Thanks for the thoughtful comments and feedback. We are pleased that most reviewers agreed that work on >2player games is needed: R1 wrote ‘this problem is important and . . . a huge step-up from . . . 2 player games’, and [R3, R4] say the >2 player domain is a strength of the paper. All reviews commented that we used multiple evaluation metrics, with positive feedback on the thoroughness of our analysis. We feel this is crucial in mixed-motive settings, where no single metric fully captures performance, so we made a special effort here - thanks for recognising this.

R3 raises an important point about whether starting with SL means results against DipNet might be misleading. For the author feedback **we ran the suggested experiment using Albert** to resolve this. Using Albert is tricky: it is slow and only available as a Windows binary using DAIDE. We replicated, within statistical error, DipNet’s result: **1 DipNet won $\mathbf{0.39 \pm 0.06}$ vs 6 Alberts** (winrate±95% CI). We only had time to test one of our methods; **FPPI-2 won much more vs 6 Alberts $\mathbf{0.75 \pm 0.05}$**. For the camera-ready paper, we’ll add results for all algorithms vs Albert to Table 1.

R2 wrote: ‘The contribution is very domain-specific, limiting the target audience at NeurIPS’. **We disagree:** (1) As R4 said, research specific to Diplomacy is still of interest to the NeurIPS community. (2) SBR, FPPI, and our evaluation methods are novel and general. (3) We present novel empirical results relevant to other domains and algorithms.

(1) Diplomacy is an **established challenge** for the AI community, with many papers in AI conferences since the 1980s, including the (very domain specific) DipNet paper at NeurIPS last year. Most of our reviewers felt this domain was of interest, so we are confident that many in the NeurIPS audience will be interested in the work.

(2) R2 wrote ‘Section 3.1-3.2 present the potentially most transferable part of the method without crisp demarcation...’. We used ideas in previous works, with new ideas to scale to Diplomacy. *NFSP and PSRO* approximate Fictitious Play (FP) using RL, applying model-free RL to produce best responses (BRs). But *DQN* (used by NFSP) requires a small action space, and *A2C* is ineffective in Diplomacy. In contrast, we use SBR to produce best responses without running an RL algorithm in the inner loop. This made FPPI-1/2 possible, the first Policy Iteration (PI) methods to approximate FP. *ExIt and AlphaZero* also use PI, but their MCTS requires sequential moves and only $\sim$ 100s of actions per turn. *CFR* (e.g. [18,81]) handles simultaneous moves, but require enumeration of the joint action space. SBR is more general as it handles simultaneous moves and larger action spaces. We’ll add a detailed discussion to highlight these points.

In summary, previous self-play RL methods cannot cope with the action space or simultaneous moves, and have rarely been studied in the general many-agent case. These features characterise **many domains** such as large scale fleet management, multi-commodity markets and multi-robot control (among the further domains that R1 pointed out). Finally, our evaluation methods, e.g. Nash based policy transitivity, are almost entirely general to many-agent settings.

(3) Challenge domains like Go, Poker, or Diplomacy are useful since they tell us what kinds of methods work. Insights from domain specific methods, e.g. AlphaGo lead to more general methods, e.g. AlphaZero. **Key takeaways** from this work are: sampling-heavy best response (BR) estimation is sufficient to tackle large-scale many-player environments; stochastic BRs improve the convergence of IBR; and that FPPI-2’s method of averaging policies is more effective than FPPI-1’s NFSP-style method. **We’ll emphasize these contributions**, of interest to the wider MARL community.

R1 suggests we study the effects of B/C values in Diplomacy. We agree this is interesting and will add ablations on it.

R1 and R3 asked about diagonal entries of Table 1. In 1v6, the ‘6’ are identical, i.e. from the same training run. So a ‘1’ player from a different training run has a slight disadvantage, being further from their training distribution. For BRPI methods, we used multiple training runs, this is why for IBR and FPPI-2 the ‘1’ player had a winrate below 1/7th.

R3 asked about **the exploitability of the SL agent**. For RL agents we use their own value function to get a very specific exploit. SL’s value net is very inaccurate, so we used a worse exploiter vs SL, using an *arbitrary* BRPI value function. Comparing these exploits between SL and BRPI confounds exploiter quality and exploitability. We have since tried different BRPI value nets to exploit SL, showing we can exploit SL by more than the RL agents. Few-shot exploitability does not have this issue, enabling better between-algorithm comparisons. We’ll add a discussion of this.

Answers to specific questions: R1: SBR, while effective, has no theoretical guarantee: there is a game with $\leq 2(B+2)$ actions where SBR with B base profiles prefers a suboptimal action vs the Nash. We can add this result with proof to an appendix. R2: We discussed in section 2.1 how different unit moves are interdependent. This implies that performance is **not smooth** over the action space. Elo *assumes* transitivity to model skill, in contrast we are *testing* whether transitivity in fact holds. R3: We agree with the comment on many-player Poker. Weighted averages in BRPI are interesting future work. R4: Yes, ‘Historical network checkpoints’ are previous versions (parameters) of the network. Meta-games are 0-sum as players’ rewards are equal to the rewards in Diplomacy, which is (ultimately) 0-sum. Appendix A.3 gives an example of bad behaviour from exact BRs in IBR, we will refer to it on line 165. We will add more detail on exploitability to the main text.

We thank the reviewers for the positive feedback on clarity, discussion of related work and broader impact section, and are grateful for the suggestions on ways to improve these further, which we will incorporate in the final version.