We would like to thank the reviewers for their valuable comments. We first address the common criticisms, then turn to each specific comments in what follows.

**Missing relevant work from the scheduling literature:** A common criticism from multiple reviewers is the lack of mentioning about the relationship of MAXREWARD with scheduling problems in the paper. Indeed, there is a strong connection between MAXREWARD and the interval scheduling problems. We have removed the description of this connection from the submitted version mainly due to lack of space (also, we decided to prefer the other related work and thus kept them instead in the paper to comply with the previous blocking bandit papers). We sincerely apologise for this mistake and will add it back to our paper in the next version. This connection is described below in more detail:

The MAXREWARD problem belongs to the class of fixed interval scheduling problems with arbitrary weight values, no preemption, and machine dependent processing time (see e.g., Kolen et al. 2007 for a comprehensive survey). This is one of the most general, and thus, hardest versions of the fixed interval scheduling literature (see, e.g., Kovalyov, Ng & Cheng 2007 for more details). In particular, MAXREWARD is a special case of this setting where for each task, the starting point of the feasible processing interval is equal to the arrival time. Note that to date, provable performance guarantees for fixed interval scheduling problems with arbitrary weight values only exist in offline, online but preemptive, or settings with some special uniformity assumptions (Erlebach & Spieksma 2000, Miyazawa & Erlebach 2004, Bender et al. 2017, Yu & Jacobson 2020). Therefore, to our best knowledge, *Theorem 2 in our paper is the first result which provides provable approximation ratio for a deterministic algorithm in an online non-preemptive setting. Note that with some modifications, our proof can also be extended to the general online non-preemptive setting, i.e., online interval scheduling with arbitrary weight values, no preemption, and machine dependent processing time.*

**R1.** Re: the presence of the reward of the Greedy algorithm in the approximation guarantee is not desirable: Indeed, we can remove the dependence on the performance of the online greedy algorithm in the approximation ratio as suggested by the reviewer. For example, when $D \in O(1)$, we have $r(\pi^*) \in \Omega(T)$. Therefore, for settings with $B_T = o(T)$ we get constant approximation ratio. Note that we also mentioned this in line 208. The reason we still used the form described in Theorem 2 is to provide a convenient way to compare the performance of the online greedy with the proposed bandit algorithm (see Appendix E for more details). We will update our paper to reflect this comment.

**R2.** Re: The hardness result for the offline problem with small blocking values: We can easily extend our current proof to the case of $T >> D$. In particular, let $T_0 = n + m = D$. We use the same proof in the paper but replace $T$ with $T_0$. Now assume that $T >> T_0$ (and thus, $T >> D$). For any $T_0 < t < T$, we set the rewards to be 0 and blocks 1 for all the arms. It is still true that the optimal solution of this instance is linked to the solution of the original 3-SAT.

**R3.** Re: more comprehensive numerical analysis needed: We indeed only focus on the theoretical analysis of the blocking bandit model. The numerical results in Appendix E is only for supporting the theoretical comparison between Greedy-BAA and RGA. In particular, Eqs (19) and (20) from Appendix E show that Greedy-BAA is significantly better than RGA when $B_T$ is small (i.e., the regret bound of RGA is $O(\sqrt{T/B_T})$-time larger). Hence the choice of $B_T = 3$.

**R4.** Re: It seems like the main idea follows the standard reduction form bandit to full feedback with some non-trivial adaptation: We agree with the reviewer that the theoretical analysis of the bandit part is a non-trival adaptation of known techniques. However, we believe that this part still has its merits, as it provides a neat analysis for a new and interesting bandit problem, laying the foundation for other adversarial blocking bandit models (e.g., contextual, combinatorial, etc). This, combined with the other contributions of the paper, make our findings novel.