H}{16} + \frac{H}{4} + \frac{H}{8} < \lfloor H/2 \rfloor$ contradicts the exhaustion of the probe budget. Hence on $A \cap C \cap \neg B$ the comparator reaches true value v, its running maximum is at least av, and it is committed. Note $\mathbb{P}(A) = \theta p = \alpha_D$.

(3) *Rewards.* On B the committed arm has running maximum at least av, hence true value at least av/b at commitment, so each of its rL committed pulls earns at least $a \cdot av/b$: the branch gives $\frac{a^2}{b} rLv$. On $A \cap C \cap \neg B$ the committed arm is $*$ and its rL committed pulls earn at least $a \sum_{j \leq rL} f_*(j) = aP_*(rL)$.

(4) *Combination.* Identical to Lemma D.9: the two events are disjoint, $\mathbb{P}(A \cap C \cap \neg B) \geq \alpha_D(\frac{1}{2} - \mathbb{P}(B))_+$, and the resulting expression is at least $\frac{1}{2} \min\{\frac{a^2}{b} rLv,\ a\alpha_D P_*(rL)\}$. □

Proof of Proposition G.2. The case $R = 0$ is trivial. For $k = 1$ the algorithm earns at least aR. Let $k \geq 2$, set $x = HL/T$, and write $D = D(k, T, \bar{\varepsilon})$.

If $L = 1$, then $H = T < 32k$ and $x = 1$. If $L > 1$, then

$$\frac{T}{32k} \leq L \leq \frac{T}{16k}, \qquad 16k \leq H \leq 32k, \qquad HL > T - L \geq \Big(1 - \frac{1}{16k}\Big)T > \frac{T}{2}.$$

Consequently, in both cases,

$$x \in (1/2, 1], \qquad H \leq \min\{T, 32k\}, \qquad D_H \leq D.$$

Also $Q_\beta(k, H) = Q_\beta(k, T)$: for $L = 1$ this follows from $H = T$, and for $L > 1$ both quantities equal q_β.

If $H < 32$, then $L > 1$ is impossible because $k \geq 2$. Thus $L = 1$ and $T = H \leq 31$, with no restriction on k. The uniform-arm fallback earns at least

$$\frac{aR}{k} \ \geq\ \frac{a\,R\,Q_\beta(k,T)}{31} \ \geq\ \frac{a^2 R\,Q_\beta(k,T)}{65536\,bD},$$

using $Q_\beta(k, T) \leq T/k \leq 31/k$. Henceforth assume $H \geq 32$.

Apply Lemma G.13 with $v = f_*(dL)/2 = \tilde{g}_*(d)/2$ for an integer $1 \leq d \leq \lfloor H/4 \rfloor$ to be chosen. Using $r \geq H/2$, $rL \geq HL/2 \geq T/4$, $R \leq mT$ and Lemma C.1(2),

$$\frac{a^2}{b} rLv \ \geq\ \frac{a^2}{b} \cdot \frac{T}{4} \cdot \frac{f_*(dL)}{2} \ \geq\ \frac{a^2}{b} \cdot \frac{R}{8} \cdot \frac{f_*(dL)}{m}, \qquad a\alpha_D P_*(rL) \ \geq\ a\alpha_D \Big(\frac{rL}{T}\Big)^2 R \ \geq\ \frac{a\alpha_D R}{16},$$

and $\alpha_D \geq \frac{1}{128} \min\{1, \frac{H}{Dkd}\}$ with $D = D(k, T, \bar{\varepsilon}) \geq D_H$. Since $a \geq a^2/b$ and $\frac{1}{8} \geq \frac{1}{2048}$,

$$\mathbb{E}[\widehat{\mathrm{ALG}}_T] \ \geq\ \frac{a^2}{b} \cdot \frac{R}{4096} \cdot \min\Big\{\frac{f_*(dL)}{m},\ 1,\ \frac{H}{Dkd}\Big\}.$$

By $\mathrm{LE}(\beta)$ and $dL/T = (d/H)x \geq (d/H)/2$ we have $f_*(dL)/m \geq (dL/T)^\beta \geq \frac{1}{2}(d/H)^\beta$ (using $\beta \leq 1$), and $\frac{H}{Dkd} \geq \frac{1}{D} \cdot \frac{H}{kd}$. Choosing the same d as in the proof of Theorem 4.1 (with H in place of T) gives $\min\{(d/H)^\beta, \frac{H}{kd}\} \geq Q_\beta(k, H)/8$, hence

$$\mathbb{E}[\widehat{\mathrm{ALG}}_T] \ \geq\ \frac{a^2}{b} \cdot \frac{R}{4096} \cdot \frac{1}{2D} \cdot \frac{Q_\beta(k,H)}{8} \ =\ \frac{a^2}{65536\,b\,D}\,R\,Q_\beta(k,T).$$

For the limit statement, fix k and take $\bar{\varepsilon} = \varepsilon$: $D(k, T, \varepsilon) \leq 1 + \frac{4\varepsilon}{1-\varepsilon^2} \cdot 32k \to 1$ and $a^2/b \to 1$ as $\varepsilon \downarrow 0$, uniformly in T and in the instance. □

H DETERMINISTIC ALGORITHMS, AND STOCHASTIC NOISE

Deterministic algorithms. Probing every arm $\lfloor T/(2k) \rfloor$ times and committing to the best probe value is deterministic, reads neither m nor β, and earns at least $R/(8k)$ for $T \geq 2k$ (by Lemma C.1(1): the committed arm has $f_c(\tau) \geq f_*(\tau) \geq m\tau/T$); conversely, against any deterministic algorithm, replacing an arm that the virtual all-decoy run (every arm $\min\{t, \lfloor T/k \rfloor\}$) pulls at most $\lfloor T/k \rfloor$ times by a linear arm leaves that run unchanged, since the two curves agree on those pulls and the algorithm is deterministic; every reward in that run is at most $\lfloor T/k \rfloor$, so $\mathrm{ALG}_T \leq T\lfloor T/k \rfloor$ against $R = T(T+1)/2$, which forces ratio $\geq k/2$. The upper bound above is proved here independently for a known horizon; the deterministic lower-bound construction follows Patil et al. (2023). We make no novelty claim for these deterministic bounds. Some horizon restriction is unavoidable, though we do not claim that $2k$ is the sharp boundary: at $T = 2k-1$, probing every arm once and committing to the best probe value already gives ratio at most $2k-1$, so the order is still $\Theta(k)$ there. For $T < k$ a deterministic algorithm leaves some arm unpulled on the all-zero trajectory; making that arm linear and the others identically zero gives $\mathrm{ALG}_T = 0$ and an unbounded ratio. Even at $T = k$ the ratio is $\Theta(k^2)$: either some arm is never visited, and the ratio is again unbounded, or every arm is pulled exactly once, in which case setting the arm visited last to $f_*(t) = t$ and the others to zero leaves the run unchanged until the final step, so $\mathrm{ALG}_T = 1$ while $R = k(k+1)/2$. Conversely, deterministically visiting every arm once gives $\mathrm{ALG}_k = \sum_i f_i(1) \geq f_*(1)$, and discrete concavity with $f_*(0) = 0$ gives $f_*(t) \leq t f_*(1)$, whence $R \leq \frac{k(k+1)}{2} f_*(1) \leq \frac{k(k+1)}{2} \mathrm{ALG}_k$; the deterministic minimax ratio at $T = k$ is therefore exactly $k(k+1)/2$. We therefore state the $\Theta(k)$ characterization on the range $T \geq 2k$, which is sufficient for every use made of it here, and make no claim about the exact horizon at which the order changes.

Additive stochastic noise. Stochastic noise on the rewards. Here the multiplicative model must be replaced with care: with additive sub-Gaussian noise of a fixed variance, no competitive guarantee of the present form can survive unconditionally. Concretely, let the observations carry independent additive Gaussian noise $N(0, \sigma^2)$ with a fixed $\sigma > 0$, compare against the uncorrupted benchmark $R = \delta T$, and place the single arm of constant mean $\delta > 0$ at a uniformly random position $U \in [k]$, all other arms having mean zero. (The position must be randomised: for a fixed designated position the algorithm that always pulls that arm has ratio 1. Nor does "sub-Gaussian" alone suffice, since zero noise is sub-Gaussian and some discrete noises make an arbitrarily small shift exactly identifiable.) Write P_0 for the law of the history under all-zero means and P_u^δ for the law when arm u has mean δ, and let N_u be the number of pulls of arm u. The chain rule for Kullback–Leibler divergence gives $\mathrm{KL}(P_0 \| P_u^\delta) = \frac{\delta^2}{2\sigma^2} \mathbb{E}_0 N_u$, so by Pinsker's inequality $\mathbb{E}_u^\delta N_u \leq \mathbb{E}_0 N_u + \frac{T\delta}{2\sigma}\sqrt{\mathbb{E}_0 N_u}$. Averaging over u and using $\sum_u \mathbb{E}_0 N_u = T$ with Cauchy–Schwarz, $\frac{1}{k}\sum_u \mathbb{E}_u^\delta N_u \leq \frac{T}{k} + \frac{T\delta}{2\sigma}\sqrt{T/k}$. Hence for every algorithm some position forces ratio at least $k - o(1)$ as $\delta \downarrow 0$, and letting k grow rules out any uniform $o(k)$ multiplicative guarantee without further signal assumptions. What is *not* ruled out is a guarantee of order k: pulling one uniformly random arm throughout always achieves that. So the casualty is the scale-free $O(k^\gamma)$-type guarantee of this paper, not competitiveness as such. The question we would like answered is therefore: under which signal-to-noise, minimum-scale or horizon conditions — or against which benchmark corrected for an unavoidable estimation loss — does a prior-free competitive guarantee exist for stochastic noise? Probe-and-commit with confidence-adjusted probe values is the natural candidate on the positive side; the increment-based prior-free algorithms would need new ideas in any case.

I Related work

Two objectives on one model. Heidari et al. (2016) introduced improving bandits, established the optimality of a single arm in hindsight, and obtained instance-dependent sublinear policy regret. Work on the stochastic *rising bandit* model studies additive regret or identification: Metelli et al. (2022) treat regret under sub-Gaussian noise, Mussi et al. (2024) study best-arm identification, Fiandri et al. (2024) give lower bounds and algorithms for the rested case, and Song et al. (2026) study combinatorial rising bandits, whose single-arm case $\mathcal{S} = \{\{1\}, \ldots, \{K\}\}$ is the rising model. We study deterministic worst-case rewards and multiplicative guarantees. This changes both the objective and the adversary; our results make no claim about the upper bounds in that line of work.

Competitive ratios. Patil et al. (2023) study deterministic horizon-unaware algorithms and state an $O(k)$ competitive guarantee. For deterministic algorithms with known T, the minimax ratio is $\Theta(k)$ for $T \geq 2k$, exactly $k(k+1)/2$ at $T = k$, and unbounded for $T < k$. These lower bounds use only determinism and therefore also apply when the horizon is unknown (Appendix H: the construction follows Patil et al. (2023), the horizon qualifications are ours, and we make no novelty claim). Appendix E quotes the precise prior statements and horizon conditions, and records our reading of the randomized extension claimed in Appendix C.1 of the arXiv v1 of Patil et al. (2023) (19 August 2022). Relative to Blum et al. (2026), Theorems D.4, 4.1 and G.1 provide per-instance worst-case guarantees adapting to β without reading it. They address one instance of the direction proposed there, both without noise and with noise when the scale is known. At fixed positive noise, Theorem 5.3 excludes constant uniform adaptation price over β and $T \geq 16k$ when the scale is also unknown, asymptotically in k at the stated hard horizons.