Causal Learning and Reasoning in Multi-Agent Reinforcement Learning

Giovanni Briglia*
Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia
Reggio Emilia, Italy
giovanni.briglia@unimore.it

ABSTRACT

Reinforcement Learning (RL) has achieved remarkable success in sequential decision-making tasks, particularly under restrictive conditions such as full observability, on-policy interaction, and online learning. However, outside these conditions, hence especially in Multi-Agent (MA) domains, policies become brittle to distribution shifts and confounding, leading to poor generalization and limited transferability. Motivated by these limitations, my PhD project aims to identify *when* and *why* causal learning and reasoning are necessary or desirable in (MA)RL, and *how* they can be systematically integrated within (MA)RL methods to improve robustness, efficiency, and interpretability of the decision-making process.

KEYWORDS

Causal Reasoning, Causal Inference, Reinforcement Learning, Multi-Agent Reinforcement Learning, Causal Reinforcement Learning

ACM Reference Format:
Giovanni Briglia. 2026. Causal Learning and Reasoning in Multi-Agent Reinforcement Learning. 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/RZTO8815

1 INTRODUCTION AND MOTIVATION

Artificial Intelligence (AI) continues to expand the capabilities of artificial agents operating in real-world environments, driven by advances in areas such as Multi-Agent Systems (MAS) [44], Planning [35], Deep Learning [13], Reinforcement Learning (RL) [43], and more recently, Agentic AI [41].

Despite these advances, it has become increasingly clear that many AI systems still lack a genuine capacity to understand and reason about complex decision-making tasks. Current approaches excel at recognizing patterns, however often fail to properly generalize and transfer their expertise across domains and tasks, as they lack the ability to robustly reason about the underlying mechanisms that generate those patterns. As a result, a growing body of work has identified *causality*—the ability to reason about cause–effect relationships—as a key scientific direction for advancing AI beyond pattern recognition towards more robust, adaptable, and interpretable systems [21, 33].

Along these lines, there is evidence that human-level intelligence fundamentally requires explicit causal reasoning, and that many of the core shortcomings of AI systems stem from the absence of principled causal understanding and reasoning[1] [30, 37]. Thus, the limitations of contemporary AI systems are not merely algorithmic, but reflect a deeper conceptual gap: the lack of mature formal and practical tools for systematically reasoning beyond statistical association and embrace causality.

Accordingly, in my PhD project the overarching goal is to exploit causal learning and reasoning to improve how software agents (learn to) behave and take decisions. Here, the notion of causality is grounded in Pearl's causal modeling framework [31], which provides a rigorous mathematical foundation for representing cause–effect relationships. Central to this framework are *Structural Causal Models* (SCMs) [29], which enable reasoning beyond correlations through explicit interventions, formalized by the *do*-operator [27]. This interventional perspective is naturally situated within Pearl's *Ladder of Causality* [32], which organizes causal reasoning into a hierarchy of increasing expressive power—*association*, *intervention*, and *counterfactuals*—thereby clarifying what it means to understand, act, and generalize in complex decision-making environments.

Within Pearl's causal hierarchy, *under specific conditions* [38], the data-generation process satisfies *interventional* causal semantics, rather than mere observational, since actions generated by the policy correspond to interventions in the environment. As a consequence, in these conditions, RL goes beyond purely statistical learning toward causal reasoning—albeit typically without an explicit representation of causal structure. Accordingly, in my PhD project (MA)RL will be the main technique considered, and complemented by causal learning and reasoning, for the construction of AI agents.

Causal Reinforcement Learning [4] arises from the need to endow RL with causal semantics, aligning policy learning and evaluation with reasoning about the underlying data-generating process, and addressing challenges such as: (i) *Generalization and Transferability:* RL policies are brittle under distribution shifts and unobserved confounders because action/value functions encode correlations rather than causal structure [10, 11, 45]; richer forms of experience are required to scale intelligence beyond current limits [42]. (ii) *Efficiency:* RL typically requires large amounts of interaction data and prolonged exploration, especially in high-dimensional or constrained environments, since most methods fail to exploit structural knowledge of the underlying agent-environment dynamics [2, 14, 16]. (iii) *Interpretability:* RL agents increasingly rely on deep neural networks, making their learnt policies difficult to explain

*Also with Department of Computer Science, University of Pisa.

[1] https://www.youtube.com/watch?v=JJ9PQTYciZU

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/RZTO8815

or certify, while post-hoc explanations rarely expose the causal mechanisms governing behavior [11, 22, 34].

Within this context, Causal RL can find the sweet spot between *model-free* and *model-based* approaches to build autonomous systems/agents, as motivated also in [12]. *Model-free* methods learn policies or value functions directly from data, offering scalability and flexibility in high-dimensional settings, but often relying on implicit representations that limit systematic reasoning and transfer. In contrast, *model-based* approaches explicitly learn or assume a model of the environment's dynamics, enabling structured planning and reasoning, albeit at the cost of stronger modeling assumptions and reduced scalability in complex domains.

Motivated by this tension, my PhD project aims at systematically investigating *where* the synergy between causal learning and reasoning and (MA)RL is most effective, especially within multi-agent settings, where these challenges are exacerbated by the presence of other agents as adaptive decision-makers whose actions shape state transitions, rewards, and observations for everyone. The final purpose of this research is to improve robustness, efficiency, and interpretability of the decision-making process, enabling agents to reason about cause–effect relationships rather than merely exploiting statistical regularities.

2 PRELIMINARY RESULTS

My PhD studies aim to focus on multi-agent RL, for which studying the single-agent setting is a necessary preparatory step to deal with the additional challenges arising in MARL.

RL. In [7], I integrated explicit causal reasoning into single-agent RL with the aim of *improving the efficiency and safety of action selection*, especially during exploration. Causal knowledge is encoded as a Bayesian Network (BN), implemented in `pgmpy` [1], and learned from state–action–reward triplets. At decision time, the agent, through the BN, conditions the current observation and performs interventional queries on all candidate actions. The resulting interventional reward posteriors are used to construct an *action mask* that prunes low-value or risky actions and prioritizes those with higher expected utility. RL is conducted online using off-policy algorithms. In contrast, the causal model is learned either online, from interactions collected in the same environment as RL, or offline, from data generated in a simplified environment distinct from the RL setting. Empirical results show a substantial improvement in sample efficiency, enable *safer exploration*, and favors *more effective exploitation.* Furthermore, transferring both the learned policy and the offline causal model to a related but more challenging environment demonstrates improved generalization.

In [9], I shifted from external causal action filtering [7] to embedding causal reasoning within the policy itself. Considering the modular structure of actor–critic framework [19], I replaced the (neural) critic with my first attempt to build an efficient SCM-based representation: a *Vectorized Bayesian Network* (VBN) of the reward function. VBN can handle continuous data and parallel queries. Experiments show *more efficient and accurate value function learning* in the online setting and *higher sample-efficiency.*

MARL. Building on the foundation of [7], my second contribution extends causal action filtering to MARL [8]. In this work, I considered cooperative and semi-cooperative settings with continuous state and reward spaces and discrete action spaces, and causal reasoning is introduced through *independent causal augmentation*: each agent learns its own causal model from local interaction data and applies the same interventional action-filtering mechanism used in the single-agent case [7]. Results show that independent causal augmentation improves sample efficiency and safety in semi-cooperative scenarios. In contrast, gains diminish in strongly cooperative tasks, where agents must explicitly reason about the causal influence of other agents' actions. This behavior is expected, as independent augmentation ignores inter-agent causal dependencies and influences. Crucially, these findings allow us to precisely identify the limitations of naive causal extensions in MARL and clarify when and why richer, explicitly multi-agent causal representations are required.

3 ON-GOING & FUTURE WORKS

Building on these contributions, open challenges emerge guiding my ongoing and future research.

First, practical **implementation**s of explicit SCMs, such as with Causal Bayesian Networks [26, 28] still represent a computational bottleneck. Existing implementations [1, 5, 6] are limited in scalability, have trouble handling continuous data efficiently, and are hardly parallelizable. To overcome these limitations, I am developing VBN—already validated in [9]—a scalable causal modeling tool that enables efficient, parallel causal inference and seamless integration with learning-based systems.

Second, the **foundations** of causal learning and reasoning in MARL are not yet as clear as for single-agent RL. Recent work started to address the issue [12, 24, 36, 37], however, my aim is to expand the limited set of scenarios analyzed in [39]. Classical RL is successful in restricted regimes where key causal assumptions are implicitly satisfied [38]; outside these conditions—such as partial observability, distribution shift, off-policy data, or multi-agent interaction—standard (MA)RL methods often struggle. Especially in MARL, the presence of multiple adaptive decision-makers introduces additional challenges that cannot be captured by independent learning alone [15], making MARL the most challenging, under-explored, and compelling regime. Therefore, characterizing *when* these implicit assumptions break down, and *why* purely associative learning becomes insufficient, is a key aspect of this study. Relevant works in this context, outside (MA)RL, are [17, 18].

The **integration** of causal learning and reasoning into (MA)RL remains an open challenge, with several existing approaches providing important reference points [3, 20, 23, 25, 40, 46]. In my early works, I studied causal reasoning as an external action-filtering mechanism [7, 8], which, while effective, limits scalability. Afterwards, I worked towards integrating causality directly into learning through *the estimation of the causal value function* [9]. Building on this direction, I plan to extend these ideas to MARL, also at the action selection level, to improve stability, generalization, and robustness.

REFERENCES

[1] Ankur Ankan and Johannes Textor. 2024. pgmpy: A python toolkit for bayesian networks. *Journal of Machine Learning Research* 25, 265 (2024), 1–8.

[2] Oron Anschel, Nir Baram, and Nahum Shimkin. 2017. Averaged-DQN: Variance Reduction and Stabilization for Deep Reinforcement Learning. In *Proceedings of the 34th International Conference on Machine Learning, ICML*, Vol. 70. PMLR, 176–185.

[3] Elias Bareinboim, Andrew Forney, and Judea Pearl. 2015. Bandits with Unobserved Confounders: A Causal Approach. In *Advances in Neural Information Processing Systems 28, NeurIPS.* 1342–1350.

[4] Elias Bareinboim, Jiji Zhang, and Sanghack Lee. 2021. An introduction to causal reinforcement learning. *arXiv preprint arXiv:2101.06498* (2021).

[5] Paul Beaumont, Ben Horsburgh, Philip Pilgerstorfer, Angel Droth, Richard Oentaryo, Steven Ler, Hiep Nguyen, Gabriel Azevedo Ferreira, Zain Patel, and Wesley Leong. 2021. *CausalNex.* https://github.com/quantumblacklabs/causalnex

[6] Patrick Blöbaum, Peter Götz, Kailash Budhathoki, Atalanti A Mastakouri, and Dominik Janzing. 2024. DoWhy-GCM: An extension of DoWhy for causal inference in graphical causal models. *Journal of Machine Learning Research* 25, 147 (2024), 1–7.

[7] Giovanni Briglia, Marco Lippi, Stefano Mariani, and Franco Zambonelli. 2024. Improving Reinforcement Learning-Based Autonomous Agents with Causal Models. In *25th International Conference on Principles and Practice of Multi-Agent Systems PRIMA (Lecture Notes in Computer Science, Vol. 15395).* Springer, 267–283.

[8] Giovanni Briglia, Stefano Mariani, and Franco Zambonelli. 2025. Towards Safe Action Policies in Multi-robot Systems with Causal Reinforcement Learning. In *Agents and Robots for reliable Engineered Autonomy.* Springer Nature Switzerland, 51–71.

[9] Giovanni Briglia, Stefano Mariani, and Franco Zambonelli. 2026. Causal Models Improve Reinforcement Learning for Pervasive and Robotic Tasks. In *2026 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops).* IEEE Computer Society. In press.

[10] Karl Cobbe, Oleg Klimov, Christopher Hesse, Taehoon Kim, and John Schulman. 2019. Quantifying Generalization in Reinforcement Learning. In *Proceedings of the 36th International Conference on Machine Learning, ICML (Proceedings of Machine Learning Research, Vol. 97).* PMLR, 1282–1289.

[11] Gabriel Dulac-Arnold, Daniel Mankowitz, and Todd Hester. 2019. Challenges of real-world reinforcement learning. *arXiv preprint arXiv:1904.12901* (2019).

[12] Samuel J Gershman. 2017. 17 Reinforcement Learning and Causal Models. *The Oxford handbook of causal reasoning* (2017), 295.

[13] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning.* MIT Press. http://www.deeplearningbook.org.

[14] Evan Greensmith, Peter L. Bartlett, and Jonathan Baxter. 2004. Variance Reduction Techniques for Gradient Estimates in Reinforcement Learning. *Journal of Machine Learning Research* 5 (2004), 1471–1530.

[15] St John Grimbly, Jonathan Shock, and Arnu Pretorius. 2021. Causal multi-agent reinforcement learning: Review and open problems. *arXiv preprint arXiv:2111.06721* (2021).

[16] David Ha and Jürgen Schmidhuber. 2018. Recurrent World Models Facilitate Policy Evolution. In *Advances in Neural Information Processing Systems 31, NeurIPS.* 2455–2467.

[17] Lewis Hammond, James Fox, Tom Everitt, Ryan Carey, Alessandro Abate, and Michael J. Wooldridge. 2024. Reasoning about Causality in Games (Abstract Reprint). In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI.* 22697.

[18] Sylvia S. Kerkhove, Natasha Alechina, and Mehdi Dastani. 2025. Causes and Strategies in Multiagent Systems. In *Proceedings of the 24th International Conference on Autonomous Agents AAMAS 2025.* 1098–1106.

[19] Vijay R. Konda and John N. Tsitsiklis. 1999. Actor-Critic Algorithms. In *Advances in Neural Information Processing Systems 12, NIPS.* The MIT Press, 1008–1014.

[20] Chaochao Lu, Bernhard Schölkopf, and José Miguel Hernández-Lobato. 2018. Deconfounding Reinforcement Learning in Observational Settings. *CoRR* abs/1812.10576 (2018).

[21] Tshilidzi Marwala. 2015. *Causality, correlation and artificial intelligence for rational decision making.* World Scientific.

[22] Stephanie Milani, Nicholay Topin, Manuela Veloso, and Fei Fang. 2024. Explainable reinforcement learning: A survey and comparative review. *Comput. Surveys* 56, 7 (2024), 1–36.

[23] Hongseok Namkoong, Ramtin Keramati, Steve Yadlowsky, and Emma Brunskill. 2020. Off-policy Policy Evaluation For Sequential Decisions Under Unobserved Confounding. In *Advances in Neural Information Processing Systems 33: Annual Conference NeurIPS.*

[24] Samer B. Nashed, Saaduddin Mahmud, Claudia V. Goldman, and Shlomo Zilberstein. 2025. Causal Explanations for Sequential Decision Making. *Journal of Artificial Intelligence Research* 83 (2025).

[25] Michael Oberst and David A. Sontag. 2019. Counterfactual Off-Policy Evaluation with Gumbel-Max Structural Causal Models. In *Proceedings of the 36th International Conference on Machine Learning (ICML) (Proceedings of Machine Learning Research, Vol. 97).* PMLR, 4881–4890.

[26] Judea Peal. 1985. Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 7.

[27] Judea Pearl. 1993. [Bayesian analysis in expert systems]: Comment: Graphical models, causality and intervention. *Statist. Sci.* 8, 3 (1993), 266–269.

[28] Judea Pearl. 1995. From Bayesian networks to causal networks. In *Mathematical models for handling partial knowledge in artificial intelligence.* Springer, 157–182.

[29] Judea Pearl. 2009. *Causality.* Cambridge university press.

[30] Judea Pearl. 2019. The seven tools of causal inference, with reflections on machine learning. *Commununication of ACM* (2019). https://doi.org/10.1145/3241036

[31] Judea Pearl et al. 2000. Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress* 19, 2 (2000), 3.

[32] Judea Pearl and Dana Mackenzie. 2018. *The book of why: the new science of cause and effect.* Basic books.

[33] Derek C Penn and Daniel J Povinelli. 2007. Causal cognition in human and nonhuman animals: A comparative, critical review. *Annu. Rev. Psychol.* 58, 1 (2007), 97–118.

[34] Erika Puiutta and Eric MSP Veith. 2020. Explainable reinforcement learning: A survey. In *International cross-domain conference for machine learning and knowledge extraction.* Springer, 77–95.

[35] Stuart Russell and Peter Norvig. 2020. *Artificial Intelligence: A Modern Approach (4th Edition).* Pearson. http://aima.cs.berkeley.edu/

[36] Bernhard Schölkopf. 2022. Causality for machine learning. In *Probabilistic and causal inference: The works of Judea Pearl.* 765–804.

[37] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. 2021. Toward Causal Representation Learning. *Proc. IEEE* 109, 5 (2021), 612–634.

[38] Oliver Schulte and Pascal Poupart. 2024. Why Online Reinforcement Learning is Causal. *CoRR* abs/2403.04221 (2024). https://doi.org/10.48550/ARXIV.2403.04221 arXiv:2403.04221

[39] Oliver Schulte and Pascal Poupart. 2025. When Should Reinforcement Learning Use Causal Reasoning? *Trans. Mach. Learn. Res.* 2025 (2025). https://openreview.net/forum?id=D1PPuk8ZBI

[40] Maximilian Seitzer, Bernhard Schölkopf, and Georg Martius. 2021. Causal Influence Detection for Improving Efficiency in Reinforcement Learning. In *Advances in Neural Information Processing Systems 34, NeurIPS.* 22905–22918.

[41] Yonadav Shavit, Sandhini Agarwal, Miles Brundage, Steven Adler, Cullen O'Keefe, Rosie Campbell, Teddy Lee, Pamela Mishkin, Tyna Eloundou, Alan Hickey, et al. 2023. Practices for governing agentic AI systems. *Research Paper, OpenAI* (2023).

[42] David Silver and Richard S Sutton. 2025. Welcome to the era of experience. *Google AI* 1 (2025).

[43] Richard S. Sutton and Andrew G. Barto. 1998. *Reinforcement learning - an introduction.* MIT Press. https://www.worldcat.org/oclc/37293240

[44] Michael J. Wooldridge. 2009. *An Introduction to MultiAgent Systems, Second Edition.* Wiley.

[45] Chiyuan Zhang, Oriol Vinyals, Remi Munos, and Samy Bengio. 2018. A study on overfitting in deep reinforcement learning. *arXiv preprint arXiv:1804.06893* (2018).

[46] Junzhe Zhang and Elias Bareinboim. 2016. Markov decision processes with unobserved confounders: A causal approach. *Purdue AI Lab, West Lafayette, IN, USA, Tech. Rep* (2016).