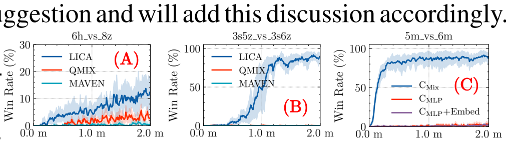
We appreciate all reviewers for their feedback! We're glad that they find our methods well presented (**R2,3,4**), motivated (**R3,4**), and contextualized (**R2,4**), novel (**R1,2,4**), simple and practical (**R3**), and experiments well-designed (**R2**).

Reviewer #1 Q1. Definition of Q. The critic aims to estimate the joint action-value based on the **action probabilities** (AP). As discussed in L121-128, our intuition is to train policies directly towards optimal cooperation with **full differentiability**, and we use sampled actions (**special cases of AP with probability 1**) for critic training because the target values (defined by action-specific rewards $r(s_t,u_t)$) over arbitrary AP are hard to estimate. In fact, similar ideas were explored for single RL settings [Wierstra,Schmidhuber,ECML'07; Weber et al, AISTATS'19] with proper justification. We'll add more discussion and citations accordingly. Q2. k iterations critic update. Yes, k is intended to give better critic estimation and tuning LR is equivalent; it was included as a **practical generalization**: training till convergence could take long and risk overfitting, and tuning k instead of LR may avoid overshooting. Q3. k for other PG baselines? Yes, e.g. k=2 for LICA/MADDPG and k=1 for others work best empirically in SC II. We'll revise to avoid confusion. Q4. Different λ for MLP vs mixing critic. We observed that the **architecture change alone resulted in more stochastic joint actions**, and as clarified in L296, the choices of λ for MLP critic ensure a fair comparison of **policy stochasticity** (Fig.2(b)) against the best LICA run (λ=0.09). We found that setting λ=0.09 for MLP critic clearly results in over-regularization and gives even worse performance. Q5. Need more runs/inconsistency with SMAC paper. We want to point out that our results on all maps except `2c_vs_64zg` are consistent with previous work (e.g. [3,20,21]); for `2c_vs_64zg` specifically, our investigation suggests that the inconsistency is due to a **mismatch in SC II gameplay version**: we base our experiments on the latest SMAC repo which uses **v4.10**, while SMAC paper seems to use **v4.6** (commit history); critically, **v4.7** added changes that made Colossi units more powerful, changing the dynamics of `2c_vs_64zg`. Nevertheless, we'll add more runs for SCII as suggested. Q6. Compare with MAVEN. As suggested, we added comparisons on 2 **Super Hard** maps in Fig.A/B. With same #iterations, **LICA performs considerably better**. Q7. Why t in s_t for Eq.2? Optimizing expected returns over different t is rather standard and often implied under various notational choices; e.g. see [4,3,28] and their implementation. Q8. Eqn for per-agent policy gradients. Due to full differentiability (L145), the PG for agent $a \propto \sum_t \nabla_{\theta_a} p_t^a \nabla_{p_t^a} Q^\pi\left(s_t, p_t^1, ..., p_t^a, ..., p_t^n\right)$ with $p_t^a = \pi_{\theta_a}^a(\cdot|z_t^a)$; we'll update accordingly. Q9. Details of MPE. For Fig.3(b,c), we use 200 steps (L214), -1 reward for every pairwise collision, and we report the **mean reward over all timesteps and agents** in each episode. We'll clarify the metrics in the paper; see also our base repos [13,28]. Q10. Add discussions for QMIX/MADDPG. Thanks! We'll update accordingly.

Reviewer #2 ***Thanks for recognizing our work!*** Q1. LICA in continuous domains. While this is a future extension, we emphasize that LICA doesn't pose extra constraints on top of previous work [4,9,13] that readily handles continuous actions.

Reviewer #3 Q1. Benefits/novelty of mixing critic. Let us consider the generalization where both MLP critic ($\mathrm{C}_{\mathrm{MLP}}$) and mixing critic ($\mathrm{C}_{\mathrm{Mix}}$) operate on representations of states and actions $f_s(s)$, $f_a(a)$. Then, in both cases, we have $\frac{\partial Q}{\partial a}=\frac{\partial Q}{\partial h}\frac{\partial h}{\partial a}$, where h=$f_s(s)$+$f_a(a)$ for $\mathrm{C}_{\mathrm{MLP}}$ and h=$f_s(s)f_a(a)$ for $\mathrm{C}_{\mathrm{Mix}}$ is the **first mixed representation** of s,a before activation (i.e. after concat+linear for $\mathrm{C}_{\mathrm{MLP}}$ and before $\sigma(\cdot)$ for $\mathrm{C}_{\mathrm{Mix}}$, Fig.1(b)). Since $g(h)$=Q is non-linear/non-interpretable in both cases, the crucial difference is thus that $\frac{\partial Q}{\partial a}=\frac{\partial Q}{\partial h}\frac{\partial h}{\partial f_a}\frac{\partial f_a}{\partial a}=\frac{\partial Q}{\partial h}\frac{\partial f_a}{\partial a}$ for $\mathrm{C}_{\mathrm{MLP}}$ and $\frac{\partial Q}{\partial a}=\frac{\partial Q}{\partial h}\frac{\partial h}{\partial f_a}\frac{\partial f_a}{\partial a}=\frac{\partial Q}{\partial h}f_s(s)\frac{\partial f_a}{\partial a}$ for $\mathrm{C}_{\mathrm{Mix}}$, i.e. $\mathrm{C}_{\mathrm{Mix}}$ **adds an extra, direct state representation**. ...do not necessarily lead to better credit assignment (CA): While better CA is not *guaranteed*, we argue **better utilization** of state provides a basis for better CA. Rightness of $\frac{\partial Q}{\partial a}$ ...determined by accuracy of $Q(s,a)$...$\mathrm{C}_{\mathrm{Mix}}$ just learns a better $Q(s,a)$? we argue that the **composition** of $\frac{\partial Q}{\partial a}$ in $\mathrm{C}_{\mathrm{Mix}}$ is the key factor, and a better $Q(s,a)$, if any, would rather be a result of it. $\mathrm{C}_{\mathrm{MLP}}$ also contains state...: We intend to convey that $\mathrm{C}_{\mathrm{Mix}}$ has a better utilization of s and will revise all inaccuracies in Sec 3.2. Discussion (3,4)...aren't contributions: We'll revise accordingly; note that they remain valid and were discussed as LICA's *properties* rather than novelties. Concat after MLP for $\mathrm{C}_{\mathrm{MLP}}$: As suggested, we ran a comparison in Fig.C where MLPs are added before concat; results confirm our earlier analysis which covers this case. Q2. Could LICA converge to stable policies? While we cannot provide a full analysis here, we emphasize that our empirical evidence across different λ's, scenarios, complexity (Fig.4(a-f)), and environments with repeated runs (Fig.3/4) suggests that policies eventually reach a stochasticity equilibrium (Fig.2(b,c)); this may in fact sustain smoother object landscapes and aid policy convergence [1]. Q3. Compare with MAAC. By design, the simplicity of the quoted **1-step** game obviates most key aspects that differentiate on/off-policy learning (future estimation, separate target/behavior nets, replay buffers) and focuses only on the **mechanism for credit assignment**. However, we appreciate your suggestion and will add this discussion accordingly.



Reviewer #4 Q1. Improvements in MPE. We stress that compared to the previous SOTA [28], our method achieved similar gains despite approaching the limits of the selected envs. Q2. Complex settings w/ uneven mix of 'individual performance' and 'cooperation'. In fact, `MMM2` (**Super Hard**, Fig.4(f), Supp L20-23, and demo) is *precisely* one such setting where our method has **sizable advantage over others**. Winning heavily relies on the performance of the 1 healer unit and cooperation of the 9 attack units. Q3. SC II: further training/more complex settings. We emphasize that many previous work mainly focuses on **Easy** maps (e.g. [3,4,20]) and lacks diversity in map choices (e.g. [3,4,20,14,ROMA ICML'20]); on our diverse maps (L252-254), **we achieved similar or significantly more gains** compared to previous work **with similar #iterations**. At **R4**'s request, we also added results on 2 extra **Super Hard** maps (`6h_vs_8z`,`3s5z_vs_3s6z`) in Fig. A/B, showing **sizable gains over previous methods**. Q4. It reads more like a report. We respectfully disagree. On top of **R2**'s recognition and our above response, we'd also highlight our comparison against SOTA in 2019 [3,25] and our extensive component studies (Sec 4.3, Supp A2, Fig.2) that are equally or more comprehensive compared to previous work (e.g. [3,4,20,28,14]).