We would like to thank all of the reviewers for their time and thoughtful comments on our paper.

To begin, we concede that the attention mechanism in SARNet may appear similar to TarMAC (***Reviewers 1, 3***) and related to Multi-Actor-Attention-Critic (MAAC) (***Reviewer 4***). However, there are important differences in SARNet that we argue contribute to its significant gains in performance: (1) **how the attention mechanism is calculated with respect to TarMAC**, and (2) **MAAC's use of critic-attention only to reduce state-space representation, not for improving communicating policies**. TarMAC and MAAC both adopt dot-product attention that *equally* sums across the query-key pairs [9]. In contrast, SARNet uses an attention scheme based on a Hadamard product followed by a linear projection, which allows the network to generate richer and more effective communicating policies by learning interactions *across* query-key pairs. To substantiate this claim, we performed an analysis of TarMAC's dot-product attention applied to SARNet's memory in Appendix C.1, showing improvement in SARNet when moving to a Hadamard/projection based attention. More importantly, SARNet's use of a dedicated memory unit and the ability to *simultaneously attend* to both *newly received information and past memories* allows SARNet to have substantial performance gains over TarMAC, as TarMAC can only attend to new messages (values). Regarding the **omission of MAAC in our baselines**, our focus was on architectures that perform explicit communication during the execution phase. **MAAC uses the attention mechanism for the centralized critic** during training and not in the action policy. Based on ***Reviewer 4***'s suggestion we will add results from **MAAC to complement MADDPG as a baseline without communication**. Since receiving the reviews, we have performed initial evaluation with the following **results for recurrent-MAAC with extended TD3 (MT2D3)**: (1) Cooperative Navigation ($N = L = 6$) resulted in an aggressive policy with lower avg. distance to landmarks, but significantly higher collisions than SARNet, with rewards -22.02$\pm$ 0.87 *vs* SARNet's -12.39$\pm$ 1.0 , (2) Predator-Prey 6 *vs* 2 with a mean score of 14.49$\pm$ 0.46 *vs* SARNet's 17.51$\pm$ 0.26 .

With regards to our **training curves and attention metrics**, we agree with the reviewers and will improve the graphs to make then more readable by **adding error bars in the training graphs** to better reflect training stability.

Our contribution of **MT2D3 has been applied to competitive scenarios in the paper, with Predator-Prey**, where the agents compete with each other. We have described it in Appendix A.1.4 and we will add further details by including figures on the design methodology. Agent training, **both for SARNet and all baselines**, was performed with **MT2D3 for the continuous action space environments**, and **REINFORCE for discrete tasks of Traffic Junction**.

Reviewer 1: We appreciate the feedback to make our paper more concise, and we **will combine the Thought and Question Unit** in a single section. Choosing to have a **maximum of 20 agents for each environment** is attributed to limits on computation and the fact that the baseline works in our paper have trained up to a maximum of 20 agents. For Predator-Prey environments, we had a maximum of 12 *vs* 4 agents as training involves two different architectures with different parameters, which heavily affects training time. We are actively working to address agent limits by introducing a scalable multi-GPU multi-agent RL library to reduce training times, which will be released in the near future.

Reviewer 2: Your suggestion on including an analysis of the memory unit is very valuable. First, the term **reasoning** is inspired from RRL [11] and NLP [24], where the authors term the interactions of query-key-value pairs as reasoning between different entities. However, to clarify the reasoning that occurs in SARNet, we will add an **analysis of the memory** through a Principal Component Analysis. Usage of **multi-step/multi-head attention** was explored, but it required the memory unit's write method to use more computation time as it would require N-memory reads/writes. SARNet can incorporate **forget/write gates for the memory unit for longer tasks** similar to that of an LSTM. However, we did not see performance gains for the tasks in the paper. We will note results with *gates* in the revision. We leave the **scalibility** of our approach for **larger tasks** for future work, through an extension with Graph Neural Networks. **Estimates on running times** for tasks are reported in Appendix A.2, and will be noted in a dedicated table. We agree with the reviewer, and will revise the manuscript to add a **descriptive analysis of IC3Net**. As the authors of IC3Net have noted, IC3Net is CommNet with gates when trained with individualized rewards. The additional complexity in training of the gating function in cooperative environments partially explains IC3Net's lower performance.

Reviewer 3: We have described **key differences between SARNet and TarMAC** in our response (lines 2-13). Additionally, SARNet is equipped with a distinct memory unit that does not rely on an RNN encoder to aggregate messages, and is thus *adaptable to non-recurrent observation encoders*. **Performance of SARNet in Traffic Junction for 6 agents** is within the standard deviations of the baselines as communication is not critical for a few agents. However, SARNet's performance is substantially better than baselines when the task becomes harder (more agents) and communication is key, a trend that can be observed across all environments.

Reviewer 4: The suggestion to include MAAC as a baseline is highly appreciated, and we will include it as part of the baselines, along with extending SARNet with MAAC. We address our original motivation for our baseline selection on lines 13-20 in our response. **Hyperparameters** were carefully chosen over 10 test runs to accommodate near-optimal learning, and originally proposed networks sizes for all architectures. Additionally, we agree with the suggestions to **improve the figures**, which is addressed in lines 21-22 in our response.