We thank the reviewers for their insightful comments. We summarize the questions from each reviewer below and address them separately. We will incorporate the feedback and suggestions into the next revision of the paper.

(Reviewer 1) Q1: Clarify what information is exchanged between the agents.
A: The messages exchanged between the agents generally convey agent status information (location, health status, etc.) that are crucial for executing group strategy. Taking 6h_vs_8z as an example, we measured run-time agent communication pattern during formation stage (Figure 1(c)). Initially (step 80), the agents require frequent communication to build the formation. Overtime, communication level gradually decreases as agents move to the right position (step 250,430). As agent conditions are almost unchanged except for their locations during formation, it can be inferred that messages contain location information. We can also design similar experiments to infer the meaning of other types of messages.

(Reviewer 1) Q2: Which real life applications would benefit most from the proposed approach?
A: VBC is most beneficial to multi-agent systems that require quick decision making and low communication overhead. An example application is *Real-Time Traffic Signal Control*, where hundreds of thousands of traffic signal devices need to make cooperative decisions within tens of seconds. VBC can drastically reduce the communication cost and decrease latency on decision-making, as quite a number of decisions can be made locally. *Swarm Robotics* that are used in rescue missions and remote sensing services can also benefit from the bandwidth reduction and improve responsiveness offered by VBC in dynamic settings. VBC can also be deployed on *autonomous driving*, *cellular base station control*, etc.

(Reviewer 1) Q3: Motivate the choice for calculating the confidence. What if two actions are quasi-optimal.
A: We have tried other methods for calculating the confidence. For example, the variance of the local action vector is a natural choice for measuring the confidence. However, we found variance is not a good criteria in many cases. For example, when local action vector is $(10, 0.5, 0, 0.5)$, its variance is relatively high, but the action can be decided confidently (first action). When the action vector is $(10, 0, 10, 0)$, the variance is high but the best action is undecidable. In both cases, our approach, which is to compute the difference between the $1st$ and $2nd$ best local action scores provides a more direct measure of the action confidence. If the difference is small (i.e. two actions are quasi-optimal), the agent will request additional information from other agents, and then select the action with the highest action value.

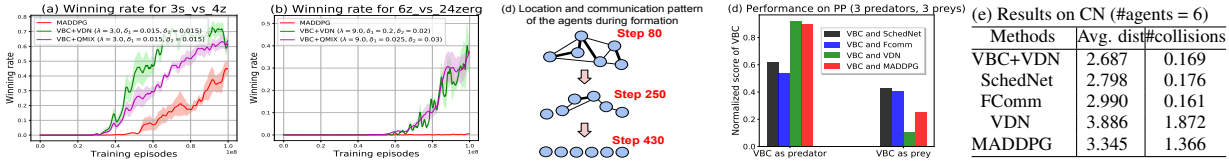
(Reviewer 1) Q4: Equation (1) is unclear: which variance is computed?
Equation (1) minimizes the square loss as well as the sum of variance of the message from each agent.

(Reviewer 2) Q1: Evaluate on a few other environments and compare with MADDPG.
A: We evaluated MADDPG and found VBC achieves higher winning rates than MADDPG for all the six scenarios. Due to space limit, we only report the results for two StarCraft scenarios (Figure 1(a,b)). Furthermore, we evaluate the algorithms for two more scenarios: (1) Cooperative Navigation (CN) which is a cooperative scenario, and (2) Predator-prey (PP) which is a competitive scenario. The game settings are the same as in MADDPG paper. We train each method until convergence and test the result models for 2000 episodes. For PP, we make the agents of VBC to compete against the agents of other methods, and report the normalized score of VBC (Figure 1(d)). For CN we report the average distance between agents and their destinations, and average number of collisions (Figure 1(e)). We notice that methods which allows communication (*i.e.*, SchedNet, FComm, VBC) outperform the others (*i.e.*, VDN, MADDPG) for both tasks, and VBC achieves the best performance. Moreover, VBC incurs a communication overhead of 10.07% and 8.80% for PP and CN respectively, which is lower than SchedNet (33% and 50%). In CN, most of the communication of VBC occurs when the agents are close to each other to prevent collisions. In PP, the communication of VBC occurs mainly to rearrange agent positions for better coordination. We will cite MADDPG in our paper later.

(Reviewer 2) Q2: I would like more discussion on the methods and introduced hyperparameters.
A: For hyperparameters used by VBC (*i.e.*, regularization constant of variance, thresholds on confidence and message variance), we first search for a coarse parameter range based on random trial, experience and message statistics. We then perform a random search within a smaller hyperparameter space. Best selections are shown in Figure 3 of the paper.

(Reviewer 4) Q1:An agent can access the global observation and history only through messages. Correct?
A: Correct, we will make it clear in the next revision of the paper.

(Reviewer 4) Q2: I assume that you are using the same communication protocol during training.
A: That's correct, we will make it explicit in the next version of the paper.



(e) Results on CN (#agents = 6)

Methods	Avg. dist	#collisions
VBC+VDN	2.687	0.169
SchedNet	2.798	0.176
FComm	2.990	0.161
VDN	3.886	1.872
MADDPG	3.345	1.366

Figure 1: (a) and (b) show the performance of MADDPG on two StarCraft scenarios. In (c) thin/bold edges represent uni/bidirectional communication respectively. (d) Results on PP with 3 predators and 3 prey. (e) shows results of CN.