Algorithmic Contract Design with Reinforcement Learning Agents

Extended Abstract

David Molina Concha
University of Toronto
Toronto, Canada
damolina@mie.utoronto.ca

Kyeonghyeon Park
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
kyeonghyeon.park@kaist.ac.kr

Hyun-Rok Lee
Inha University
Incheon, South Korea
hyunrok.lee@inha.ac.kr

Taesik Lee
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
taesik.lee@kaist.edu

Chi-Guhn Lee
University of Toronto
Toronto, Canada
cglee@mie.utoronto.ca

ABSTRACT

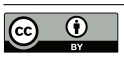
Designing incentive mechanisms for multi-agent systems in stochastic and dynamic environments is a critical challenge, as system outcomes emerge from the complex interplay of agent learning and environmental uncertainty. Existing principal–multi-agent contract design methods often assume static settings or ignore learning dynamics, limiting their applicability in multi-agent reinforcement learning (MARL). Furthermore, the contract design space is highly constrained by feasibility requirements, such as individual rationality and incentive compatibility, making it difficult to explore.

We introduce the principal-MARL contract design problem, where a principal optimizes both recruitment and incentive contracts evaluated via MARL. To address this problem, we propose Constrained Pareto Maximum Entropy Search (cPMES), a multi-objective Bayesian optimization framework that treats feasibility as an explicit objective and selects designs based on information gain over the Pareto front. Experiments in social dilemma environments demonstrate that cPMES efficiently identifies feasible, high-performing contracts, significantly improving coordination and system-level rewards.

KEYWORDS

Multi-agent systems; Reinforcement Learning; Contract Design

ACM Reference Format:
David Molina Concha, Kyeonghyeon Park, Hyun-Rok Lee, Taesik Lee, and Chi-Guhn Lee. 2026. Algorithmic Contract Design with Reinforcement Learning Agents: 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/TBTZ3566



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/TBTZ3566

1 INTRODUCTION

In multi-agent contract design, a principal aligns self-interested agents with a system objective through incentives. Classical models assume static environments, whereas in many settings outcomes emerge from agents learning in dynamic environments [12].

Recent dynamic contract approaches [4, 13, 14] do not explicitly model learning dynamics and must handle highly constrained feasibility requirements, including individual rationality and incentive compatibility.

When contracts are evaluated in Markov games (MGs), each design requires solving a multi-agent reinforcement learning (MARL) problem, making evaluation expensive and stochastic [9–11]. This raises a central challenge: how can a principal efficiently search for feasible, high-performing contracts under learning dynamics?

We (i) formalize contract evaluation as a MG with learning agents, (ii) cast contract design as constrained black-box optimization, and (iii) propose a data-efficient multi-objective Bayesian optimization framework.

2 METHODOLOGY

We formulate the principal–MARL contract design problem as a constrained optimization problem over a learning-based MG. The principal chooses (i) linear incentive weights $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]$ for the set of entire agents $\mathcal{N}$ and (ii) the set of additional agents $\mathcal{N}_a$ to recruit, while the set of baseline agents $\mathcal{N}_b$ cannot be removed [1, 5]. Thus, $\mathcal{N}_b \subseteq \mathcal{N}$, $\mathcal{N}_a \subseteq \mathcal{N}$, $\mathcal{N}_a \cap \mathcal{N}_b = \emptyset$, $\mathcal{N}_a \cup \mathcal{N}_b = \mathcal{N}$, where $N = |\mathcal{N}|, N_a = |\mathcal{N}_a|$ and $N_b = |\mathcal{N}_b|$.

The principal solves:

$$\max_{\boldsymbol{\alpha}, N_a} \mathcal{G}(\boldsymbol{\alpha}, N_a) = \mathbb{E}_{\boldsymbol{\pi}^{\boldsymbol{\alpha}, N_a}}[R^{\boldsymbol{\alpha}, N_a}]$$

$$s.t.\ \mathbb{E}_{\pi_i^{\boldsymbol{\alpha}, N_a}, \pi_{-i}^{\boldsymbol{\alpha}, N_a}}[r_i^{\alpha_i}] \geq \mathbb{E}_{\pi'^{\boldsymbol{\alpha}, N_a}_i, \pi_{-i}^{\boldsymbol{\alpha}, N_a}}[r_i^{\alpha_i}],$$

$$\mathbb{E}_{\pi_j^{\boldsymbol{\alpha}, N_a}, \pi_{-j}^{\boldsymbol{\alpha}, N_a}}[r_j^{\alpha_j}] - \mathbb{E}_{\pi_j, \pi_{-j}}[r_j] \geq 0$$

$$\mathbb{E}_{\pi_k^{\boldsymbol{\alpha}, N_a}, \pi_{-k}^{\boldsymbol{\alpha}, N_a}}[r_k^{\alpha_k}] - c \geq 0$$

$$\forall i \in \{1, \cdots, N\}, \forall j \in \{1, \cdots, N_b\}, \forall k \in \{1, \cdots, N_a\},$$

$$\forall s \in S, \forall \pi_i, \pi_i' \in \Pi_i, \pi_{-i} \in \Pi_{-i},$$

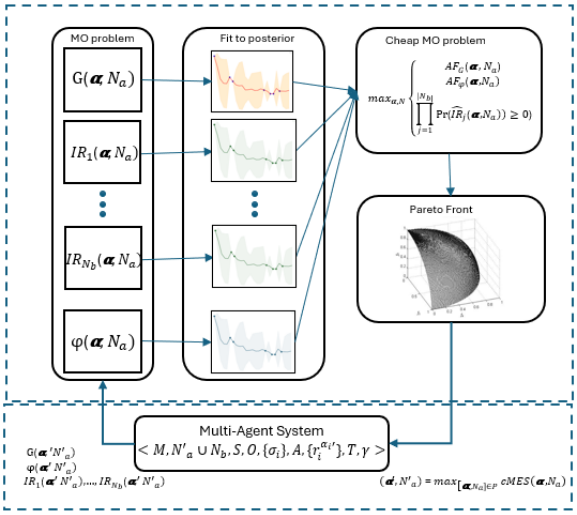


Figure 1: cPMES architecture using MOBO to obtain candidates designs at the contract optimization level and the joint policy π for the given $[\alpha, N_a]$ at the MARL level.

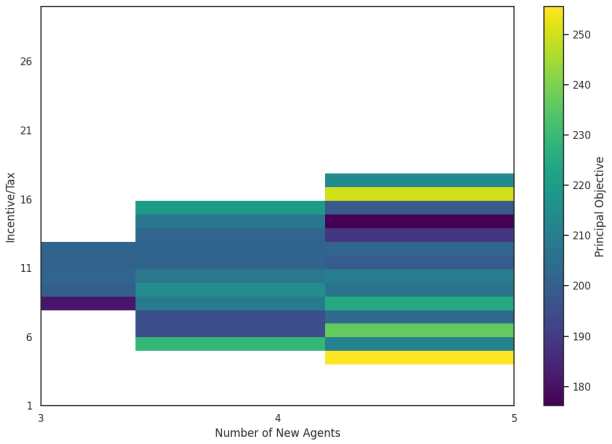


Figure 2: Principal's objective and recommended feasible contracts.

where $\mathcal{G}$ is the principal's objective, R^{α,N_a} denotes the system return under contract $[\alpha, N_a]$, and π^{α,N_a} is the joint policy learned by agents. The first constraints enforce best response conditions. The second and third correspond to Individual Rationality (IR) constraints [3] for baseline and recruited agents, respectively, where c is the minimum acceptable expected return for new agents.

For a given contract $[\alpha, N_a]$, agents learn via MARL in a partially observable MG [7], making both the objective and constraints expensive, stochastic, and black-box.

To efficiently explore this design space, we propose Constrained Pareto Maximum Entropy Search (cPMES), a multi-objective Bayesian optimization (MOBO) framework. The principal's objective and IR constraints are modeled using independent Gaussian process (GP) surrogates [2].

To handle the varying number of IR constraints induced by recruitment, we introduce a feasibility indicator $\phi(\alpha, N_a)$ that captures whether all recruited agents satisfy their minimum return. This yields a compact multi-objective problem balancing predicted performance and feasibility:

$$\max_{\alpha, N_a} \mathrm{AF}_{\mathcal{G}}(\alpha, N_a), \mathrm{AF}_{\phi}(\alpha, N_a), \prod_{j=1}^{N_b} Pr(\hat{IR}_j(\alpha, N_a) \geq 0), \quad (1)$$

where $\mathrm{AF}_{\mathcal{G}}$ and AF_{ϕ} denote acquisition functions for system performance and recruitment feasibility, $\hat{IR}_j(\alpha, N_a)$ is the surrogate of the IR constraint of the baseline agent j, and $Pr(\hat{IR}_j(\alpha, N_a) \geq 0)$ is the predicted probability of feasibility of the IR constraint.

Figure 1 illustrates the entire architecture of the proposed framework. At each iteration, candidate contracts are selected via an entropy-based acquisition rule, evaluated through MARL, and used to update the surrogates, enabling data-efficient exploration of constrained contract spaces.

3 EMPIRICAL RESULTS

We evaluate cPMES on the SSD Clean-up Markov game [6, 8], a social dilemma environment with five baseline harvester agents. The principal may recruit up to five cleaner agents. The contract variables are the number of recruited agents N_a and a tax parameter α that redistributes harvester rewards to cleaners. Recruited agents must satisfy a minimum expected return of zero.

Each contract is evaluated by solving the induced Markov game via MARL, making evaluations expensive and stochastic. We run cPMES for 20 evaluations (10 initial designs) across five random seeds.

Figure 2 shows the feasible contract designs identified by cPMES and their corresponding system-level performance. Under strict individual rationality constraints, cPMES consistently identifies feasible contracts with high recruitment and low tax levels. The best contract ($N_a = 5$, $\alpha = 0.04$) achieves a principal objective of 255.53, a cleaner utility of 0.28, and a harvester return of 49.96, substantially outperforming the no-recruitment baseline (harvester utility: 30.27).

These results show that cPMES efficiently explores constrained contract spaces and discovers incentive structures that improve welfare in learning-based multi-agent systems.

4 CONCLUSIONS

We introduced the principal–MARL contract design problem, a new formulation of algorithmic contract design in which incentives and recruitment decisions are evaluated through Markov games. We proposed cPMES, a constrained multi-objective Bayesian optimization framework that efficiently explores feasible contract designs under learning-induced uncertainty. Experiments in a social dilemma environment demonstrate that cPMES identifies incentive structures that improve coordination and system-level performance while satisfying individual rationality constraints. These results highlight the importance of learning-aware contract design for multi-agent systems.

REFERENCES

[1] Moshe Babaioff, Michal Feldman, Noam Nisan, and Eyal Winter. 2012. Combinatorial agency. *Journal of Economic Theory* 147, 3 (2012), 999–1034.

[2] Syrine Belakaria, Aryan Deshwal, Nitthilan Kannappan Jayakodi, and Janardhan Rao Doppa. 2020. Uncertainty-Aware Search Framework for Multi-Objective Bayesian Optimization. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 06 (2020), 10044–10052.

[3] Matteo Castiglioni, Alberto Marchesi, and Nicola Gatti. 2023. Multi-agent contract design: How to commission multiple agents with individual outcomes. In *Proceedings of the 24th ACM Conference on Economics and Computation*. 412–448.

[4] Che Chen, Shimin Gong, Wenjie Zhang, Yifeng Zheng, and Yeo Chai Kiat. 2022. Deep Reinforcement Learning based Contract Incentive for UAVs and Energy Harvest Assisted Computing. In *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2224–2229.

[5] Paul Dütting, Tomer Ezra, Michal Feldman, and Thomas Kesselheim. 2023. Multi-agent contracts. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. 1311–1324.

[6] Edward Hughes, Joel Z Leibo, Matthew Phillips, Karl Tuyls, Edgar Dueñez-Guzman, Antonio García Castañeda, Iain Dunning, Tina Zhu, Kevin McKee, Raphael Koster, et al. 2018. Inequity aversion improves cooperation in intertemporal social dilemmas. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. 3330–3340.

[7] Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence* 101, 1 (1998), 99–134.

[8] Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. 2017. Multi-agent Reinforcement Learning in Sequential Social Dilemmas. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. 464–473.

[9] David Mguni, Joel Jennings, Emilio Sison, Sergio Valcarcel Macua, Sofia Ceppi, and Enrique Munoz de Cote. 2019. Coordinating the Crowd: Inducing Desirable Equilibria in Non-Cooperative Systems. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. 386–394.

[10] Kyeonghyeon Park, David Molina Concha, Hyun-Rok Lee, Taesik Lee, and Chi-Guhn Lee. 2025. Reward design in multi-agent systems using successor features and multi-information source bayesian optimization. *International Journal of Machine Learning and Cybernetics* (2025), 1–22.

[11] Zhenyu Shou and Xuan Di. 2020. Reward design for driver repositioning using multi-agent reinforcement learning. *Transportation research part C: emerging technologies* 119 (2020), 102738.

[12] Thomas A. Weber and Hongxia Xiong. 2006. Efficient contract design in multi-principal multi-agent supply chains. *SSRN Electronic Journal* (2006).

[13] Yimei Xie, Chuan Ding, Yang Li, and Kaihong Wang. 2023. Optimal incentive contract in continuous time with different behavior relationships between agents. *International Review of Financial Analysis* 86 (2023), 102521.

[14] Nan Zhao, Yiyang Pei, Ying-Chang Liang, and Dusit Niyato. 2023. A Deep Reinforcement Learning-Based Contract Incentive Mechanism for Mobile Crowdsourcing Networks. *IEEE Transactions on Vehicular Technology* (2023).