

# Abstract: A Practical Guide to Multi-Objective Reinforcement Learning and Planning<sup>\*</sup>

Conor F. Hayes, Roxana Rădulescu, Eugenio Bargiacchi, Johan Källström, Matthew Macfarlane, Mathieu Reymond, Timothy Verstraeten, Luisa M. Zintgraf, Richard Dazeley, Fredrik Heintz, Enda Howley, Athirai A. Irissappane, Patrick Mannion, Ann Nowé, Gabriel Ramos, Marcello Restelli, Peter Vamplew, and Diederik M. Roijers

NUI Galway (IE), Vrije Universiteit Brussel (BE), Linköping University (SE), University of Amsterdam (NL), University of Oxford (GB), Deakin University (AU), University of Washington (US), Universidade do Vale do Rio dos Sinos (BR), Politecnico di Milano (IT), Federation University (AU), and HU University of Applied Sciences Utrecht (NL)

Corresponding: [diederik.roijers@vub.be](mailto:diederik.roijers@vub.be)

**Keywords:** Multi-objective decision making · Multi-objective reinforcement learning · Multi-objective planning · Multi-objective multi-agent systems

**Abstract.** Real-world sequential decision-making tasks are usually complex, and require trade-offs between multiple – often conflicting – objectives. However, the majority of research in reinforcement learning (RL) and decision-theoretic planning assumes a single objective, or that multiple objectives can be handled via a predefined weighted sum over the objectives. Such approaches may oversimplify the underlying problem, and produce suboptimal results.

This article serves as a guide to the application of explicitly multi-objective methods to difficult problems, and is aimed at researchers who are already familiar with single-objective RL and planning methods who may wish to adopt a multi-objective perspective, as well as practitioners who may encounter multi-objective decision problems. It identifies the factors that may influence the nature of the desired solution, and illustrates by example how these influence the design of multi-objective decision-making systems for complex problems.

## 1 Introduction

Real-world sequential decision-making tasks can have multiple, often conflicting, objectives [3,5,6,14,1]. Reinforcement learning (RL) and decision-theoretic planning have been used extensively to solve sequential decision making problems by maximising a scalar reward signal [9]. In settings with multiple objectives, RL

---

<sup>\*</sup> Original article published in JAAMAS, April 13, 2022 [4].

approaches assume the objectives can be combined into a single scalar reward using a predefined weighted sum. An iterative process is used to tune the weights for each objective. During learning the algorithm is tuned, turned on, then the reward function is re-engineered until the behaviour is satisfactory. However, such an approach may produce suboptimal results in practical settings [10,2,13].

In this work, we argue an iterative process is problematic for a number of reasons: (a) it is a semi-blind manual process, (b) it prevents people who should take the decisions from making well-informed trade-offs, (c) it damages the explainability of the decision-making process, (d) it cannot handle different types of preferences that human decision makers might actually have, and finally (e) preferences between the objectives may change over time and a single objective agent will have to be retrained when this happens.

Let’s briefly consider reason (b), given the reward function needs to be engineered a priori, there is uncertainty as to the effects a reward function may have on the policy. For example, when training an agent in a power production system, we may wish to double the average power output. However, even if the objectives are linearly weighted in the reward function, simply doubling the reward associated with power output may not be sufficient to achieve the desired behaviour, given the relationship between the reward weights and the objective outcomes may be nonlinear [11]. Generally, an AI engineer is tasked with adjusting the associated weights for the reward function. Therefore, the decision power is put where it does not belong: with the AI engineers. By leaving such decisions up to AI engineers, they are effectively making assumptions about the preferences of the actual decision makers. In practical settings this is not a responsibility that can be left to AI engineers. By taking an explicitly multi-objective approach it is possible to remove such responsibilities from the AI engineer. Multi-objective algorithms can be used to compute all possibly optimal policies [7,12,15,8], where the computed policies can be inspected by a system expert before making a decision.

Considering the reasons outlined above, a multi-objective approach to decision making is necessary in many practical settings. In this work we outline why taking an explicitly multi-objective approach to planning and learning may be essential to deploying AI in decision problems. Moreover, this article provides a detailed introduction to multi-objective decision making and guides the reader through getting started with modelling and solving such decision problems.

**Acknowledgements** This research was supported by the following organisations and grants: FWO #1SA2820N, the Flemish Government “Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen”, the National University of Ireland Galway Hardiman Scholarship, FAPERGS grant 19/2551-0001277-2, FAPESP grant 2020/05165-1, the Swedish Governmental Agency for Innovation Systems NFFP7/2017-04885, the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program (WASP) (the Knut and Alice Wallenberg Foundation), NWO LIFT-project 019.011, the 2017 Microsoft Research PhD Scholarship Program, and the 2020 Microsoft Research EMEA PhD Award.

## References

1. Bryce, D., Cushing, W., Kambhampati, S.: Probabilistic planning is multi-objective. Arizona State University, Tech. Rep. ASU-CSE-07-006 (2007)
2. Brys, T., Van Moffaert, K., Van Vaerenbergh, K., Nowé, A.: On the behaviour of scalarization methods for the engagement of a wet clutch. In: 2013 12th International Conference on Machine Learning and Applications. vol. 1, pp. 258–263. IEEE (2013)
3. Castelletti, A., Pianosi, F., Restelli, M.: A multiobjective reinforcement learning approach to water resources systems operation: Pareto frontier approximation in a single run. *Water Resources Research* **49**(6), 3476–3486 (2013)
4. Hayes, C.F., Rădulescu, R., Bargiacchi, E., Källström, J., Macfarlane, M., Raymond, M., Verstraeten, T., Zintgraf, L.M., Dazeley, R., Heintz, F., Howley, E., Irissappane, A.A., Mannion, P., Nowé, A., Ramos, G., Restelli, M., Vamplew, P., Roijers, D.M.: A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems* **36**(1), 1–59 (2022)
5. Jalalimanesh, A., Haghighi, H.S., Ahmadi, A., Hejzian, H., Soltani, M.: Multi-objective optimization of radiotherapy: distributed q-learning and agent-based simulation. *Journal of Experimental & Theoretical artificial intelligence* **29**(5), 1071–1086 (2017)
6. Pianosi, F., Castelletti, A., Restelli, M.: Tree-based fitted Q-iteration for multi-objective Markov decision processes in water resource management. *Journal of Hydroinformatics* **15**(2), 258–270 (2013)
7. Roijers, D.M., Vamplew, P., Whiteson, S., Dazeley, R.: A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research* **48**, 67–113 (2013)
8. Roijers, D.M., Whiteson, S., Oliehoek, F.A.: Computing convex coverage sets for faster multi-objective coordination. *Journal of Artificial Intelligence Research* **52**, 399–443 (2015)
9. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT press (2018)
10. Vamplew, P., Yearwood, J., Dazeley, R., Berry, A.: On the limitations of scalarisation for multi-objective reinforcement learning of Pareto fronts. In: Australasian Joint Conference on Artificial Intelligence. pp. 372–378. Springer (2008)
11. Van Moffaert, K., Brys, T., Chandra, A., Esterle, L., Lewis, P.R., Nowé, A.: A novel adaptive weight selection algorithm for multi-objective multi-agent reinforcement learning. In: 2014 International joint conference on neural networks (IJCNN). pp. 2306–2314. IEEE (2014)
12. Van Moffaert, K., Nowé, A.: Multi-objective reinforcement learning using sets of pareto dominating policies. *The Journal of Machine Learning Research* **15**(1), 3483–3512 (2014)
13. Van Vaerenbergh, K., Rodríguez, A., Gagliolo, M., Vrancx, P., Nowé, A., Stoev, J., Goossens, S., Pinte, G., Symens, W.: Improving wet clutch engagement with reinforcement learning. In: The 2012 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2012)
14. Verstraeten, T., Nowé, A., Keller, J., Guo, Y., Sheng, S., Helsen, J.: Fleetwide data-enabled reliability improvement of wind turbines. *Renewable and Sustainable Energy Reviews* **109**, 428–437 (2019). <https://doi.org/https://doi.org/10.1016/j.rser.2019.03.019>
15. White, D.: Multi-objective infinite-horizon discounted markov decision processes. *Journal of mathematical analysis and applications* **89**(2), 639–647 (1982)