Opponent Modelling in the Iterated Prisoner's Dilemma: A Literature Review

Faqih Mahardika
Department of Computer Science and Electronics
Gadjah Mada University
Yogyakarta, Indonesia
fm.faqihmahardika@gmail.com

***Abstract*—Lorem ipsum dolor sit amet, consectetur adipiscing elit. In tristique ipsum id lectus euismod molestie. In eget nunc vel nisi venenatis sagittis non at ante. Ut mattis tortor et libero dignissim maximus. Nunc vitae ante sit amet tellus fringilla tincidunt at quis elit. Fusce a laoreet elit, in scelerisque felis. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Fusce eu nibh varius, mollis turpis sed, vehicula metus. Nullam rhoncus euismod odio vel ullamcorper. Quisque venenatis id massa quis tempus. Morbi luctus enim lacinia finibus convallis.**

Phasellus varius congue tortor sit amet porttitor. Integer velit turpis, facilisis vitae imperdiet id, egestas eu turpis. Proin feugiat a erat eget vehicula. Sed ligula libero, sollicitudin sed nibh iaculis, fermentum feugiat lacus. Donec lectus risus, semper ac posuere non, hendrerit non nisl. Donec at laoreet dolor. Nullam efficitur auctor nisl, vel tincidunt neque lobortis eget. Phasellus elementum justo nec ante malesuada tristique. Curabitur sed magna condimentum, rutrum turpis sed, vehicula lacus. Vivamus ultricies nulla id nulla dictum sollicitudin. Nullam varius, lorem at tincidunt auctor, diam lectus varius mauris, ut pretium nisl arcu a massa. Sed et nisi a mauris eleifend fringilla. Pellentesque facilisis nisl sit amet dictum pellentesque.

***Index Terms*—keyword1, keyword2, keyword3**

I. Introduction

Cooperation and conflict are fundamental features of interaction in human societies, biological systems, and artificial-agent environments. Game theory provides a rigorous analytical framework for studying such strategic interdependence, and foundational work on the evolution of cooperation demonstrates how behavioural patterns emerge when agents repeatedly face dilemmas that require balancing self-interest with mutual benefit [2], [3]. Among these models, the Prisoner's Dilemma (PD) has become the canonical representation of situations in which individual rationality conflicts with collective welfare. In the one-shot PD, mutual cooperation yields the highest joint payoff, yet the incentive structure leads rational players to defect, exposing a core tension that recurs in many social, economic, and political contexts.

When the game is repeated as the Iterated Prisoner's Dilemma (IPD), the strategic landscape changes significantly. Repetition enables agents to condition their actions on previous interactions, giving rise to behaviours such as reciprocity, retaliation, forgiveness, and trust-building. Axelrod's influential IPD tournaments demonstrated that simple contingent strategies most notably Tit-for-Tat can foster stable cooperation even in competitive environments, illustrating how long-term interaction and memory contribute to the emergence of cooperative norms without centralised enforcement [2], [3]. These results established IPD as a fundamental tool for analysing adaptive, history-dependent decision-making.

The relevance of IPD dynamics extends across multiple domains. In biology, theories of reciprocal altruism describe how repeated encounters favour contingent cooperation, outlining conditions under which helping behaviour can evolve among self-interested organisms [4]. Evolutionary game theory formalises these mechanisms through the concept of evolutionarily stable strategies, showing how certain behavioural rules can persist under selective pressures [7]. In the social sciences, repeated-game frameworks explain human cooperation, norm enforcement, and punishment systems in public-goods settings [5]. In economics and resource management, problems such as commercial fishing demonstrate how strategic interdependence in shared environments mirrors the structure of repeated dilemmas [6]. Across these fields, the IPD serves as a versatile lens for understanding how cooperation emerges, stabilises, or collapses depending on incentives, information structures, and institutional conditions.

Although classical analyses of the IPD provide essential insights, their simplifying assumptions often diverge from the complexities of real strategic interaction. Real agents either human, biological, or artificial frequently operate under uncertainty, possess limited observability, and adapt their behaviour dynamically. These realities have motivated growing interest in opponent modelling, which equips an agent with the ability to infer an opponent's strategy, behavioural tendencies, or intentions from observed actions. Contemporary surveys highlight a broad spectrum of opponent-modelling techniques across adversarial and multi-agent domains, underscoring their importance for enabling adaptive decision-making in repeated games, negotiation, autonomous systems, and competitive learning environments [1]. Within the IPD, such capabilities are critical: sustaining cooperation depends on recognising cooperative partners, detecting exploiters, and adapting effectively to non-stationary or deceptive strategies.

Therefore, this study aims to systematically review the existing

literature on opponent modelling in the Iterated Prisoner's Dilemma. The review is guided by the following research questions:

- **RQ1:** What environmental settings and monitoring mechanisms are used in opponent modelling for the Iterated Prisoner's Dilemma?
- **RQ2:** What methodological approaches to opponent modelling have been applied in the Iterated Prisoner's Dilemma?
- **RQ3:** How is the effectiveness of opponent-modelling strategies evaluated in the Iterated Prisoner's Dilemma (e.g., cooperation rate, speed of cooperation, robustness)?

To address these questions, Section 2 outlines the Systematic Literature Review (SLR) methodology used to identify, screen, and synthesise relevant studies. Section 3 presents the results of the review from three analytical perspectives: (i) modelling-based comparison, (ii) environment-based comparison, and (iii) evaluation-based comparison. Section 4 integrates findings across these perspectives to identify persistent limitations and articulate the research gap. Section 5 proposes the methodological framework that this thesis develops to address the identified gap, followed by concluding remarks in Section 6.

II. Methodology

A systematic review will be employed to address the research questions. This method applies a transparent and structured procedure to collect relevant studies, appraise their quality, and integrate their findings in order to answer a well-defined research question [8].

A. Inclusion and Exclusion Criteria

Studies were included in the review if they met all of the following conditions:

- **Topic relevance:** The study presents a substantive discussion, method, or empirical implementation of opponent modelling specifically within the Iterated Prisoner's Dilemma (IPD) or a directly comparable repeated social dilemma.
- **Scholarly quality:** The work is published in a peer-reviewed journal or international conference proceedings.
- **Publication window:** The study was published within the last five years.
- **Accessibility:** The full text is available in English and accessible for analysis.

Studies were excluded if they met any of the following conditions:

- **Irrelevant focus:** The study does not address opponent modelling or does not involve the Iterated Prisoner's Dilemma (e.g., focuses only on one-shot PD, general multi-agent reinforcement learning without opponent inference, or unrelated game-theoretic models).
- **Theoretical only:** The paper provides purely theoretical or conceptual discussion without proposing, analysing, or evaluating an opponent-modelling method in the IPD context.
- **Outside the publication window:** Published more than five years before the search date.
- **Non-English or inaccessible:** The full text is not available in English or cannot be accessed.

B. Information Sources and Search Strategy

The literature search was conducted using four major academic databases selected for their comprehensive coverage of computer science, artificial intelligence, and multi-agent systems research: IEEE Xplore, ScienceDirect, Scopus, and SpringerLink. The search strategy employed a structured set of keywords designed to capture studies related to opponent modelling and the Iterated Prisoner's Dilemma (IPD). The following query was applied across all databases:

```
("opponent modelling" OR "opponent
 modeling" OR "opponent model" OR
        "agent modeling")
              AND
("prisoner's dilemma" OR "iterated
 prisoner's dilemma" OR "IPD" OR
"repeated game" OR "social dilemma")
```

Database-specific filters were applied to ensure the relevance and accessibility of retrieved publications.

1) **IEEE Xplore** (04–11–2025): Search applied to all fields.
2) **ScienceDirect** (04–11–2025): Search applied to all fields; restricted to Open Access publications.
3) **Scopus** (04–11–2025): Search applied to all fields.
4) **SpringerLink** (04–11–2025): Search applied to all fields; restricted to Open Access publications.

C. Research Procedure

The literature search was conducted following the defined search strategy, with keywords adapted to the specific query capabilities of each database. All identified records were collected and managed using a reference manager to remove duplicates and ineligible entries during the initial screening. The remaining studies were then screened based on their titles and abstracts to determine relevance. Full-text versions of studies that passed this stage were retrieved and assessed for eligibility as part of the systematic review.

Data extraction focused on gathering information relevant to Opponent Modelling in the Iterated Prisoner's Dilemma, including environmental settings, monitoring methods, modelling approaches, and evaluation metrics. The extracted data were subsequently analyzed and compared to identify patterns, advantages, and limitations of the different approaches. These findings are then used to address the research questions outlined for this review.

III. Results

At each stage of the selection process illustrated in Figure 1, n denotes the number of studies considered. During the

identification phase, 10 records were removed as duplicates based on their titles and abstracts. An additional 48 records were excluded for failing to meet the inclusion criteria. At the screening stage, 26 papers were removed because they were not relevant to this study (for example, they did not involve opponent modelling or did not use the IPD as the research context), and 6 papers were excluded because they were non-primary studies, resulting in a total of 32 removals. One record could not be retrieved due to access limitations. Consequently, 20 records were selected for full-text review.

Among the 20 full-text articles obtained, 2 were excluded because they focused on IPD population simulations or did not present any opponent modelling framework. Another 2 were excluded because they were non-primary studies and therefore did not propose their own opponent modelling approach. One additional paper was excluded for focusing solely on theoretical aspects of the IPD without providing an opponent modelling framework.

After applying all exclusion criteria, 15 studies were included in the final systematic review. The titles and characteristics of these studies are presented in Table I.

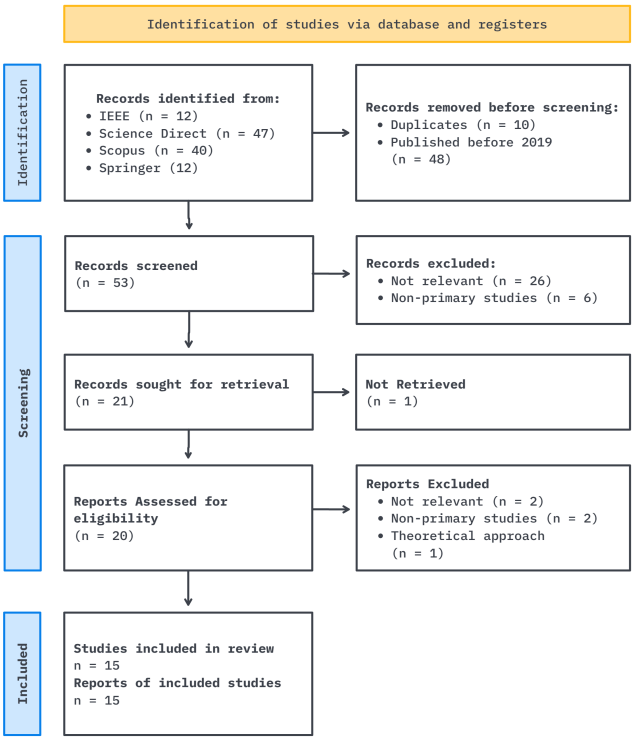

Fig. 1. Flow diagram for study selection.

TABLE I
CHARACTERISTICS OF INCLUDED STUDIES

Ref.	Paper	Characteristics
[9]	O2M: Online Opponent Modeling in Online General-Sum Matrix Games	Presents O2M, an online opponent-modeling algorithm for dynamic general-sum games.
[10]	Evolutionary Dynamics of Cooperation and Extortion on Networks With Fitness-Dependent Rules	Analyzes zero-determinant strategies and derives criteria for cooperation on networks.
[11]	Inducing Coordination in Multi-Agent Repeated Game through Hierarchical Gifting Policies	Proposes a hierarchical gifting mechanism that drives long-term coordination.
[12]	Recursively modeling other agents for decision making: A research perspective	Reviews recursive modeling frameworks for reasoning about others' intentions.
[13]	Grounded predictions of teamwork as a one-shot game: A multiagent multi-armed bandits approach	Models voluntary teamwork and shows how agents learn equilibrium collaboration.
[14]	Delegation and strategic altruism: A theoretical approach	Introduces quasi-stable and stable equilibria enabling cooperation via altruistic delegates.
[15]	The coupling effect between the environment and strategies drives the emergence of group cooperation	Shows that memory–environment coupling causes cycles of cooperation and defection.
[16]	Learning to cooperate against ensembles of diverse opponents	Proposes a scalable RL approach that fosters cooperation without explicit opponent models.
[17]	Exploiting a No-Regret Opponent in Repeated Zero-Sum Games	Presents a framework for modeling and exploiting no-regret opponents using RNNs.
[18]	Modeling Theory of Mind in Dyadic Games Using Adaptive Feedback Control	Proposes control-theoretic ToM agents that outperform standard RL strategies.
[19]	Modeling opponent learning in multiagent repeated games	Introduces a Stackelberg-based method that finds stable, high-reward equilibria.
[20]	The effects of population size and information update rates on the emergent patterns of cooperative clusters in a large-scale social particle swarm model	Shows that faster information updates in SNS networks increase cooperative clustering.
[21]	Achieving cooperation through deep multiagent reinforcement learning in sequential prisoner's dilemmas	Introduces a sequential PD and a deep RL method that adapts to detected cooperation.
[22]	Achieving collective welfare in multi-agent reinforcement learning via suggestion sharing	Introduces a MARL method where sharing action suggestions enables collective cooperation.
[23]	Higher-order theory of mind is especially useful in unpredictable negotiations	Finds that higher-order ToM is most beneficial in unpredictable environments.

IV. SYNTHESIS / DISCUSSION

integrates findings across these perspectives to identify persistent limitations and articulate the research gap. While a broad set of attributes was extracted, only methodologically discriminative dimensions are used for direct comparison; environment and evaluation attributes are summarized separately.

To operationalize *opponent behavior assumptions* across heterogeneous formulations, we summarize observable behavioral properties rather than internal learning mechanisms. The table below provides a descriptive categorization based on whether an opponent's actions (i) depend on environmental conditions,

(ii) are mediated by population-level interactions, (iii) respond directly to the focal agent's actions, and (iv) exhibit action divergence under equivalent recent interaction histories. These criteria are intentionally behavior-centric and evaluated at the level of interaction traces, without assuming access to the opponent's internal model, update rules, or optimization objectives.

This categorization is not intended as a formal or exhaustive taxonomy of opponent modeling methods. Instead, it serves as an analytical aid to support comparison across studies that adopt different game settings, learning paradigms, and modeling abstractions. In cases where a paper does not explicitly define opponent assumptions, classifications are derived conservatively from the described experimental setup and observed interaction dynamics, following standard practice in systematic literature reviews.

TABLE II
OPPONENT BEHAVIOR ASSUMPTION BASED ON OBSERVABLE BEHAVIORAL DEPENDENCIES.

Env.	Pop.	Agent	Div.	Opponent Behaviour Category	Paper
–	–	–	–	Stationary	[14]
–	–	✓	–	Reactive	[22]
–	✓	–	–	Population-Conformist	[13]
–	✓	✓	–	Contextual Reactive	[20]
–	–	✓	✓	Learning Opponent	[9], [11], [12], [17]–[19], [21], [23]
–	✓	✓	✓	Population-Contextual Strategic	[16]
✓	✓	–	✓	Heterogeneous Collective Behavior	[10]
✓	✓	✓	✓	Environment-Conditioned Strategic	[15]

Note:
Env. — behavior varies across environment within the same game;
Pop. — behavior is mediated by population-level interactions;
Agent — behavior responds directly to the focal agent's actions;
Div. — action divergence occurs on equivalent recent interaction histories.
Checkmarks indicate the presence of a dependency.

Several consistent patterns emerge from the behavioral categorization in Table II. Early and analytically oriented works predominantly assume stationary or purely reactive opponents, where behavior does not diverge under equivalent interaction histories [14]. In contrast, a substantial portion of recent studies model opponents whose actions are explicitly agent-responsive and exhibit action divergence [9]–[12], [15]–[19], [21], [23], indicating an assumption of learning or strategic adaptation. This shift reflects a growing research emphasis on robustness against non-stationary and adaptive opponents rather than optimization against fixed strategies.

Moreover, population-mediated behavior appears primarily in studies of evolutionary or networked environments, where opponent actions are shaped indirectly through aggregate dynamics rather than direct bilateral adaptation [10], [13], [15], [16], [20]. These settings often decouple individual agent responsiveness from population-level change, leading to behavioral patterns that differ qualitatively from pairwise learning scenarios. Notably, only a subset of works combine environment-dependent, population-mediated, and agent-responsive behaviors [15], suggesting that fully context-conditioned strategic opponents remain comparatively underexplored.

The distribution of opponent behavior assumptions in Table II reflects a broader methodological shift in opponent modeling research, from controlled and stationary assumptions toward behaviorally rich, adaptive, and heterogeneous opponents. While this characterization highlights the nature of the opponents considered, it does not capture how agents are designed to cope with such complexity. Therefore, in Table III reorganize prior work from the perspective of agent-side modelling, categorizing approaches by their dominant opponent modelling paradigm and corresponding learning mechanism.

TABLE III
COMPARISON OF OPPONENT MODELLING APPROACHES BY DOMINANT MODELLING PARADIGM AND LEARNING MECHANISM.

Modelling Paradigm	**Learning / Update Mechanism**	**Papers**
Reactive reinforcement learning	Value-based RL (DQN)	[11]
Gradient-based opponent shaping	Gradient descent / opponent-aware updates	[9], [19], [21]
Recursive belief reasoning	Bayesian belief update / cognitive hierarchy	[12], [18], [23]
Population-based training	Policy-gradient MARL (population sampling)	[16]
Evolutionary population dynamics	Fitness-based strategy imitation	[10], [15], [20]
Bandit-based learning-in-games	Multi-armed bandit / smooth best response	[13]
System identification	Autoregressive model with exogenous inputs (NARX)	[17]
Communication-driven coordination	Policy-gradient with communication objective	[22]
Analytical game-theoretic modelling	Equilibrium analysis (no learning)	[14]

While Table III categorizes prior work by modelling paradigm and learning mechanism, the listed approaches also differ in whether opponent adaptation is handled explicitly or implicitly. Explicit opponent modelling approaches, such as gradient-based opponent shaping [9], [19], [21], recursive belief reasoning [12], [18], [23], and system identification [17], construct internal representations of opponent behavior or learning dynamics. In contrast, implicit approaches—including reactive reinforcement learning [11] , population-based training [16], evolutionary dynamics [10], [15], [20] and Bandit-based learning-in-games [13] adapt to opponents without maintaining an explicit opponent model. This distinction highlights alternative design philosophies for addressing opponent non-stationarity, independent of the specific learning mechanisms

employed.

V. Proposed framework

my thesis proposed framework.

VI. Conclusion

Your conclusions and future work.

References

[1] S. B. Nashed and S. Zilberstein, "A Survey on Opponent Modeling in Adversarial Domains" unpublished.

[2] R. Axelrod, "The Evolution of Cooperation" unpublished.

[3] R. Axelrod and W. D. Hamilton, "The Evolution of Cooperation" *Science*, 1981.

[4] R. L. Trivers, "The Evolution of Reciprocal Altruism" *The Quarterly Review of Biology*, vol. 46, no. 1, pp. 35–57, 1971. Available: https://www.journals.uchicago.edu/doi/10.1086/406755

[5] Y. Chen, "Cooperation and punishment in public goods experiments (by Ernst Fehr and Simon Gächter)," in *The Art of Experimental Economics*, G. Charness and M. Pingle, Eds., London: Routledge, 2021, pp. 134–143. Available: taylorfrancis.com/books/9781003019121/chapters/10.4324/9781003019121-12

[6] A. Leung and A.-Y. Wang, "Analysis of Models for Commercial Fishing: Mathematical and Economical Aspects," *Econometrica*, vol. 44, no. 2, p. 295, 1976. Available: https://www.jstor.org/stable/1912725

[7] J. M. Smith, *Evolution and the Theory of Games*. (Publication year not provided.)

[8] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, Mar. 2021.

[9] X. Qiao, C. Han, and T. Guo, "O2M: Online Opponent Modeling in Online General-Sum Matrix Games," in *2024 4th International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC)*, Dec. 2024, pp. 358–361. [Online]. Available: https://ieeexplore.ieee.org/document/10899996/. doi: 10.1109/ICAIRC64177.2024.10899996.

[10] L. Zhu, Y. Zhu, and C. Xia, "Evolutionary Dynamics of Cooperation and Extortion on Networks With Fitness-Dependent Rules," in *2025 Joint International Conference on Automation-Intelligence-Safety (ICAIS) & International Symposium on Autonomous Systems (ISAS)*, May 2025, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/11051564/. doi: 10.1109/ICAISISAS64483.2025.11051564.

[11] M. Lv, J. Liu, B. Guo, Y. Ding, Y. Zhang, and Z. Yu, "Inducing Coordination in Multi-Agent Repeated Game through Hierarchical Gifting Policies," in *2023 IEEE 20th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*, Sep. 2023, pp. 279–287. [Online]. Available: https://ieeexplore.ieee.org/document/10298392/. doi: 10.1109/MASS58611.2023.00041.

[12] P. Doshi, P. Gmytrasiewicz, and E. Durfee, "Recursively modeling other agents for decision making: A research perspective," *Artificial Intelligence*, vol. 279, p. 103202, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000437021930027X. doi: 10.1016/j.artint.2019.103202.

[13] A. L. de A. Gómez, C. Sierra, and J. Sabater-Mir, "Grounded predictions of teamwork as a one-shot game: A multiagent multi-armed bandits approach," *Artificial Intelligence*, vol. 341, p. 104307, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0004370225000268. doi: 10.1016/j.artint.2025.104307.

[14] L. Méndez-Naya, "Delegation and strategic altruism: A theoretical approach," *Mathematical Social Sciences*, vol. 138, 2025, pp. 102465. doi: 10.1016/j.mathsocsci.2025.102465.

[15] C. Di, Q. Zhou, J. Shen, J. Wang, R. Zhou, and T. Wang, "The coupling effect between the environment and strategies drives the emergence of group cooperation," *Chaos, Solitons & Fractals*, vol. 176, 2023, pp. 114138. doi: 10.1016/j.chaos.2023.114138.

[16] I. Perera, F. de Nijs, and J. Garcia, "Learning to cooperate against ensembles of diverse opponents," *Neural Computing and Applications*, vol. 37, no. 23, 2025, pp. 18835–18849. doi: 10.1007/s00521-024-10511-9.

[17] K. Li, W. Huang, C. Li, and X. Deng, "Exploiting a No-Regret Opponent in Repeated Zero-Sum Games" *Journal of Shanghai Jiaotong University (Science)*, vol. 30, no. 2, 2025, pp. 385–398. doi: 10.1007/s12204-023-2610-2.

[18] I. T. Freire, X. D. Arsiwalla, J. Y. Puigbò, and P. Verschure, "Modeling Theory of Mind in Dyadic Games Using Adaptive Feedback Control," *Information*, vol. 14, no. 8, 2023. doi: 10.3390/info14080441.

[19] Y. Hu, C. Han, H. Li, and T. Guo, "Modeling opponent learning in multiagent repeated games," *Applied Intelligence*, vol. 53, no. 13, 2023, pp. 17194–17210. doi: 10.1007/s10489-022-04249-x.

[20] Z. Elhamer, R. Suzuki, and T. Arita, "The effects of population size and information update rates on the emergent patterns of cooperative clusters in a large-scale social particle swarm model," *Artificial Life and Robotics*, vol. 25, no. 1, 2020, pp. 149–158. doi: 10.1007/s10015-019-00558-6.

[21] W. Wang, Y. Wang, J. Hao, and M. E. Taylor, "Achieving cooperation through deep multiagent reinforcement learning in sequential prisoner's dilemmas," in *ACM International Conference Proceeding Series*, 2019. doi: 10.1145/3356464.3357712.

[22] Y. Jin, S. Wei, and G. Montana, "Achieving collective welfare in multi-agent reinforcement learning via suggestion sharing," *Machine Learning*, vol. 114, no. 8, 2025, p. 190. doi: 10.1007/s10994-025-06823-z.

[23] H. De Weerd, R. Verbrugge, and B. Verheij, "Higher-order theory of mind is especially useful in unpredictable negotiations," *Autonomous Agents and Multi-Agent Systems*, vol. 36, no. 2, 2022, p. 30. doi: 10.1007/s10458-022-09558-6.