We thank the reviewers for insightful comments and suggestions. We adopted most of your suggestions and added clarifications in the paper. Here, we highlight some of the more significant modifications so far.

Updated Remark 2.1 on Batching

Our fixed confidence algorithm, given by combining Alg. 1 with Alg. 2 as above, requires only $O(1)$ **batches** in expectation. Here a *batched* algorithm operates in a small number of batched phases. At the start of each phase, such an algorithm chooses b arms to sample exactly s times each, where b, s can both depend adaptively on the previous feedback, but cannot be changed during the current phase. Minimizing the number of required batches is often desirable, see e.g. [PRCS16, GHRZ19]. In particular Alg. 1 uses a single batch with $s_1 = \frac{C\log(1/\eta)}{\varepsilon^2}$ samples of $b_1 = \frac{C\log(1/\delta)}{\eta}$ arms. Then Alg. 2 can be implemented in a batched way with $s_2 = \frac{C\log(1/\eta\delta)}{\varepsilon^2}$ and a sequence of batch sizes $b_{2,i} = \frac{2^i}{\eta}$ for $1 \le i \le \log_2\left(C\log(1/\delta)\right)$. (In the latter phase, one stops after finding an arm to accept.) While this construction uses $O(1)$ batches **in expectation**, it could be interesting to explore pure exploration with an exactly fixed number of batches, which is more analogous to the fixed budget setting.

Added some intuition on the independence on η in the fixed budget setting

To succeed in pure exploration, one should have sampled the eventually outputted arm at least $\Omega(\log 1/\delta)$ times. The main obstacle to success in the fixed budget case is that any arm we obtain many samples of might gradually degrade over time. The probability of this degradation is essentially given by small probabilities coming from Chernoff-bound type events, which dominate the prior probability that the arm is in a top quantile.

Updated Proof for Lemma 4

Proof. At each step t, let the random variable E_t denote the minimum possible conditional probability of the event $p_{i^*} \le \beta$ for *any* algorithm (in the algorithm's jointly Bayesian filtration). We claim that conditioned on $\mathbb{A}$'s declarations holding, for any $\mathcal{A}$ the quantity $M_t \equiv E_t \prod_{s \le t} P_s$ evolves as a supermartingale in this filtration. This suffices because it implies $E_0 = M_0 \ge \mathbb{E}[M_T] \ge e^{-\mathsf{Cost}}$. (Here T is the random number of total batches used.)

Indeed suppose we are at time t and the next batch has been declared but not sampled. (Identical arguments apply when the adversary restricts p_{i^*} in the last stage.) Let the σ-field $\mathcal{F}_t$ denote all information up to this point including the declaration of the next batch. Let $\mathbb{E}$ denote an expectation where the samples from the next batch is distributed according to $\boldsymbol{\mu}^t$. Let $\widetilde{\mathbb{E}}$ denote an expectation where $\mathbb{A}$'s declaration is conditioned to hold on the next batch. The dynamic programming principle implies that

$$\mathbb{E}[E_{t+1} \mid \mathcal{F}_t] \le E_t$$

for any $\mathcal{A}$ (with equality for the optimal $\mathcal{A}$). Moreover since $\mathbb{A}$'s declaration has $\boldsymbol{\mu}^t$-probability P_t,

$$\widetilde{\mathbb{E}}[E_{t+1} \mid \mathcal{F}_t] \le \mathbb{E}[E_{t+1} \mid \mathcal{F}_t]/P_t \le E_t/P_t.$$

P_t is $\mathcal{F}_t$ measurable, so $\widetilde{\mathbb{E}}[P_t E_{t+1} \mid \mathcal{F}_t] \le E_t$. This establishes the claim and ends the proof. □

Added Conclusion

Our aim in this paper was to understand the sample complexity of pure exploration with infinitely many arms. We showed that, surprisingly, the behavior of fixed confidence and fixed budget problems is provably very different. In the former setting, there is a nearly optimal algorithm which precisely balances between sampling enough distinct arms (to estimate the quantile) and obtaining enough samples of a single arm (to be output). In the latter, the optimal algorithm must repeatedly decide whether to continue with the current arm or switch to a fresh one, via a gradually decreasing sequence of rejection thresholds.

Several interesting questions remain. One is that our fixed budget analysis is tailored to the $\delta \to 0$ setting, and does not apply if $\varepsilon, \eta, \delta$ all tend to zero at comparable rates. Hence other behaviors could be present in such parameter regimes. Additionally, a key conceptual feature of infinite-armed bandits is the possibility that no "good" arms are among those sampled by the algorithm. By definition, this simply cannot happen in K-armed bandits. It would be interesting to identify a natural problem setting that interpolates between them. Finally high probability bounds on the fixed confidence sample complexity would be another way to interpolate between the two settings we studied.