**Rev. 1** **1)** Yes, the formulation of "robust Markov game" originated from [12]. But our contribution is to model the "robust MARL" problem as such a game and develop algorithms to solve it, both of which are original. Introducing the "nature" agent is exactly one contribution that makes the computation tractable and easier to understand in MARL context. It is not simply a "trick", but our effort to transfer [12] to robust MARL. We will add more clarification on this. **2)** With general function approximation, PG is solving a non-convex optimization problem, the theoretical guarantee of its optimality is very challenging. In fact, the optimality guarantee of PG methods, even in the single-agent setting, can only be established in limited cases until recently. Note that the equilibrium guarantee of the original MADDPG is non-existent yet. As our work is the first one that dealt with robust MARL under the framework of robust Markov games, many other aspects, e.g., the algorithm design, solution existence, Q-learning convergence, and empirical validation, need to be done first . **3)** We had put experiment details in the supplement due to page limit. We will re-organize the experiment session to add those details and include results for the cooperative navigation in the main content. We have updated all the tables with 95% confidence intervals across 5 trials, 1000 evaluation episodes for each. Below is an example for keep-away. The figures for the cooperative navigation in the supplement have also shown 95% CI across five runs. We hope our response can help address your concerns on the novelty, and re-evaluate our submission.

**Rev. 2** Thanks for the very positive feedback, we will address the grammatical typos and the notations in Table 2, and add more clarification on the nature agent. **1)** We will re-iterate the "distribution-free" point and add clarification to avoid the possible confusion with "distributional uncertainties". **2)** For Keep-away, we have run student's t tests. All calculated $p$-values are below 0.01, suggesting that the improvements are statistically significant.

**Rev. 3** **1)** We are not sure what was meant by "as I mentioned above", as there was only the reviewer's summary above it . **2)** The "nature" policy is approximated by neural networks, as for other agents. We are confused about the comment "eq 3.5 only minimizes the reward", since Eq. 3.5 has no minimization. There is no "original reward" in our robust MARL setting, as the reward function is "uncertain". Instead, the nature agent is always "playing against" all agents, while the agents want to find the policy that account for this adversarial uncertainty. The MSE in line 13 is some heuristic to regularize the nature policy's output, with $\eta$ being the weight coefficient. We have mentioned the choice of $\eta$ in Sec. F (appendix), and will add more clarification in the main paper.

| Models | | Uncertainty level ($\lambda$) | | | |
|---|---|---|---|---|---|
| AG | ADV | 0.0 | 1.0 | 2.0 | 3.0 |
| M3 | M3 | **13.23 (0.12)** | 6.34 (0.18) | 3.01 (0.14) | 1.34 (0.09) |
| M3 | MA | 8.07 (0.18) | **7.51 (0.17)** | 3.47 (0.14) | 2.86 (0.13) |
| M3 | RM | 13.15 (0.12) | 7.23 (0.18) | **5.63 (0.17)** | **4.31 (0.16)** |
| MA | M3 | **13.57 (0.11)** | **8.12 (0.18)** | 2.58 (0.13) | 1.66 (0.09) |
| MA | MA | 10.91 (0.16) | 6.04 (0.17) | 3.89 (0.15) | 1.44 (0.09) |
| MA | RM | 12.52 (0.13) | 6.96 (0.17) | **4.05 (0.15)** | **3.47 (0.14)** |
| RM | M3 | **11.87 (0.15)** | 6.98 (0.18) | **5.65 (0.16)** | 2.09 (0.11) |
| RM | MA | 8.83 (0.17) | 7.38 (0.18) | 4.37 (0.16) | 3.25 (0.14) |
| RM | RM | 7.89 (0.18) | **8.21 (0.18)** | 4.32 (0.16) | **3.61 (0.14)** |

Table 1 in the paper updated with 95% confidence interval.

**3)** Clarity: We will address the typos. The "robust Markov games" is fundamentally different from "Markov games with stochastic reward/transition". The latter is nothing but the standard MG model, since it allows the reward/transition to be stochastic. However, our robust MG model accounts for the "model uncertainty" explicitly, like the single-agent setting in [1]. The mathematical description of uncertainty is explicitly provided in the robust MG formulation. **4)** Related work: We note that they are not that relevant, as non-stationary MDP usually deals with an online/adaptive setting, with the focus of reducing the accumulated regret, while we account for the "model uncertainty" from the beginning, and our algorithms are model-free . Most importantly, our main focus is on the "multi-agent" side. **5)** The MADDPG paper [19] has already shown that MADDPG outperforms DDPG in all these particle environments. There is no added value to compare to DDPG again. **6)** About the variance, please check our response for reviewer 1, answer 3. Since we run evaluation over 1000 episodes for each trial, the variance, i.e., CI, is fairly acceptable. **7)** We have the same intuition that the confidence interval for high noise level (7.0) should be larger than 5.0/6.0, we think we get the narrower CI only by chance. **8)** We have started running the suggested ablation experiments that the nature always selects the minimum reward in an uncertainty set. But the training takes much longer for convergence and we could not obtain the final results before deadline. We will add it to the final version. We notice that most of your comments are regarding the clarification, and the "weakness" part seems not that fatal. We sincerely hope our response can help address your concerns, and re-evaluate our paper.

**Rev. 4** **1)** Clarity and framing: our focus was "the agents may not have perfect information of the model" from each agent's perspective, including both *its own* and *others'* model. Sorry for the confusion in the traffic network example, we will clarify on this point. **2)** When there is no reward noise, non-R-MADDPG methods can obtain the exact reward, while R-MADDPG still uses the nature agent to approximate the certain reward, which leads to extra approximation errors. However, R-MADDPG does not always "perform notably worse" than the baselines. E.g., in the keep-away experiments ($\lambda = 0$ or 1), R-MADDPG adversary performs better than MADDPG most of the time, and slightly worse than M3DDPG. In practice, when entering a new environment, we suggest running both algorithms and compare. **3)** Please check our response to reviewer 1, answer 3 and reviewer 3, answer 6 on the variance concern. **4)** We will explain the "Without loss of generality..." sentence with more details in revision. **5)** We will provide the proof of Proposition 2.2 for completeness. As the main weakness was on clarify and framing, which can be addressed as per your suggestion easily, we really appreciate the reviewer to re-evaluate our contribution, especially considering that our work provides the first framework that accounts for "model-uncertainty" in MARL, with both theory and simulations.