

# A Novel Reinforcement Learning Approach in Detecting Non-Collaborators for Trust Game with a Large Number of Participants <sup>\*</sup>

Ziyu Huang

Duke Kunshan University, Kunshan, Jiangsu 215316, China  
zh147@duke.edu

**Abstract.** This paper explores the integration of Generative AI (GAI) and multi-agent reinforcement learning (MARL) to advance traditional game theory research, particularly in simulating complex human-like decision-making processes. By examining the "good and bad collaborators" game, the study aims to model behaviors in trust and reputation scenarios using GAI to generate profiles for simulation. The innovative use of GAI facilitates the creation of dynamic, realistic profiles, which are then tested in MARL environments to analyze behavior patterns and improve detection methods for uncooperative participants. This approach not only enhances the understanding of strategic interactions in controlled settings but also addresses the limitations of traditional game theory in handling dynamic and incomplete information scenarios. The research indicates potential for developing more sophisticated game theory models that can adapt to and accurately reflect evolving gameplay and strategies, thus providing deeper insights into the foundational principles of human strategic behavior and decision-making.

*Notes: In submission to Problem Set 2 for COMPSI/ECON 206 Computational Microeconomics, 2024 Spring Term (Seven Week - Second) instructed by Prof. Luyao Zhang at Duke Kunshan University.*

**Keywords:** computational economics · game theory · innovative education · generative AI · multi-agent reinforcement learning

## 1 Introduction

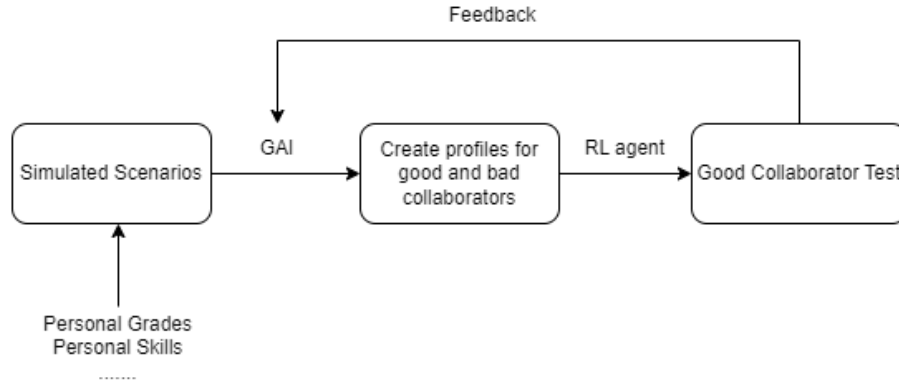
Traditional game theory is principally concerned with the negotiations between people. Within the context of research, it faces challenges due to a lack of data and the difficulty of conducting experiments. Recently, generative AI (GAI) has illuminated the path to generating human-like decision-making processes [1]. This could potentially provide convenience for experimental endeavors in traditional game theory.

---

<sup>\*</sup> **Acknowledgments:** I am deeply grateful to my instructor, Luyao Zhang, for her exceptional guidance during the development of this computational economics paper. Additionally, I want to thank my colleagues for their support and collaboration, and my classmates for their stimulating discussions and valuable contributions.

Consider the example of the "good and bad collaborators" game, which is a variation of trust and reputation games [2]; without controls in place, the good collaborators tend to eventually exit the game. Currently, a prospective avenue for enhancement is to construct a system of mutual selection that fosters equitable cooperation. Nevertheless, an inherent concern arises from our inability to curb the production of fraudulent resumes—a problem traditionally mitigated by more stringent regulatory measures.

To address this issue, I intend to incorporate GAI to model good and bad collaborators in different scenarios and utilize MARL to learn these results. As shown in Figure 1, we may first simulate various scenarios (like personal performance and skills), followed by employing Generative AI to create profiles for good and bad collaborators, all striving to succeed in a "good collaborator test." We can then appoint multiple agents to repeatedly execute the "good collaborator detection" tasks using these profiles and relay the outcomes back to the Generative AI for its learning process. Ultimately, we aspire to decipher the conduct paradigms of uncooperative collaborators, which will assist in advancing game theory research in such contexts.

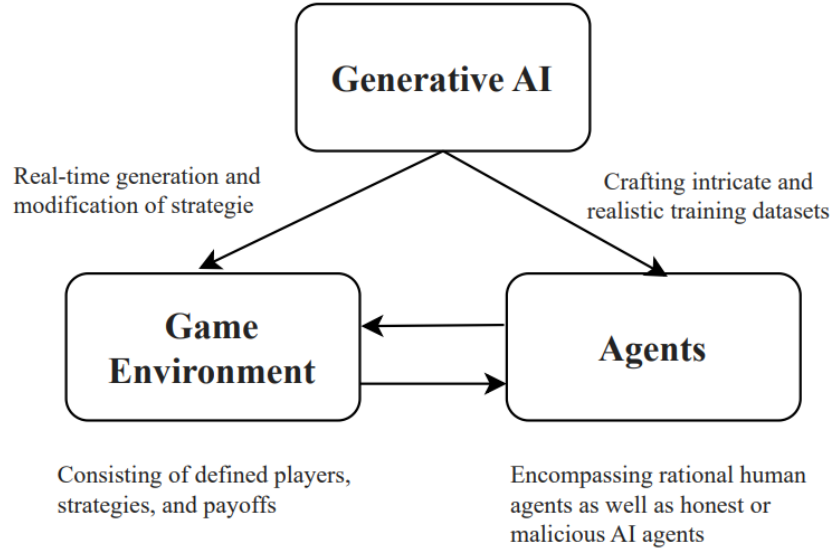


**Fig. 1.** Flowchart of how GAI working in the collaborator game

## 2 Background

As shown in Figure 2, in the conventional frameworks of game theory, models usually encompass the gaming environment and the human or AI agents on two distinct levels. The present-day constraints revolve around the often too static gaming circumstances that resist dynamic fine-tuning. By integrating Generative AI, we gain the capability to craft intricate and realistic training datasets, particularly advantageous in scenarios with a deficiency or inaccessibility of real-world data. Additionally, Generative AI can be utilized for the real-time generation

and modification of strategies to navigate the unpredictable shifts within the game setting. This methodology equips AI proxies with enhanced agility in their responses to uncertain and continuously evolving adversary tactics.



**Fig. 2.** Flowchart of how GAI cooperate in current game theory structure

### 3 An Illustration Example

#### 3.1 Research Question

Our research aims to address several crucial questions: How can natural language processing combined with reinforcement learning effectively predict and classify cooperative behaviors in generated resumes? Can a model that analyzes historical interaction patterns accurately predict the transition from cooperative to non-cooperative behaviors over time? And what are the key indicators in textual and behavioral data that distinguish good collaborators from bad collaborators, particularly those who initially disguise their intentions? These questions are significant because accurately predicting and understanding cooperative behaviors can lead to more effective team formations, enhance AI-human interactions, and ensure safety in trust-critical contexts. Traditional game theory literature, which often focuses on static scenarios with fixed strategies and lacks empirical data, does not sufficiently address these dynamic and complex behavior patterns, nor

does it incorporate advanced AI technologies like natural language processing and machine learning. Our research seeks to fill these gaps by using modern AI techniques to provide a dynamic, context-sensitive analysis of cooperative behaviors, potentially enhancing both theoretical and practical applications in game theory and related fields.

### 3.2 Methodology

**Simulating Participants with Different Cooperation Patterns** In the overall architecture, I aim to simulate various forms of non-cooperative participants, including those who are completely uncooperative, occasionally cooperative, and gradual defectors [3]. Completely uncooperative actors choose not to cooperate in all circumstances, regardless of the behavior of other participants. These participants typically pursue short-term interests without considering the damage to long-term relationships or collective welfare. In the prisoner's dilemma game, these participants always choose to defect. We can simulate this behavior by setting fixed strategy parameters, such as a cooperation probability of zero. Occasionally cooperative participants (Conditional Cooperators) base their cooperative behavior on specific conditions, such as the prior actions of others, environmental incentives, or positive results from historical interactions. They may show cooperation when a cooperative environment is detected but quickly switch to non-cooperation if they feel exploited or if the other party does not cooperate. We can use a Tit-for-Tat strategy, which starts with cooperation and continues if the other party cooperates, but defects in the next round if the other party defects. Alternatively, we can adjust the cooperation probability based on the ratio of cooperation in historical interactions. Gradual Defectors show cooperative behavior in the initial stages of interaction, but as time passes and conditions change, they gradually shift to a non-cooperative strategy. This behavior pattern is usually strategic, aimed at maximizing personal benefits by leveraging the advantages of cooperation after trust is established, then opting out at an opportune moment to avoid long-term costs or severing ties after maximizing benefits. We can set the cooperation probability with a function based on time or the number of interactions, where the cooperation probability gradually decreases as time progresses or as the number of interactions increases. Algorithm 1 shows the simple simulation process for different collaborator behaviors.

**Simulating Human Cooperation Behavior with ChatGPT** As previously mentioned, we plan to use ChatGPT, a large language model equipped with a rich implicit knowledge base, to simulate human-like decision-making processes. Specifically, we will construct two background databases: one for "Good Collaborators" (Table 1) and another for "Bad Collaborators." (Table 2) These databases will contain detailed background information and behavioral characteristics of each type of collaborator. In our experiments, the ChatGPT agent will randomly sample from these databases to generate representative resumes of both good and bad collaborators. This process aims to explore and identify

**Algorithm 1** Simulate Interactions

---

```

1: procedure SIMULATEINTERACTION(participants, rounds)
2:   history  $\leftarrow$  list()
3:   for i  $\leftarrow$  1 to rounds do
4:     actions  $\leftarrow$  dictionary()
5:     for all p in participants do
6:       actions[p]  $\leftarrow$  p.will_cooperate()
7:     end for
8:     for all p in participants do
9:       if p is instance of ConditionalCooperator then
10:        p.update_cooperation(actions[not p])
11:       else if p is instance of GradualDefector then
12:        p.update_cooperation()
13:       end if
14:     end for
15:     history.append(actions)
16:   end for
17:   return history
18: end procedure

```

---

key decision-making factors and behavior patterns in cooperation by simulating different cooperative behaviors. The sample ChatGPT response is shown in Figure 3.

**Table 1.** Sample characteristics table for good collaborators.

Name	Characteristics
John Doe	Highly cooperative, strong communication skills, always seeks a win-win solution, has a history of successful team projects.
Jane Smith	Trustworthy, always meets deadlines, positive attitude, frequently praised for her empathetic leadership.

**Implementing Dual Reinforcement Learning Networks for Simulating and Distinguishing Collaborative Behaviors** To deeply simulate and distinguish between different types of collaborators, our model incorporates two functional reinforcement learning networks. These networks are tasked with learning and processing distinct types of data to effectively identify and categorize "good collaborators" and "bad collaborators." The first type is a language information learning network, which focuses on processing and analyzing resume texts generated by ChatGPT. This network employs natural language processing techniques to recognize key words and phrases in resumes that may indicate the candidate's cooperative tendencies and behavior patterns. For example, by analyzing achievements and past behavior descriptions in the resume,

**Table 2.** Sample characteristics table for bad collaborators.

Name	Characteristics
Mike Ross	Often puts own interests first, has been involved in conflicts with team members, tends to miss deadlines.
Sarah Lee	Poor communication skills, often unresponsive to team communication, displays a lack of commitment to project goals.

### ChatGPT

Integrating these collaborator databases into ChatGPT can be achieved through two primary methods: preset scenarios and dynamic role selection. Here's how you can implement each method:

#### 1. Preset Scenarios

In preset scenarios, you prepare a set of decision-making situations and potential responses based on the characteristics of "Good Collaborators" and "Bad Collaborators." For each scenario, you would script responses that reflect the typical behavior of each type of collaborator. Here is an example of how you might set this up:

- **Scenario:** The team needs to decide whether to extend a project deadline.
- **Good Collaborator Response** (e.g., Jane Smith): "I think extending the deadline would allow us to thoroughly complete all aspects of the project and deliver a higher quality product. Let's discuss how we can adjust our workflow efficiently."
- **Bad Collaborator Response** (e.g., Mike Ross): "I don't see why I need to spend more time on this. I'd rather finish it as planned and move on to other things."

#### 2. Dynamic Role Selection

For dynamic role selection, ChatGPT uses the information from the collaborator databases during the conversation to adapt its responses according to the selected role (good or bad collaborator). Here's a breakdown of how this could work:

- **Initialization:** At the start of the interaction, ChatGPT randomly selects a role based on the databases. For example, it might adopt the persona of "John Doe," a good collaborator, or "Sarah Lee," a bad collaborator.
- **Conversation Handling:** During the dialogue, ChatGPT maintains the selected role. When faced with questions or decisions, it generates responses that are consistent with the collaborator's profile:
  - If acting as "John Doe," ChatGPT would provide cooperative and constructive responses.
  - If acting as "Sarah Lee," it might exhibit poor communication and a lack of engagement.

**Fig. 3.** GPT response to generate collaborators behavior

the network can preliminarily classify candidates as "good collaborators" or "bad collaborators." We plan to use Recurrent Neural Networks (RNN) or Long Short-Term Memory networks (LSTM) to handle sequence data, capturing long-distance dependencies in language. The second type is a historical information learning network, considering that some "bad collaborators" may initially exhibit cooperative behaviors and then switch to non-cooperation. This network focuses on analyzing participants' behavior history. By learning the interaction patterns and decision changes of participants, the network can identify individuals who may initially hide their true intentions. As in Figure 4, we will utilize reinforcement learning networks such as Q-learning or Deep Q Networks (DQN) to assess and optimize decision-making strategies based on historical behavior, implementing mechanisms to recognize transition patterns from cooperation to non-cooperation, such as the traits of gradual defectors.

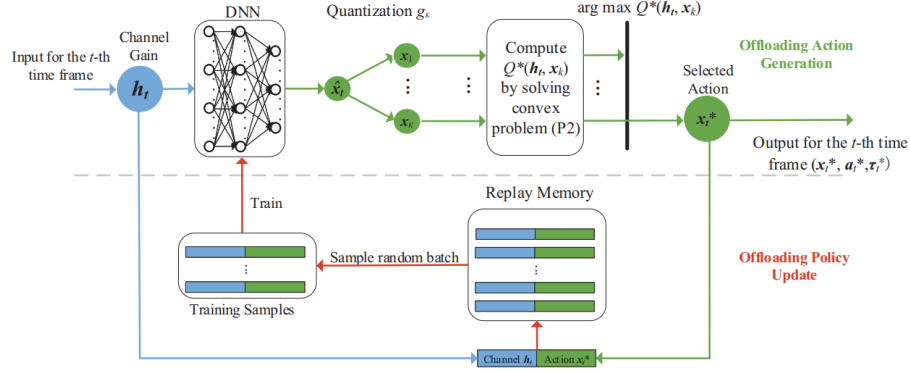


Fig. 4. Sample architecture for trust game reinforcement learning [4]

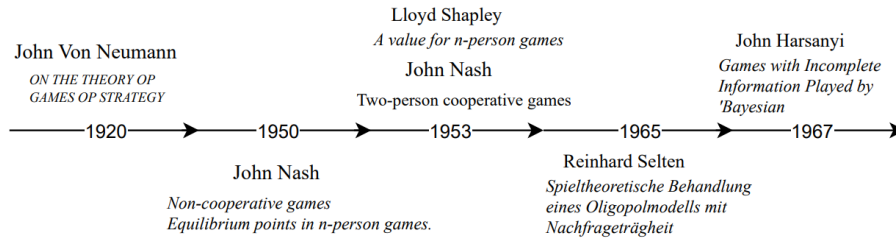
### 3.3 Application Scenario

Our newly proposed solution, leveraging natural language processing and reinforcement learning to predict and classify cooperative behaviors, is applicable in various real-world scenarios, such as corporate human resources for analyzing potential hires' resumes [5], social robots and virtual assistants for improving interactions with humans [6], and autonomous vehicles to predict human drivers' behaviors for safer and more efficient decision-making [7]. To ground our approach in behavioral science, we can draw from psychological theories such as Social Exchange Theory [8], which analyzes relationships through cost-benefit analysis; Theory of Planned Behavior, which links intentions to attitudes, norms, and perceived control; Cognitive Dissonance Theory, explaining behavior contrary to beliefs due to social pressures; and research on trust and reciprocity,

crucial for understanding the dynamics of cooperation. Integrating these psychological insights with our AI-driven methods will enhance our model's predictive accuracy and deepen the interdisciplinary connection between AI, game theory, and psychology, thus providing a robust foundation for applying game theoretical concepts in practical, impactful ways.

## 4 The Pioneers in the History of Game Theory

Figure 5 illustrates the timeline for this section.



**Fig. 5.** Timeline of the development of Game theory

**Transition from decision theory to game theory:** Von Neumann, John. "ON THE THEORY OF GAMES OF STRATEGY1."

In his influential paper, John Von Neumann uniquely diverged from the traditional study of individual optimum decision-making that characterizes decision theory [9]. He innovatively introduced the strategy of minimizing the maximum possible loss, specifically within the context of zero-sum games. This pivotal notion marked the emergence of game theory as a significant and distinct subdivision within decision theory.

**Evolution from pure-strategy Nash Equilibrium to mixed-strategy Nash Equilibrium:** Nash, John F. "Non-cooperative games." and Nash Jr, John F. "Equilibrium points in n-person games."

In 1950, John Nash introduced the concept of Nash equilibrium in his doctoral dissertation entitled "Non-cooperative Games." [10]. He initially discussed what is known as the pure Nash equilibrium, where each player's strategy is fixed and singular, akin to discrete events in probability theory. However, it became clear that considering only discrete scenarios was insufficient. Consequently, Nash went on to formulate the mixed strategy Nash equilibrium, incorporating probabilities into each decision-maker's process, making it uncertain



which strategy each player would adopt [11]. Nash further refined this concept of mixed strategy Nash equilibrium in his paper published the same year, "Equilibrium Points in N-person Games."

**Differentiation between non-cooperative games and cooperative games:** Nash, John. "Two-person cooperative games." and Shapley, Lloyd S. "A value for n-person games."

As previously mentioned, John Nash delineated a model of non-cooperative games in 1950 that highlighted the pursuit of individual profit maximization. With his ongoing research into the field, by 1953, Nash had developed his work on cooperative games with his article "Two-person cooperative games," exploring the prospects of participants cooperating to achieve an optimal allocation of resources, contrasting with the earlier theories centered on non-cooperation [12]. In addition to Nash's contributions, Lloyd Shapley, in his seminal work "A Value for n-person Games" published the same year, introduced the Shapley value as a quantitative measure for cooperative games [13].

**Progression from static games to dynamic games:** Selten, Reinhard. "Spieltheoretische behandlung eines oligopolmodells mit nachfrageträgheit: Teil i: Bestimmung des dynamischen preisgleichgewichts."

The game-theoretic models previously discussed are static in nature, implying that our considerations are confined to a fixed moment in time, without any regard for historical states. To address this issue, Reinhard Selten published "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit" in 1965, presenting a dynamic game theory model within the context of an oligopolistic economy [14]. Dynamic games allow players to incorporate the past decisions of their opponents when determining their own strategies.

**Shift from games with perfect information to games with imperfect information:** Harsanyi, John C. "Games with incomplete information played by "Bayesian" players, I-III Part I. The basic model."

In the context of games of complete information, it is assumed that all players possess full knowledge of the game environment, including the potential strategies and payoffs of other players. Yet, this is often not the case in reality. To tackle this, John Harsanyi, in 1967, published "Games with Incomplete Information Played by 'Bayesian' Players," proposing the notion of games with incomplete information [15]. This represents a category of intricate models that are commonly encountered in everyday life, thereby holding substantial real-world relevance.

## 5 Review Classic Games, Nash Equilibrium and the Analytical Tools

### 5.1 Bayesian Nash Equilibrium and Subgame Perfect Nash Equilibrium

In a game with incomplete information, let  $N$  be the set of players, each player  $i$  has a private type  $t_i$  drawn from the type set  $T_i$ . Each player  $i$  adopts a strategy, which is a function  $s_i : T_i \rightarrow A_i$ , where  $A_i$  is the action set for player  $i$ . The utility function for player  $i$  is denoted as  $u_i(a, t_i, t_{-i})$ , where  $a$  is the action profile and  $t_{-i}$  represents the type profile of all players except  $i$ .

**Definition 1 (Bayesian Nash Equilibrium).** *A Bayesian Nash Equilibrium in this setting is a strategy profile  $s^* = (s_1^*, s_2^*, \dots, s_N^*)$  such that for all players  $i$  and all possible types  $t_i$ , the following inequality holds:*

$$\mathbb{E}[u_i(s_i^*(t_i), s_{-i}^*(t_{-i}), t_i) \mid t_i] \geq \mathbb{E}[u_i(s_i(t_i), s_{-i}^*(t_{-i}), t_i) \mid t_i]$$

for all  $s_i : T_i \rightarrow A_i$ . The expectation is taken over  $t_{-i}$ , considering the probability distribution of  $t_{-i}$  conditional on  $t_i$  being known. [16]

**Theorem 1 (Bayesian Nash Equilibrium).** *Suppose each player  $i$  in a two-player auction independently values an item between 0 and 1 and bids their valuation  $b_i = t_i$ . The highest bidder wins and pays their bid. This bidding strategy forms a Bayesian Nash Equilibrium. [16]*

*Proof.* To verify that the strategy  $b_i = t_i$  is a Bayesian Nash Equilibrium, consider each player  $i$ 's utility given by:

$$u_i = \begin{cases} t_i - b_i & \text{if } b_i > b_{-i} \\ 0 & \text{if } b_i \leq b_{-i} \end{cases}$$

Under the strategy  $b_i = t_i$ , the probability of winning is  $t_i$ , as  $t_{-i}$  is less than  $t_i$ . Therefore:

$$\mathbb{E}[u_i \mid t_i] = (t_i - t_i) \times t_i + 0 \times (1 - t_i) = 0$$

If player  $i$  deviates to  $b'_i \neq t_i$ , bidding  $b'_i > t_i$  results in negative utility if  $b'_i > t_{-i}$ , and bidding  $b'_i < t_i$  reduces the probability of winning, hence not improving utility. Therefore, no player has an incentive to deviate, establishing the strategy as a Bayesian Nash Equilibrium. [16]

**Definition 2 (Subgame Perfect Nash Equilibrium).** *A strategy profile  $(s_1^*, s_2^*, \dots, s_N^*)$  is a Subgame Perfect Nash Equilibrium (SPNE) in an extensive-form game if, for every subgame, the strategy profile  $(s_1^*, s_2^*, \dots, s_N^*)$  forms a Nash Equilibrium. Formally, for any subgame starting at a decision node  $n$ , if  $s_i^*(h)$  denotes the action taken by player  $i$  at a history  $h$  leading to  $n$ , then:*

$$\mathbb{E}[u_i(s_i^*(h), s_{-i}^*(h)) \mid h] \geq \mathbb{E}[u_i(s_i(h), s_{-i}^*(h)) \mid h]$$

for all  $s_i$  alternative strategies of player  $i$  and all histories  $h$  within the subgame. [17]

**Theorem 2.** *In a sequential game with perfect information, backward induction yields a strategy profile that constitutes a Subgame Perfect Nash Equilibrium. [17]*

*Proof.* Consider a sequential game where players make decisions one after another with full knowledge of previous actions.

**Base Case:** At the final decision node  $n$  of the game, the action  $a_i^*$  chosen by the player maximizes their payoff  $u_i$ :

$$u_i(a_i^*) = \max_{a_i \in A_i} u_i(a_i)$$

**Inductive Step:** Assume that for any decision node  $k$  steps from the end, players select actions forming Nash Equilibria in the subsequent subgames. At a node  $k + 1$  steps from the end, the strategy  $s_i^*$  chosen by the acting player maximizes their expected payoff, given the strategies  $s_{-i}^*$  of other players:

$$\mathbb{E}[u_i(s_i^*, s_{-i}^*)] = \max_{s_i \in S_i} \mathbb{E}[u_i(s_i, s_{-i}^*)]$$

Each player's choice ensures that the overall strategy profile  $(s_1^*, s_2^*, \dots, s_N^*)$  forms a Nash Equilibrium at every node, including  $n$ , thus satisfying the conditions for a SPNE. [17]

From above definition and theorem, Bayesian Nash Equilibrium and Subgame Perfect Nash Equilibrium are adapted for specific scenarios in strategic interactions. BNE deals with the complexity of incomplete information by focusing on players' probabilistic beliefs about other participants, aiming for a strategy profile that maximizes expected utility amidst uncertainty. Meanwhile, SPNE ensures rationality through the entire structure of a sequential game, requiring strategies to constitute Nash Equilibria within every conceivable subgame. Although BNE involves conjectural dynamics playing with private information, SPNE mandates consistent optimal strategies throughout the possible unfolding of the game's events, reconciling foresight with strategic action.

## 5.2 Exploring Inspirational Games in Strategic or Normal Form

I'm particularly interested in the first-price auction within auction theory. Essentially, a first-price auction is similar to a bidding process we see in daily life where each participant is unaware of the other bidders' strategies beforehand [18]. As shown in Figure 6, all bids are revealed simultaneously, and the highest bidder wins the item. Such a game can lead to outcomes where the final offer significantly exceeds the actual value of the item, thereby reducing the utility for all bidders. I find this auction format intriguing because the lack of transparent information parallels many real-life challenges, such as competition in the job market or applying to schools. Thus, finding an efficient approach to this can offer meaningful insights and direction in my everyday life. Below is my simplify problem setting of first-price auction problem.

**Problem setting:** We can simplify this game-theoretic problem using the format of a bid matrix. Let's assume there are two bidders participating in this game, each with the option to submit a low (L), medium (M), or high (H) bid.

**Assumption:** To further simplify this problem, I assume that the true value of the item,  $V$ , and the bid from M are the most similar, and this can bring a profit of 500. As for the bid from H, it would result in a profit of 100. Additionally, I assume that if both bids are the same, the auction will fail to result in a sale, leading to two possible reactions from the participants: 1) Loss aversion, which equates to a utility of -500; 2) Acceptance of a no-win situation, which equates to a utility of 0.

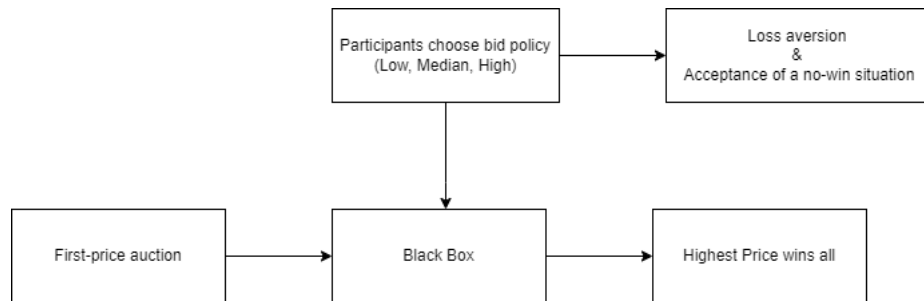
**Implementation:** Building on the aforementioned assumptions and definitions, I have constructed a bidding matrix as shown in Figure 7. In order to calculate its Nash equilibrium, I replicated it using Nashpy, with the link provided as follows:

[Colab](#)

**Result:** As shown in Figure 8, the Nash equilibrium for this problem seems to be reached when both participants choose the high price strategy.

**Discussion:** The results indicate that when bidders are completely unaware of each other's strategies and display a marked aversion to loss, they tend to be uncooperative and drive up prices, ultimately resulting in a detrimental outcome for all parties involved. Consequently, in practice, we need to devise smart policies to prevent such scenarios. Numerous scholars have conducted research on relevant strategies and their intersections with other fields: Wilson [19] talks about how second-price auctions can be effective in enhancing rationality for all participants; Milgrom and Robert [20] design a generalized auction model to control for information asymmetry among bidders and the impact of expectations.

**Extention:** Furthermore, auction issues are also widely applied in interdisciplinary research. For instance, Klemperer [21] discussed the impact of auction rules on overall market efficiency and fairness; Lazar and Nemo [22] addressed the influence of auction issues on network resource allocation; and Cason and Charles [23] explored the effects of auction issues on pollution emission rights.



**Fig. 6.** Flowchart of first-price auction structure

	L	M	H
L	(0, 0)	(-500, 500)	(-500, 100)
M	(500, -500)	(0, 0)	(-500, 100)
H	(100, -500)	(100, -500)	(0, 0)

**Fig. 7.** Bid matrix under simplify assumption for first-price auction problem

```
[34] # Find the Nash Equilibrium with Support Enumeration
      equilibria = game2.support_enumeration()
      for eq in equilibria:
          print(eq)

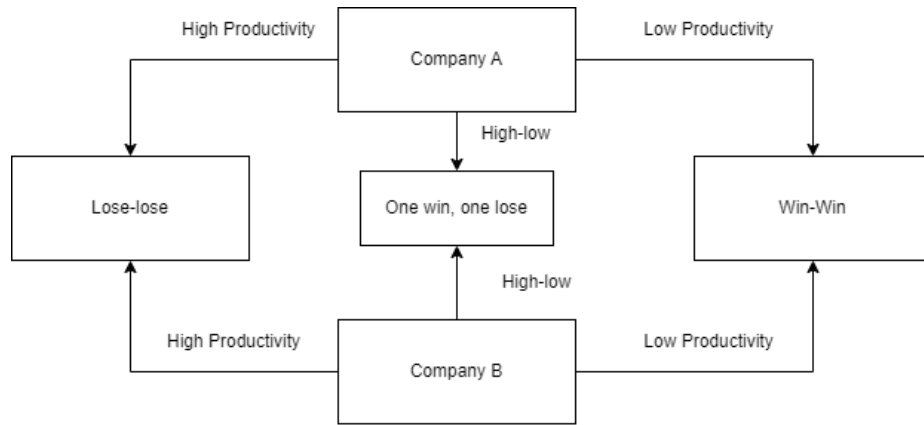
(array([0., 0., 1.]), array([0., 0., 1.]))
```

**Fig. 8.** Result for Nashpy implettation

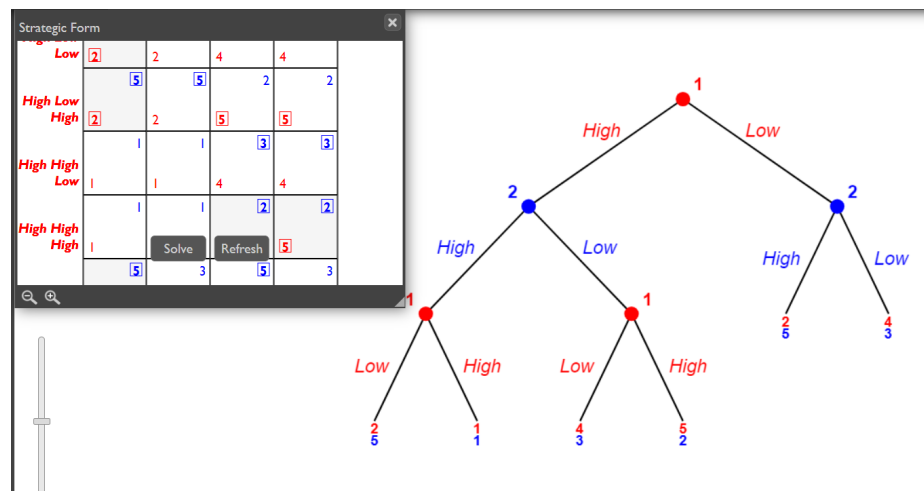
### 5.3 Delving into Extensive-Form Games

An example I use for Extensive-Form Games is Stackelberg competition. This competition involves a power play between two firms regarding their production capacities. When one firm chooses to produce at a higher capacity while the other opts for lower output, the firm that goes for the higher production enjoys a profit; conversely, when both firms select high production, neither benefits [14]. As shown in Figure 9, Firm B is aware of Firm A's initial strategic choice in advance, while Firm A has the possibility to alter its strategy based on Firm B's actions. What makes this issue fascinating is that it involves the concept of infinite games, meaning that their range of actions is theoretically limitless.

Figure 10 displays the simulated game dynamics I conducted using the Game Theory Explorer software. The left branch of the tree encapsulates Firm A's strategic adjustments in response to Firm B's decisions, whereas the right branch showcases a single-layered game scenario. In navigating this process, I encountered the intricate complexities at play. For instance, Firm A can preemptively signal a willingness to cooperate before deciding on a high production level; similarly, Firm B might employ such a feint as well. It becomes apparent that in a system devoid of constraints, both Firm A and Firm B are likely to invariably select high production capacities, irrespective of their adversary's choices.



**Fig. 9.** Flowchart of Stackelberg competition



**Fig. 10.** Game Theory Explorer illustration of Stackelberg competition

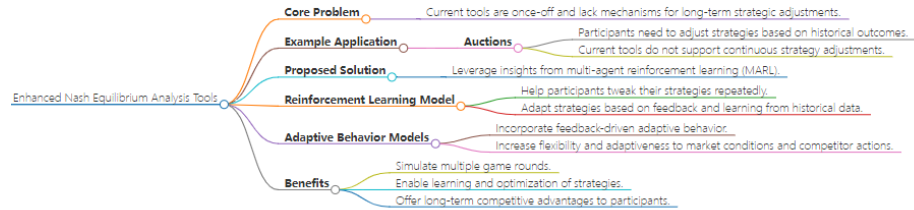
#### 5.4 Critiquing Nash Equilibrium and Envisioning Innovations:

Figure 11 shows the mindmap for this section.

I believe a major problem with current tools for analyzing Nash equilibrium is that they are typically static and do not accommodate long-term strategic adjustments. As discussed by Leibo et al. [24], applying multi-agent reinforcement learning (MARL) can be effective in adapting to complex social dilemmas, demonstrating the need for tools that allow for dynamic strategy evolution. Taking the auction example mentioned earlier, if participants could continuously adjust their strategies based on historical information, it could yield far more interpretable outcomes. Unfortunately, current tools lack this capability.

To address this, insights from previous MARL research, such as those presented by Silver et al. [25], which showcase the power of reinforcement learning in developing evolving strategies, could be leveraged to develop a reinforcement learning model that aids participants in repeatedly tweaking their strategies. Furthermore, incorporating feedback-driven adaptive behavior models, akin to those explored by Cardaliaguet and Hadikhanloo [26], could significantly improve the utility of this framework.

Models of this nature would allow strategies to exhibit increased flexibility and adaptiveness in response to fluctuating market conditions and rival behaviors. By simulating multiple game rounds, this method could empower participants to learn and optimize their strategies for long-term competitive advantage, a concept supported by Nguyen et al. [27] in their review of MARL challenges and applications. Additionally, the optimal regret bounds and policy adaptations discussed by Syrgkanis et al. [28] further underscore the potential of adaptive learning models in continuous games, providing a strong foundation for this approach.



**Fig. 11.** Mindmap of the Critique of Nash Equilibrium and Envisioning Innovations

## 6 Game Theory Glossary Tables

**Table 3.** Game Theory glossary tables.

<b>Glossary</b>	<b>Definition</b>	<b>Sources</b>
Nash Equilibrium	A set of strategies where no player can benefit by changing their strategy while the other players keep theirs unchanged.	[10]
Mixed strategy Nash equilibrium	A strategy where a player chooses between possible moves according to a probability distribution.	[11]
Cooperative game	In cooperative game theory, players can form binding agreements that allow them to share strategies and benefits.	[12]
Non-cooperative game	A non-cooperative game is one where players make decisions independently, without the possibility of forming binding agreements. Text follows	[13]
Dynamic game	Dynamic games are characterized by the sequence of decisions made over time, where earlier choices can influence future opportunities and payoffs. Text follows	[14]



## Bibliography

- [1] E. Shapira, O. Madmon, R. Reichart, and M. Tennenholtz, “Can large language models replace economic choice prediction labs?” *arXiv preprint arXiv:2401.17435*, 2024. [Online]. Available: <https://arxiv.org/abs/2401.17435>
- [2] J. Berg, J. Dickhaut, and K. McCabe, “Trust, reciprocity, and social history,” *Games and economic behavior*, vol. 10, no. 1, pp. 122–142, 1995. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0899825685710275>
- [3] R. Axelrod and W. D. Hamilton, “The evolution of cooperation,” *science*, vol. 211, no. 4489, pp. 1390–1396, 1981. [Online]. Available: <https://www.journals.uchicago.edu/doi/abs/10.1086/383541>
- [4] L. Huang, S. Bi, and Y.-J. A. Zhang, “Deep reinforcement learning for online computation offloading in wireless powered mobile-edge computing networks,” *IEEE Transactions on Mobile Computing*, vol. 19, no. 11, pp. 2581–2593, 2019. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8771176/>
- [5] P. Cappelli and J. Keller, “Talent management: Conceptual approaches and practical challenges,” *Annu. Rev. Organ. Psychol. Organ. Behav.*, vol. 1, no. 1, pp. 305–331, 2014. [Online]. Available: <https://www.annualreviews.org/doi/abs/10.1146/annurev-orgpsych-031413-091314>
- [6] C. Breazeal, *Designing sociable robots*. MIT press, 2004. [Online]. Available: <https://books.google.com/books?hl=en&lr=&id=402dquhxSTQC&oi=fnd&pg=PR15&dq=Designing+sociable+robots&ots=oCRmwZy6xh&sig=X0ojg3nchPo7iYT2Tr-xdWzLXNM>
- [7] M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, *Autonomous driving: technical, legal and social aspects*. Springer Nature, 2016. [Online]. Available: <https://library.oapen.org/bitstream/handle/20.500.12657/27811/1/1002194.pdf>
- [8] G. C. Homans, “Social behavior as exchange,” *American journal of sociology*, vol. 63, no. 6, pp. 597–606, 1958. [Online]. Available: <https://www.journals.uchicago.edu/doi/abs/10.1086/222355>
- [9] J. Von Neumann, “On the theory of games of strategy,” *Mathematische Annalen*, vol. 100, pp. 295–320, 1928. [Online]. Available: <https://www.degruyter.com/document/doi/10.1515/9781400882168-003/pdf?licenseType=restricted>
- [10] J. F. Nash *et al.*, “Non-cooperative games,” 1950. [Online]. Available: <https://www.elgaronline.com/downloadpdf/monochap/book/9781781956298/book-part-9781781956298-9.pdf>
- [11] J. F. Nash Jr, “Equilibrium points in n-person games,” *Proceedings of the national academy of sciences*, vol. 36, no. 1, pp. 48–49, 1950. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.36.1.48>

- [12] J. Nash, "Two-person cooperative games," *Econometrica: Journal of the Econometric Society*, pp. 128–140, 1953. [Online]. Available: <https://www.jstor.org/stable/1906951>
- [13] L. S. Shapley *et al.*, "A value for n-person games," 1953. [Online]. Available: <https://www.degruyter.com/document/doi/10.1515/9781400829156-012/pdf?licenseType=restricted>
- [14] R. Selten, "Spieltheoretische behandlung eines oligopolmodells mit nachfrageträgheit: Teil i: Bestimmung des dynamischen preisgleichgewichts," *Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics*, no. H. 2, pp. 301–324, 1965. [Online]. Available: <https://www.jstor.org/stable/40748884>
- [15] J. C. Harsanyi, "Games with incomplete information played by "bayesian" players, i–iii part i. the basic model," *Management science*, vol. 14, no. 3, pp. 159–182, 1967. [Online]. Available: <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.14.3.159>
- [16] D. Fudenberg and J. Tirole, *Game theory*. MIT press, 1991. [Online]. Available: <https://books.google.com/books?hl=en&lr=&id=pFPHKwXro3QC&oi=fnd&pg=PR15&dq=Game+Theory&ots=P46KrpvjKx&sig=UDu8tzvyPLvIA3WGAcsS0N8H4s>
- [17] S. Tadelis, *Game theory: an introduction*. Princeton university press, 2013. [Online]. Available: [https://books.google.com/books?hl=en&lr=&id=\\_4OqAaITAWMC&oi=fnd&pg=PP1&dq=Game+Theory:+An+Introduction&ots=oMS7Q4NmQt&sig=kEzeRDv54YMuH1KodDOZN\\_ZyJ84](https://books.google.com/books?hl=en&lr=&id=_4OqAaITAWMC&oi=fnd&pg=PP1&dq=Game+Theory:+An+Introduction&ots=oMS7Q4NmQt&sig=kEzeRDv54YMuH1KodDOZN_ZyJ84)
- [18] W. Vickrey, "Counterspeculation, auctions, and competitive sealed tenders," *The Journal of finance*, vol. 16, no. 1, pp. 8–37, 1961. [Online]. Available: <https://www.jstor.org/stable/2977633>
- [19] R. Wilson, "Auctions of shares," *The Quarterly Journal of Economics*, vol. 93, no. 4, pp. 675–689, 1979. [Online]. Available: <https://academic.oup.com/qje/article-abstract/93/4/675/1932634>
- [20] P. R. Milgrom and R. J. Weber, "A theory of auctions and competitive bidding," *Econometrica: Journal of the Econometric Society*, pp. 1089–1122, 1982. [Online]. Available: <https://www.jstor.org/stable/1911865>
- [21] P. Klemperer, "What really matters in auction design," *Journal of economic perspectives*, vol. 16, no. 1, pp. 169–189, 2002. [Online]. Available: <https://www.aeaweb.org/articles?id=10.1257/0895330027166>
- [22] A. A. Lazar and N. Semret, "The progressive second price auction mechanism for network resource sharing," in *8th International Symposium on Dynamic Games, Maastricht, The Netherlands*, 1998. [Online]. Available: [https://www.researchgate.net/profile/Aurel-Lazar-2/publication/2408597\\_The\\_Progressive\\_Second\\_Price\\_Auction\\_Mechanism\\_for\\_Network\\_Resource\\_Sharing/links/5437d8280cf2590375c55b68/The-Progressive-Second-Price-Auction-Mechanism-for-Network-Resource-Sharing.pdf](https://www.researchgate.net/profile/Aurel-Lazar-2/publication/2408597_The_Progressive_Second_Price_Auction_Mechanism_for_Network_Resource_Sharing/links/5437d8280cf2590375c55b68/The-Progressive-Second-Price-Auction-Mechanism-for-Network-Resource-Sharing.pdf)
- [23] T. N. Cason and C. R. Plott, "Epa's new emissions trading mechanism: a laboratory evaluation," *Journal of environmental economics and*

- management*, vol. 30, no. 2, pp. 133–160, 1996. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0095069696900108>
- [24] J. Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki, and T. Graepel, “Multi-agent reinforcement learning in sequential social dilemmas,” *arXiv preprint arXiv:1702.03037*, 2017. [Online]. Available: <https://arxiv.org/abs/1702.03037>
- [25] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel *et al.*, “A general reinforcement learning algorithm that masters chess, shogi, and go through self-play,” *Science*, vol. 362, no. 6419, pp. 1140–1144, 2018. [Online]. Available: <https://www.science.org/doi/abs/10.1126/science.aar6404>
- [26] P. Cardaliaguet and S. Hadikhanloo, “Learning in mean field games: the fictitious play,” *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 23, no. 2, pp. 569–591, 2017. [Online]. Available: <https://www.esaim-cocv.org/articles/cocv/abs/2017/02/cocv160004-s/cocv160004-s.html>
- [27] T. T. Nguyen, N. D. Nguyen, and S. Nahavandi, “Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications,” *IEEE transactions on cybernetics*, vol. 50, no. 9, pp. 3826–3839, 2020. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9043893/>
- [28] V. Syrgkanis, V. Lei, M. Oprescu, M. Hei, K. Battocchi, and G. Lewis, “Machine learning estimation of heterogeneous treatment effects with instruments,” *Advances in Neural Information Processing Systems*, vol. 32, 2019. [Online]. Available: <https://proceedings.neurips.cc/paper/2019/hash/3b2acfe2e38102074656ed938abf4ac3-Abstract.html>