Fair Contracts in Principal-Agent Games with Heterogeneous Types

Extended Abstract

Jakub Tłuczek
Université de Neuchâtel
Neuchâtel, Switzerland
jakub.tluczek@unine.ch

Victor Villin
Université de Neuchâtel
Neuchâtel, Switzerland
victor.villin@unine.ch

Christos Dimitrakakis
Université de Neuchâtel
Neuchâtel, Switzerland
christos.dimitrakakis@unine.ch

ABSTRACT

Fairness is desirable yet challenging to achieve within multi-agent systems, especially when agents differ in latent traits that affect their abilities. This hidden heterogeneity often leads to unequal distributions of wealth, even when agents operate under the same rules. Motivated by real-world examples, we propose a framework based on repeated principal-agent games, where a principal, learns to offer adaptive contracts to agents. By leveraging a simple yet powerful contract structure, we show that a fairness-aware principal can learn homogeneous linear contracts that equalize outcomes across agents in a sequential social dilemma. Importantly, this fairness does not come at the cost of efficiency: our results demonstrate that it is possible to promote equity and stability in the system while preserving overall performance.

KEYWORDS

Multi Agent Reinforcement Learning; Contract Theory; Incentive Design; Fairness

ACM Reference Format:
Jakub Tłuczek, Victor Villin, and Christos Dimitrakakis. 2026. Fair Contracts in Principal-Agent Games with Heterogeneous Types: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/IAXU7159

1 INTRODUCTION

Modern economies, at both macro and micro levels, constitute complex multi-agent systems. Interactions between agents often involve contracts [1, 2, 7, 10, 12] designed to align incentives between parties, fitting the principal-agent model: one party (the principal) offers a reward to another (the agent) in exchange for a specific outcome. This framework applies to a wide range of real-world scenarios, from employment agreements to government subsidies.

The parties in the principal-agent model are often assumed to be greedy with respect to their own wealth, whereas in real-life settings, it has been shown that fairness considerations play an important role in contract design [4]. The principal might be inclined to offer more lucrative contracts to ensure that the agent stays motivated to take the actions that lead to rewarding outcomes.

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/IAXU7159

We extend this insight to Multi-Agent Reinforcement Learning (MARL), where the principal and the agents independently seek to maximize their rewards through strategic decision making. By shaping the reward structure through contract design, the principal can steer multiple agents towards more cooperative and equitable behavior, ultimately promoting system-wide efficiency and fairness.

We consider a setting with heterogeneous agents, with the magnitude of rewards being dependent on their intrinsic type, unknown to the principal. The principal's task is to offer contracts that achieve equitable wealth among all agents despite their differences. Nevertheless, the principal must still consider her own wealth. We propose two objectives for the contract design: The first adapts prior work [8] by incorporating *wealth* (sum of rewards) of the agents into the principal's objective. The second explicitly integrates the fairness factor into the principal's objective, by penalizing large variance between wealths of the agents and the principal.

2 LEARNING CONTRACTS

We model the problem as an extension of a finite horizon n-player Markov Game, where we introduce an additional player who performs the role of the principal. At each timestep t, the principal proposes a homogenous linear contract α_t [3] to all agents, which represents a share in the potential reward collected that the agent would earn. To take an action, the agents have to bear the costs, therefore the contracts have to be designed so that the resulting payments offset the costs under expectation.

To capture heterogeneity among agents, each agent i is endowed with a type $\theta^i \in \Theta$, with $\boldsymbol{\theta}$ the vector of all agent types. These are not observable by the principal and can represent differences in skill, efficiency, or preferences. Following Zheng et al. [14], an agent's type simply scales their effective contributions to $\theta^i r^i_t$.

The learning aspect of the problem is addressed by the *policy gradient* methods. Each player (both the principal and the agents) optimizes their own policy independently, having observed the current state and (in the case of the agents only) a contract. Since the result is a bi-level optimization problem with the principal's objective being outer, and the agent's an inner optimization loop, we adjust the learning rate so that the principal learns significantly slower. It gives the agents the opportunity to learn to respond to the principal's policy before it significantly changes.

3 REGULARIZATION FOR FAIR CONTRACTS

Rather than solely maximizing its own wealth, the principal can incorporate fairness considerations into its objective. Promoting fairness requires some degree of altruism from the principal, which

Table 1: Comparison of training metrics. *For NoP, metrics were computed without the principal.

	NoP*	Greedy	Fix	Regularized						
				Welfare				Wealth Variance		
λ				1	9	12	15	0.75	1	1.25
1 - Gini	0.952	0.200	0.947	0.704	0.834	0.852	0.957	0.978	**0.992**	0.986
Welfare	44.6	3.4	45.5	28.6	34.4	43.0	21.2	44.6	**45.5**	44.7
Rawlsian	**18.0**	-0.2	11.2	0.9	5.9	6.9	5.7	13.3	14.7	14.1
AIE	42.5	2.2	43.1	20.3	29.9	36.8	20.3	43.6	**45.1**	44.1

can be quantified by an altruism parameter λ [13]. This captures the extent to which it values the overall welfare, or fairness of the system, relative to the profits.

It has been shown that directly tying the principal's objective to agent wealth can encourage more cooperative behavior [8]. Following this insight, we incorporate system welfare into the principal's reward by augmenting it with the agents' collected rewards. This leads to a modified reward function: $R_{\text{p}}^{\text{welfare}}(s_t, \mathbf{a}_t, s_{t+1}, b_t) = \sum_{i=1}^{n} \left((1-\alpha_t)\theta^i r_t^i\right) + \lambda \sum_{i=1}^{n} \theta^i r_t^i$.

Here, the principal observes the rewards $\theta^i r_t^i$ for each agent but cannot infer the agent's underlying type. The coefficient λ plays a central role by modulating the degree of altruism in the principal's behavior. When $\lambda \to 0$, the principal becomes greedy, which in turn leads to a situation where the contracts offered exploit agents and let them learn to collect amounts just barely enough to offset the costs. As λ grows, the principal becomes altruistic to the point of becoming a central planner, who disregards its own wealth and seeks to only maximize the welfare of agents interacting with the environment. The result is thus sensitive to the choice of λ, having risks of the system either converging to unfair, exploitative contracts (low λ) or to overly altruistic behavior that ignores the principal's wealth (high λ). This highlights the need for a more nuanced regularization method that embeds fairness more explicitly and robustly into the contract design objective.

An alternative approach is to regularize the principal's objective based on the fairness over parties' wealth, rather than overall welfare. Specifically, we suggest penalizing the principal for creating wealth inequalities across all parties, thereby encouraging contract offers that promote a more balanced distribution over an episode.

Let $\mathcal{W}_t$ be the cumulative wealth of all parties up to time t, and let $F(\mathcal{W}_t)$ be a fairness metric applied to that distribution. The principal fairness-aware rewards becomes $R_{\text{p}}^{\text{fairness}}(s_t, \mathbf{a}_t, s_{t+1}, b_t) = \sum_{i=1}^{n} \left((1-\alpha_t)\theta^i r_t^i\right) + \lambda F(\mathcal{W}_t)$, where the principal's wealth $w^p \in \mathcal{W}_t$ is calculated with its non-regularized rewards R_{p}. A simple equity-promoting concept would be to penalise wealth variance: $F(\mathcal{W}) = -\text{Var}[\mathcal{W}]$. While simple and computationally tractable, this formulation does not explicitly incorporate agent heterogeneity and may not ensure proportional fairness [9], where wealth appropriately reflects differences in ability or effort.

We include the variance of wealth of all parties in the principal's objective, since having a penalty of only agents' wealth variance entails a trivial drawback. Namely, in the case when each agent collected nothing (i.e. $\forall_a w_a = 0$) the variance penalty would also equal zero. Having the principal's wealth incorporated into the variance term creates a dual objective, where the principal wants to offer contracts maximizing its wealth, while at the same time ensuring it is not radically different from the agent's endowments.

4 EXPERIMENTS

We test our framework on a modified version of Coin Game [5], a sequential social dilemma with two agents, red and blue, on a square grid. Their goal is to collect a red or blue coin. The coin is always present on an empty grid spot. Upon collection, a new coin is generated randomly. The agents are rewarded with one point if they collect a coin matching their color, and 0.2 points for a coin not matching their color. Reward is then scaled by their type. Selfish agents will give no attention to the wealth of another agent, seeking to maximize their own. It is detrimental to the welfare of the entire system, with less skilled agents having fewer incentives to move.

While agents seek to maximize their individual welfare under the contract dynamics, the principal's goal, when regularized, is to design contracts that promote welfare for agents (WR) or equitable welfare distribution (VR). Since multiple metrics for measuring such equity have been proposed in the literature, we include several of them in our study. Based on the widely used parity metric of Gini index [6], its complement, the *1 - Gini*, maps equality in wealth to the range $[0, 1]$, from most unequal to most equal. *Welfare*, or the total wealth sum over players, provides information on whether fairness comes at the price of welfare compared to other solutions. Rawlsian index [11, 14], provides the wealth of the poorest agent, while the product of 1 - Gini and Welfare, a metric used in the AI economist [14] (thus *AIE*), captures both equality and productivity.

Using variance regularization and $\lambda = 1$ consistently outperformed other benchmarks over several seeds (see Table 1), each time achieving near equal distribution of the wealth as measured by 1 - Gini, while at the same time excelling in total accumulated wealth, as captured by welfare and AIE.

5 CONCLUSION

We have proposed a framework for designing fair contracts in SSD games with heterogeneous agents and verified the results empirically. Crucially, we have shown that simple linear fairness-aware contracts are able to induce cooperative behavior between agents, ensuring the fairness of the entire system, without sacrificing welfare in comparison to the baselines.

ACKNOWLEDGMENTS

This research was partially supported by Swiss National Science Foundation (grant no. 219499).

REFERENCES

[1] Patrick Bolton and Mathias Dewatripont. 2005. *Contract theory.* MIT Press, Cambridge, MA Londin, England.

[2] Paul Dütting, Michal Feldman, and Inbal Talgam-Cohen. 2024. Algorithmic Contract Theory: A Survey. https://doi.org/10.48550/arXiv.2412.16384 arXiv:2412.16384 [cs].

[3] Paul Dütting, Tim Roughgarden, and Inbal Talgam-Cohen. 2019. Simple versus Optimal Contracts. In *Proceedings of the 2019 ACM Conference on Economics and Computation*. ACM, Phoenix AZ USA, 369–387. https://doi.org/10.1145/3328526.3329591

[4] Ernst Fehr, Alexander Klein, and Klaus M Schmidt. 2007. Fairness and Contract Design. *Econometrica* 75, 1 (Jan. 2007), 121–154. https://doi.org/10.1111/j.1468-0262.2007.00734.x

[5] Jakob N. Foerster, Richard Y. Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. 2018. Learning with Opponent-Learning Awareness. https://doi.org/10.48550/arXiv.1709.04326 arXiv:1709.04326 [cs].

[6] Corrado Gini. 1912. *Variabilità e mutabilità.* Tipogr. di P. Cuppini.

[7] Bengt Holmstrom and Paul Milgrom. 1987. Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica* 55, 2 (March 1987), 303. https://doi.org/10.2307/1913238

[8] Dima Ivanov, Paul Dütting, Inbal Talgam-Cohen, Tonghan Wang, and David C. Parkes. 2024. Principal-Agent Reinforcement Learning. http://arxiv.org/abs/2407.18074 arXiv:2407.18074 [cs].

[9] Peizhong Ju, Arnob Ghosh, and Ness B. Shroff. 2023. Achieving Fairness in Multi-Agent Markov Decision Processes Using Reinforcement Learning. https://doi.org/10.48550/arXiv.2306.00324 arXiv:2306.00324 [cs].

[10] Jean-Jacques Laffont and David Martimort. 2009. *The Theory of Incentives: The Principal-Agent Model.* Princeton University Press. https://doi.org/10.2307/j.ctv7h0rwr

[11] John Rawls. 1971. *A Theory of Justice: Original Edition.* Harvard University Press. http://www.jstor.org/stable/j.ctvjf9z6v

[12] Bernard Salanié. 2005. *The economics of contracts: a primer* (second edition ed.). The MIT Press, Cambridge, Massachusetts London, England.

[13] Jack Sawyer. 1966. The Altruism Scale: A Measure of Co-operative, Individualistic, and Competitive Interpersonal Orientation. *Amer. J. Sociology* 71, 4 (1966), 407–416. http://www.jstor.org/stable/2774936

[14] Stephan Zheng, Alexander Trott, Sunil Srinivasa, Nikhil Naik, Melvin Gruesbeck, David C. Parkes, and Richard Socher. 2020. The AI Economist: Improving Equality and Productivity with AI-Driven Tax Policies. https://doi.org/10.48550/arXiv.2004.13332 arXiv:2004.13332 [econ].