# When blockchain meets game theory: A mechanism that eliminates the necessity of trust in cooperations ⋆

Jiaqi Yao[1]

Duke Kunshan University, Kunshan, Jiangsu 215316, China
jy336@duke.edu
*LinkedIn*

**Abstract.** This article presents a comprehensive review and critique of contemporary game theory, with a particular focus on its application within computational economics and emerging technological platforms like blockchain and the metaverse. Traditional game theory models often assume perfect information and rational behavior, which do not fully capture the nuances of human decision-making characterized by bounded rationality and strategic complexity in real-world scenarios. This paper analyzed several games as examples, introducing their basic logic under Nash Equilibrium and offering innovative insight into better solving the game. By integrating case studies and theoretical insights, the paper demonstrates how this mechanism can foster more equitable and competitive market environments, particularly in digital and decentralized systems

*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.*

⋆ **Acknowledgments**: Thanks for the guidance and advice from the course instructor: Luyao Zhang. Also, thanks to Xinagnan Hu for writing me a useful peer evaluation and my classmates for offering solid ideas in class discussions and offering help when needed.

**Keywords:** computational economics · game theory · innovative education · blockchain

## 1 Introduction

Past game theory has already mentioned imperfect information, which applies to most of the realistic market cases nowadays but lacks a solution for the market considering humanity's bounded rationality and ability to strategically play games. The main point we usually found was that rational players don't have trust in each other in games. The game that buyers entering a second-hand car market may be an instance, where buyers always don't have the perfect information and cannot utilize their benefits as long as they don't want to leave the market. Even if we hold a system rating the performance and provide enforcement and punishment policies for sellers, there can always be strategy-playing sellers that perform well for a while and then cheat and leave the market, with themselves utilized. Other games like public good games [1] and centipede games[2] hold this problem more seriously, that for rational players this game will only end at the starting point, even if all of us know that continuing the game will result in higher benefits for both parties.

Therefore, I propose to create a mechanism for the third party to solve this game, by offering smart contracts to both parties, the benefit will be maximized at the end, though the center (third party) will take an amount of prize away. This solution will always make games with imperfect information transform into perfect information, and provide both parties with the same profits that even rational players will choose to continue the game. This mechanism will provide a perfect competition environment to help solve the information gap and achieve an equilibrium for each good in this market, applied to blockchain platforms or metaverse.

## 2 Background

With the development of technology, the progress of AI may reshape human behavior, strategic decision-making, and other factors from some perspectives suggested below.

The evolution from AI into GAI shows opportunities for collaborative interactions between humans and AI, suggesting a new environment for traditional game theory. Rahwan et al. [3] discussed the social and ethical implications of human-AI interaction, emphasizing that we need a collaborative system for humans and AI agents to work together for common goals. (see figure 1) The strong data processing ability with human creativity and strategic thinking, will lead to a larger population-based and more complex game theory.

Current AI systems also create dynamic and adaptive game environments that evolve. With growing capability and deep learning mechanisms, better

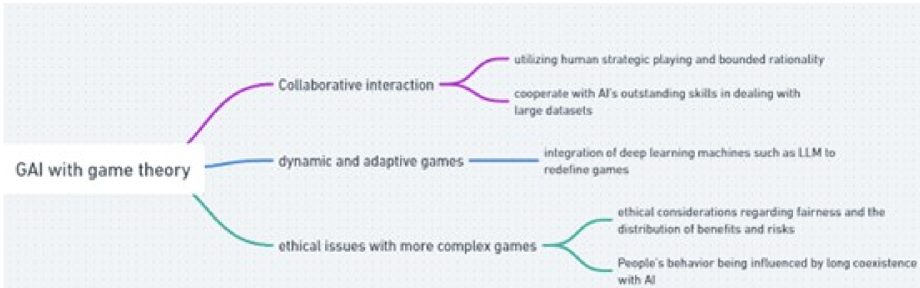



**Fig. 1.** In the times with GAI

strategies, payoffs, and agent characteristics may be defined by AI. Silver et al. [4] demonstrate the capabilities of AI in dynamic environments through AlphaGo's success. This shows the potential for AI to redefine game dynamics and outcomes.

Furthermore, the integration of GAI within societal frameworks raises ethical considerations regarding fairness and the distribution of benefits and risks. Bostrom and Yudkowsky [5] discussed the ethical implications of superintelligent AI systems, suggesting the importance of aligning AI objectives with human values. This is a way that AI-human interaction may cause some differences between human agents. The newly developed game theory should these ethical dimensions and find a manner that maximizes societal welfare and minimizes harm.

If we keep the current evolution for GAI and human interaction, future developments in AI may continue to challenge the boundaries of traditional game theory. In current society, we always need to find innovative approaches to understand and guide the complex interactions between humans and AI agents.

## 3 An Illustration Example

Figure 2 shows how the traditional Centipede game was played, under Nash equilibrium (1950), rational players always end the game at the start of the game. I used a simple matrix to represent an easy centipede game, see Figure 3.

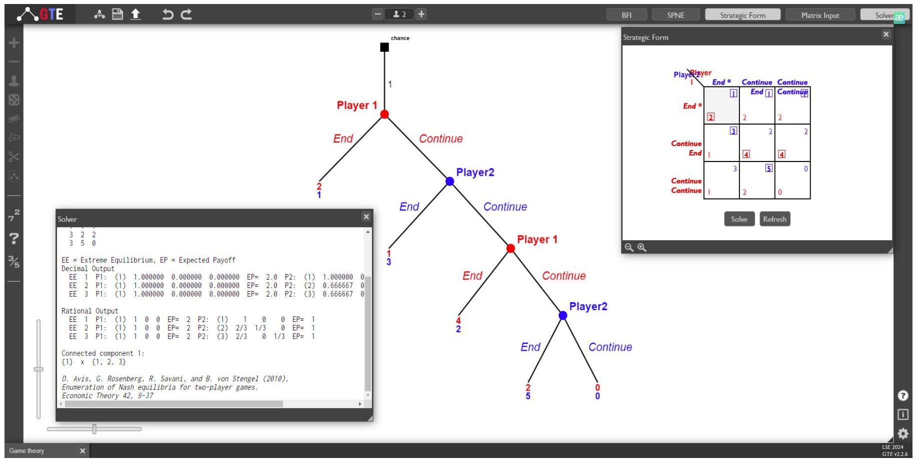



**Fig. 2.** Illustration for the Centipede game

| | B:T | B:C |
|---|---|---|
| A:T | (3,1) | (3,1) |
| A:C | (2,4) | (5,5) |

**Fig. 3.** matrix for the Centipede game

As Figure 3 shows, T represents take and C represents continue, even though it is apparent that both parties choose to continue will maximize the profit, by interpreting a QuantEcon package for Nash Equilibrium, the array outcome indicates that Player 1 has full probability to take at the beginning. See the link below:Colab QuantEcon The first array indicates that Player A takes with certainty (effectively the same as the first profile but represented with floating point precision). The second array indicates that Player B takes approximately 67 probability and continues with approximately 33 probability. However, it will be impossible for the second player to choose if player one has already taken.

### 3.1 Current research

There hasn't been much research on the centipede game, but one inspiring and notable research was from Brams and Kilgour (2020), they revised the rule of the centipede game in a way that when player chooses to cooperate in each turn, the reward will be exchanged to each other, as shown in Figure 4 The result is that for rational players they naturally have the motivation to cooperate in the game since their reward for continue is always larger than the reward for taking. However, it is limited to its application scenario since in real-life examples there may not be so many cases that players can find apparent higher rewards for profit, otherwise, no betraying and competition will happen. What is more, by changing the rules of the game, the original idea of asking players to make a choice between short-term quick benefits and long-term cooperation was meaningless since normal players would just choose to cooperate in this mode.

Therefore, I would like to design a method that does not change the game rule, but a new type to play the game that makes all parties engaged better off compared to before this methodology was integrated.

| | B:T | B:C |
|---|---|---|
| A:T | (1,3) | (1,3) |
| A:C | (4,2) | (5,5) |

**Fig. 4.** Illustration for current research

### 3.2 methodology on solving the game

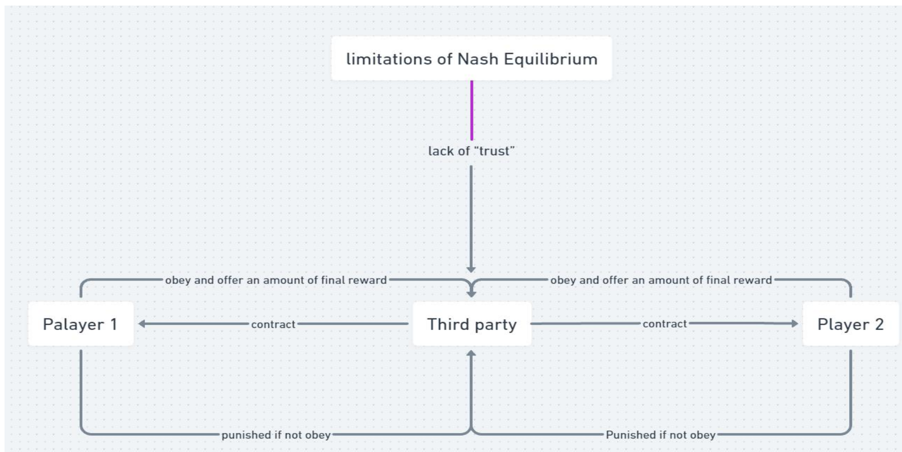



**Fig. 5.** suggestion for revised Centipede game

| | B:T | B:C |
|---|---|---|
| A:T | (t,t) | (t,t) |
| A:C | (t,t) | (4.9,4.9) |

**Fig. 6.** revised matrix with blockchain centipede game

To solve this, I proposed a web3 (see Figure 5) that includes the idea of a smart contract in the game, and by agreeing to sign the contract, players are suggested to play until the end of the game, otherwise, they will be worse off if they choose to take. The blockchain platform will offer a decentralized system with higher safety and lower transaction costs comparing to traditional third parties, and smart contract make sure that all parties adhere to the rules at first. The web3 will take 2 percent of the final reward as a service fee, while the contract will automatically distribute the rest equally to the parties. Under this condition, the newly revised matrix would be as Figure 6 shows. In this case, even for rational players there is only a probability to cooperate in the game, and the issue of ”trust” is solved.

### 3.3 My innovation on output

Table 1 shows the result of the original centipede game and my design to illustrate how I make the players in the game better off. Rational players, care more about their short-term benefits in the past and will not be willing to cooperate in the game, and players have no choice but to lose a lot in the game.

| | Player1 | Player2 | Web |
|---|---|---|---|
| before | 3 | 1 | 0 |
| after | 4.9 | 4.9 | 0.2 |

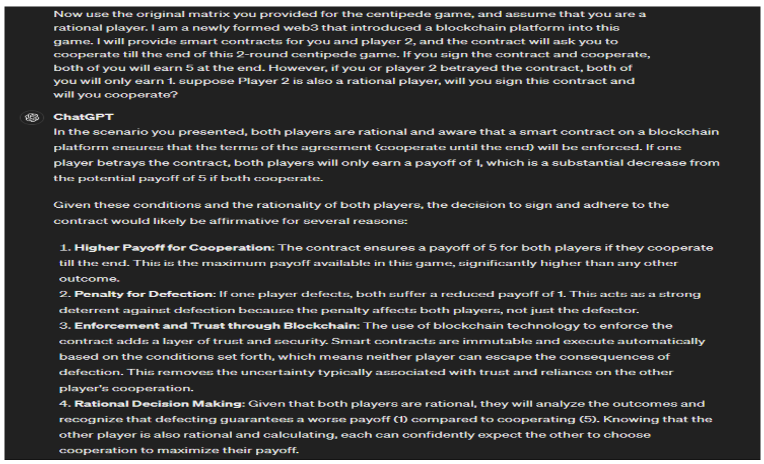
**Table 1.** Payoffs before and after the change.

However, after my implementation of the blockchain platform, although both parties can choose to sign the smart contract or not in the game, as long as they choose to sign, the smart contract will make sure that each party cooperates for long-term benefits in this game, or they will receive a penalty that makes them worse off. This idea can be applied to lots of negotiation cases that are similar to the centipede game. For instance, when making cooperations in the supply chain, suppliers are player 1 that can first choose to take (meaning providing unqualified goods for short-term benefits) or continue( provide qualified goods that promote long-term cooperation that maximize the benefit of both suppliers and buyers). If they use the web system I proposed, unlike traditional third-party enforcement, which may rely on conventional legal systems, blockchain provides a decentralized and transparent way to enforce contracts and ensure compliance without a central authority. This can reduce the risk of corruption and increase trust among players. Under the case of supply chain, the web will also gather the income pool and only release that to player 1 (supplier) as long as some evidence is shown that they provided qualified goods to seek cooperation.

While third-party involvement typically introduces fees for mediation or enforcement, using blockchain technology can significantly reduce these costs. Since blockchain reduces the need for traditional intermediaries by allowing secure, peer-to-peer transactions, it can lower the overall cost of third-party services. Just as I proposed in 1, the 2 percent intermediate fee would save the budget for those wishing to create long-term cooperation.

Finally, the idea of a smart contract can ensure that all parties adhere to the rules set at the beginning of the game, with penalties automatically applied if any party deviates from the agreed-upon strategy.

### 3.4 Practically playing the game with ChatGPT


Now use the original matrix you provided for the centipede game, and assume that you are a rational player. I am a newly formed web3 that introduced a blockchain platform into this game. I will provide smart contracts for you and player 2, and the contract will ask you to cooperate till the end of this 2-round centipede game. If you sign the contract and cooperate, both of you will earn 5 at the end. However, if you or player 2 betrayed the contract, both of you will only earn 1. suppose Player 2 is also a rational player, will you sign this contract and will you cooperate?

**ChatGPT**

In the scenario you presented, both players are rational and aware that a smart contract on a blockchain platform ensures that the terms of the agreement (cooperate until the end) will be enforced. If one player betrays the contract, both players will only earn a payoff of 1, which is a substantial decrease from the potential payoff of 5 if both cooperate.

Given these conditions and the rationality of both players, the decision to sign and adhere to the contract would likely be affirmative for several reasons:

1. **Higher Payoff for Cooperation:** The contract ensures a payoff of 5 for both players if they cooperate till the end. This is the maximum payoff available in this game, significantly higher than any other outcome.
2. **Penalty for Defection:** If one player defects, both suffer a reduced payoff of 1. This acts as a strong deterrent against defection because the penalty affects both players, not just the defector.
3. **Enforcement and Trust through Blockchain:** The use of blockchain technology to enforce the contract adds a layer of trust and security. Smart contracts are immutable and execute automatically based on the conditions set forth, which means neither player can escape the consequences of defection. This removes the uncertainty typically associated with trust and reliance on the other player's cooperation.
4. **Rational Decision Making:** Given that both players are rational, they will analyze the outcomes and recognize that defecting guarantees a worse payoff (1) compared to cooperating (5). Knowing that the other player is also rational and calculating, each can confidently expect the other to choose cooperation to maximize their payoff.

**Fig. 7.** Chatgpt's answer for the new methodology

Just as illustrated in Figure 7, ChatGPT 4 totally agreed with the new idea provided for cooperation in the centipede game, and provided several solid reasons for why rational players will absolutely cooperate in this game system. Through the imitation of AI agents, we have the confidence to clarify that this methodology solves the problem of trust in the traditional centipede game without changing the rules of the game, with a result that makes all parties better off.

Therefore, I propose a hypothesis that in the AI agents have the ability to design certain smart contracts according to different negotiation situations automatically in the future because they are fully aware of how the benefits are contributed for rational players in games. With my methodology, and I believe through deep machine learning on current negotiation cases, AI agents will become the third party in this blockchain platform with a decrease in the cost of labor force, and future research can work on this area to better improve the system that increase the social benefits

# A The Pioneers in the History of Game Theory

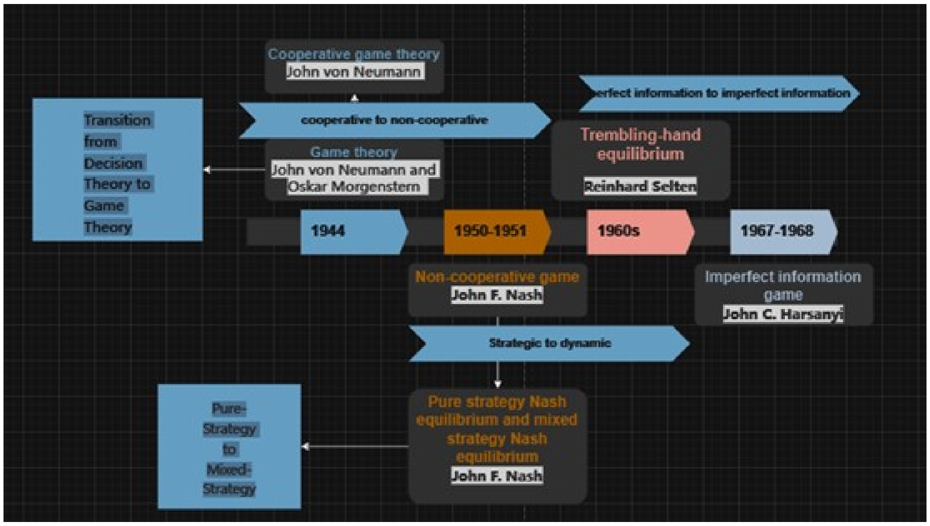



**Fig. 8.** The history of evolution of Game Theory

## A.1 From decision theory to game theory

1947: John von Neumann and Oskar Morgenstern publish "Theory of Games and Economic Behavior", marking the official birth of a game theory. (See Figure (8)) This work diverged from decision theory by introducing a framework for analyzing strategic interactions among multiple decision-makers, rather than only focusing on individual decisions.[8]

## A.2 Form pure strategy to mixed strategy

1950: John Nash introduces the concept of the Nash Equilibrium in his doctoral dissertation, "Non-Cooperative Games," at Princeton University. Nash extends the equilibrium concept to a broader class of games beyond the zero-sum games analyzed by von Neumann and Morgenstern.(See Figure (8)) Notably, Nash's definition encompasses both pure and mixed strategies. [6]

Pure-Strategy Nash Equilibrium (1950) occurs when players choose a specific action with certainty, and no player has an incentive to deviate from their chosen strategy, given the strategies of all other players.(See Figure (8)) [6]

Mixed-strategy Nash Equilibrium (1950), on the other hand, involves players randomizing over multiple actions. This occurs in situations where no pure strategy exists that can serve as the best response to the strategies of other players. [6]

### A.3 From cooperative to non-cooperative

Oskar Morgenstern Introduced the concept of cooperative games, emphasizing the possibility of binding agreements among players in 1947, and later John F. Nash founded a non-cooperative game in which binding agreements were not possible to happen in the game in 1950. (See Figure (8))

### A.4 Progression from Static Games to Dynamic Games

1965: Selten refined the Nash Equilibrium for dynamic games. He introduced the concept of "trembling-hand" equilibrium, a pivotal development for analyzing games where the timing of moves and the sequence of actions are crucial. [9] (See Figure (8))

### A.5 Shift from Perfect Information to Imperfect Information

Through John C. Harsanyi's work on games with incomplete information played by Bayesian players, published in three parts in 1967, Harsanyi laid the groundwork for analyzing games where players do not have complete knowledge about each other's payoffs or strategies.[10] (See Figure (8))

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

### B.1 Exploring Inspirational Games in Strategic or Normal Form

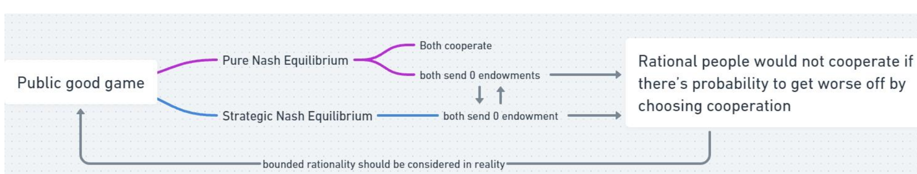



**Fig. 9.** A basic illustration of the public good game

The public good game introduced by Samuelson (1954) was designed for testing behavioral economics, where players are given a set of endowments, and in each round, they can choose how much of their endowments they are willing to contribute to the common good. At the end of the round, the endowments in the public good pool will be added by bonus and it will be separated to all players at the end. (see figure 9) I was inspired by this game mainly for two reasons. First, although the matrix created for this game looks very much like the stag hunt game introduced by Rousseau(1754)), the outcomes of Nash Equilibrium strategies were different. The second is that this anonymous type of game is similar to the blockchain platform and research on these kinds of behaviors will better help me come up with mechanisms on new technology platforms.

Below is how to set up a public good game with Bayesian Nash Equilibrium with its definitions, theorems, and proof from the textbook.

### B.1.1 Definition, theorem, and proof for playing a public game

#### Definition of Bayesian Nash Equilibrium

*Refer to Textbook: Osborne, Martin J. and Ariel Rubinstein.* 1994. A Course in Game Theory (Chapter 2, Page 25, DEFINITION 25.1)

**Definition 1.** *A Bayesian Nash Equilibrium (BNE) is a strategy profile in a game with incomplete information where each player's strategy maximizes their expected utility, given their private information (type) and their beliefs about other players' types and strategies.*

*For a game with players $i = 1, ..., n$, types $\theta_i$ drawn from a probability distribution $\Theta_i$, and actions $a_i$ from a set $A_i$, a strategy profile $(\sigma_1^*, ..., \sigma_n^*)$ is a Bayesian Nash Equilibrium if for each player $i$, the strategy $\sigma_i^*$ solves:*

$$\sigma_i^*(\theta_i) = \arg\max_{a_i \in A_i} \mathbb{E}[u_i(a_i, a_{-i}; \theta_i)|\theta_i]$$

*where $a_{-i} = \sigma_{-i}^*(\theta_{-i})$ and $u_i$ is the utility function of player $i$, which depends on their own action, the actions of others, and their type.*

**Theorem 1.** *In any Bayesian game with a finite number of players, each having a finite number of types and actions, and continuous utility functions, there exists at least one Bayesian Nash Equilibrium.*

1. **Game Representation**:
   - Consider a public good game with $n$ players.
   - Each player $i$ has a private type $\theta_i$ which influences their valuation of the public good and is drawn from a known distribution $\Theta_i$.

  - The strategy $\sigma_i : \Theta_i \to A_i$ maps player $i$'s type to their action (contribution level).
2. **Expected Utility Function**:
   The utility of player depends on their contribution the contributions of others and their type.
3. **Strategy Optimization**:
   Each player chooses to maximize their expected utility given their type and beliefs about other players' actions, which depend on types:

$$\mathbb{E}[u_i(a_i, \sigma^*_{-i}(\theta_{-i}); \theta_i)|\theta_i]$$

4. **Fixed Point Argument**:
   - Define a correspondence $F_i$ for each player $i$ that maps a profile of strategies $\sigma_{-i}$ to the set of best responses $\sigma_i$.
   - Using a fixed-point theorem (such as Kakutani's or Brouwer's, depending on the continuity and compactness of the strategy sets and payoff functions), demonstrate that there is a fixed point in the product of these correspondences, $(\sigma^*_1, ..., \sigma^*_n)$, where each $\sigma^*_i$ is the best response to $\sigma^*_{-i}$.
5. **Conclusion**:
   The fixed point represents a Bayesian Nash Equilibrium, where no player can improve their expected utility by unilaterally deviating from their strategy given their type and the strategies/types of others.

**B.1.2 practice in playing the game** To play the game simply, I formed a two-times-two matrix with the following:

**Assumptions**

*Endowment:* Each player starts with an endowment of 10 units.

*Cost of Contributing:* The cost to contribute to the public good is set at 4 units.

*Benefits from Public Good:* The public good provides a benefit calculated as 1.5 times the total contributions made by all players. This total benefit is then evenly split between both players.

*Non-Contribution Scenario:* If neither player contributes, each player retains their initial endowment of 10 units.

**Game Setup and Payoff Calculation**

**Parameters**

*Contribution Multiplier (Benefit Factor):* The benefit from the public good is multiplied by a factor of 1.5.

*Cost of Contributing:* Players incur a cost of 4 units when they decide to contribute.

*Endowment:* Each player is endowed with 10 units at the beginning of the game.

#### Matrix Setup

*Strategy Choices:* Players have two strategic options: Contribute (C) or Not Contribute (N).

*Payoff Matrix:* Payoffs are calculated based on the combined choices of the players regarding their contributions. The matrix setup involves calculating the net benefit to each player, after accounting for the cost of contribution and the distribution of benefits from the public good.

### B.2 Delving into Extensive-Form Games

The extensive game I choose is the Centipede Game [2]. This is inspiring for me since even with the fact that the decisions right now are made sequentially rather than simultaneously, and both players share the same information in the game, the result can still be largely diversified due to the personality, risk premium, strategy playing, and social relationships. (as shown in figure 4) As Rosenthal (1981) proposed, in this game, two players take turns deciding whether to continue or stop the game. With each move, the potential payoff increases, but if a player stops the game, this player will capture a larger share of the increasing pot while the other gets a smaller share. This may be even more challenging to the rationality Nash(1950) proposed, since rational people should end the game at the first opportunity, but under the observed condition the game may often continue.

#### B.2.1 Definition, theorem, and proof for playing a Centipede game

*Refer to Textbook:* *Osborne, Martin J. and Ariel Rubinstein.* 1994. A Course in Game Theory (Chapter 2, Page 25, DEFINITION 25.2)

Since the definition has been shown above, I will start with the specific theorem for the centipede game.

**Theorem 2.** *In the Centipede game, under Bayesian settings, a Nash Equilibrium exists where players may choose to stop earlier than the subgame perfect Nash equilibrium predicts, depending on their beliefs about the opponent's likelihood of continuing.*

**Proof Concept**

**Model Description**

– Let the game have $n$ rounds.
– Players have types $\theta_i$ which affect their strategies and possibly their payoffs.
– Let $p_i(\theta_i)$ be the probability distribution over player $i$'s types.

**Utility Functions**

– Denote the utility of player 1 and player 2 as $u_1$ and $u_2$ respectively.
– If player 1 decides to stop at their turn, their utility might be $u_1(x, \theta_1)$, where $x$ is the size of the pot.

**Strategy Formulation**

– A strategy for player 1, $\sigma_1(\theta_1, t)$, is a function from their type and the turn number to {stop, continue}.
– Similarly, $\sigma_2(\theta_2, t)$ for player 2.

**Belief Updates**

– After each action, players update their beliefs about the other player's type based on the observed actions.
– Use Bayes' Rule to update beliefs. If player 1 continues at turn $t$, then player 2 updates their belief about player 1's type.

**Expected Utility Calculation**

$$EU_1(\text{stop}, t) = u_1(x, \theta_1)$$

$$EU_1(\text{continue}, t) = \mathbb{E}_{\theta_2}[u_1(x', \theta_1) \mid \text{continue}, \mu_1(\theta_2)]$$

where $x'$ is the pot size if the game continues to the next round.

**Equilibrium Strategy**

– At each turn $t$, for each player $i$ with type $\theta_i$, choose the action that maximizes their expected utility given their current beliefs about the other player's type and strategy.
– This forms the Bayesian Nash Equilibrium strategies $\sigma_1^*$ and $\sigma_2^*$.

**Termination Condition**

– The game ends either when $t = n$ (the last round) or when a player chooses to stop.
– Players incorporate the probability of the game continuing based on their strategies and beliefs into their utility calculations.

### B.3 Critiquing Nash Equilibrium and Envisioning Innovations:

The former examples I provided separately illustrated a significant limitation of the current Nash Equilibrium and related tools. In the public good case, it is shown that for rational players in the Nash Equilibrium, there will be no choice of cooperation if the choice can make the player worse off. “Trust” may be the keyword representing what rational players lack. In the case of Centipede Game, we can also find similar things happening, that all solutions lead to ending the game as soon as possible, although if the game continues for 5 more rounds, both players will benefit a lot more than the amount they achieve when they stop the game at the first. The idea Brans(2020) proposed may be a solution, that by adjusting the benefits we can still motivate rational players to cooperate in an updated game. However, my perspective was to admit the limitation of rationality in the game theory and improve those games by providing a third party that can also benefit from the game. In each round, the third party would provide a ‘legally effective’ contract for the willingness of each player to cooperate or not, but they will be punished even worse if they violate the contract. For games like Centipede Game, the third party has the right to finally receive the gap between two players as a reward and let the payoff be assigned equally. For originally equally paid cooperation games, the third party will receive a fixed percentage of the final payoff from both players. In this way, we can make this game more related to reality situations. My background in economics provides me with more knowledge in market equilibriums, which can better help me design a mechanism that can be utilized for platforms like blockchain or metaverse in the future.

## C Game Theory Glossary Tables

| Glossary | Definition | Sources |
| --- | --- | --- |
| Decision theory | The mathematical study of strategies for optimal decision-making under conditions of uncertainty and interdependent decisions among rational agents. | Neumann and Morgenstern 1947 |
| Static game | Static games are defined as games where players make simultaneous decisions without any knowledge of the other's choices, often analyzed using normal form representations. | Nash Jr 1950 |
| Dynamic Games | These are games where players' decisions are made in a sequence, with each decision potentially affecting subsequent choices, characteristically analyzed using extensive form. | Selten 1965 |
| Perfect Information Game | A game where all players have complete information about the entire game history at every point in the game. Chess is a classic example. | Nash Jr 1950 |
| Imperfect Information Game | This refers to games where at least some players lack complete information about the actions previously taken by other players, typically represented in the form of information sets in game trees. | Harsanyi 1967 |
| Public Good Game | A game analyzing situations where individual contributions to a common pool result in benefits distributed across all players, regardless of individual contribution. | Samuelson 1954 |
| Centipede game | Originated by Rosenthal in 1981, this is an extensive form game famous for its counterintuitive implications for backward induction and rational decision-making. | Rosenthal 1981 |

**Table 2.** Game Theory Glossary Table

# Bibliography


[1] P. A. Samuelson, "The pure theory of public expenditure," *The review of economics and statistics*, pp. 387–389, 1954.

[2] 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. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a7552bd3302f6dee7c0b74dea6fc486f794db0b8

[3] I. Rahwan, M. Cebrian, N. Obradovich, J. Bongard, J.-F. Bonnefon, C. Breazeal, J. W. Crandall, N. A. Christakis, I. D. Couzin, M. O. Jackson *et al.*, "Machine behaviour," *Nature*, vol. 568, no. 7753, pp. 477–486, 2019. [Online]. Available: https://www.nature.com/articles/s41586-019-1138-y

[4] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton *et al.*, "Mastering the game of go without human knowledge," *nature*, vol. 550, no. 7676, pp. 354–359, 2017. [Online]. Available: https://www.nature.com/articles/nature24270;

[5] N. Bostrom and E. Yudkowsky, "The ethics of artificial intelligence," in *Artificial intelligence safety and security*. Chapman and Hall/CRC, 2018, pp. 57–69. [Online]. Available: https://www.taylorfrancis.com/chapters/edit/10.1201/9781351251389-4/ethics-artificial-intelligence-nick-bostrom-eliezer-yudkowsky

[6] 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

[7] S. J. Brams and D. M. Kilgour, "A note on stabilizing cooperation in the centipede game," *Games*, vol. 11, no. 3, 2020. [Online]. Available: https://www.mdpi.com/2073-4336/11/3/35

[8] V. Neumann and O. Morgenstern, *Theory of Games and Economic Behavior. (Second edition.)*. Princeton University Press, 1947. [Online]. Available: https://psycnet.apa.org/record/1947-03159-000

[9] 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

[10] J. C. Harsanyi, "Games with incomplete information played by "bayesian" players," *Management Science*, vol. 14, pp. 159–182, 11 1967. [Online]. Available: http://www.dklevine.com/archive/refs41175.pdf

[11] J.-J. Rousseau, "Discourse on the origin and basis of inequality among men," *Rousseau J.-J. Traktaty*, pp. 31–109, 1754. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Discourse+on+the+Origin+and+Basis+of+Inequality+Among+Men&btnG=

[12] M. J. Osborne and A. Rubinstein, *A course in game theory*. MIT press, 1994. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0899825699907236

[13] S. J. Brams and D. M. Kilgour, "A note on stabilizing cooperation in the centipede game," *Games*, vol. 11, no. 3, 2020. [Online]. Available: https://www.mdpi.com/2073-4336/11/3/35