Multi-Agent Multi-Objective Planning with Contextual Lexicographic Reward Preferences

Doctoral Consortium

Pulkit Rustagi
Oregon State University
Corvallis, USA
rustagip@oregonstate.edu

ABSTRACT

Autonomous agents often operate in environments with multiple, conflicting objectives, where preference orderings vary with context. Existing multi-objective planning approaches assume a single preference ordering across the state space, making them unsuitable for context-dependent priorities. In multi-agent systems, this complexity increases as agents may operate under different contexts and must coordinate for safe joint operation. Current methods focus on static, single-agent settings or rely on centralized approaches that do not scale. My dissertation develops scalable, efficient techniques for multi-agent decision-making with context-dependent objective preferences. To that end, I developed a context-based planning framework for multi-objective settings, with theoretical guarantees for computing valid, cycle-free policies. I extended this framework to multi-agent systems, where agents complete tasks independently while mitigating negative side effects (NSEs) of joint actions. Future research will focus on contextual planning for cooperative multi-agent systems under partial observability and non-stationarity, enabling lifelong autonomy in dynamic environments.

KEYWORDS

Contextual Planning, Multi-Objective Planning, Multi-Agent Systems.

ACM Reference Format:
Pulkit Rustagi. 2025. Multi-Agent Multi-Objective Planning with Contextual Lexicographic Reward Preferences: Doctoral Consortium. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

In many real-world applications, such as navigation [5, 19] and warehouse management [6, 14], autonomous agents must optimize multiple, often competing objectives, with preferences that vary by *context*. These preferences follow a lexicographic ordering, where context dictates the priority of objectives and associated reward functions. For instance, a semi-autonomous car in a construction zone prioritizes avoiding uneven road surfaces over pedestrian safety and speed, whereas in urban areas, pedestrian safety takes precedence. In multi-agent systems, context becomes even more critical due to interdependencies among agents. For example, in a warehouse, robots must balance individual goals, such as minimizing travel distance, with joint goals dictated by context, such as avoiding multiple robots in regions with high human traffic.

Using context for decision-making enables agents to adapt to diverse and dynamic environments. Contextual information has been applied in contextual bandits [12], context-aware planning [9], and information retrieval [7, 21]. In existing literature, context is typically defined as a parameter influencing environment dynamics, rewards [3, 9, 13], obstacle positions [8, 20], or areas of operation [2]. However, most approaches assume static preferences, overlooking the need to adjust objective orderings based on context. We define context as a set of *exogenous features* that dictate objective priorities and rewards, enabling flexibility to handle interdependencies and shifting priorities in multi-agent, multi-objective planning.

Efficient and tractable planning in multi-agent, multi-objective settings requires (1) frameworks that seamlessly switch between objective orderings based on context, (2) optimizing individual and team objectives in a safe and efficient manner, and (3) adapting behaviors in settings with non-stationary contexts that change over time. This paper presents my contributions to multi-objective planning under contextual preferences (Section 2), extends the framework to multi-agent settings for mitigating negative side effects of joint actions (Section 3), and outlines future work (Section 4) on advancing contextual planning in dynamic multi-agent systems.

2 MULTI-OBJECTIVE PLANNING UNDER CONTEXTUAL REWARD PREFERENCES

In a recent work [15], we address the challenge of planning in environments with multiple, context-dependent preferences over objectives. Contexts define the lexicographic ordering of objectives and their associated reward functions, with multiple contexts co-existing and influencing agent behavior based on the state. Existing approaches to multi-objective planning typically assume a single, static preference ordering over objectives across the state space, limiting their applicability in environments where priorities vary. To overcome these limitations, we introduce the *Contextual Lexicographic Markov Decision Process (CLMDP)* [15], along with algorithms that computes valid, cycle-free policies tailored to contextual preferences, with theoretical guarantees for goal reachability.

Traditional methods for multi-objective decision-making rely on static lexicographic orderings [17, 22] or scalarization techniques [23] to balance objectives. While these approaches can represent fixed objective preferences effectively, they are not designed to

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

accommodate context-dependent variations. Extensions that partition the state space into regions with different preferences [22] lack principled methods for defining partitions or ensuring consistency across them. Our framework addresses these gaps by computing independent policies for each context and integrating them into a global policy that adapts to the state-to-context mapping, ensuring that the final policy is free of conflicts that prevent goal-reachability.

Our solution approach for CLMDP involves three steps. First, policies are computed independently for each context under its lexicographic ordering and then compiled into a global policy by mapping actions to states based on their contexts. Second, the global policy is analyzed using *ConflictChecker* to identify cycles that prevent goal reachability. Finally, conflicts are resolved with *ConflictResolver*, which iteratively updates lower-priority context policies while preserving the actions of higher-priority contexts. This ensures a conflict-free global policy that respects context-specific preferences and guarantees goal reachability.

The *ConflictChecker* detects inconsistencies in the global policy by evaluating goal reachability from every state. A conflict is flagged if a state lacks a valid path to the goal due to conflicting actions across contexts. Detected conflicts and the involved states are passed to the *ConflictResolver*, which resolves conflicts by iteratively updating policies. Starting with the lowest-priority context in conflict, higher-priority actions are fixed to constrain updates. Policies in the update set are sequentially revised, from highest to lowest priority, with resolved actions fixed after each update. If conflicts persist, higher-priority contexts are added, and the process repeats until all conflicts are resolved. This ensures a conflict-free policy that respects context-specific preference orderings.

We provide theoretical guarantees for both algorithms, ensuring that the resulting policies are conflict-free and enable goal reachability [15]. Our framework is validated through simulations across three domains and hardware experiments with mobile robots. Results show that the approach not only performs well across all objectives but also consistently outperforms baseline methods, achieving the highest value for the worst-performing objective. In all trials, the framework resolves conflicts reliably and without failure, demonstrating its robustness and applicability in complex multi-objective planning scenarios.

3 MITIGATING SIDE EFFECTS IN MULTI-AGENT SETTINGS

In multi-agent settings, context can be used to plan for inter-agent dependencies, particularly when addressing safety concerns such as negative side effects (NSEs). NSEs, unintended and undesirable outcomes of collective system behavior, often emerge due to incomplete or overly simplified decision-making models [1, 17, 18]. In multi-agent settings, NSEs are unintended consequences of joint actions [16]. For instance, high-traffic areas or overlapping tasks can increase the likelihood of NSEs that are typically unknown prior to deployment and jointly reported for all agents [16], making mitigation challenging. Such dependencies increase the complexity of planning, especially as the number of agents grows, making centralized methods for mitigating NSEs [4] computationally infeasible and difficult to scale. Furthermore, prior approaches focus on single-agent scenarios [10, 11, 17, 18, 24] or treat other agents as static elements of the environment [1], failing to address the dynamic interdependencies that give rise to NSEs.

In settings where NSEs may occur, the agent must plan to mitigate NSEs in addition to optimizing its task performance. The problem is formulated as a bi-objective problem, where the first objective prioritizes the efficient execution of assigned tasks, and the second objective minimizes penalties associated with negative side effects. Our work in [16] frames this as a context-based decision-making problem by introducing an additional *NSE occurrence* context alongside the existing *task completion* context. By introducing this additional context, agents can explicitly plan for conditions that lead to NSEs, such as interdependencies among agents or shared resources, and incorporate this information into their decision-making. This approach allows agents to adjust their behavior to optimize task performance while actively mitigating NSEs, ensuring that both objectives are addressed without compromising system efficiency or scalability.

The solution operates in three stages, with context playing a central role throughout. First, agents compute task completion policies independently, guided by the *task completion* context that ignores NSEs. Next, an *NSE monitor* module evaluates the joint behavior of agents to compute jointly reported NSE penalties. These penalties are analyzed using a blame resolver, which simulates counterfactual scenarios for each agent to assign blame values estimating their individual contributions to the NSE penalty. The assigned blame is incorporated into agent-specific penalty functions, forming the basis of the *NSE occurrence* context. Finally, agents recompute their policies to balance the preferences of both contexts, ensuring that safety considerations are integrated alongside task performance.

Our experiments in simulation and hardware validate the effectiveness of our context-based approach to mitigating NSE penalties. By selectively updating the policies of agents contributing most to NSEs, significant penalty reduction is achieved without requiring updates for all agents. Results further show that our method scales efficiently and outperforms baselines across simulated domains and real-world multi-agent environments.

4 FUTURE WORK

As part of future work, I have started to explore ways to handle non-stationarity, where contexts change over time and are partially observable. In such settings, agents must coordinate to infer underlying contexts and adapt their policies accordingly. This introduces challenges in balancing exploration for context inference with exploitation for task optimization, particularly in multi-agent environments where coordination is critical for accurate context detection and effective operation. Another direction involves scenarios where agents operating under different contexts create conflicts that hinder each other's task completion. Resolving such conflicts requires coordination to ensure efficiency and task success while also opening opportunities to explore credit assignment across contexts. This would allow agents to attribute responsibility for inter-agent conflicts and optimize collective behavior. These extensions aim to advance long-term autonomy in multi-agent systems.

ACKNOWLEDGMENTS

This work was supported in part by ONR grant N00014-23-1-2171.

REFERENCES

[1] Parand A. Alamdari, Toryn Q. Klassen, Rodrigo T. Icarte, and Sheila A. McIlraith. 2022. Be considerate: Avoiding negative side effects in reinforcement learning. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 18–26.

[2] Ali Bahrani, Jun Yuan, Chukwuemeka David Emele, Daniele Masato, Timothy J Norman, and David Mott. 2008. Collaborative and context-aware planning. In *MILCOM 2008-2008 IEEE Military Communications Conference*. IEEE, 1–7.

[3] Carolin Benjamins, Theresa Eimer, Frederik Schubert, Aditya Mohan, Sebastian Döhler, André Biedenkapp, Bodo Rosenhahn, Frank Hutter, and Marius Lindauer. 2022. Contextualize Me–The Case for Context in Reinforcement Learning. *arXiv preprint arXiv:2202.04500* (2022).

[4] Ferdinando Fioretto, Enrico Pontelli, and William Yeoh. 2018. Distributed constraint optimization problems and applications: A survey. *Journal of Artificial Intelligence Research (JAIR)* 61 (2018), 623–698.

[5] Kikuo Fujimura. 1996. Path planning with multiple objectives. *IEEE Robotics & Automation Magazine* 3, 1 (1996), 33–38.

[6] Jianqi Gao, Yanjie Li, Yunhong Xu, and Shaohua Lv. 2022. A two-objective ILP model of OP-MATSP for the multi-robot task assignment in an intelligent warehouse. *Applied Sciences* 12, 10 (2022), 4843.

[7] Kailash A Hambarde and Hugo Proenca. 2023. Information retrieval: recent advances and beyond. *IEEE Access* (2023).

[8] Jakub Hvězda, Tomáš Rybecký, Miroslav Kulich, and Libor Přeučil. 2018. Context-aware route planning for automated warehouses. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2955–2960.

[9] Byeonghwi Kim, Jinyeon Kim, Yuyeong Kim, Cheolhong Min, and Jonghyun Choi. 2023. Context-aware planning and environment-aware memory for instruction following embodied agents. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10936–10946.

[10] Toryn Q. Klassen, Sheila A. McIlraith, Christian Muise, and Jarvis Xu. 2022. Planning to avoid side effects. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, Vol. 36. 9830–9839.

[11] Victoria Krakovna, Laurent Orseau, Richard Ngo, Miljan Martic, and Shane Legg. 2020. Avoiding side effects by considering future tasks. *Advances in Neural Information Processing Systems (NeurIPS)* 33 (2020), 19064–19074.

[12] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*. 661–670.

[13] Aditya Modi, Nan Jiang, Satinder Singh, and Ambuj Tewari. 2018. Markov decision processes with continuous side information. In *Algorithmic Learning Theory*. PMLR, 597–618.

[14] Edgar Reehuis and Thomas Bäck. 2010. Mixed-integer evolution strategy using multiobjective selection applied to warehouse design optimization. In *Proceedings of the 12th annual conference on Genetic and evolutionary computation*. 1187–1194.

[15] Pulkit Rustagi, Yashwanthi Anand, and Sandhya Saisubramanian. 2025. Multi-Objective Planning with Contextual Lexicographic Reward Preferences. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025)*.

[16] Pulkit Rustagi and Sandhya Saisubramanian. 2024. Mitigating Negative Side Effects in Multi-Agent Systems Using Blame Assignment. *arXiv preprint arXiv:2405.04702* (2024).

[17] Sandhya Saisubramanian, Ece Kamar, and Shlomo Zilberstein. 2020. A multi-objective approach to mitigate negative side effects. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence (JAIR)*.

[18] Sandhya Saisubramanian, Ece Kamar, and Shlomo Zilberstein. 2022. Avoiding negative side effects of autonomous systems in the open world. *Journal of Artificial Intelligence Research (JAIR)* 74 (2022), 143–177.

[19] Earl B Smith and Reza Langari. 2003. Fuzzy multiobjective decision making for navigation of mobile robots in dynamic, unstructured environments. *Journal of Intelligent & Fuzzy Systems* 14, 2 (2003), 95–108.

[20] Adriaan W Ter Mors, Cees Witteveen, Jonne Zutt, and Fernando A Kuipers. 2010. Context-aware route planning. In *Multiagent System Technologies: 8th German Conference, MATES 2010, Leipzig, Germany, September 27-29, 2010. Proceedings 8*. Springer, 138–149.

[21] Jiajia Wang, Jimmy Xiangji Huang, Xinhui Tu, Junmei Wang, Angela Jennifer Huang, Md Tahmid Rahman Laskar, and Amran Bhuiyan. 2024. Utilizing BERT for Information Retrieval: Survey, Applications, Resources, and Challenges. *Comput. Surveys* 56, 7 (2024), 1–33.

[22] Kyle Wray, Shlomo Zilberstein, and Abdel-Illah Mouaddib. 2015. Multi-objective MDPs with conditional lexicographic reward preferences. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, Vol. 29.

[23] Runzhe Yang, Xingyuan Sun, and Karthik Narasimhan. 2019. A generalized algorithm for multi-objective reinforcement learning and policy adaptation. *Advances in neural information processing systems* 32 (2019).

[24] Shun Zhang, Edmund H. Durfee, and Satinder Singh. 2018. Minimax-Regret Querying on Side Effects for Safe Optimality in Factored Markov Decision Processes. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*.