We would like to thank the reviewers for their time and helpful comments. We will clarify/fix the paper as suggested.

Reviewer #1

> *the computational complexity is not studied or evaluated so the practicality of this approach might look questionable.*

Thank you for pointing that out. Note that all algorithms we proposed run in quasi-linear time in the number states, actions, samples, and $1/\gamma$ (particularly solving the RMDPs). We unintentionally omitted this discussion from the paper and we will rectify the omission. Also, the Batch-RL setup is constrained by samples and not computational complexity.

> *It would be great to improve the readability by having intuitive explanations and illustrative examples.*

We agree. There was a tradeoff in writing and explaining the ideas while satisfying the page limit constraints. We will add a very simple clarifying example as suggested.

Reviewer #2

> *sets $\mathcal{K}$ and $\mathcal{L}$ are formally laid out, but the prose does not inform the reader of why they must be defined this way.*

Thank you for pointing out this omission. We will add the intuition for these sets. Briefly, the set $\mathcal{K}$ is the set of probability distributions which, if contained in the ambiguity set, are sufficient to guarantee the safety of the solution.

> *Figures more directly tied to the proposed algorithm would be nice as well; for example, a visualization of how the POV might evolve throughout the training.*

Figure 3 is an attempt to demonstrate just that. We agree that a better visualization could help to better explain it, but we were not able to add one due to the page constraints. We will add such an illustration in the supplementary material.

> *experiments consist of only some very basic grid-world experiments against simple baselines*

We agree, we also think that future work should obtain results on bigger domains. The paper focuses on simple domains to avoid compounding errors from value function approximation and modeling.

> *The approach performs uniformly worse than a generic non-safe baseline on all tasks, which means that the experiments are inadequate for showing effectiveness.*

The non-safe baseline is not comparable to the other methods since it does not provide any guarantees. We concur that including the non-safe method in the same plot is confusing and will remove it.

> *But if we have an accurate prior, and the ability to do efficient posterior updates (or at least efficient posterior sampling), then we can just solve for the Bayes-optimal policy on the distribution of MDPs sampled from the posterior, and get optimality guarantees for that: no need to bother with robustness. Is this interpretation correct? If so, what do robustness methods add in this setting?*

That is an interesting point. The problem is that the Bayes-optimal solution requires solving a POMDP (difficult) and does not offer the required performance guarantees. The robust solution is easier to compute and provides the requisite performance guarantees. Our work can also be seen as a risk-averse solution for the Bayes-MDP but that is a topic for a different paper.

Reviewer #3

> *The computational complexity is also not established, pushing the assessment further into speculations.*

We agree that the analysis of computational complexity is important. Note that all algorithms we proposed run in quasi-linear time in the number states, actions, samples, and $1/\gamma$ (particularly solving the RMDPs). The only unknown is the number of iterations necessary, but in all our experiments the optimization converges in less than 10 iterations. We will emphasize this point in the paper.

> *The paper is dense in parts and the language is sometimes confusing. The notation is not easy to follow.*

We will make the language more precise and add a notation table as suggested.