We would like to thank the reviewers for thoughtful and constructive comments. Minor suggestions and typos will be reflected in the final version, and we will cite Degenne et al, 2020.

Computational complexity: Reviewers have pointed out the computational complexity of the algorithm. Indeed, the computational complexity is not great and is left as future work. Our main contribution is on adapting the sample complexity to finite regret and improving the dependency on K with a new algorithmic idea. More importantly, we reveal important subclasses of hypotheses in $\mathcal{F}$ that must be dealt with care, which was all hidden under forced exploration or the inflation added to the optimization problem, as researchers have done. That said, the structured bandit problem with a generic $\mathcal{F}$ covers a large variety of problems, and finding one algorithm that is optimal for all is very challenging. There are improvements to be made for popular structures like linear models; we address some of them below in response to the reviewers' comments.

R1: Item 2: When $P_1 = 0$, we have $\gamma^* = 0$, and those f's in the equivalence class also has $\gamma^*(f) = 0$ by definition. This makes the constraint set of the optimization problem $\phi(f^*)$ empty, thus $P_2 = 0$. Item 3: The constant c_1 is 4. One can replace β_t so to the order $\log(1 + t\log^2(t))$ and apply the usual analysis technique (Section 8 of Lattimore&Szepesvári, "Bandit Algorithms", 2020) to get the exact asymptotic optimality, we believe. However, this would make the analysis longer. We will make this clear in the final version. Regarding the linear case, please see our response to R3 below.

R2: For the confidence set for infinite hypothesis class, tools from empirical processes can help. We suggest trying out without the conflict case; we speculate that the conflict case may be automatically resolved by arm pulls made by ψ for simple structures. Also, one may construct a better definition for ψ as we discussed above.

R3: We agree that extending it to the continuous case is not immediately obvious. This is beyond the scope of the paper that we are currently working on. The algorithm may need some modifications, but the pessimism principle can still be useful, we believe. Also, the reviewer is correct that it may end up entering the conflict case a lot. However, P_2 may not be that large as it is related to $\mathcal{E}^*$ only, the equivalence class of f^*; further, many f's in the disc may support the same arm. Some more detail: entering the conflict cases when $\overline{f}_t \notin \mathcal{E}^*$ will contribute to P_3 – these happen for a finite number of times. We believe more work is needed to deal with cases like Figure 2 while mitigating the side effect mentioned by the reviewer.

On the linear case: Thank you for the examples! While $K_\psi = O(\log(K))$ for the cheating code example, it is true that for some problems K_ψ can be close to K (including linear bandits as pointed out by R3). This brings up an interesting point. Clearly, our regret upper bound is far from being optimal for the example R3 provided ($K = 2^d$) since we know that the regret of OFUL does not scale like 2^d. But then, what is the optimal finite-time instance-dependent regret here, and how does it scale with K? Studying finite-time lower bounds may help and is an interesting open problem.

R4: The optimization is often convex (e.g., linear bandits) and can be solved in a reasonable amount of time with popular software. For finite classes, they scale with the number of hypotheses, and can easily be inefficient for a very large number of hypotheses. New algorithms that do not require solving the optimization problem in the first place is an interesting open problem.

The final question refers to optimization 5, but we assume it was Eq. (4). The two issues the reviewer mentioned are indeed the reason for the change. For example, in Figure 2, imagine adding $f_5 = (.97, 1, .99, .25, .25)$. The pessimism at the beginning is f_3, and we will keep pulling arm A5 (the only action in the support of $\gamma(f_3)$) but f_5 will not be eliminated.

To answer the question on the ERM, note that the example above shows the weakness of tracking the ERM *without* forced sampling; in earlier rounds, the ERM can be f_3 with nontrivial probability, and it will get stuck for the same reason. We believe the ERM can still work if we properly leverage ψ and ϕ in similar ways as CROP, which deals with docile hypotheses and conflicting hypotheses. However, we are not sure if this will make the algorithm any simpler. The forced exploration is a simple and naive solution that can replace the role of ψ and ϕ, but it brings many issues mentioned in the paper. Overall, we believe the pessimism is just a simple choice that works.