We thank the reviewers for their time and thoughtful feedback! We are encouraged that they found our method intuitive (**R1**,**R2**), simple (**R1**,**R4**), novel (**R2**,**R3**,**R4**), well positioned with respect to past work (**R1**,**R2**,**R4**), and overall clearly presented (**R1**,**R2**,**R4**). We were furthermore heartened that reviewers found our empirical evaluations were a highlight (**R3**), extensive (**R4**), convincing (**R2**), and surprising (**R1**). We found the criticisms overall very constructive and appreciate them greatly regardless of whether our submission is accepted!

More baselines The most common criticism from the reviewers was that we only experimentally compare our method to purely selfish agents (**R1**,**R3**,**R4**). We present the first method that shows emergence of *both* reciprocity and team formation, and because there are no other methods that have claimed as such, we felt there were no obvious choices with which to directly compare. High social welfare, robust equilibria have been so elusive that MARL social dilemma research often investigates whether these behaviors can emerge *at all* rather than comparing efficiency in obtaining those behaviors. Finally, none of the reviewers suggested specific methods with which to compare, and comparing against the multitude of prior methods that achieve only a subset of the behaviors emergent with RUSP would be very labor and compute intensive to the extent it may deserve its own independent publication and so we chose to leave this to future work.

R1- **More intuition on sustained cooperation** Thank you for these great questions! We are happy to add more discussion around this in the paper. Our intuition on why cooperation persists to evaluations with selfish, certain preferences is that there are cases during training where agents with selfish preferences but asymmetric uncertainty (such that one has selfish, certain preferences but the other selfish, uncertain preferences), allowing the agent to experience the requisite variance over cooperative and defective strategies. In cases where agents learn without uncertainty during training, we believe it may be from the smooth transition over the threshold where cooperation is directly incentivized (high reward sharing) and where it is less clear if it is beneficial (low reward sharing). This hypothesis is somewhat supported by the results in Figure 4 where we see the cases with hard teams and no uncertainty failing to learn cooperative behavior.

R2- **More discussion of method limitations** We are happy to discuss limitations more in the paper. We already mention the potential credit assignment issue with reward sharing methods in Section 6; we also found that past policy play was necessary and would like to investigate more in future work how this and other methods that induce variance in agent play interact with RUSP.

R3- **"I feel the work is closer to game theory literature than MARL literature."** Works like ours that focus on learning in higher complexity environments such as harvest, cleanup, and oasis have historically been submitted to ML journals rather than game theory journals. The focus of this paper is providing pressures for MARL methods to converge to higher social welfare equilibria so we feel it belongs in the MARL camp.

R3- **"Showing computation of reward transformation matrix for one of the games used for experiments will be helpful."** We already give an example of this in Figure 2(b).

R4- **"In proposing this RUSP environment augmentation, the justification for this approach is not fully convincing."**, **"The authors do not offer a theoretical grounding for their work."** Our motivation (in Section 3 of the paper) is based around the intuition that RL agents will learn adaptive strategies in the face of uncertainty or partial observability. In RUSP, we induce uncertainty over social preferences which we hypothesize and validate experimentally leads to social adaptability and robust cooperation. That being said, we would love to see future work linking RUSP to biological mechanisms or more game theoretic justifications for the method, and we will mention this as an interesting avenue for future research in our discussion section.

R4- **"The empirical evaluation is also limited"** Evaluating in multiple IPD matrices is not standard practice to our knowledge (e.g. see LOLA); in general modifying the payouts should simply move the threshold on the discount factor at which some cooperative strategies such as Grim Trigger become Nash. We did not cherry-pick or finetune this matrix – it was simply the first one we tried, and we are happy to say as such in the paper.

R4- **"the authors do not describe their RL methods; it is unclear which RL algorithms they leveraged"** We describe them in the Appendix (C) and say as such on L147. Our RL algorithm (distributed PPO with omniscient value function) is standard so we did not think it important enough to describe in detail in the main text but will add more.

Improving Paper Clarity In the final submission we will be happy to include algorithm pseudo-code (**R3**), add introductory sentences to the sections to set expectations (**R3**), expound upon the problem definitions in the Preliminaries section (**R3**), divide the Method section into subsections "motivation" and "method" (**R3**), add more discussion about the Oasis environment and results (**R1**,**R2**), clarify the explanation of the indirect reciprocity evaluation setups (**R1**,**R2**), soften language around uncertainty's efficacy for Prisoner's Buddy (**R1**), and fix any typos + minor clarifications (**R1**,**R2**,**R3**,**R4**). We thank the reviewers again for the their time and thoughtful comments, and we hope that with these (and those listed above) improvements to the paper, **R3** and **R4** will consider increasing their score!