We appreciate all reviewers for their feedback! We’re glad that they find our methods well presented. 

**[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 $r(s_t,a_t)$) are non-linear, limited, and often 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 $\lambda$ for MLP vs mixing critic. We observed that the architecture change alone resulted in more stochastic joint actions, and as clarified in L296, the choice of $\lambda$ for MLP critic ensure a fair comparison of policy stochasticity (Fig.2(b)) against the best LICA run ($\lambda=0.09$). We found that setting $\lambda=0.09$ 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 SCII 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 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 [4,328] and their implementation. Q8. Eqn for per-agent policy gradients. Due to full differentiability (L145), the PG for agent $\alpha$ is $\sum \nabla_{\theta_{\alpha t}} Q(s_t,a_t)$, where $h=f(s_t)+f(a_t)$ for $\text{CMLP}$ and $h=f(s_t)+f(a_t)$ for $\text{CMix}$, is the first mixed representation of $s,a$ before activation (i.e. after concat-linear for $\text{CMLP}$ and before $\sigma(\cdot)$ for $\text{CMix}$, Fig.1(b)). Since $g(h)=Q$ is non-linear/non-interpretable in both cases, the crucial difference is that $\frac{\partial Q}{\partial h}$ for $\text{CMLP}$ and $\frac{\partial Q}{\partial h}$ for $\text{CMLP}$ and $\frac{\partial Q}{\partial h}$ for $\text{CMix}$, $s_{t_{\alpha}}$ for $\text{CMix}$, i.e. $\text{CMix}$ 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. Rightside of $\frac{\partial Q}{\partial h}$...determined by...of $\text{CMix}$ just learns a better $Q(s,a)$! We argue that the composition of $\frac{\partial Q}{\partial h}$ in $\text{CMix}$ is the key factor, and a better $Q(s,a)$, if any, would rather be a result of it. $\text{CMix}$ also contains state... We intend to convey that $\text{CMix}$ has a better utilization of state 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 $\text{CMix}$: 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 $\lambda$’s, scenarios, complexity (Fig.4(a)-f), and environments with repeated runs (Fig.3/4) suggest that policies eventually reach a stochasticity equilibrium (Fig.2(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. Q4. Improvements in MEP. 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 (SuperHard, 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 heater 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. [3,20,14,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]).