A General Incentives-Based Framework for Fairness in Multi-agent Resource Allocation

Ashwin Kumar
Washington University in St. Louis
Saint Louis, MO, USA
ashwinkumar@wustl.edu

William Yeoh
Washington University in St. Louis
Saint Louis, MO, USA
wyeoh@wustl.edu

ABSTRACT

We introduce the General Incentives-based Framework for Fairness (GIFF), a novel approach for fair multi-agent resource allocation that infers fair decision-making from standard value functions. In resource-constrained settings, agents optimizing for efficiency often create inequitable outcomes. Our approach leverages the action-value (Q-)function to balance efficiency and fairness without requiring additional training. Specifically, our method computes a local fairness gain for each action and introduces a counterfactual advantage correction term to discourage over-allocation to already well-off agents. This approach is formalized within a centralized control setting, where an arbitrator uses the GIFF-modified Q-values to solve an allocation problem.

Empirical evaluations across diverse domains—including dynamic ridesharing, homelessness prevention, and a complex job allocation task—demonstrate that our framework consistently outperforms strong baselines and can discover far-sighted, equitable policies. The framework's effectiveness is supported by a theoretical foundation; we prove its fairness surrogate is a principled lower bound on the true fairness improvement and that its trade-off parameter offers monotonic tuning. Our findings establish GIFF as a robust and principled framework for leveraging standard reinforcement learning components to achieve more equitable outcomes in complex multi-agent systems.

KEYWORDS

Fairness; Resource Allocation; Multiagent Systems; Multiagent Planning

ACM Reference Format:
Ashwin Kumar and William Yeoh. 2026. A General Incentives-Based Framework for Fairness in Multi-agent Resource Allocation. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 9 pages. https://doi.org/10.65109/GQAG8531

1 INTRODUCTION

In many real-world applications, ranging from ridesharing [2, 21] to allocating homelessness resources [8], multiple agents vie for limited resources, often leading to significant disparities in outcomes. Traditional RL-based allocation methods focus on maximizing individual or aggregate utility, but they typically overlook fairness, resulting in inequitable resource distributions that can undermine both system-wide performance and societal acceptance. In this work, we address this gap by proposing a novel framework for fair multi-agent resource allocation that leverages the existing action-value functions to *infer* fairness improvements, without any additional learning.

Existing approaches to fair RL generally incorporate fairness directly into the reward structure or modify the learning process to produce value functions that reflect fairness considerations. However, these strategies can introduce complexities such as non-stationarity and may require extensive retraining, which is impractical in many dynamic and high-stakes environments. Moreover, such methods typically assume that agents are intrinsically motivated to be fair, that is, they are willing to sacrifice their own utility for collective equity. In many real-world scenarios, agents may have competing objectives or operate independently without centralized coordination, making it unreasonable to expect them to adopt fairness as an inherent goal. Our work takes a different path: Rather than embedding fairness into the value function, we infer fair decisions by analyzing the long-term Q-values computed by agents. This allows us to extract information about the potential fairness gains associated with different actions and adjust the allocation process accordingly.

The core idea of our approach, named General Incentives-based Framework for Fairness (GIFF), is to combine the standard utility-driven Q-values with additional fairness-related signals. Specifically, we decompose the fairness gain into a component that captures the marginal improvement in fairness resulting from an action, and we introduce an advantage correction term that incentivizes better-off agents to moderate their resource consumption in favor of those who are disadvantaged. This decomposition not only translates to a diverse set of fairness metrics but also offers a transparent mechanism to balance efficiency and fairness during the resource allocation process.

We formalize our framework within a centralized control setting, where a central arbitrator uses the GIFF-modified Q-values to solve an optimization problem subject to resource constraints, and wishes to incorporate fairness into the decision-making. Through extensive experiments in domains such as ridesharing, homelessness, and job allocation, we demonstrate that our framework achieves competitive fairness-utility trade-offs compared to state-of-the-art methods.[1] Our evaluations also highlight the critical role of the advantage correction term, particularly in environments where independent evaluations of fairness improvement fail to capture the benefits of inter-agent cooperation.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/GQAG8531

[1] The supplementary material and code can be accessed at: https://github.com/YODA-Lab/General-Incentives-based-Framework-for-Fairness/

In summary, our contributions are as follows:

(1) We develop a **General Incentives-based Framework for Fairness (GIFF)** that can be used to improve diverse fairness metrics in multi-agent resource allocation problems without the need for additional learning. This framework infers fairness from standard Q-values without altering the underlying RL models.
(2) We propose a principled mechanism—comprising fairness gain estimation and counterfactual advantage correction—to adjust allocation decisions in favor of fairer outcomes, with only two hyperparameters.
(3) We derive instantiations of GIFF for various fairness measures like variance, α-fairness and Generalized Gini Functions (GGF).
(4) We provide theoretical bounds on the actual fairness improvement in relation to the locally estimated fairness gain when using GIFF, for multiple fairness functions. We also provide a Pareto-improvement property over the total local fairness.
(5) We provide experimental evidence across multiple domains and fairness metrics to illustrate the effectiveness of our approach.

By rethinking how fairness can be integrated into multi-agent systems, our work opens up promising new directions for achieving equitable resource distribution in complex, dynamic environments—a crucial step toward deploying RL in socially sensitive applications.

2 RELATED WORK

Many domains employ multi-agent resource allocation with centralized decision-makers, like ridesharing [2, 21], homelessness prevention [8, 9], satellite allocation [16], and wireless networks [24]. In our work, we focus on fair decision making in multi-agent systems that have a similar structure. There has been recent work on developing methods to learn fair policies in multi-agent RL [7, 13, 22, 25], using policy optimization [19] with modified rewards or hierarchical policies to elicit fair behavior. Alamdari et al. [1] look at fair resource allocation at different time horizons, learning to be fair by constructing counterfactual experiences. DECAF [12] is a similar learning-time algorithm that trains Q-functions to add fairness considerations during policy learning. In contrast, GIFF is a strictly post-training mechanism. It assumes that Q-values (from any source—RL, heuristics, simulators) are already available and modifies them online through a closed-form correction. GIFF therefore requires no additional learning, making it lightweight and computationally inexpensive. Simple Incentives [11] is a related approach looking at fairness in ridesharing systems, employing domain and metric specific fairness post-processing. Our approach is much more general and adaptable to different domains and fairness metrics, as we show in our experimental results.

Beyond reinforcement learning and resource allocation, the machine learning community has broadly investigated bias and fairness. The survey by Mehrabi et al. [14] provides comprehensive overviews of fairness metrics and mitigation strategies, while foundational works [5, 6] have laid the groundwork for understanding fairness in supervised settings. In parallel, economic and social welfare research has long provided robust formulations of equity, giving rise to measures like α-fairness and the Generalized Gini Function—which are rooted in social welfare theory [3, 15, 18, 20]. By drawing on these diverse strands of literature, our work bridges multi-agent reinforcement learning, dynamic resource allocation, and fairness, contributing a novel framework that infers fairness directly from long-term Q-value estimates.

3 PRELIMINARIES

We now formally describe the resource allocation problem that tackled in this paper as well as describe several key fairness concepts that motivates our work and used later in empirical evaluations.

3.1 Resource Allocation

A typical resource allocation problem consists of n agents $i \in \alpha$ with diverse preferences over K types of resources, with a total set of resources $\mathcal{R}$ being available. Each time-step, agents attempt to take resources they want, following which new resources appear and the global state changes according to some transition dynamics.

In the centralized control setting, an arbitrator aggregates agent preferences and allocates actions while ensuring no two agents try to take the same resource. This constrained Multi-agent MDP model [4] is described by the tuple $\mathcal{M}$ with the following components:

$$\mathcal{M} = \langle \alpha, S, \boldsymbol{O}, \{A_i\}_{i\in\alpha}, T, R, \gamma, c \rangle \tag{1}$$

- α is the set of agents indexed by i (n agents).
- S is the global state space.
- $\boldsymbol{O} : S \to O_1 \times O_2 \times \ldots \times O_n$ is the observation function that maps the true state to agent observations.
- A_i is the action space for agent i, where an action a includes allocation of a set of resources.
- $T : S \times A_1 \times A_2 \times \ldots \times A_n \times S \to [0, 1]$ represents the joint transition probabilities.
- $R : S \times A_1 \times A_2 \times \ldots \times A_n \to \mathbb{R}^n$ denotes the (utility) reward function, which returns a vector of rewards, one for each agent.
- γ is the discount factor for future rewards.
- $c : A_1 \cup A_2 \cup \ldots \cup A_n \to \mathbb{R}^K$ maps each action to its resource consumption for K types of resources.

Each agent can then learn an action-value function $Q(o_i, a)$, which estimates the expected long-term return from taking action a given the local observation o_i, without knowledge of other agents' current actions. Formally, this is defined as:

$$Q(o_i, a) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_i^{(t)} \,\middle|\, o_i^{(0)} = o_i, a_i^{(0)} = a\right], \tag{2}$$

where $r_i^{(t)}$ is the reward received by agent i at time-step t, and $\gamma \in [0, 1)$ is the discount factor. The expectation is taken over the trajectories induced by the environment dynamics and policies of all agents, conditioned only on agent i's observation and action. In any state s, agents compute the Q-values for all of their available actions and communicate them to the central arbitrator, who can then use them to compute a utilitarian allocation.

Let $\mathcal{A}$ denote the allocation of actions decided by the central allocator such that $\mathcal{A}_i$ is the action assigned to agent i. Let $x_i(a) \in \{0, 1\}$ be a binary decision variable that indicates whether

agent i is assigned action $a \in A_i$. Let $\mathcal{R} \in \mathbb{R}^K$ denote the vector of available resources, with $\mathcal{R}_k$ representing the quantity of resource type $k \in \{1, 2, \ldots, K\}$. Let $c(a) \in \mathbb{R}^K$ denote the resource consumption vector for action a, where $c(a)_k$ is the amount of resource k consumed by action a. This gives us the following optimization:

$$\max_{x_i(a)\in\{0,1\}} \quad \sum_{i\in\alpha}\sum_{a\in A_i} x_i(a)\cdot Q(o_i,a) \tag{3}$$

$$\text{subject to} \quad \sum_{a\in A_i} x_i(a) = 1, \quad \forall i \in \alpha \tag{4}$$

$$\sum_{i\in\alpha}\sum_{a\in A_i} x_i(a)\cdot c(a)_k \le \mathcal{R}_k, \quad \forall k \in \{1,\ldots,K\} \tag{5}$$

These constraints ensure that each agent is assigned exactly one action and that total resource usage does not exceed available supplies. This is a general formulation, and many real-world problems follow this approach [2, 8, 16, 21, 24]. This problem can be formulated as an integer linear program (ILP), but more efficient algorithms and distributed approaches exist for allocation problems with stricter constraints. For instance, if the constraints boil down to a bipartite matching between agents and resources, the Hungarian algorithm [10] can solve the allocation problem in polynomial time.

Alternatively, some systems assume agents act independently without a central arbitrator or consensus-based decision making. In this case, methods like first-come-first-served or random tie breaks are used to decide who gets contested resources. This approach is much more commonly seen in multi-agent RL [7, 23, 25]. When using a policy optimization based approach, agents express preferences over actions which maximize their chances of getting good resources in the form of a policy π rather than expressing valuations over bundles of resources.

In this paper, we restrict ourselves to using Q-functions. Further, we select the centralized control setting, where the decision-maker has the ability to enforce fairness constraints by providing incentives to agents. Thus, in this work, we assume agents bid for actions by communicating their Q-values to the central decision maker, and each action is associated with the allocation of some resources and the corresponding gain in utility is the reward. The Q-function then captures the long-term expected utility for each agent.

Payoff-vector (Z): We are interested in resource allocation problems that have a temporal aspect to them, i.e., after each allocation, new resources may arrive in the system, and resources and agents may enter or exit the allocation pool. We consider two cases: 1) A fixed number of agents, and 2) An arbitrary number of agents that belong to a fixed number of groups. In both cases, we consider a total of n groups or agents. Given this, we can construct a vector of payoffs $\mathbf{Z} = [z_1, z_2, \ldots, z_n]$ that captures the accumulated value of all resources allocated to each agent/group over time.

Unless specified otherwise, we use z_i to denote the cumulative payoff for agent or group i at the current time-step. When referring to a specific time, we will write $z_i^{(t)}$ to indicate the value at time-step t. The cumulative reward is given by:

$$z_i = \sum_{\tau=0}^{t} r_i^{(\tau)}, \tag{6}$$

where $r_i^{(\tau)}$ is the reward received by agent i at time-step τ. In some settings, an average payoff is used instead:

$$\bar{z}_i = \frac{1}{t+1}\sum_{\tau=0}^{t} r_i^{(\tau)}. \tag{7}$$

This payoff vector serves as a temporal record of how resources have been distributed across agents and provides a foundation for incorporating fairness criteria into future allocation decisions.

3.2 Fairness Concepts

To capture fairness, we consider a fairness function $F : \mathbb{R}^n \to \mathbb{R}$, which maps any payoff vector $\mathbf{Z} = [z_1, \ldots, z_n]$ to a numerical value such that larger values correspond to fairer distributions. In other words, for any two payoff vectors $\mathbf{Z}_1$ and $\mathbf{Z}_2$, we say that $\mathbf{Z}_1$ is fairer than $\mathbf{Z}_2$ if and only if $F(\mathbf{Z}_1) > F(\mathbf{Z}_2)$.

In the literature on fair multi-agent resource allocation, two main schools of thought emerge:

- **Social Welfare Function Approaches.** In these methods, fairness is directly embedded into a social welfare function that aggregates individual utilities [22, 25]. Such functions may be designed to satisfy three desirable properties: They exhibit **impartiality**, remaining invariant under any permutation of agents to ensure equal treatment; they promote **equity** by rewarding reallocations that transfer resources from better-off agents to those worse-off, consistent with the Pigou-Dalton principle; and they ensure **efficiency** by assigning a higher value to allocations in which every agent receives higher or equal utility compared to alternative allocations.
  These metrics usually take the form of a summation over transformations of agent utilities, for example:
  - α**-Fairness:** For a given payoff vector $\mathbf{Z} = [z_1, \ldots, z_n]$, define the per-agent α-fair utility as:
  $$U_\alpha(z) = \begin{cases} \frac{z^{1-\alpha}}{1-\alpha} & \text{if } \alpha \neq 1, \\ \log(z) & \text{if } \alpha = 1, \end{cases} \tag{8}$$
  and the overall fairness measure as: $F_\alpha(\mathbf{Z}) = \sum_{i=1}^n U_\alpha(z_i)$.
  - **Generalized Gini Function (GGF):** Order the components of $\mathbf{Z}$ as $z_{(1)} \le z_{(2)} \le \cdots \le z_{(n)}$. Then, the GGF function is defined as
  $$F_{GGF}(\mathbf{Z}) = \sum_{i=1}^{n} w_i\, z_{(i)}, \tag{9}$$
  with weights satisfying $w_1 \ge w_2 \ge \cdots \ge w_n$ and $\sum_{i=1}^n w_i = 1$.
  By varying α or w_i, these metrics can transition between utilitarian and egalitarian fairness. Specifically, α-fairness with $\alpha = 1$ is equivalent to the popular log Nash Welfare metric.
- **Distributional Approaches.** Alternatively, fairness may be measured separately via distributional metrics—such as variance, the Gini index, or Jain's fairness index—and then combined with total utility [11, 17]. These metrics can capture non-linear relationships among agents' utilities and provide a distinct measure of fairness that is later interpolated with the overall efficiency.

Having outlined these approaches, our work aims to balance efficiency and fairness via a joint objective. Specifically, for a given time horizon T with payoff vector $\mathbf{Z}_T$, we seek an allocation policy

that maximizes:

$$\max \quad (1-\beta)U_T + \beta F(\mathbf{Z}_T), \tag{10}$$

$$U_T = \sum_{i=1}^{n} z_i, \tag{11}$$

Here, $\beta \in [0, 1]$ is a trade-off parameter that controls the relative importance of efficiency (total utility) versus fairness.

In our evaluations, we show results using both kinds of fairness metrics. In particular, for the social welfare function approaches, we incorporate a specialized term that enables agents to locally assess the benefits of their actions for others. Notably, our incentives-based framework operates entirely online with zero additional training.

4 GENERAL INCENTIVES-BASED FRAMEWORK FOR FAIRNESS (GIFF)

The key insight in our work is that the Q-function, which captures the long-term utilitarian effects of actions, can also be used to guide the decision maker towards a fair allocation. Previous work has considered directly learning to optimize for fairness [7, 25] or using knowledge of the true reward function to myopically improve variance using domain-specific post-processing [11]. Instead, we provide a General Incentives-based Framework for Fairness (GIFF) that improves fairness without the need for additional learning for a variety of domains and fairness functions.

To achieve this, we develop an approach that takes advantage of the instantaneous pre-decision payoffs $\mathbf{Z}_t$ at the current time t to guide the present decision towards a fairer outcome by improving the perceived value of fairer allocations. We do this by computing the estimated improvement to fairness, the *fairness gain* of actions, and augmenting the pre-trained Q-values to reflect our objective (Eq. 10).

First, we assume an idealized scenario. Let $\mathcal{A} = [a_i]_{i\in\alpha}$ denote an allocation that contains one action for each agent. We overload the notation for the reward function to let $R(\mathcal{A}_i)$ be a shorthand for the true reward received by agent i under allocation $\mathcal{A}$. The updated payoff vector $\mathbf{Z}_{t+1}|\mathcal{A}$ can be computed using $\mathbf{Z}_t$ and $\mathcal{A}$, by updating the payoffs using the true rewards. Then, the fairness gain for any allocation can be defined as:

$$\Delta F(\mathcal{A}) = F(\mathbf{Z}_{t+1}) - F(\mathbf{Z}_t) \tag{12}$$

$$z_i^{t+1} = z_i^t + R(\mathcal{A}_i) \qquad \forall i \tag{13}$$

$$\mathcal{A}_f^* = \underset{\mathcal{A}}{\operatorname{argmax}}\, \Delta F(\mathcal{A}) \tag{14}$$

Here, $\mathcal{A}_f^*$ is the allocation that improves fairness the most in the current step. We can also conceive of a similar search which maximizes over an entire sequence of allocations to maximize long-term fairness. However, even for the one-step allocation, the search space is combinatorial in the number of agents and their respective action spaces, as we have to consider all possible joint actions. This makes the global optimization intractable, necessitating alternate methods for computing a fair allocation.

4.1 Using Q-values to Estimate Fairness

We observe that it is much easier to reason about fair actions if we can decompose the fairness gain across agents. To achieve this, we reason only over the locally conditioned updated payoff vector $\mathbf{Z}_t^{a^i}$, updating all accumulated utilities based only on a single agent's action a^i, keeping everything else unaffected.

$$z_j^{t+1} = z_j^t + \mathbb{I}\{j = i\}\, R(a^i) \qquad \forall j \tag{15}$$

In many real problems, having access to the true reward function $R(a)$ is unlikely. Further, in dynamic environments, agents may take a critical action earlier, which leads to a payoff after multiple steps; however, a critical decision towards getting to it may happen much earlier. If the agent is a Q-learner, we can leverage the fact that the Q-values encode the long-term value of taking certain actions, and the difference in Q-values will be small in states where all routes lead to similar payoffs. Thus, if we use the Q-value as a proxy for the reward function, we can account for long-term returns without knowing the reward function or the environment dynamics.

$$\Delta F(a^i) = F(\mathbf{Z}_{t+1}^{a^i}) - F(\mathbf{Z}_t) \tag{16}$$

$$z_j^{t+1} = z_j^t + \mathbb{I}\{j = i\}\, Q(o_i, a^i) \qquad \forall j \tag{17}$$

This quantity $\Delta F(a^i)$, termed the *fairness gain* measures the marginal (local) impact of agent i's action on the overall fairness of the allocation given the current distribution of resources $\mathbf{Z}_t$. Note that computing this for all agent actions is linear in the size of the action space for each agent. This has two benefits. First, we do not need to have access to the true reward function (which is not available in many cases) and, second, we capture more than just the immediate return, allowing us to capture long-term effects of certain allocations. However, this has the drawback that it does not capture inter-agent interactions very well. Thus, we introduce an additional mechanism that can provide this information.

4.2 Advantage Correction: Incorporating Counterfactual Fairness Gains

Fairness concerns in dynamic resource allocation may also require that agents who are already well-off (i.e., have a high accumulated utility) are discouraged from taking resources that can help worse-off agents improve their return, thus allowing disadvantaged agents to catch up. This is also a desirable property of social welfare functions, termed **equity**, used to capture the notion of accumulated wealth: Moving resources from better-off agents to worse-off agents should improve the fairness function's value. This is also known as the Pigou-Dalton principle [25] and has been discussed in previous works in fair multi-agent RL.

However, in practice, the local fairness gain $\Delta F(a^i)$—which reflects only the immediate change in an agent's own utility—does not capture the broader altruistic impact of reallocating resources. In many fairness metrics, this counterfactual update considers solely the acting agent's payoff, thereby ignoring the significant improvement in fairness that would occur if the resources were allocated to a disadvantaged agent. Consequently, even when an agent's decision to forgo an action yields $\Delta F = 0$, it may still lead to substantial overall fairness gains if the resource were reallocated. This limitation motivates the inclusion of a method to more accurately account for these counterfactual benefits.

To this end, we introduce an *advantage correction* term that incentivizes better-off agents to give up their top preferences to benefit other disadvantaged agents. Recall that in our setting, when

agent i takes action a, its local fairness gain $\Delta F(a)$ is computed by updating the agent's payoff z_i with $Q(o_i, a)$ (Eqs. 16 and 17). To measure the counterfactual benefit of this action, we consider allocating this resource to any other agent that can take this action $j \neq i, a \in A_j$, and compute the counterfactual benefit $\Delta F^{(j)}$, using $Q(o_j, a)$ to update agent j's payoff:

$$\Delta F^{(j)} = F\big(\mathbf{Z}_{t+1}^{(j)}\big) - F(\mathbf{Z}_t), \tag{18}$$

$$z_j^{t+1} = z_j^t + Q(o_j, a) \tag{19}$$

Let us define the set of these candidate counterfactual agents as $\alpha_c(a) = \{j \in \alpha : j \neq i, a \in A_j\}$. We can then compute the average counterfactual fairness improvement:

$$\Delta F_{\text{avg}}(a) = \frac{1}{|\alpha_c(a)|} \sum_{j \in \alpha_c(a)} \Delta F^{(j)} \tag{20}$$

To capture the benefit of allocating this resource to agent i, we can calculate the advantage function:

$$F_{\text{adv}}(a) = \Delta F(a) - \Delta F_{\text{avg}}(a) \tag{21}$$

A negative F_{adv} suggests that another agent would benefit more from the resource allocation, whereas a positive F_{adv} indicates that the current agent can better improve fairness. If the fairness metric follows the principle of equity, then we expect agents with higher fairness gain to be the disadvantaged agents. Instead of ΔF_{avg}, alternate baselines like the maximum fairness improvement may also be considered.

To integrate this counterfactual fairness measure with the action's inherent quality, we weigh the fairness advantage by the relative Q-value gap:

$$\Delta Q(a) = Q(o_i, a) - \min_{a' \in A_i} Q(o_i, a'), \tag{22}$$

which reflects how much better action a is compared to the worst option for agent i. This is helpful in preventing disproportional changes because of Q-value overestimation.

Finally, we define the counterfactual advantage correction term as:

$$\Delta Q_{\text{adv}}(a) = F_{\text{adv}}(a)\, \Delta Q(a). \tag{23}$$

This formulation has the following intuitive implications:

- If the fairness gain $\Delta F(a)$ is lower than the mean counterfactual gain $\Delta F_{\text{avg}}(a)$, then $F_{\text{adv}}(a)$ is negative, leading to a negative $\Delta Q_{\text{adv}}(a)$. This reduces the attractiveness of action a, discouraging further accumulation by already advantaged agents.
- Conversely, if $\Delta F(a)$ exceeds $\Delta F_{\text{avg}}(a)$, then $F_{\text{adv}}(a)$ is positive, and $\Delta Q_{\text{adv}}(a)$ is positive. This boosts the value of actions that help a disadvantaged agent catch up.

4.3 GIFF-modified Q-values

We combine the original Q-value estimate with the fairness gain and the counterfactual advantage correction to obtain the GIFF-modified Q-value, which we can use in the optimization in Equation 3 to compute allocations:

$$Q_f(a) = \Delta F(a) + \delta\, \Delta Q_{\text{adv}}(a)$$

$$\boxed{Q^{\text{GIFF}}(o_i, a, \beta, \delta) = (1 - \beta)\, Q(o_i, a) + \beta\, Q_f(a)} \tag{24}$$

- $\beta \in [0, 1]$ controls the trade-off between efficiency (standard Q-values) and fairness, with $\beta = 1$ leading to allocations based purely on the fairness gain.
- $\delta \geq 0$ controls the degree of advantage correction. Empirically, we observed that small positive values (< 0.5) lead to good results, but this is dependent on the environment and fairness function being used.

In practice, by adjusting β and δ, the central optimizer can be nudged toward actions that balance both overall system utility and fairness, as measured by $F(\mathbf{Z})$.

5 THEORETICAL RESULTS

In this section, we provide theoretical guarantees for the core mechanism of GIFF. The algorithm's fairness-aware Q-value, Q^{GIFF}, relies on a tractable surrogate for the true, combinatorial fairness improvement of a joint allocation. Specifically, GIFF approximates the true joint gain by summing the individual, local fairness gains of each agent's action (or Q-value), a quantity we formally define as the surrogate, S. We now prove three key properties of this design: (1) This surrogate S is a principled lower bound on the true fairness improvement for several canonical fairness functions; (2) Maximizing this surrogate is guaranteed to improve as the fairness weight β increases; and (3) How these guarantees on our surrogate translate to guarantees on the realized, real-world fairness.

Throughout, let $\mathbf{Z} = (z_1, \ldots, z_n) \in \mathbb{R}^n$ be the current payoff vector and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n_{\geq 0}$ be the vector of utility increments from a feasible allocation. For a fairness function $F : \mathbb{R}^n \to \mathbb{R}$, we define the *realized fairness improvement* as $\Delta_{\text{joint}} := F(\mathbf{Z}+y) - F(\mathbf{Z})$ and the *local fairness gain* for agent i as $\Delta_i^{\text{local}} := F(\mathbf{Z}+y_i e_i) - F(\mathbf{Z})$, where e_i is the i-th unit vector. GIFF's objective uses the sum of local gains, $S := \sum_{i=1}^n \Delta_i^{\text{local}}$, as a surrogate for the true joint improvement.

Our proofs rely on the following assumptions:

Assumption 1 (Nonnegative increments)**.** *Any agent's change in utility from an allocation is nonnegative: $y_i \geq 0$ for all i.*

In resource allocation domains, receiving additional resources does not decrease an agent's utility. This is satisfied by virtually all real-world applications (e.g., more compute time, bandwidth, or energy increases utility). This assumption enables monotonicity arguments in the fairness correction.

Assumption 2 (Q-value correctness)**.** *Q-values are perfectly accurate predictors of utility increments.*

GIFF relies on Q-values being accurate predictors of utility increments. This is typical in centralized allocation domains where Q-values come from either (i) sufficiently trained RL models, (ii) well-validated domain simulators, or (iii) engineered scoring functions. The assumption mirrors standard requirements for preference-based optimization: the fairness mechanism must operate on accurate inputs.

5.1 Local-Gain Lower Bound

Our first result establishes that the surrogate S is a conservative lower bound on the realized fairness improvement Δ_{joint} for four canonical fairness metrics.

Theorem 1 (Local–Gain Lower Bound). *Let $\mathbf{Z} \in \mathbb{R}^n$ be a payoff vector and $y \in \mathbb{R}^n_{\geq 0}$ be a nonnegative increment vector. For each fairness function $F \in \{F_\alpha, F_{GGF}, F_{var}, F_{min}\}$, the realized joint gain dominates the sum of local gains:*

$$\Delta_{\text{joint}} \;\geq\; S = \sum_{i=1}^{n} \Delta_i^{\text{local}},$$

Proof Sketch. The result for α-fairness follows from the separability of the function, which yields exact equality. For negative variance, the inequality arises from the non-negative cross-term $\frac{2}{n^2}\sum_{i<j} y_i y_j$ in the expression for $\Delta_{\text{joint}} - S$. The proofs for GGF and maximin rely on the properties of order statistics and case analysis on the minimum-achieving agents, respectively. □

5.2 Monotonicity of Surrogate Fairness in β

Next, we show that GIFF's surrogate objective is guaranteed to be nondecreasing as the fairness weight β increases. This provides a reliable mechanism for tuning the fairness-utility trade-off.

Theorem 2 (Monotone Surrogate Fairness). *Fix a decision round and let $\mathcal{A}$ be the finite set of feasible joint allocations. Let $U(A)$ be the total utility and $S(A)$ be the sum of local fairness gains for an allocation $A \in \mathcal{A}$. Let $A^*(\beta)$ be the allocation chosen by GIFF for a given fairness weight $\beta \in [0, 1)$. For any $0 \leq \beta_1 < \beta_2 < 1$, the corresponding surrogate fairness values are ordered:*

$$S(A^*(\beta_2)) \;\geq\; S(A^*(\beta_1)).$$

Proof Sketch. Let $\theta = \beta/(1-\beta)$. The GIFF objective is to maximize $G_\theta(A) = U(A) + \theta S(A)$. By optimality, we have $G_{\theta_1}(A_1) \geq G_{\theta_1}(A_2)$ and $G_{\theta_2}(A_2) \geq G_{\theta_2}(A_1)$. Subtracting these two inequalities yields $(\theta_2 - \theta_1)(S(A_2) - S(A_1)) \geq 0$. Since $\theta_2 > \theta_1$, we must have $S(A_2) \geq S(A_1)$. □

Corollary 1 (Strict Increase at a Switch). *If the maximizer of the GIFF objective is unique at β_1 and β_2, and $A^*(\beta_1) \neq A^*(\beta_2)$, then $S(A^*(\beta_2)) > S(A^*(\beta_1))$.*

5.3 From Surrogate to Realized Fairness Guarantees

Finally, we connect the surrogate guarantees to realized fairness by bounding the *slack*, defined as slack := $\Delta_{\text{joint}} - S$. By Theorem 1, the slack is always non-negative for the metrics considered. We can derive exact expressions or tight bounds for it:

- **α-fairness:** The surrogate is exact, so slack = 0.
- **Negative Variance:** The slack is a precisely computable quadratic term, slack = $\frac{2}{n^2}\sum_{i<j} y_i y_j$.
- **GGF & Maximin:** The slack depends on whether utility increments cause agents to change ranks (for GGF) or on the uniqueness of the minimum-payoff agent (for maximin).

These bounds allow us to translate the monotonicity of the surrogate into a guarantee on realized fairness. When an increase in β triggers a switch to a new allocation, if the surrogate S increases by more than the maximum possible slack of the previous allocation, the realized fairness Δ_{joint} is guaranteed to strictly increase. For α-fairness, any switch to a different allocation with a higher surrogate value implies a strict increase in realized fairness.

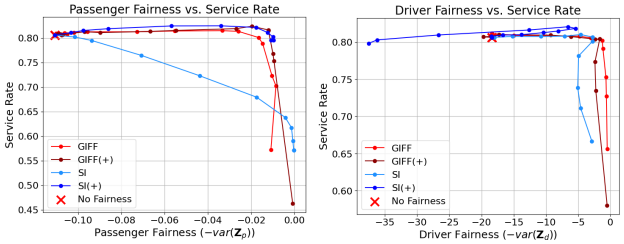


Figure 1: Comparison of fairness versus system utility. Each line is plotted in order of increasing fairness tradeoff weight β, starting from the red X ($\beta = 0$) Top: Passenger fairness (measured as $-var(\mathbf{Z}_p)$) versus overall service rate. Bottom: Driver fairness (measured as $-var(\mathbf{Z}_d)$) versus overall service rate. The red X indicates the baseline performance (raw Q-values without fairness adjustments).

In summary, our theoretical results provide a firm foundation for GIFF. We have shown that its core mechanism of using a sum-of-local-gains surrogate is principled (Theorem 1), predictable (Theorem 2), and directly translatable into guarantees on realized fairness (Section 5.3). This analysis confirms that GIFF is not just a heuristic but a structured framework for improving fairness in complex allocation problems.

6 EMPIRICAL RESULTS

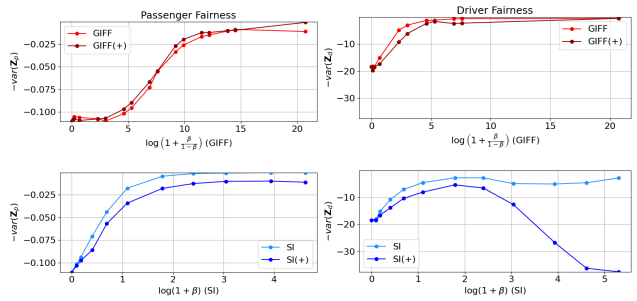
We show results from two experiments. First, we compare GIFF to an existing domain-specific method for ridesharing, optimizing variance (a distributional metric). Then, we extend this to a new domain of homelessness, minimizing the Gini index. Finally, we look at SWF-based metrics in a domain highlighting the need for the counterfactual advantage correction.

6.1 Baseline Comparisons in Real-World Domains: Ridesharing

We test our approach in the complicated ridesharing domain [11, 21], where passengers are allocated to drivers in a dynamic matching environment. Part of the complexity also arises from the fact that more than one passenger can be allocated to the same driver, sharing the trip, leading to a huge combinatorial search space that is difficult to directly optimize. Each vehicle estimates the Q-values based on groups of passengers, and the central allocation maximizes this Q-value for all drivers, subject to passenger constraints.

We compare our results to SI [11], a recent approach for myopic fairness designed for the ridesharing application. SI's fairness objective is to reduce variance in groups of passengers (measured in terms of service rate) and in drivers (measured as differences in trips assigned). Their approach operates in a centrally constrained resource allocation environment by augmenting the base model with an additive fairness term. In our experiments, we compare both the original SI and its heuristic variant SI(+), which clips negative fairness incentives to focus solely on improvements. We also implement a corresponding heuristic in our method, denoted GIFF(+).

Both GIFF and SI start from the same base model that predicts raw Q-values (indicated by the red X's in Figure 1, which represents the



(a) Passenger Fairness (b) Driver Fairness

Figure 2: Variance vs. Fairness Weight. Top row: GIFF results using $\log\left(1+\frac{\beta}{1-\beta}\right)$. Bottom row: SI results using $\log(1+\beta)$.

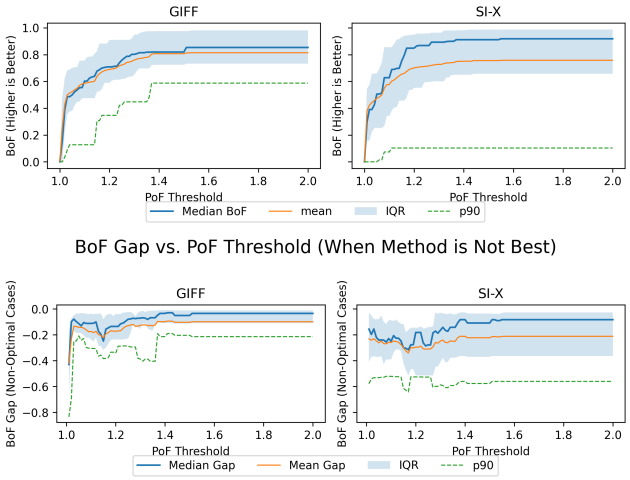


Figure 3: Results for the homelessness dataset. Top: Comparison of the benefit of fairness (BoF) distribution as the price of fairness (PoF) threshold is increased. Bottom: The BoF gap compared to the best method, excluding BoF=0. Each vertical slice corresponds to the distribution over all 38 features.

baseline with no fairness adjustments). Our goal is to demonstrate that GIFF not only improves fairness over SI, but also avoids the drawbacks of SI(+) at high fairness weights. For passengers, fairness is measured as the variance in service rate for groups traveling between source and destination neighborhoods ($\mathbf{Z}_p$). For drivers, the objective is to minimize variance in driver income, measured by the number of trips per driver ($\mathbf{Z}_d$). In our simulation, 1000 vehicles are deployed on the island of Manhattan, using a real-world dataset from New York City, capturing passenger requests between 8am and 12pm during the busy morning hours.

Our results, shown in Figure 1, demonstrate that GIFF consistently achieves a better fairness-utility tradeoff than SI for both passengers and drivers. For passengers, GIFF outperforms the base SI method. While the heuristic variant SI(+) provides a marginal improvement, applying the same heuristic to our method (GIFF(+)) matches its performance. This shows that GIFF's core formulation is strong and can be enhanced with simple heuristics when long-term values are unavailable.

The superiority of GIFF is most pronounced for drivers. As the fairness weight β increases, GIFF maintains a stable and favorable tradeoff. In contrast, SI(+) becomes counterproductive, eventually degrading fairness to a level **worse than the baseline** with no fairness adjustments (indicated by the red X). This instability at high fairness weights highlights a critical drawback of the SI(+) heuristic, whereas our method, GIFF, remains robust and effective across the full range of fairness weights.

Fairness with changing β: In SI, the objective is additive ($U+\beta F$), as opposed to GIFF's weighted combination (Eq. 10). To compare the effect of this tradeoff weight on fairness, we transform the weights for GIFF as $\frac{\beta}{1+\beta}$, and plot the change in fairness for both SI and GIFF with this hyperparameter on a logarithmic scale. Figure 2 shows the tradeoff between fairness weight and the variance in utilities $\mathbf{Z}_p$ and $\mathbf{Z}_d$.

For passengers, both GIFF and SI keep improving fairness as β is increased, as shown in the left panels. In the top right panel, GIFF continues to lower the variance for drivers as the fairness weight increases. However, SI(+) (bottom right) eventually worsens fairness, even falling below the baseline performance. This demonstrates that GIFF achieves a more stable fairness–utilty tradeoff.

6.2 Generalization to Homelessness Prevention

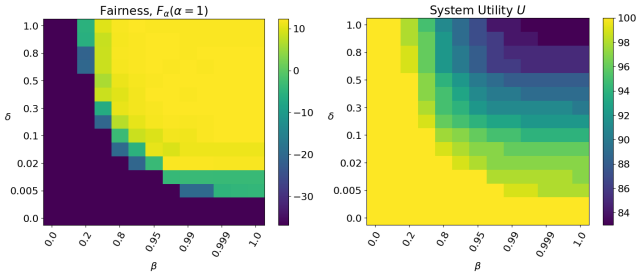
To demonstrate the versatility of our approach beyond dynamic, Q-value-based environments, we test GIFF in the domain of homelessness prevention using a real-world dataset [8]. Here, the task is to assign households to one of four interventions to minimize the total probability of re-entry into homelessness. The "cost" of assigning household h to intervention a is given by a pre-computed counterfactual probability, $\Pr(h,a)$.

Our framework is adapted to this new context by treating it as a cost-minimization problem. We use the negative of the re-entry probabilities as utility values, so that $Q^{\text{GIFF}}(h,a,\beta,\delta) = (1-\beta)\,(-\Pr(h,a)) + \beta\, Q_f(a)$. To further highlight GIFF's flexibility, we move beyond variance reduction and adopt the **Gini coefficient** as the fairness metric, $F_{\text{gini}} = -\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|z_i - z_j|}{2n\sum_{k=1}^{n} z_k}$, where $z_i \in [0,1]$ is the average re-entry probability for a given demographic group.

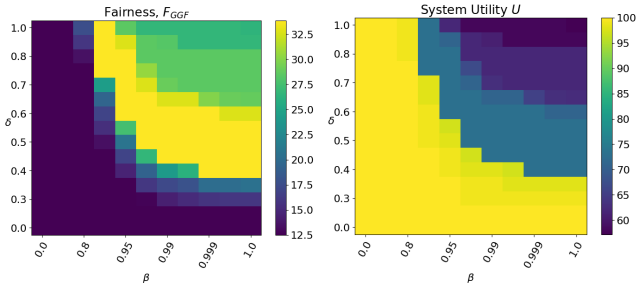
The dataset includes 38 household features (e.g., race, gender, family size), and we run 38 independent experiments, each defining fairness groups based on one feature. This allows us to assess the robustness of each method across a wide variety of fairness definitions. Since the original SI method is not applicable directly, we developed a competitive baseline, Simple Incentives - Extended (SI-X). Further details about the dataset, implementation and SI-X formulation are included in the supplement.

To evaluate performance across these 38 experiments, we introduce two metrics:

- **Price of Fairness (PoF):** The ratio of the total re-entry probability with fairness adjustments to the baseline (fairness-unaware) total probability. A PoF of 1.05 means a 5% increase in the overall re-entry rate.
- **Benefit of Fairness (BoF):** The percentage reduction in the Gini coefficient compared to the baseline, calculated as $1 - \frac{\text{Gini(new)}}{\text{Gini(base)}}$.



(a) α-fair results



(b) GGF results

Figure 4: Fairness and utility for the Job Allocation environment as functions of β and δ.

Figure 3 summarizes the results by plotting the distribution of BoF achieved for a given PoF threshold. Each vertical slice represents the BoF distribution over all 38 feature-based groupings. The top panel shows that GIFF is a more effective and reliable method. On 90% of the features, GIFF is able to get 60% improvement in F_{gini} compared to the baseline. GIFF consistently yields a higher **mean** BoF. More importantly, GIFF demonstrates superior worst-case performance; its 90th percentile is substantially higher than that of SI-X, indicating that GIFF avoids the severe fairness failures that SI-X is prone to on certain groups. The bottom panel reinforces this conclusion by analyzing the **BoF Gap**, which measures how much a method underperforms *only on the tasks where it was not the best*. The gap for GIFF is minimal and concentrated near zero, showing that even when it is not the top performer, it is a close second. In contrast, SI-X exhibits a wide gap distribution, with its 90th percentile exceeding 0.4. This means that when SI-X fails, it fails badly, achieving fairness outcomes that are dramatically worse than what is possible with GIFF.

Overall, these results in a distinct problem domain with a non-linear fairness metric confirm that GIFF is a robust and broadly applicable framework for integrating fairness into resource allocation systems.

6.3 Job Allocation and Advantage Correction

We evaluate our method in a challenging Job Allocation environment where 4 agents compete for a single job over 100 time steps. An agent occupying the job earns reward but must `forfeit` it (earning no reward for that step) to allow another agent to take over. This setup creates a tension between a greedy strategy (one agent gets 100 utility), a simple turn-taking strategy (50 total utility, fairly split), and a hard-to-compute **oracle solution** that achieves 96 total utility (24 per agent). Our goal is to show that our method can discover this oracle solution via a simple hyperparameter search over β and δ, without needing to plan over the full time horizon.

Our approach relies on the advantage correction term, ΔQ_{adv}. We test two **Efficient** fairness metrics (α-fair and GGF), where the fairness gain ΔF is always positive. Consequently, without the correction term, an agent would never forfeit the job, and even a fully fairness-focused policy ($\beta = 1$) would produce the same unfair outcome as a purely utilitarian one ($\beta = 0$). The advantage correction solves this by reducing the value of actions that offer below-average fairness gains to agents that are already better off, thereby encouraging equitable turn-taking.

Our results, shown in Figure 4, confirm that this mechanism successfully balances utility and fairness:

- With the **α-fair** metric (Figure 4a), the policy achieves near-optimal performance with a moderate $\delta \approx 0.1$. As δ increases, fair behavior emerges at lower β values.
- With the **GGF** metric (Figure 4b), the impact of advantage correction becomes noticeable around $\delta \approx 0.3$. This metric reveals a distinct band of high-utility, high-fairness solutions before the policy converges to a simple turn-taking strategy when both β and δ are high. This difference occurs because the logarithmic nature of α-fairness is less sensitive to utility changes once all agents are doing reasonably well.

In both experiments, we observe that as δ increases, the transition from utilitarian to fair behavior happens at a smaller β. Crucially, our method identifies the optimal long-term strategy through a simple evaluation-only grid search, without any information about the time horizon. This demonstrates that the advantage correction term is a powerful and essential mechanism for achieving complex, far-sighted fairness without explicit planning.

7 DISCUSSION AND CONCLUSION

In this work, we introduced GIFF, a general and lightweight framework that integrates fairness into multi-agent resource allocation systems. By modifying pre-trained Q-values with a local fairness gain and a crucial counterfactual advantage correction, our method adjusts allocations to be more equitable without requiring any additional training. Our empirical evaluation showcased GIFF's power and versatility across diverse and challenging domains. The practical strengths of GIFF are underpinned by a solid theoretical foundation. Its design is simple, with minimal computational overhead and just two interpretable hyperparameters (β, δ) that allow for easy tuning of the fairness-utility trade-off at deployment. We formally proved that this is not merely a heuristic approach; GIFF's fairness surrogate is a principled lower bound on the true, realized fairness improvement for several canonical metrics. Furthermore, we showed that the fairness weight β provides a monotonic guarantee, ensuring that increasing the emphasis on fairness predictably improves the surrogate objective. In conclusion, GIFF provides a practical, powerful, and principled bridge between the efficiency of reinforcement learning and the critical societal need for equity, representing a significant step toward creating multi-agent systems that are not only optimal but also just.

REFERENCES

[1] Parand Alizadeh Alamdari, Toryn Q. Klassen, Elliot Creager, and Sheila A. McIlraith. 2024. Remembering to be fair: On non-Markovian fairness in sequential decision making. In *Proceedings of the International Conference on Machine Learning*.

[2] Javier Alonso-Mora, Samitha Samaranayake, Alex Wallar, Emilio Frazzoli, and Daniela Rus. 2017. On-demand High-capacity Ride-sharing via Dynamic Trip-vehicle Assignment. *Proceedings of the National Academy of Sciences* 114 (2017), 462–467.

[3] Ioannis Caragiannis, David Kurokawa, Hervé Moulin, Ariel D Procaccia, Nisarg Shah, and Junxing Wang. 2019. The unreasonable fairness of maximum Nash welfare. *ACM Transactions on Economics and Computation* 7, 3 (2019), 1–32.

[4] Frits De Nijs, Erwin Walraven, Mathijs De Weerdt, and Matthijs Spaan. 2021. Constrained Multiagent Markov Decision Processes: A Taxonomy of Problems and Algorithms. *Journal of Artificial Intelligence Research* 70 (2021), 955–1001.

[5] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the Conference on Innovations in Theoretical Computer Science*. 214–226.

[6] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of opportunity in supervised learning. In *Proceedings of the Conference on Neural Information Processing Systems*. 3323–3331.

[7] Jiechuan Jiang and Zongqing Lu. 2019. Learning fairness in multi-agent systems. In *Proceedings of the Conference on Neural Information Processing Systems*. 13854–13865.

[8] Amanda R. Kube, Sanmay Das, and Patrick J. Fowler. 2019. Allocating interventions based on predicted outcomes: A case study on homelessness services. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 622–629.

[9] Amanda R. Kube, Sanmay Das, and Patrick J. Fowler. 2023. Community-and data-driven homelessness prevention and service delivery: optimizing for equity. *Journal of the American Medical Informatics Association* 30, 6 (2023), 1032–1041.

[10] Harold W Kuhn. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 2, 1-2 (1955), 83–97.

[11] Ashwin Kumar, Yevgeniy Vorobeychik, and William Yeoh. 2023. Using simple incentives to improve two-sided fairness in ridesharing systems. In *Proceedings of the International Conference on Automated Planning and Scheduling*. 227–235.

[12] Ashwin Kumar and William Yeoh. 2025. DECAF: Learning to be Fair in Multi-agent Resource Allocation. *arXiv preprint arXiv:2502.04281* (2025).

[13] Ashwin Kumar and William Yeoh. 2025. Remember, but also, Forget: Bridging Myopic and Perfect Recall Fairness with Past-Discounting. *arXiv preprint arXiv:2504.01154* (2025).

[14] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *Comput. Surveys* 54, 6 (2021), 1–35.

[15] John F Nash et al. 1950. The bargaining problem. *Econometrica* 18, 2 (1950), 155–162.

[16] Ali Nauman, Haya Mesfer Alshahrani, Nadhem Nemri, Kamal M Othman, Nojood O Aljehane, Mashael Maashi, Ashit Kumar Dutta, Mohammed Assiri, and Wali Ullah Khan. 2024. Dynamic resource management in integrated NOMA terrestrial–satellite networks using multi-agent reinforcement learning. *Journal of Network and Computer Applications* 221 (2024), 103770.

[17] Naveen Raman, Sanket Shah, and John Dickerson. 2021. Data-driven methods for balancing fairness and efficiency in ride-pooling. In *Proceedings of the International Joint Conference on Artificial Intelligence*. 363–369.

[18] John Rawls. 1971. A Theory of justice. *Cambridge (Mass.)* (1971).

[19] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347* (2017).

[20] Amartya Sen. 2017. *Collective Choice and Social Welfare: Expanded Edition*. Penguin UK.

[21] Sanket Shah, Meghna Lowalekar, and Pradeep Varakantham. 2020. Neural approximate dynamic programming for on-demand ride-pooling. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 507–515.

[22] Umer Siddique, Paul Weng, and Matthieu Zimmer. 2020. Learning fair policies in multi-objective (deep) reinforcement learning with average and discounted rewards. In *Proceedings of the International Conference on Machine Learning*. 8905–8915.

[23] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z. Leibo, Karl Tuyls, and Thore Graepel. 2018. Value-Decomposition Networks For Cooperative Multi-Agent Learning Based On Team Reward. In *Proceedings of the Conference on Autonomous Agents and Multiagent Systems*. 2085–2087.

[24] Yuan Xue, Baochun Li, and Klara Nahrstedt. 2003. Price-based resource allocation in wireless ad hoc networks. In *Proceedings of the International Workshop on Quality of Service*. 79–96.

[25] Matthieu Zimmer, Claire Glanois, Umer Siddique, and Paul Weng. 2021. Learning fair policies in decentralized cooperative multi-agent reinforcement learning. In *Proceedings of the International Conference on Machine Learning*. 12967–12978.