ECAI 2024
U. Endriss et al. (Eds.)

doi:10.3233/FAIA240464

The World is a Multi-Objective Multi-Agent System: Now What?

Roxana Rădulescu[a,b,*]

[a]Intelligent Systems Group, Utrecht University, the Netherlands
[b]Artificial Intelligence Lab, Vrije Universiteit Brussel, Belgium
ORCID (Roxana Rădulescu): https://orcid.org/0000-0003-1446-5514

Abstract. Most complex problems of social relevance, such as climate change mitigation, traffic management, taxation policy design, or infrastructure management, involve both multiple stakeholders and multiple potentially conflicting objectives. In a nutshell, the majority of real world problems are multi-agent and multi-objective in nature. Artificial intelligence (AI) is a pivotal tool in designing solutions for such critical domains that come with high impact and ramifications across many dimensions, from societal and economic well-being, to ethical, political, and legal levels. Given the current theoretical and algorithmic developments in AI, it is an opportune moment to take a holistic approach and design decision-support tools that: (i) tackle all the prominent challenges of such problems and consider both the multi-agent and multi-objective aspects; (ii) exhibit vital characteristics, such as explainability and transparency, in order to enhance user agency and alignment. These are the challenges that I will discuss during the Frontiers in AI session at ECAI 2024, together with a brief overview of my work and next steps for this field. This paper summarises my contribution to the session.

1 Introduction

Our world is a highly complex environment that requires great effort to process and navigate. For example, humans are remarkably capable at operating under uncertainty and incomplete information in order to achieve a varied set of goals. We are also rarely isolated from each other; every day we make decisions and work towards our goals while interacting with others, either cooperating (e.g., a job in the development team of a software company), competing (e.g., participating in a chess tournament) or some combination of the two (e.g., driving home and participating in city traffic). Additionally, our coexistence with artificial intelligence (AI)-powered systems [44] is slowly transforming many aspects of both physical and online environments, from our workplace [27], to education [46], to healthcare [7], to our communication paradigms [13]. From an AI perspective, every entity that is present and able to act in the world is considered to be an *agent*, be it a human, a robot, or a piece of software [43]; we can therefore say that most real-world decision problems are inherently *multi-agent*.

In addition to this *multi-agent* aspect, when zooming in on real-world decision making problems, we notice that usually more than one objective should be taken into consideration. For example, travelling from one point to another involves evaluating options in terms of cost, travel time, congestion levels, environmental impact, and comfort [33]; in medical treatment we want to maximise effectiveness, while minimising side effects [49]; when devising pandemic mitigation strategies we need to balance between mortality, morbidity, economic costs, and social wellness [37]. In other words, most real-world decision problems are also inherently *multi-objective*.

The increased prevalence of artificial agents in our world makes it crucial to explicitly consider objectives such as safety, fairness, reliability, and transparency [18, 22], while ensuring satisfactory levels of task performance. Figure 1 illustrates a high-level sketch of hybrid human-AI collectives, where artificial agents and human exist, observe, and act in the same environment. To ensure a seamless interaction, artificial agents might need to possess capabilities such as communication, theory of mind, and social intelligence. Furthermore, artificial agents can represent different stakeholders through decision delegation [2, 47]. In this case, transparency, together with the ability to understand and continuously model the preferences of users [51], becomes paramount to achieve trust in the system, enhance human agency and facilitate oversight [8].

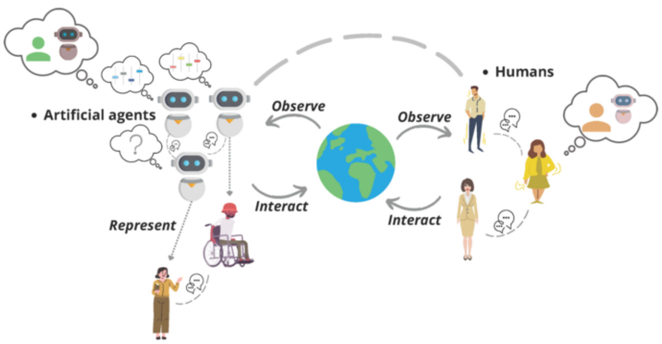

Figure 1. High-level sketch of hybrid collectives of humans and artificial agents. Elements such as communication, decision delegation, preference modelling, social intelligence, theory of mind, etc. are crucial ingredients for a seamless interaction.

Multi-objective multi-agent systems (MOMAS) are thus more holistic and realistic models that capture the complexity of interactions among agents and the multiple dimensions of their objectives. Multi-objective multi-agent decision making (MOMADM) can potentially provide solutions that explore the possible trade-offs among the objectives, as well as the dependencies between the agents.

* Corresponding Author. Email: r.t.radulescu@uu.nl

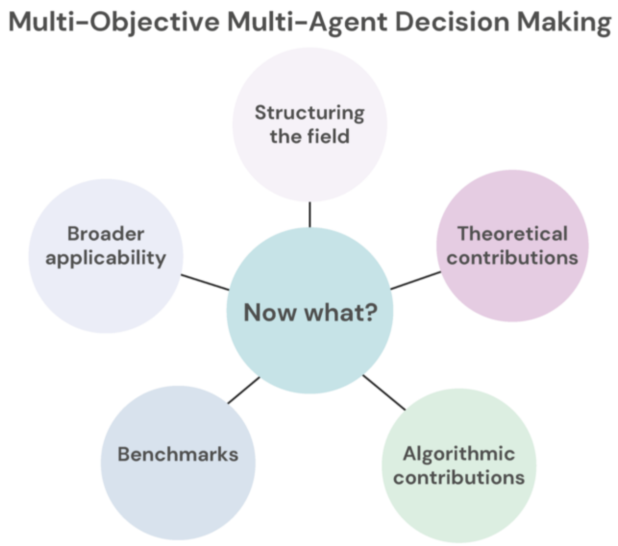

Figure 2. Overview of directions of past and future developments in the field of multi-objective multi-agent decision making

In the following sections I provide an overview of developments in MOMADM and identify directions for future research in the field. Figure 2 illustrates the organisation of this discussion around five themes: (i) structuring the field, (ii) theoretical and (iii) algorithmic contributions, (iv) benchmarks, and (v) broader applicability.

2 Structuring the Field

A general framework for modelling multi-objective multi-agent decision-making settings is the multi-objective partially observable stochastic game (MOPOSG). MOPOSGs extend Markov decision processes [29] to both multiple agents and multiple objectives. Notable extensions introduced by MOPOSGs are: (i) a vectorial reward function $\boldsymbol{R}_i$, for every agent i, taking values in $\mathbb{R}^d$, with d the number of considered objectives; (ii) the agents are in general unable to observe the full state of the environment, and receive instead partial information in the form of observations.

By making additional assumptions on the MOPOSG model, regarding observability, the structure of the reward function, or whether the problem is sequential or not, we can derive a subset of models such as the multi-objective stochastic game (MOSG), multi-objective decentralised partially observable Markov decision process (MODec-POMDP), multi-objective Bayesian game (MOBG), multi-objective cooperative Bayesian game (MOCBG), multi-objective multi-agent Markov decision process (MOMMDP), multi-objective normal form game (MONFG), or multi-objective multi-agent multi-armed bandit (MOMAMAB), as illustrated in Figure 3 [30].

Agents usually aim to optimise their individual expected discounted return, under a joint policy $\boldsymbol{\pi}$. Since an agent only directly controls its own policy π_i, we arrive at the same challenges present in multi-agent learning settings, such as credit assignment (i.e., disentangling the individual agents' contributions to the resulting reward signal), and non-stationarity (i.e, agents simultaneously learning in the environment creates a moving target problem [50]).

Moreover, as a consequence of the fact that the reward and hence the value function is a vector, $\boldsymbol{v}_i^{\boldsymbol{\pi}} \in \mathbb{R}^d$, only a partial ordering over the policy space is available. Determining a single optimal policy for the execution phase [14] requires additional information on how agents, or their corresponding users, prioritise the objectives or what

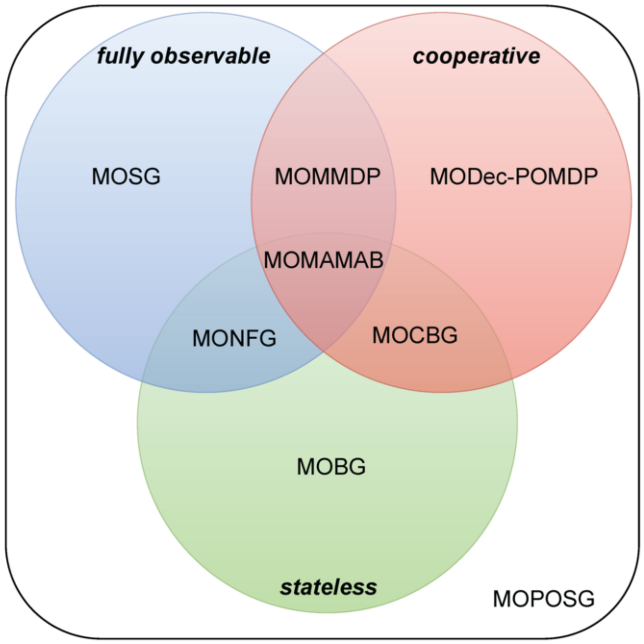

Figure 3. Multi-objective multi-agent decision-making models characterised along three dimensions: (i) observability; (ii) cooperativeness; (iii) statefulness [30].

their preferences over the objectives are. We can capture such a trade-off choice using a *utility function*, $u_i : \mathbb{R}^d \to \mathbb{R}$, that maps the vector to a scalar value.

In order to offer a structured view of the field, we introduced the taxonomy illustrated in Figure 4, along the reward and utility axes [30], differentiating between *individual* and *team rewards*, as well as *individual*, *team* and *social choice utility*. By characterising problems along both these dimensions, we can obtain a better understanding of the appropriate methods and solution concepts.

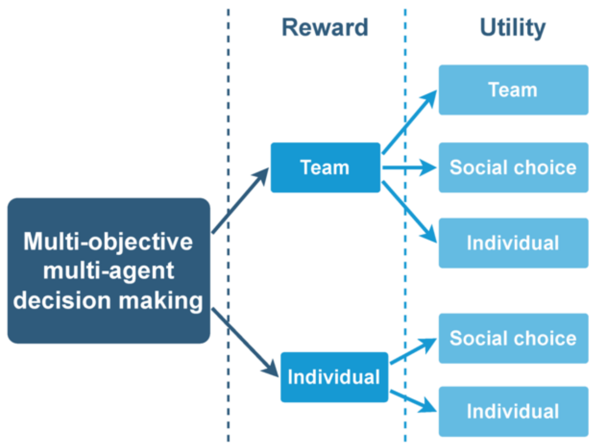

Figure 4. Multi-objective multi-agent decision-making taxonomy, in terms of the reward and utility functions [30].

The multi-objective decision-making literature [14, 38] discusses two distinct perspectives for defining solutions in multi-objective settings: the axiomatic and the utility-based paradigms. We briefly discuss each perspective below, tailored to the multi-objective multi-agent setting.

Axiomatic approach

The axiomatic approach assumes Pareto dominance as the optimality criterion and designates the Pareto set (PS)[1] as the optimal solution set, under the minimal assumption that the utility function is a monotonically increasing function. Informally, Pareto dominance introduces a partial ordering over vectors, where one vector is preferred over another if it is at least equal on all objectives and strictly better on at least one.

Team reward setting When agents cooperate, they often share a team reward and value, i.e. $\boldsymbol{v}_1^{\boldsymbol{\pi}} = \boldsymbol{v}_2^{\boldsymbol{\pi}} = \ldots = \boldsymbol{v}_n^{\boldsymbol{\pi}}$, denoted as $\boldsymbol{v}^{\boldsymbol{\pi}}$. Given this shared value, Pareto dominance can be straightforwardly applied and we can subsequently define the set of all joint policies which are not Pareto dominated as the *Pareto set*.

Individual reward setting While the Pareto set and Pareto front are natural solutions in cooperative settings, extending this to settings where each agent receives a different reward vector is non-trivial. We note that there is little work so far on the *individual reward* setting with *unknown utility functions*, so this more general setting remains an important open challenge in MOMARL.

Utility-based approach

The utility-based approach advocates for exploiting any additional domain knowledge that might be available for deriving the optimality criterion. This involves knowledge regarding the user's utility function and preferences with respect to the solution characteristics (i.e., stochastic or deterministic policy, single- or multi- policy output) [38]. Such additional knowledge can lead to smaller coverage sets (e.g., if the utility function is known to be linear the convex coverage set can be used, instead of the PS), or less time spent on exploring regions of the objective space that are not of interest to the user (e.g., when the user requires some minimum value for a certain objective). When no additional knowledge is available, the utility-based approach falls back on the axiomatic approach.

Roijers et al. [38] define two optimisation criteria in multi-objective decision-making when applying the utility function to the vector-valued outcomes. One can compute the expected value of the payoffs of a policy first and then apply the utility function, leading to the *scalarised expected returns (SER)* optimisation criterion:

$$v_{u_i}^{\boldsymbol{\pi}} = u_i \left(\mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t \boldsymbol{R}_i(s_t, \mathbf{a}_t, s_{t+1}) \mid \boldsymbol{\pi} \right] \right) \tag{1}$$

where $\boldsymbol{R}_i(s_t, \mathbf{a}_t, s_{t+1})$ is agent i's reward at timestep t, state s_t, joint action $\mathbf{a}_t$ and next state s_{t+1}, γ is the discount factor and u_i is agent i's utility function. $v_{u_i}^{\boldsymbol{\pi}}$ is the scalarised return derived by agent i. Alternatively, under the *expected scalarised returns (ESR)* optimisation criterion [15, 36], the utility function is applied before computing the expectation:

$$v_{u_i}^{\boldsymbol{\pi}} = \mathbb{E} \left[u_i \left(\sum_{t=0}^{\infty} \gamma^t \boldsymbol{R}_i(s_t, \mathbf{a}_t, s_{t+1}) \right) \mid \boldsymbol{\pi} \right] \tag{2}$$

Semantically, the two optimisation criteria distinguish between settings in which users are interested in optimising the utility over multiple policy executions (SER), or over each policy application (ESR). SER and ESR coincide under linear utility functions.

[1] The Pareto front (PF) contains the value vectors corresponding to all policies in the PS.

Taking a utility-based perspective, in Rădulescu et al. [30] we also introduce a mapping of potential solution concepts to each of the identified categories of the MOMADM taxonomy (Figure 5).

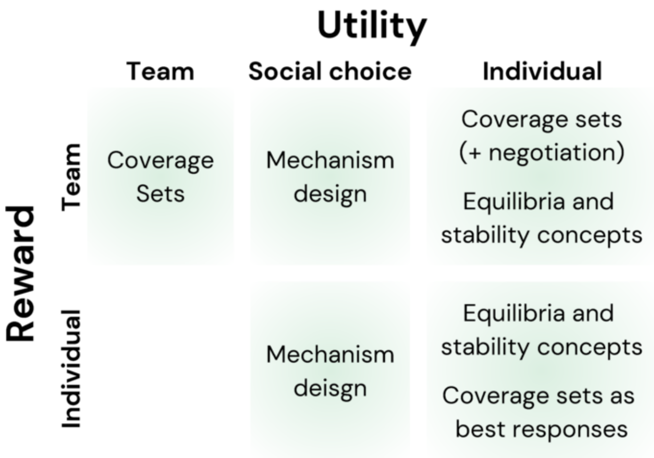

Figure 5. Mapping of solution concepts for the multi-objective multi-agent decision-making taxonomy [30].

Coverage sets Coverage sets (CS) are the appropriate solution concept in cooperative settings, when all agents share the same reward and utility function. In team reward individual utility settings, coverage sets can still be used, with the added step of agents agreeing on the joint policy selection procedure from the CS. In fully individual settings, agents could construct (approximate) coverage set as possible best responses to the behaviours of the others.

Mechanism design The goal here is to guide the system towards solutions that are (approximately) optimal under a social welfare function. This is achieved by designing additional payment mechanisms that incentivise agents to truthfully reveal their utilities.

Equilibria and stability concepts Other suitable solution concepts for fully individual settings are game theoretic equilibria or, if binding agreements are possible, concepts such as coalition formation from cooperative game theory can also apply. These elements need to however be extended to the multi-objective case.

3 Theoretical Contributions

The interplay between the multi-agent and multi-objective dimensions in MOMADM instils additional complexity in comparison to each composing field. As a consequence, the majority of our initial theoretical contributions adopt a utility-based perspective, where we assume each agent possesses a known, non-linear utility function that dictates their preferred trade-offs among objectives.

Equilibria Analysis Using the framework of multi-objective normal-form games, in Rădulescu et al. [31] we explored the impact of SER and ESR on two game theoretic solution concepts, namely Nash equilibria [23] and correlated equilibria [4]. We re-iterate below the extension of the Nash equilibrium (NE) to MOMADM. We define $\boldsymbol{\pi}_{-i} = (\pi_1, \ldots, \pi_{i-1}, \pi_{i+1}, \ldots, \pi_n)$ to be a joint policy without agent's i policy. We can thus write $\boldsymbol{\pi} = (\pi_i, \boldsymbol{\pi}_{-i})$.

Definition 1 (Nash equilibrium). *A joint policy $\boldsymbol{\pi}^{NE}$ is a Nash equilibrium if, for each agent $i \in \{1, \ldots, n\}$ and for any alternative policy π_i, no agent can improve its scalarised return by unilaterally changing its policy:*

$$v_{u_i}^{(\pi_i^{NE}, \boldsymbol{\pi}_{-i}^{NE})} \geq v_{u_i}^{(\pi_i, \boldsymbol{\pi}_{-i}^{NE})}. \tag{3}$$

Note that $v^{\pi}_{u_i}$ can be derived under either SER (Equation 1) or ESR (Equation 2). This initial study showed that the choice of optimisation criterion can radically alter the set of equilibria, under non-linear utility functions. In the ESR case, any MONFG can be reduced to a corresponding single-objective NFG. This guarantees, for example, the existence of at least one NE [24]. In the SER case, on the other hand, Nash equilibria need not exist. This is an important result in MOMADM, signalling that there are cases in which a strong disagreement in preferences among users impedes reaching stable outcomes without additional mechanisms (e.g., negotiation, contractual agreements).

In Röpke et al. [40] we continued this line of work, and investigated the sufficient conditions under which Nash equilibria are guaranteed to exist, under SER with non-linear utilities. Looking at relations between SER and ESR, we also showed that no equilibrium needs to be shared between the two criteria, when NE do exist. This result highlights the importance of choosing the appropriate criterion for the planning or learning phase, since each setting can output different solutions.

Pareto versus utility-based solutions In Mannion and Rădulescu [21] we contrasted the utility-based and axiomatic approaches and investigated the relationship between their corresponding solution sets in multi-objective normal form games (MONFGs), for the setting with team rewards and individual utility. We demonstrated that there are cases in which the set of NE and the PS are disjoint, implying that in these situations Pareto-based approaches will not find stable solutions in the joint strategy space.

Connections to other models Discovering connections or equivalences between seemingly unrelated models can bridge research communities and fill gaps in each field. In Röpke et al. [42] we introduced a novel equivalence class between continuous games and multi-objective normal-form games, and we showed that a pure strategy Nash equilibrium in a continuous game is a mixed strategy Nash equilibrium in an equivalent MONFG. We demonstrated potential algorithmic transfers by learning NE in two continuous games utilising a multi-objective fictitious play algorithm.

From a theoretical perspective, MONFGs stand to benefit from the extensive knowledge on continuous games. For example, this equivalence implies that some MONFGs fall in the category of continuous game without pure strategy NE [12] and thus have no mixed-strategy NE themselves, strengthening the result of [31]. In the context of Stackelberg games, it was showed that commitment can be worse in infinite (single-objective) NFGs than the utility from any NE [52]. Relying on the equivalence relationship, we transferred this results to MONFGs [39].

Next steps The results so far for the known utility case emphasise the importance of studying the alignment between the preferences of the users in multi-objective settings, as this strongly impacts the number and even the existence of stable solutions, as well as the capacity of learning approaches to converge to Pareto optimal outcomes. This will play a critical role in the advancement of decision-support systems across a variety of sectors, such as smart grids, logistics, epidemiology or resource management.

So far we have adopted the stateless MONFG framework for the analyses. Studies should be extended to sequential settings, e.g., multi-objective stochastic games (MOSGs), to better reflect and translate to real-world problem domains. Extending methods such as empirical game theoretic analysis [53] to multi-objective settings will enable insights into more complex interactions.

4 Algorithmic Contributions

Multi-objective multi-agent reinforcement learning (MOMARL) targets complex decision-making tasks that must balance multiple conflicting objectives and coordinate the actions of various (independent) decision-makers. Figure 6 illustrates the MOMARL interaction loop, where a set of agents observe (part of) the environment's state, take a joint action and receive their corresponding vectorial reward functions, under one of the frameworks described in Figure 3 (e.g., in the most general case, the MOPOSG; in team reward and fully observable settings, the MOMMDP). Each component of the reward vector represents the feedback signal for a different objective. In our work, MOMARL represents a core framework for enabling autonomous agents to learn policies in multi-objective multi-agent decision making settings.

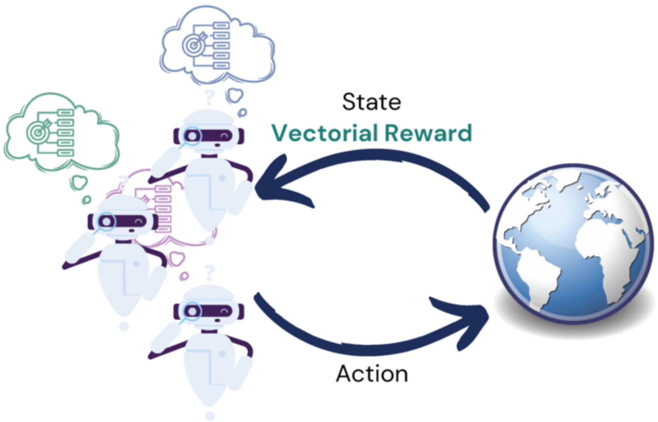

Figure 6. Illustration of the multi-objective multi-agent reinforcement learning interaction loop.

Our initial algorithmic approaches for MOMARL were designed to empirically validate the theoretical contributions. For example, Röpke et al. [39] uses independent multi-objective actor-critic learners, where all agents have non-linear utilities and optimise for the SER criterion. We also incorporate mechanisms such as communication, either by receiving action recommendations to allow agents to coordinate their strategies and reach correlated equilibria [31], or by allowing players to communicate preferences over their actions [41].

In Rădulescu et al. [32] we investigate the impact of opponent modelling in MONFGs, under SER, with non-linear utilities. We build on advances from the multi-agent learning literature, more specifically, on the idea of learning with opponent learning awareness [11] (i.e., anticipate one's impact on the opponent's learning step). Modelling the opponents' learning step is not straightforward in multi-objective settings, since the learning direction is defined by the opponents' utility, which is private information. The key idea behind our opponent learning awareness method is to train a Gaussian process [34] as an estimator for the opponents' learning step, removing the need to have access to the opponent's policy or utility.

Finally, in Felten et al. [10] we introduce MOMAPPO, an extension of the multi-agent proximal policy optimisation algorithm [54] to return a Pareto set of multi-agent policies in team reward settings.

Next steps From an algorithmic standpoint, few solving methods address both dimensions of MOMARL in complex settings (e.g., high dimensional action or state spaces, sequential settings). Using developments from multi-objective and multi-agent reinforcement learning will offer strong foundations for creating novel approaches. Research on the axiomatic approach to MOMADM (i.e., in cases in which the utility function is not available) is also in early stages, with work mainly considering the team reward case (e.g., [10, 16]). Defining solution concepts for the individual reward setting and designing methods to identify them remains an open challenge.

As mentioned in Section 1, and sketched in Figure 1, MOMAS is a holistic model for capturing interactions in hybrid human-AI collectives. I believe MOMARL methods will become an important component for tackling decision making tasks in these settings, when focusing on interactions among autonomous agents. Nevertheless, additional inter- and transdisciplinary efforts are necessary to fill in gaps, especially in terms of human - AI interaction [6]. For example, in the case of decision delegation, we can outline at least two important phases: preference elicitation, and, in situations in which decision oversight is necessary, solution selection. Each of these phases will require many components ranging from interface design and bi-directional communication protocols [5], efficiently capturing user preferences, rapid adjustment in the case of dynamic preferences, ability to identify missing or overlooked dimensions, clear visual and verbal expositions of identified solutions. Each component should be crafted according to responsible AI principles [3], such as transparency, explainability, fairness, and privacy, in order to augment the user's knowledge and agency [19], and allow for better informed decisions.

5 Benchmarks

Benchmarks are crucial for sustaining progress, enabling evaluation, and ensuring reproducibility of reinforcement learning methods. To support the advancement of the MOMARL field, in Felten et al. [10] we introduce MOMALAND[2], the first collection of standardised environments for multi-objective multi-agent reinforcement learning, with problems varying in the number of agents, state and action spaces, reward structures, and utility considerations.

MOMALAND is developed within the library ecosystem of the Farama Foundation (Figure 7). Through its scalarisation wrappers, MOMALAND enables the conversion of multi-objective environments into single-objective ones under the standard PettingZoo API [48]. This adaptation allows the usage of multi-agent RL algorithms to learn policies for a designated trade-off. The centralisation wrapper provides a direct conversion to the MO-Gymnasium API [1]. This adaptation enables learning using multi-objective single-agent algorithms, such as those featured in MORL-Baselines [9]. MOMALAND currently provides over a dozen environments, covering the majority of frameworks presented in Figure 3.

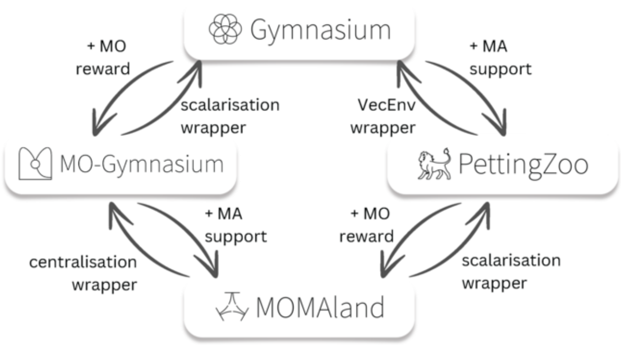

Figure 7. MOMAland within the Farama Foundation library ecosystem [10].

[2] Documentation: https://momaland.farama.org/

Next steps MOMALAND represents an important milestone for establishing community-wide standards in terms of evaluation and progress tracking for MOMARL. But it is also an invitation for additional problem proposals, especially from real-world or industrial-inspired settings.

MOMALAND is currently restricted to problems involving interactions among autonomous agents. We note that some of the environments would lend themselves well to hybrid or human-in-the-loop settings [35], so extensions and studies in this directions are another avenue for future work. Coupled with the research agenda outlined in Section 4, we hope that MOMALAND will provide prolific support for MOMADM, as well as for the development of hybrid human-AI interactions within the MOMADM framework.

6 Broader Applicability

Significant opportunities exist to re-examine problems that were initially modelled as single-objective multi-agent decision problems using a multi-objective perspective. This could for example provide richer solution sets for cooperative multi-agent system, provide more insight into the collective versus individual tensions in social dilemmas, or improve performance by considering additional objectives to represent sub-tasks explicitly, for example, using concepts such as curiosity or intrinsic rewards in MARL [45].

For example, in Orzan et al. [25] we take a multi-objective perspective on a specific class of social dilemmas, namely Public Goods Games under uncertainty [26]. In this setting, the payoffs of the players can naturally be separated in two components: the payoff obtained by equally splitting the total resulting common good (denoted as the collective payoff – this component is characterised by uncertainty due to noisy observations and dependence on the behaviour of other agents), and the payoff resulting from the remaining personal endowment after the player contributes or not to the common good (denoted as the individual payoff). Traditionally, these components are simply summed up to form the final payoff. We refined the decision-making process by explicitly considering the collective and the individual payoffs as separate objectives for each agent. This allows us to model the valuation of the riskier component – the collective payoff – at an individual level, creating *risk-seeking*, *risk-averse* or *risk-neutral* agents. We model these attitudes using a parametric non-linear utility function. We show that the presence of risk-seeking agents can increase cooperation, while a population with heterogeneous risk attitudes can fail to reach cooperation even in cooperative settings.

Another potential direction is to use multi-objectivisation to improve team behaviour through social welfare. Additionally, multi-objective RL techniques can be used to develop agents capable of adopting a range of different behaviours during deployment (e.g. co-operative, competitive) [17], or to create populations of agents that develop effective behaviours against a large range of opponents [20, 28].

7 Conclusions

I started by advocating for multi-objective multi-agent systems as more realistic and encompassing models for decision-making settings that need to explore trade-offs among multiple objectives, and involve the interaction among multiple stakeholders. I gave a brief overview of the field, discussed our theoretical and algorithmic contributions, and outlined potential next steps.

The field of multi-objective multi-agent decision making presents numerous open challenges across many disciplines. This offers the unique opportunity to create a multidimensional research agenda, that balances between efficient methods for highly-performing systems and responsible design, with a careful consideration toward positive societal impact.

Acknowledgements

Part of the presented research was supported by the Research Foundation – Flanders (FWO), grant number 1286223N.

References

[1] L. N. Alegre, F. Felten, E.-G. Talbi, G. Danoy, A. Nowé, A. L. Bazzan, and B. C. da Silva. MO-Gym: A Library of Multi-Objective Reinforcement Learning Environments. In *Proceedings of the 34th Benelux Conference on Artificial Intelligence BNAIC/Benelearn*, 2022.

[2] T. Araujo, N. Helberger, S. Kruikemeier, and C. H. De Vreese. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & society*, 35(3):611–623, 2020.

[3] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58: 82–115, 2020.

[4] R. J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics*, 1(1):67–96, 1974.

[5] H. Baier and M. Kaisers. Explainable search. In *IJCAI-PRICAI Workshop on Explainable Artificial Intelligence (Vol. 178)*, 2020.

[6] D. Dell'Anna, P. K. Murukannaiah, B. Dudzik, D. Grossi, C. M. Jonker, C. Oertel, and P. Yolum. Toward a quality model for hybrid intelligence teams. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, pages 434–443, 2024.

[7] P. Esmaeilzadeh. Use of ai-based tools for healthcare purposes: a survey study from consumers' perspectives. *BMC medical informatics and decision making*, 20:1–19, 2020.

[8] S. European Commission et al. Ethics guidelines for trustworthy AI. *Publications Office*, 2019.

[9] F. Felten, L. N. Alegre, A. Nowe, A. L. C. Bazzan, E. G. Talbi, G. Danoy, and B. C. d. Silva. A Toolkit for Reliable Benchmarking and Research in Multi-Objective Reinforcement Learning. In *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

[10] F. Felten, U. Ucak, H. Azmani, G. Peng, W. Röpke, H. Baier, P. Mannion, D. M. Roijers, J. K. Terry, E.-G. Talbi, G. Danoy, A. Nowé, and R. Rădulescu. MOMAland: A set of benchmarks for multi-objective multi-agent reinforcement learning, 2024. URL https://arxiv.org/abs/2407.16312.

[11] J. Foerster, R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch. Learning with opponent-learning awareness. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 122–130, 2018.

[12] D. Fudenberg and J. Tirole. *Game theory*. MIT press, 1991.

[13] A. L. Guzman and S. C. Lewis. Artificial intelligence and communication: A human–machine communication research agenda. *New media & society*, 22(1):70–86, 2020.

[14] C. Hayes, R. Rădulescu, E. Bargiacchi, J. Källström, M. Macfarlane, M. Reymond, T. Verstraeten, L. Zintgraf, R. Dazeley, F. Heintz, E. Howley, A. Irissappane, P. Mannion, A. Nowe, G. Ramos, M. Restelli, P. Vamplew, and D. Roijers. A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems*, 36, Apr. 2022. doi: 10.1007/s10458-022-09552-y.

[15] C. F. Hayes, M. Reymond, D. M. Roijers, E. Howley, and P. Mannion. Distributional monte carlo tree search for risk-aware and multi-objective reinforcement learning. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1530–1532, 2021.

[16] T. Hu, B. Luo, C. Yang, and T. Huang. MO-MIX: Multi-Objective Multi-Agent Cooperative Decision-Making With Deep Reinforcement Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):12098–12112, Oct. 2023. ISSN 1939-3539. doi: 10.1109/TPAMI.2023.3283537. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[17] J. Källström and F. Heintz. Tunable dynamics in agent-based simulation using multi-objective reinforcement learning. In *Proceedings of the Adaptive and Learning Agents Workshop (ALA-19) at AAMAS*, May 2019.

[18] S. Larsson and F. Heintz. Transparency in artificial intelligence. *Internet policy review*, 9(2):1–16, 2020.

[19] R. Legaspi, W. Xu, T. Konishi, S. Wada, N. Kobayashi, Y. Naruse, and Y. Ishikawa. The sense of agency in human–ai interactions. *Knowledge-Based Systems*, 286:111298, 2024.

[20] X. Liu, H. Jia, Y. Wen, Y. Hu, Y. Chen, C. Fan, Z. Hu, and Y. Yang. Towards unifying behavioral and response diversity for open-ended learning in zero-sum games. *Advances in Neural Information Processing Systems*, 34:941–952, 2021.

[21] P. Mannion and R. Rădulescu. Comparing utility-based and pareto-based solution sets in multi-objective normal form games. In *Multi-Objective Decision Making Workshop 2023*, 2023.

[22] G. Marcus and E. Davis. *Rebooting AI: Building artificial intelligence we can trust*. Vintage, 2019.

[23] J. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950. ISSN 0027-8424.

[24] J. Nash. Non-cooperative games. *Annals of Mathematics*, 54(2):286–295, 1951.

[25] N. Orzan, E. Acar, D. Grossi, P. Mannion, and R. Rădulescu. Learning in public goods games with non-linear utilities: a multi-objective approach. In *Proceedings of the 27th European Conference on Artificial Intelligence (ECAI 2024)*, 2024.

[26] N. Orzan, E. Acar, D. Grossi, and R. Rădulescu. Emergent cooperation under uncertain incentive alignment. In *Proceedings of the 2024 International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024)*, 2024.

[27] V. Pereira, E. Hadjielias, M. Christofi, and D. Vrontis. A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective. *Human Resource Management Review*, 33(1):100857, 2023.

[28] N. Perez-Nieves, Y. Yang, O. Slumbers, D. H. Mguni, Y. Wen, and J. Wang. Modelling behavioural diversity for learning in open-ended games. In *International conference on machine learning*, pages 8514–8524. PMLR, 2021.

[29] M. L. Puterman. Markov decision processes. *Handbooks in operations research and management science*, 2:331–434, 1990.

[30] R. Rădulescu, P. Mannion, D. M. Roijers, and A. Nowé. Multi-objective multi-agent decision making: a utility-based analysis and survey. *Autonomous Agents and Multi-Agent Systems*, 34(1):1–52, 2020.

[31] R. Rădulescu, P. Mannion, Y. Zhang, D. M. Roijers, and A. Nowé. A utility-based analysis of equilibria in multi-objective normal-form games. *The Knowledge Engineering Review*, 35:e32, 2020.

[32] R. Rădulescu, T. Verstraeten, Y. Zhang, P. Mannion, D. M. Roijers, and A. Nowé. Opponent learning awareness and modelling in multi-objective normal form games. *Neural Computing and Applications*, pages 1–23, 2021.

[33] G. de. O. Ramos, R. Rădulescu, A. Nowé, and A. R. Tavares. Toll-based learning for minimising congestion under heterogeneous preferences. In B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, and G. Sukthankar, editors, *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020)*, pages 1098–1106, Auckland, New Zealand, May 2020. IFAAMAS.

[34] C. E. Rasmussen and M. Kuss. Gaussian processes in reinforcement learning. *Advances in Neural Information Processing Systems*, 16:751–758, 2003.

[35] C. O. Retzlaff, S. Das, C. Wayllace, P. Mousavi, M. Afshari, T. Yang, A. Saranti, A. Angerschmid, M. E. Taylor, and A. Holzinger. Human-in-the-loop reinforcement learning: A survey and position on requirements, challenges, and opportunities. *Journal of Artificial Intelligence Research*, 79:359–415, 2024.

[36] M. Reymond, C. F. Hayes, D. Steckelmacher, D. M. Roijers, and A. Nowé. Actor-critic multi-objective reinforcement learning for non-linear utility functions. *Autonomous Agents and Multi-Agent Systems*, 37 (2):23, Apr. 2023. ISSN 1573-7454. doi: 10.1007/s10458-023-09604-x. URL https://doi.org/10.1007/s10458-023-09604-x.

[37] M. Reymond, C. F. Hayes, L. Willem, R. Rădulescu, S. Abrams, D. M. Roijers, E. Howley, P. Mannion, N. Hens, A. Nowé, et al. Exploring the pareto front of multi-objective covid-19 mitigation policies using reinforcement learning. *Expert Systems with Applications*, 249:123686, 2024.

[38] D. M. Roijers, P. Vamplew, S. Whiteson, and R. Dazeley. A Survey of Multi-Objective Sequential Decision-Making. *Journal of Artificial Intelligence Research*, 48:67–113, Oct. 2013. ISSN 1076-9757. doi: 10.1613/jair.3987. URL http://arxiv.org/abs/1402.0590. arXiv: 1402.0590.

[39] W. Röpke, R. Radulescu, A. Nowé, and D. M. Roijers. Commitment and cyclic strategies in multi-objective games. In *Proceedings of the Adaptive and Learning Agents Workshop (ALA 2022), Francisco Cruz, Conor F. Hayes, Felipe Leno da Silva, and Fernando P. Santos (Eds.). Online,\url {https://ala2022. github. io/}*, volume 9, 2022.

[40] W. Röpke, D. M. Roijers, A. Nowé, and R. Rădulescu. On nash equilibria

in normal-form games with vectorial payoffs. *Autonomous Agents and Multi-Agent Systems*, 36(2):53, Oct. 2022. ISSN 1573-7454. doi: 10.1007/s10458-022-09582-6.

[41] W. Röpke, D. M. Roijers, A. Nowé, and R. Rădulescu. Preference communication in multi-objective normal-form games. *Neural Computing and Applications*, pages 1–26, 2022.

[42] W. Röpke, C. Groenland, R. Rădulescu, A. Nowé, and D. M. Roijers. Bridging the gap between single and multi objective games. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, pages 224–232, 2023.

[43] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach (Third edition)*. Pearson Education, Inc., 2010.

[44] F. P. Santos. Prosocial dynamics in multiagent systems. *AI Magazine*, 45(1):131–138, 2024.

[45] L. Schäfer and S. Albrecht. *Curiosity in Multi-Agent Reinforcement Learning*. PhD thesis, University of Edinburgh, 08 2019.

[46] K. Seo, J. Tang, I. Roll, S. Fels, and D. Yoon. The impact of artificial intelligence on learner–instructor interaction in online learning. *International journal of educational technology in higher education*, 18:1–23, 2021.

[47] I. Terrucha, E. Fernández Domingos, P. Simoens, and T. Lenaerts. Committing to the wrong artificial delegate in a collective-risk dilemma is better than directly committing mistakes. *Scientific reports*, 14(1):10460, 2024.

[48] J. Terry, B. Black, N. Grammel, M. Jayakumar, A. Hari, R. Sullivan, L. S. Santos, C. Dieffendahl, C. Horsch, R. Perez-Vicente, et al. Pettingzoo: Gym for multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 34:15032–15043, 2021.

[49] Y. Tian, Y. Feng, C. Wang, R. Cao, X. Zhang, X. Pei, K. C. Tan, and Y. Jin. A large-scale combinatorial many-objective evolutionary algorithm for intensity-modulated radiotherapy planning. *IEEE transactions on evolutionary computation*, 26(6):1511–1525, 2022.

[50] K. Tuyls and G. Weiss. Multiagent learning: Basics, challenges, and prospects. *Ai Magazine*, 33(3):41–41, 2012.

[51] P. Vamplew, R. Dazeley, C. Foale, S. Firmin, and J. Mummery. Human-aligned artificial intelligence is a multiobjective problem. *Ethics and Information Technology*, 20(1):27–40, 2018.

[52] B. Von Stengel and S. Zamir. Leadership games with convex strategy sets. *Games and Economic Behavior*, 69(2):446–457, 2010.

[53] M. P. Wellman. Methods for empirical game-theoretic analysis. In *AAAI*, volume 980, pages 1552–1556, 2006.

[54] C. Yu, A. Velu, E. Vinitsky, J. Gao, Y. Wang, A. Bayen, and Y. Wu. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, 35:24611–24624, 2022.