Comparison to existing literature (R1,R3): As for **R3**'s major comment, our setting is fundamentally more general than [3], which assumes stochastic and i.i.d. delays, while our delays can be arbitrary. For other literature, [1] assumes a constant delay parameter d. [10] considers stochastic rewards and delays. [11] considers the full information case and not the bandit feedback case. It also assumes that all feedback is received before T, and that $\sum_{t=1}^{T} d_t$ is known - which we do not assume. **Our paper is the first to address adversarial (arbitrary) delays and costs with bandit feedback**. Additionally, none of them consider zero-sum games with delays, that we show are surprisingly more robust against delays than the single-agent setting. We now include this discussion, with more details.

Choosing the step size η_t when $\sum_{t=1}^{T} d_t$ is unknown (R1,R3): We provide Algorithm 2 as an adaptive algorithm that does not require prior knowledge of $\sum_{t=1}^{T} d_t$ and T. As shown by the counterexample of **R3**, standard doubling trick epochs are not enough. We now address this issue in detail, fixing Algorithm 2 and providing a full proof that a regret of $O\left(\sqrt{\ln K\left(K^2 T + \sum_{t=1}^{T} d_t\right)}\right)$ is achievable even when $\sum_{t=1}^{T} d_t$ (and T) is unknown, using a novel doubling trick. Let m_t be the number of missing feedback samples at time t, (including the t-th feedback). The idea is to start a new epoch every time $\sum_{\tau=1}^{t} m_\tau$, that tracks $\sum_{\tau=1}^{t} d_\tau$, doubles. Define the e-th epoch as $\mathcal{T}_e = \left\{t \,|\, 2^{e-1} \le \sum_{\tau=1}^{t} m_\tau < 2^e\right\}$, with step size $\eta_e = \sqrt{\frac{\ln K}{2^e}}$. Define by $\mathcal{M}_e$ the set of feedback samples for costs in epoch e that are not received within epoch e. These feedback samples are discarded once received, and the strategy $\boldsymbol{p}_t$ is initialized at the beginning of every epoch. A compact version of the proof is provided next. The K^2 replacing K, which has no affect when $d_t \ge K$, can be improved with a more careful computation. To answer **R3**, Lemma 3 is a general version of Theorem 1 for any arbitrary non-increasing η_t, in particular for any constant η.

Define $T_e = \max \mathcal{T}_e$, and note that $\mathcal{T}_e = [T_{e-1}+1, T_e]$. Applying Lemma 3 on epoch e yields

$$R_e \triangleq E^{\boldsymbol{a}}\left\{\sum_{t\in\mathcal{T}_e}\langle \boldsymbol{l}_t, \boldsymbol{p}_t\rangle - \min_i \sum_{t\in\mathcal{T}_e} l_t^{(i)}\right\} \le \frac{\ln K}{\eta_e} + \eta_e\left(\frac{K}{2}\left|\mathcal{T}_e\right| + 2\sum_{t\in\mathcal{T}_e, t\notin\mathcal{M}_e} d_t\right) + 2\left|\mathcal{M}_e\right|. \tag{1}$$

Now we want to find the maximal $|\mathcal{M}_e|$ such that $\sum_{\tau=T_{e-1}+1}^{T_e} m_\tau \le 2^{e-1}$ is still possible. The "cheapest" way to increase $|\mathcal{M}_e|$ is when the feedback from round T_e is delayed by one (contributes 1 to $\sum_{\tau=T_{e-1}+1}^{T_e} m_\tau$), the feedback from round $T_e - 1$ is delayed by two (contributes 2 to $\sum_{\tau=T_{e-1}+1}^{T_e} m_\tau$) and so on, which gives $\sum_{i=1}^{|\mathcal{M}_e|} i = \frac{|\mathcal{M}_e|(|\mathcal{M}_e|+1)}{2} \le 2^{e-1} \implies |\mathcal{M}_e| \le 2^{\frac{e}{2}}$. Hence, by choosing $\eta_e = \sqrt{\frac{\ln K}{2^e}}$ we obtain

$$R_e \le \sqrt{\ln K}\left(2^{\frac{e}{2}} + 2^{-\frac{e}{2}}\left(\frac{K}{2}\left|\mathcal{T}_e\right| + 2\sum_{t\in\mathcal{T}_e, t\notin\mathcal{M}_e} d_t\right)\right) + 2^{\frac{e}{2}+1} \underset{(a)}{\le} 2^{\frac{e}{2}+1}\sqrt{\ln K} + 2^{-\frac{e}{2}-1}\left|\mathcal{T}_e\right| K\sqrt{\ln K} + 2^{\frac{e}{2}+1} \tag{2}$$

where (a) follows since every $t \in \mathcal{T}_e$ s.t. $t \notin \mathcal{M}_e$ contributes d_t to $\sum_{\tau=T_{e-1}+1}^{T_e} m_\tau$ (the t-th feedback is missing for d_t rounds between $T_{e-1}+1$ and T_e). Therefore $\sum_{t\in\mathcal{T}_e, t\notin\mathcal{M}_e} d_t \le \sum_{\tau=T_{e-1}+1}^{T_e} m_\tau \le 2^{e-1}$. We conclude that

$$E\left\{R(T)\right\} = \sum_{e=1}^{E} R_e \le 2\left(\sqrt{\ln K}+1\right)\sum_{e=1}^{E} 2^{\frac{e}{2}} + \frac{K}{2}\sqrt{\ln K}\sum_{e=1}^{E}\left|\mathcal{T}_e\right| 2^{-\frac{e}{2}} \le 2\sqrt{2}\left(\sqrt{\ln K}+1\right)\frac{2^{\frac{E}{2}}-1}{\sqrt{2}-1} +$$
$$K\sqrt{\ln K}\sum_{e=1}^{E}\left|\mathcal{T}_e\right| 2^{-\frac{e}{2}} \underset{(a)}{\le} 10\left(\sqrt{\ln K}+1\right)\sqrt{\sum_{t=1}^{T} d_t} + 5K\sqrt{T\ln K} = O\left(\sqrt{\ln K\left(K^2 T + \sum_{t=1}^{T} d_t\right)}\right) \tag{3}$$

where E is the last epoch and in (a) we used that $\sum_{t=1}^{T} d_t \ge \sum_{t=1}^{T}\min\left\{d_t, T-t+1\right\} = \sum_{t=1}^{T} m_t \ge \sum_{\tau=1}^{T_E} m_\tau \ge 2^{E-1}$, and also that $\sum_{e=1}^{E}\left|\mathcal{T}_e\right| 2^{-\frac{e}{2}}$ subject to $\sum_{e=1}^{E}\left|\mathcal{T}_e\right| = T$ is maximized when $E = \lceil \log_2 T\rceil$, with maximal length 2^e for epoch e, so $\sum_{e=1}^{E}\left|\mathcal{T}_e\right| 2^{-\frac{e}{2}} \le \sum_{e=1}^{\lceil \log_2 T\rceil} 2^{\frac{e}{2}} \le \sqrt{2}\frac{2^{\frac{\lceil \log_2 T\rceil}{2}}-1}{\sqrt{2}-1} \le 5\sqrt{T}$.

Unbounded delays (R1): We mean that Theorem 2 holds even for **some** unbounded delays s.t. $d_t \le f(t)$ for increasing $f(t)$ (e.g., $f(t) = t\log t$). $f(t) = e^t$, or even $f(t) = t^2$ grow too fast. This is better explained now.

Ergodic Average (R1): This is a weighted average that coincides with the standard average for $\eta_t = \frac{1}{T}$. Its importance is mostly just being computable, so a computation of a NE is still possible by sampling/simulating the game even with superlinear delays. When $d_t = 0$, $\eta_t = \frac{1}{T}$ is a valid choice for Theorem 2 which gives the classical result.

No Exploration Term (R1): It was shown that the exploration term of the original EXP3 is not necessary (see "Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems" by Bubeck & Cesa-Bianchi). In any case, our self sufficient proof independently shows that no exploration term is needed. We have now clarified this issue.

Minor Comments (R1,R3): We have fixed all minor issues (a-e for **R1**, reorganization and line 118 for **R3**). With some effort, the results are extendable to the continuous case, which is exactly the subject of our current work.