We gratefully appreciate the efforts made by all the reviewers. Thanks to **Reviewers #1** and **#2** for bringing up the missing references. Hughes et al. [2018] extend the inequity aversion model and define a shaped reward $r_i - \frac{\alpha}{N-1}\sum\max(r_j - r_i, 0) - \frac{\beta}{N-1}\sum\max(r_i - r_j, 0)$. Wang et al. [2018] design a reward network generating intrinsic reward, evolved based on the group's collective reward. Peysakhovich et al. [2018] propose a shaped reward $r_i = \alpha r_i + (1-\alpha)r_j$ for two-player Stag Hunts. The baseline `Avg` can be seen as its multi-player version as the authors claim. These works aim to improve cooperation but cannot guarantee fairness. We compare against Hughes et al. [2018], `Inequity Aversion`, in job scheduling. Table 1 shows the CV of `Inequity Aversion` is better than `Independent` but still much worse than `FEN`, and the resource utilization is much lower. That shows `Inequity Aversion` cannot solve job scheduling fairly and efficiently. More details will be included in the final version. We will also include the review of the missing references in the final version.

Table 1: Job scheduling

	resource utilization	CV
`Independent`	96% ±11%	1.57 ±0.26
`FEN`	**90%** ±5%	**0.17** ±0.05
`Inequity Aversion`	72% ±9%	0.69 ±0.17

Table 2: Hierarchy

	resource utilization	CV
`Min w/ Hierarchy`	62% ±9%	0.31 ±0.11
`Avg w/ Hierarchy`	84% ±6%	0.61 ±0.14
`Min+`α`Avg w/ Hierarchy`	71% ±8%	0.28 ±0.09
`FEN`	**90%** ±5%	**0.17** ±0.05

Reviewer #1 To verify the effectiveness of the hierarchy, we use the hierarchy with other baselines in job scheduling. Table 2 shows their performance has a certain degree of improvement, especially the resource utilizations of `Min` and `Min+`α`Avg` raise greatly and the CV of `Min+`α`Avg` reduces significantly. That demonstrates the effect of the hierarchy. However, these baselines with the hierarchy are still worse than FEN in both resource utilization and CV, verifying the effectiveness of the fair-efficient reward.

The intuition of the fair-efficient reward is to maximize the resource utilization while punish the agent's utility deviation from the average, taking both fairness and efficiency into consideration. Also, the fair-efficient reward is suitable for decentralized training, which can be easily coordinated by $\bar{u}$. We design the fair-efficient reward, prove it satisfies the criteria of Propositions 1 and 2, and empirically verify it really works well.

The main hyperparameters are contained in the Appendix, we will make a further supplement in the final version. All the results are obtained by five runs with different random seeds and presented with standard deviation (line 234), and we will make it clearer.

Reviewer #2 It is really a constructive suggestion to analyze the behavior of the controller. Figure 1 visualizes the probability of selecting sub-policy ϕ_1 and other sub-policies in terms of the utility deviation from average, $(u_i - \bar{u})/\bar{u}$. It shows when the agent's utility is below average, the controller is more likely to select ϕ_1 to occupy the resources, and when the agent's utility is above average, the controller tends to select other sub-policies to improve fairness. Thus, it can be seen that the balance of these two kinds of sub-policies depends on the current fair-efficient reward.

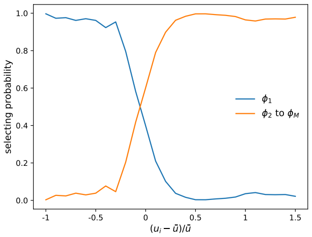

Figure 1: Selecting probability over $(u_i - \bar{u})/\bar{u}$.

Sorry for the confusion induced by "decentralized training." By that we mean the training of each agent requires only limited information exchanging with neighboring agents. We will use more precise expression to replace that in the final version. At test time, it is the same as training where the controller chooses one sub-policy every T timesteps.

Reviewer #3 We explain the necessity of the hierarchy from three aspects. First, the hierarchy reduces the difficulty of learning both efficiency and fairness. Since the problem is a multi-objective optimization, the learning difficulty for a single neural network cannot be neglected. In the hierarchy, each sub-policy focuses on its own easy objective and there is no conflict; the controller focuses on the fair-efficient reward by selecting the sub-policies, without directly interacting with the environment. Second, the fair-efficient reward changes slowly since it is slightly affected by the sub-policy's action in one timestep. Thus, the controller can plan over a long-time horizon to optimize both objectives. Third, in all the three experiments, FEN with hierarchy outperforms the version without hierarchy, verifying the hierarchy helps the learning greatly. We also use the hierarchy with other baselines and their performance improves as shown in Table 2, which also shows the effectiveness of the hierarchy.

Although the training is decentralized, the agent can obtain the average utility $\bar{u}$ by Gossip, and hence each agent knows the utility deviation from the average. To make its own fair-efficient reward higher, the agent with lower u_i must occupy more resources and the agent with higher u_i than $\bar{u}$ will choose to not occupy resources. That is why the policies can be coordinated. Each agent only focuses on its own fair-efficient reward and the fairness could be achieved.

This paper focuses on achieving both fairness and efficiency in the context of *common resources* which is one of important fields in MARL. Propositions 1 and 2 have proved that the resources will be *fully* occupied and *equally* allocated *in infinite-horizon sequential decision-making*, thus the CV is also minimized according to its definition.