Reviewer 1 Q: Novelty against [3]: The differences: (1) They do reinforcement learning, while we do imitation learning, which is much harder. (2) We established a Bayesian formulation for MAIL to achieve sample efficiency. (3) The approximation form is different. We do both within-type and cross-type mean field, while they only do the former.

Q: typos, and how the error bars were calculated Figures 3 & 4: Thanks for the suggestion! Will do. As shown in appendix, we run each algorithm multiple times and plot the mean and standard deviation (shaded area) of the results.

Q: the clarity of the paper. For example, why to assume it is only necessary to model interactions within one type of agent (lines 151-152): Thank you for the suggestion. Will do. A quick answer is that agents of the same type have similar effects on another agent, and agents of different types show different effects (Eqs 6 and 7, lines 158-161).

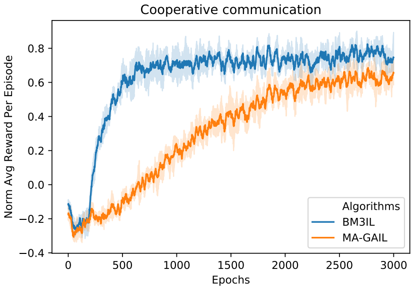
Q: MA-DAAC not scalable: The computation becomes expensive as number of agents increases. In our implementation, the computational complexity depends on the number of types, which is much smaller than the number of agents.

Reviewer 2 Q: why such a bayesian framework provides special advantages in the context of multi-agent learning..: Sample efficiency is a key issue in multi-agent imitation learning, because collecting samples in the real world is expensive. Bayesian parameter estimator can take full account of the uncertainties related the cost-function parameters ϕ compared to a point estimator, and is shown in Section 5.1 to improve sample efficiency by enhancing exploration.

Q: The connection between the theory of Nash equilibrium and the method of the paper: Our paper is motivated by applying MAIL in real-world, where we focus on the challenge of sample efficiency and scalability. Multi-agent mean field is one method we introduce towards solving the scalability challenge. What we can show now is that with our approximation we can converge to Nash equilibrium, and other theoretical development will be future work.

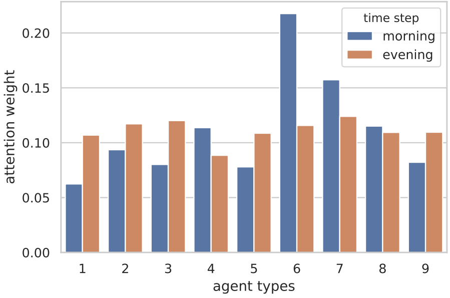
Q: missing a detailed experiment setup (environments illustration, tasks description, evaluation metrics, etc): Thanks for pointing out. We will give more details in the revision. Details for the Rover Tower can be found in line 249, citation [6]. Details for transportation environment can be found in [17]. Appendix 8.4 gives more details for both experiments.

Reviewer 3 Q: I am open to increasing my review score if the authors are able to add comparative results across additional new domains (not transportation): Thanks for your suggestions. The right figure shows the performance comparison in Cooperative communication environment (Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments (Ryan et al., 2017)). We are also investigating other more complex environments — we haven't reached conclusions yet due to the extremely short time for rebuttal but will incorporate them in the final version.



Q: not compared against "Coordinated multi-agent imitation learning" (Le et al., 2017), "Data-Driven Ghosting using Deep Imitation Learning" (Le et al., 2017): Many thanks for pointing out the two papers! We will include them in the discussion. Actually one paper, "Coordinated multi-agent imitation learning", has been discussed in one of our benchmarking algorithm, "Multi-Agent Generative Adversarial Imitation Learning" (MA-GAIL), Section 6 citation 40.

Q: More insights into the connection among types, attention, and interactions: As shown in the right figure, where we plot the attention weight that agent type 0 put on other types in transportation environment with 10 types of vehicles. It shows that the agent learns to put different attention at different time step. In the morning (8 am) it puts more attention on 1-2 types of vehicles. In the evening (6 pm) it spreads it attention to more types of vehicles since more types of vehicles are moving around in the road network at this time.



Q: [Equation 1] typo: Thanks, it is a typo.

Q: [Section 3.1.1] Whether the formulation is a SG or POSG since it introduces the notion of binary observations (which are a function of rewards here): We are solving a Markov game, where each agent has full observation to the system state. We introduce the binary observations to delegate the reward, for the purpose of drawing the connection between imitation learning and probabilistic graphical models, and formulating as a Bayesian model. How we delegate the reward has nothing to do with the problem formulation.

Q: why performance for the baselines drops when the number of types increases: One type refers to a subset of agents who have the same state and action spaces, and the same goal. Type is a virtual concept, which we introduced in our algorithm to improve scalability. Some of the baseline algorithms, such as MA-GAIL and MA-DAAC do not have this component. They work on each individual agents as suggested in their original paper. In Figure 4, as the number of types increase, there are more diverse behaviors of the agents (since different types have different goals), potentially more interactions, and hence more difficult to learn for the baseline algorithms.

Reviewer 4 Q: study more complex scenario where even the baseline struggles to learn: Thanks for your suggestions. We are investigating in more details and more complex environments, due to the limited short period of time in rebuttal, we are not able to get the results, but will make sure to incorporate more in the final version.