Integrating GAI and Human Strategy: A New Paradigm in Game Theory*

Jiayang Hong[1]

Duke Kunshan University, Kunshan, Jiangsu 215316, China
jh794@duke.edu
Github

Abstract. This paper proposes a novel framework for game theory, integrating Artificial Intelligence (AI) and human strategic interactions. It addresses the limitations of traditional models by introducing adaptive strategies and evolutionary dynamics to better reflect the complexity of modern strategic environments. This approach not only enhances my understanding of strategic decisions in varied contexts but also paves the way for a dynamic interplay between AI and human decision-making, leading to more resilient and efficient systems. The proposed framework is demonstrated through a hypothetical evolving chess game model, illustrating the potential for AI-driven innovation in game theory.
Notes: In submission to Problem Set 2 for COMPSCI/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 · Artificial Intelligence · Strategic Decision-Making · Adaptive Strategies

1 Introduction

This research proposes a novel integrative framework that addresses the dynamic complexities of game theory where artificial intelligence and human decision-making converge. This paradigm introduces adaptive strategies, evolutionary dynamics, and the potential for AI-driven innovation in strategic thinking, which are crucial for systems exhibiting unpredictability and complex interdependencies.

* **Acknowledgments**: I extend my deepest gratitude to Professor Luyao Zhang, whose guidance was invaluable throughout this research project. I am also grateful to my classmates for their stimulating exchanges and collaborative spirit that greatly enhanced our collective learning experience. Lastly, I must express my heartfelt appreciation to my family and friends, whose encouragement and belief in my abilities sustained me through this academic journey. Their support has been a cornerstone of my success.

Fig. 1. Evolving Chess Game Theory Model

Through the metaphor of an evolving chess game, where pieces learn from each move and the board changes after every game, this study illustrates the inadequacies of traditional game theory's static rules. By creating a model that learns and adapts, this work lays the foundation for systems where AI and humans collaborate, reflecting a true understanding of the strategic complexities faced in modern decision-making environments. This approach is not merely theoretical; it serves as a practical blueprint for designing systems in various domains, from market strategies to resource management, capturing the dynamic interplay of strategic decisions that shape future outcomes.

2 Background

Reimagining game theory in the age of AI, especially considering the advent of General Artificial Intelligence (GAI) and the complexity of modern systems, calls for an expansive framework that transcends traditional boundaries. The rise of AI and GAI challenges the conventional game environment, which has historically been marked by clearly defined players, strategies, and payoffs, and has considered agents as either rational humans or AI with honest or malicious intents.

In the reimagined game theory shown in figure 2:

Game Environment: The game environment would become more fluid and adaptive. With GAI, we would expect to see games where the rules and strategies evolve autonomously as the game progresses, much like learning algorithms evolve through experience [1]. This evolution would be akin to the real world, where the 'rules of the game' are constantly changing, and the strategies are not fixed but dynamically adjust based on the environment and the agents' learning processes.

Agents: The concept of agency would expand to include not just rational or irrational agents but also self-learning and adapting GAIs capable of exceeding human cognitive capabilities [2]. These agents would be able to learn from the environment, other agents, and their own previous actions. The distinction between human and AI agents would blur as augmented intelligence becomes commonplace, and collaborative human-AI interaction becomes a critical focus of strategic decision-making

Human-AI Interaction: The progress in AI and GAI is likely to revolutionize the way humans connect, behave, and interact. We would expect to see a symbiosis of human and artificial intelligence, where strategic interactions involve not only competition but also cooperation and collective intelligence [3]. Games in this context would not be zero-sum but rather would emphasize shared benefits and the global optimum, encouraging behaviors that lead to mutual gain and system-wide resilience [4].

Future of Game Theory: In this advanced milieu, game theory would encompass more complex scenarios, such as those involving network effects, where an agent's payoff depends not only on their strategy but also on the strategies of all other agents in their network [5]. The field would also have to consider the implications of superintelligent systems, where GAIs could potentially manipulate the game environment or strategies of other agents.

In conclusion, the progression in AI and human intelligence is set to redefine the foundational constructs of game theory, steering it toward a future where the lines between humans and AI are increasingly interwoven. This shift will demand new theoretical models that can accommodate the complexity, adaptability, and collaborative potential of advanced intelligent systems.

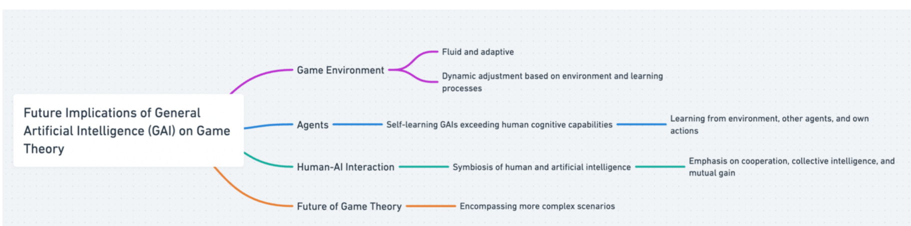

Fig. 2. Future Implications of GAI on Game Theory

3 An Illustration Example: Adaptive AI in the Centipede Game

This section presents a practical illustration of how our integrative game theory framework—incorporating both Artificial Intelligence and adaptive strategies—enhances decision-making in a complex game environment. The example chosen for this demonstration is the Centipede Game, a classic in game theory literature known for its sequential decision-making and strategic depth.

3.1 Game Description

The Centipede Game(shown in figure 3) involves two players alternately deciding whether to "continue" the game or "take" the pot, which results in the game ending with the pot being split, favoring the player who chose to "take." If a player decides to "continue," the total value of the pot increases and the decision shifts to the other player. This continues until a set number of rounds are reached or a player decides to take the pot.

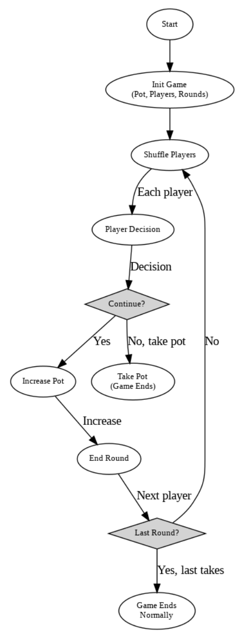

Fig. 3. Centipede game description

3.2 Incorporating AI

To incorporate adaptive AI, particularly leveraging ChatGPT-4, we integrate a predictive model trained on historical gameplay data. This AI aids players by suggesting optimal strategies based on current game state, previous moves, and an analysis of potential future moves.

AI Integration Mechanism 1. **Data Collection**: Gather extensive gameplay data to train the AI, ensuring it understands various strategies and outcomes based on different game states. 2. **Model Training**: Use the collected data to train ChatGPT-4 that can predict outcomes based on different decisions. 3. **Real-Time Decision Support**: During gameplay, input the current state of the game into ChatGPT-4, which uses the trained model to provide strategy advice in real-time. For the detailed mechanism, see this Colab notebook.

3.3 Dynamic Strategy Adaptation

The AI system dynamically adapts its recommendations based on the flow of the game, learning from each move to update its strategy predictions. This process involves: - Analyzing the potential benefits of "continuing" versus "taking" the pot at each step. - Estimating the opponent's likely response to different actions based on historical behavior patterns.

Example AI Interaction *Player's Query*: "The pot is currently at $100. If I continue, the pot increases to $150. Should I continue or take the pot?" *AI Response*: "Based on the analysis of past games and current strategies, continuing is recommended. Historical data indicates that in similar situations, continuing maximizes potential future gains, especially considering the observed patterns of your opponent's responses."

3.4 Outcome and Benefits

Utilizing AI in this manner not only enhances the strategic depth of the game but also aids players in making more informed decisions, potentially leading to higher overall gains and a deeper understanding of strategic interactions in variable scenarios. This approach exemplifies how AI can transform traditional game theory into a more dynamic and predictive science, aligned with the realities of modern strategic environments.

3.5 Human Welfare Implications

This dynamic version of the Centipede Game serves as a concrete illustration of how integrating GAI into game theory can significantly improve human welfare. It does so by fostering cooperation and trust in situations where short-term incentives might traditionally lead to suboptimal outcomes. In broader applications, such as international negotiations or business strategies, this approach could lead to more collaborative and mutually beneficial outcomes, reflecting a significant advance over existing models that lack adaptability. This example not only illustrates the integration of AI into traditional game theory but also highlights its potential to transform strategic interactions in various fields, suggesting a new paradigm where AI and human decision-making coevolve to meet complex challenges.

3.6 Conclusion

The integration of AI into the Centipede Game demonstrates a significant advancement in game theory applications, highlighting the shift towards more dynamic, informed, and strategic decision-making processes in games and real-world scenarios alike.

A The Pioneers in the History of Game Theory

This section traces the pivotal milestones in the evolution of game theory, highlighting significant theoretical advancements and the contributions of key figures whose insights have shaped the landscape of economic and strategic thinking. Figure 4 encapsulates these developments, providing a visual representation of the critical junctures and thematic shifts within game theory's history.

- 1928: **Transition from Decision Theory to Game Theory**
  von Neumann [6] pioneers game theory, which extends beyond individual decision-making in decision theory, introducing a framework for strategic interactions among rational players.

- **1950: Evolution from Pure-Strategy Nash Equilibrium to Mixed-Strategy Nash Equilibrium**
  Nash develops the Nash Equilibrium, proving that every game with a finite set of strategies has an equilibrium, not just in pure strategies but also in mixed strategies where players randomize their strategy selections. [7].

- **1953: Differentiation between Non-Cooperative and Cooperative Games**
  Nash further categorized game theory into non-cooperative games, involving independent decision-making, and cooperative games, permitting coalition formation and enforceable agreements [8].

- **1965: Progression from Static to Dynamic Games**
  Selten [9] introduced the subgame perfect equilibrium, broadening the scope of game theory to include dynamic games where strategies evolve over time.

- **1967: From Perfect to Imperfect Information Games**
  Harsanyi addresses the complexities of games where players do not have complete knowledge about other players' actions or preferences, leading to the development of Bayesian games to model these scenarios with incomplete information [10].

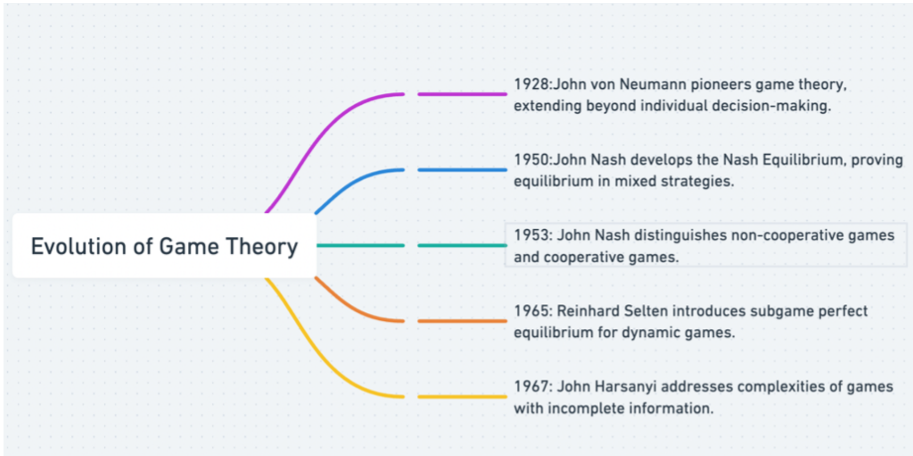

Fig. 4. Evolution of Game Theory

B Review Classic Games, Nash Equilibrium and the Analytical Tools

B.1 Bayesian Nash Equilibrium: Definition and Application

Theoretical Framework

Definition from Textbook Referencing Osborne and Rubinstein [11], Bayesian Nash Equilibrium incorporates players' private information into their strategies. Unlike the standard Nash Equilibrium, this concept assumes that players have incomplete information but hold beliefs about other players' types (private information).

Definition 1 (Bayesian Nash Equilibrium). *A strategy profile σ^* is a Bayesian Nash Equilibrium if, for each player i and any type t_i, the strategy $\sigma_i^*(t_i)$ maximizes i's expected payoff given the strategies σ_{-i}^* of the other players, taking into account the probability distribution over types.*

$$E[u_i(\sigma_i(t_i), \sigma^*_{-i}(t_{-i}))|t_i] \geq E[u_i(\sigma_i, \sigma^*_{-i}(t_{-i}))|t_i], \forall \sigma_i \in S_i$$

Analytical Frameworks From CS Perspective

From Classical to Bayesian Interpretations Expanding upon the framework detailed in Shoham and Leyton-Brown [12], the Bayesian Nash Equilibrium is pivotal for games where players consider not only strategies but also the types of other participants. The set of agent i 's best responses to mixed-strategy profile s_{-i} are given by

$$BR_i\left(s_{-i}\right) = \arg\max_{s_i' \in S_i} EU_i\left(s_i', s_{-i}\right).$$

A Bayes-Nash equilibrium is a mixedstrategy profile s that satisfies

$$\forall i \quad s_i \in BR_i\left(s_{-i}\right)$$

This equilibrium occurs when each player's strategy maximizes their expected utility based on their knowledge of other players' types, thus adapting the classic Nash Equilibrium to settings with incomplete information. Each player assesses potential outcomes by considering every possible type scenario of their opponents, ensuring their strategy remains optimal regardless of uncertainties.

Theorem 1. *Every finite game with incomplete information, characterized by a type space and a payoff function conditional on types, admits at least one Bayesian Nash Equilibrium.*

Proof. For each player i, let S_i be the set of strategies. Define a continuous mapping $f : S \to S$ by selecting the best response for each type and player. By applying the Fixed Point Theorem, such as Kakutani's, there exists at least one fixed point in S, corresponding to a Bayesian Nash Equilibrium, as no player can improve their expected payoff by unilaterally changing their strategy given their belief about other players' types.

Discussion Comparing classical Nash Equilibrium and Bayesian Nash Equilibrium highlights the extension from perfect to imperfect information scenarios, allowing for a more realistic modeling of strategic interactions in diverse environments such as economics and political science. This comparison illustrates the adaptability and depth of game theory in addressing more complex decision-making processes where the information is asymmetrical.

B.2 Exploring Inspirational Games in Strategic or Normal Form

One game from game theory literature that stands out is the Prisoner's Dilemma, a fundamental example in the study of strategic interactions used across various fields, including economics, political science, and evolutionary biology [13].The Prisoner's Dilemma is a twoplayer, nonzerosum game that demonstrates why two rational individuals might not cooperate, even if it appears that it is in their best interests to do so. The game is set up with the following payoffs:

- Cooperate-Cooperate (CC): Both players cooperate and receive a reward (R).
- Defect-Defect (DD): Both players defect and get a punishment (P).
- Cooperate-Defect (CD) and Defect-Cooperate (DC): The defector receives a temptation payoff (T), while the cooperator receives the sucker's payoff (S).

Typical payoff values satisfying $T > R > P > S$ and $2R > T+S$ ensure defecting is always a dominant strategy, yet mutual defection leads to worse outcomes than mutual cooperation [14]. For a detailed example, see this Colab notebook.

Why the Prisoner's Dilemma is Inspiring The elements of the Prisoner's Dilemma that capture my interest are its simplicity and the profound implications it has on understanding human behavior and cooperation. Despite the straightforward setup, it illustrates complex concepts such as trust, betrayal, and the conflict between individual and collective rationality. The game serves as a fundamental model for exploring the conditions under which cooperation can emerge in competitive environments.

Impact and Significance The Prisoner's Dilemma is not just a theoretical construct in game theory; it serves as a powerful explanatory tool across various fields, vividly illustrating the complexities of human and strategic interactions. Personally, I find the application of the Prisoner's Dilemma in international relations particularly compelling. It sheds light on why nations, much like individuals in the classic dilemma, often escalate arms races despite the mutual benefits of disarmament. This game theorizes the underlying tension between cooperation and competition, helping explain critical decisions on national and global scales. The game fundamentally enhances our understanding of strategic interactions by demonstrating how the structure of incentives can decisively influence behavior. It eloquently captures the paradox where individually rational decisions often culminate in collectively irrational outcomes. This key insight has reshaped my perspective on the dynamics of decision-making in scenarios where the interests of individuals and the collective are at odds. In real-world situations—from corporate negotiations to environmental agreements—the Prisoner's Dilemma provides a critical lens through which we can evaluate the efficacy and morality of various strategic choices [15]. In sum,as shown in figure 5, the Prisoner's Dilemma is not just a model of conflict or cooperation; it is a lens that magnifies the underlying mechanics of interaction, whether they be among cells, individuals, corporations, or nations. Its theoretical simplicity and practical depth make it an indispensable tool in the intellectual toolkit for anyone interested in the profound complexities of strategic decisions.

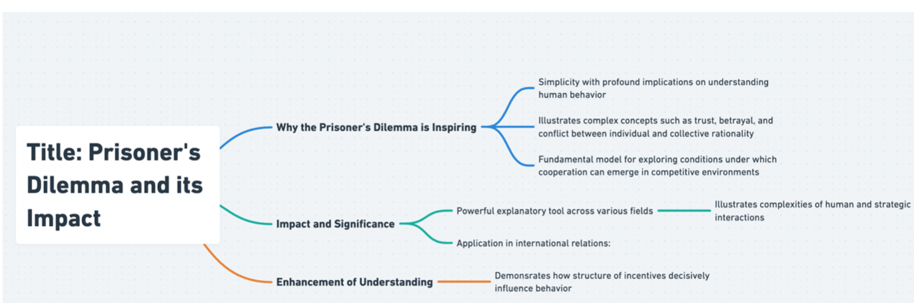

Fig. 5. Prisoner's Dilemma and its impact

B.3 Delving into Extensive-Form Games

Delving into the realm of extensive-form games within game theory allows for a deeper exploration of decision processes that unfold over time, where the order of moves is crucial and information can be perfect or imperfect. One extensive-form game that stands out in this context is the Centipede Game. This game captivates with its exploration of rationality, trust, and foresight among players, presenting a striking departure from simpler, simultaneous-move games.

Description of the Centipede Game The Centipede Game is typically played by two players who take turns deciding whether to continue or stop the game. The game is structured such that each turn involves a decision that increases the potential overall payoff for both players if they continue, but each player's immediate payoff is higher if they choose to stop the game and take the current payoff. If a player chooses to stop the game, they receive a larger share of the current payoff, and the other player receives a smaller share. The game usually ends after a pre-defined number of moves unless a player stops it earlier.

Originally proposed by Rosenthal in 1981, the Centipede Game challenges the standard predictions of backward induction in game theory. If players are purely rational and selfish, predicting that their opponent will also play rationally, the game should end on the first move shown in figure 6. However, empirical studies often show that players continue past the first move, suggesting deviations from purely rational behavior [16, 17].

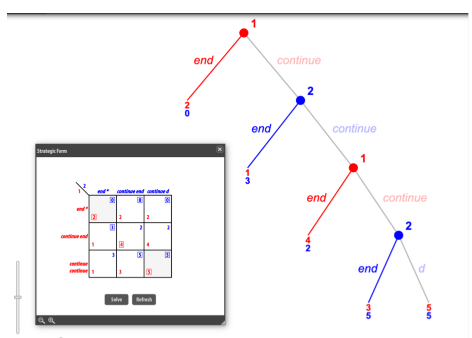

Fig. 6. Solution of Centipede Game

Personal Engagement and Influence Engaging with the Centipede Game has profoundly impacted my understanding of strategic thinking and decision-making processes in complex scenarios. Concluded from figure 8, the game illustrates that real-world decision-making often deviates from theoretical predictions due to factors like trust, risk tolerance, and expectations about others' behavior. It challenges the assumption that individuals always act on purely rational calculations, highlighting the role of psychological and social factors in strategic interactions.

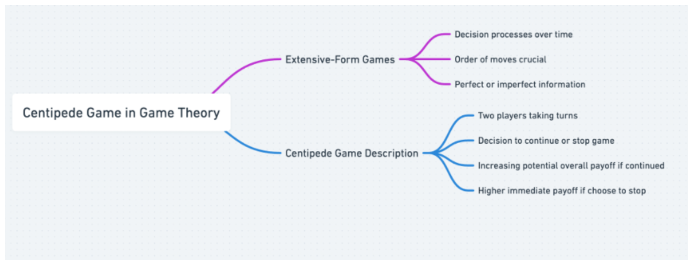

Fig. 7. Centipede Game

B.4 Critiquing Nash Equilibrium and Envisioning Innovations:

Figure 7 indicates that Nash Equilibrium assumes complete rationality and knowledge among players, which rarely aligns with real-world scenarios. One of its key limitations is the multiplicity of equilibria, making it challenging to predict which equilibrium will be selected without additional criteria. NE primarily addresses non-cooperative games and often overlooks the dynamic nature of strategic interactions, failing to account for changes over time or the sequence of moves leading to an equilibrium.

Current tools like Nashpy, QuantEcon, Gambit, and Game Theory Explorer are invaluable for calculating NE but inherit these conceptual limitations. They often focus on static analysis and can be complex, presenting barriers for those without a programming background.The Centipede Game vividly illustrates these limitations. Theoretically, rational players should stop at the first opportunity to secure a higher payoff, contradicting empirical observations where players frequently continue, influenced by factors like trust and risk aversion. This discrepancy highlights the gap between theoretical predictions of NE and actual human behavior.

Proposed Innovation: Dynamic Game Theory Analyzer (DGTA) To address these issues, I propose developing the Dynamic Game Theory Analyzer (DGTA) illustrated in figure 8 , a tool designed to model the dynamic aspects of game theory, focusing on how strategies evolve over time under varying conditions.DGTA would feature a dynamic simulation engine to model strategy evolution and incorporate behavioral models to account for non-rational decisions. It would offer interactive learning modules to help users grasp complex concepts and provide multi-scenario analysis capabilities to understand strategic differences across game variations. Importantly, it would extend to cooperative games, analyzing coalition formation and stability.With a background in psychology, economics, and computer science, I am uniquely positioned to contribute to DGTA's development. My experience integrates insights from behavioral economics into game theory, enhanced by a passion for software development and AI. This interdisciplinary approach would help create a tool that makes sophisticated game theory analysis more accessible and relevant.

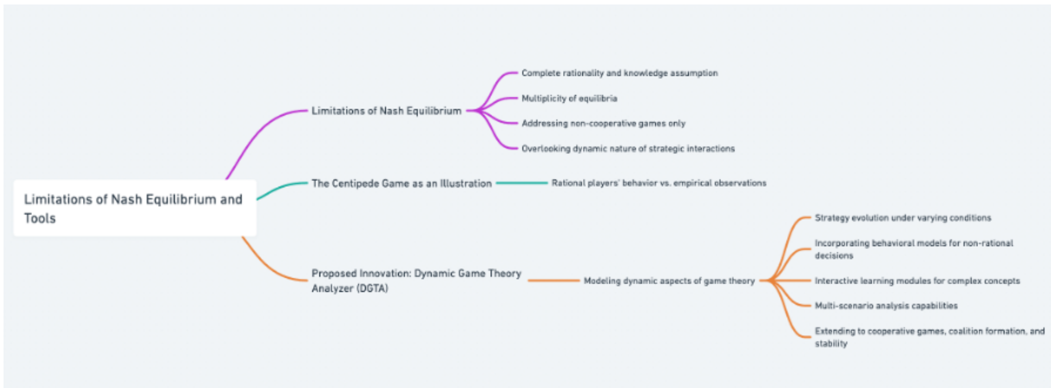

Fig. 8. Limitations of Nash Equlibrium and Tools

C Game Theory Glossary Tables

Table 1. Glossary of Basic Game Theory Terms

Term	Definition	Source
Zero-sum Game	A situation in game theory in which one participant's gain or loss is exactly balanced by the losses or gains of the other participants.	von Neumann [18]
Pareto Efficiency	A state of allocation of resources in which it is impossible to make any one individual better off without making at least one individual worse off.	Pareto [19]
Correlated Equilibrium	A solution concept that generalizes Nash Equilibrium by allowing players to coordinate strategies through shared signals.	Aumann [20]
Evolutionary Stable Strategy	A strategy which, if adopted by a population in a given environment, cannot be invaded by any alternative strategy that is initially rare.	Smith and Price [21]
Mechanism Design	A field in economics and game theory that takes an engineering approach to designing economic mechanisms or incentives, toward desired objectives, in strategic settings where players act rationally.	Hurwicz [22]

Bibliography

[1] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Pearson, 2016.

[2] N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

[3] J. Hawkins, *A Thousand Brains: A New Theory of Intelligence*. Basic Books, 2021.

[4] M. O. Jackson, *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors*. Pantheon, 2020.

[5] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[6] J. von Neumann, "Zur theorie der gesellschaftsspiele," *Mathematische Annalen*, vol. 100, pp. 295–320, 1928.

[7] J. Nash, "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences*, vol. 36, no. 1, pp. 48–49, 1950.

[8] ——, "Two-person cooperative games," *Econometrica*, vol. 21, no. 1, pp. 128–140, 1953.

[9] R. Selten, "Spieltheoretische behandlung eines oligopolmodells mit nachfrageträgheit," *Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics*, vol. 121, pp. 301–324, 667–689, 1965.

[10] J. Harsanyi, "Games with incomplete information played by 'bayesian' players," *Management Science*, vol. 14, no. 3, pp. 159–182, 320–334, 486–502, 1967.

[11] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. Cambridge, MA: MIT Press, 1994. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0899825699907236

[12] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge, UK: Cambridge University Press, 2008. [Online]. Available: http://www.masfoundations.org/mas.pdf

[13] D. Ross, "Prisoner's dilemma," *Stanford Encyclopedia of Philosophy*, 2014. [Online]. Available: https://plato.stanford.edu/archives/win2014/entries/prisoner-dilemma/

[14] S. Kuhn. (2019) Prisoner's dilemma. Accessed: 2020-01-09. [Online]. Available: https://plato.stanford.edu/entries/prisoner-dilemma/

[15] J. Chappelow. (2019) Understanding the prisoner's dilemma. [Online]. Available: https://www.investopedia.com/terms/p/prisoners-dilemma.asp

[16] R. W. Rosenthal, "Games of perfect information, predatory pricing and the chain-store paradox," *Journal of Economic Theory*, vol. 25, no. 1, pp. 92–100, 1981.

[17] R. D. McKelvey and T. R. Palfrey, "An experimental test of the centipede game," *Econometrica*, vol. 60, no. 4, pp. 803–836, 1992.

[18] J. von Neumann, "Zur theorie der gesellschaftsspiele," *Mathematische Annalen*, vol. 100, pp. 295–320, 1928.
[19] V. Pareto, *Cours d'Économie Politique*. F. Rouge, 1896, vol. 1.
[20] R. J. Aumann, "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics*, vol. 1, no. 1, pp. 67–96, 1974.
[21] J. M. Smith and G. R. Price, "The logic of animal conflict," *Nature*, vol. 246, no. 5427, pp. 15–18, 1973.
[22] L. Hurwicz, "Designing economic mechanisms," 2006.