



**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

AY2024/25 Semester 2

SC4003 INTELLIGENT AGENTS

Assignment 2: Repeated Prisoners Dilemma

Tan Choon Wee

U2120106H

COLLEGE OF COMPUTING AND DATA SCIENCE

NANYANG TECHNOLOGY UNIVERSITY

2025

Contents

Contents	i
1. Introduction	1
2. Background and Theory	2
2.1 The Iterated Prisoner's Dilemma	2
2.2 Axelrod's Tournament	2
2.3 Recipes for Success in Axelrod's Tournament	3
3. Methodology	4
3.1 Game Setup	4
3.2 Strategy Implementation	5
3.3 Tournament Simulation	5
3.4 Evaluation Criteria	6
4. Results	6
4.1 Performance of Different Defection Rate Thresholds	6
4.2 Analysis and Justification	6
4.3 Overall Tournament Performance	7
4.4 Analysis of Results	7
5. Discussion	8
5.1 Key Rules in BalancedTitForTatPlayer	8
5.2 Why BalancedTitForTat Outperformed Other Strategies	9
6. Conclusion	10
References	11

1. Introduction

The Prisoner's Dilemma is a well-known problem in game theory that demonstrates how rational agents may fail to cooperate, even when it is in their best interest. In its classical two-player version, each player independently chooses to either cooperate or defect, with defection being the dominant strategy, leading to a suboptimal equilibrium. However, in a repeated setting, where the game is played multiple times, strategies can evolve to encourage cooperation over time [1].

This assignment extends the traditional Prisoner's Dilemma to a three-player repeated game, where three agents interact across multiple rounds, with their payoffs determined by their combined choices. The repeated nature of the game allows for a broader range of strategic behaviours, including conditional cooperation, retaliation, and adaptation. Strategies such as Tit-for-Tat (TfT), Trigger Strategy, Pavlovian (Win-Stay, Lose-Shift), and BalancedTfT may emerge as potential approaches to maximize individual and collective payoffs [2,3].

The goal of this assignment is to develop a strategy for an agent in the three-player repeated Prisoner's Dilemma. The strategy must decide whether to cooperate or defect based on the previous moves of all players. The agent will compete in matches consisting of approximately 100 rounds, with the final performance evaluated based on the average payoff per round. Previous studies have shown that strategies like Win-Stay, Lose-Shift can outperform Tit-for-Tat under certain conditions [3]. Additionally, recent research has explored how certain dominant strategies can emerge in iterated settings [4].

This report explores different strategic approaches, analyses the implications of cooperation and defection in a three-player setting, and discusses the performance of the implemented strategy against various opponents. The findings contribute to a broader understanding of strategic decision-making in multi-agent interactions and the role of memory and adaptation in fostering cooperation.

2. Background and Theory

(reference from Lecture 8.1 Multiagent Interaction)

2.1 The Iterated Prisoner's Dilemma

The Iterated Prisoner's Dilemma (IPD) extends the classical two-player Prisoner's Dilemma by having the game played multiple times. This repeated interaction allows for the possibility of cooperation to emerge, as players can adjust their strategies based on the outcomes of previous rounds. In an infinitely repeated IPD, cooperation can become the rational choice, as players seek to maximize their long-term payoffs by fostering mutual cooperation.

However, when the number of rounds is finite and known to both players, the incentive to defect in the final rounds can lead to a breakdown of cooperation. This phenomenon is known as the backwards induction problem. In a fixed, finite IPD, defection becomes the dominant strategy as players anticipate the end of the game and seek to maximize their immediate payoffs.

2.2 Axelrod's Tournament

In 1984, Robert Axelrod conducted a computer tournament to investigate strategies for the IPD. Participants submitted programs that played the IPD against each other, with the goal of maximizing their overall payoffs. Axelrod's tournament revealed several key insights into successful strategies:

- **ALLD (Always Defect):** This strategy always defects, seeking to exploit cooperative opponents.
- **TIT-FOR-TAT (TfT):** This strategy starts by cooperating and then mimics the opponent's previous action. It is simple, retaliatory, and forgiving.
- **TESTER:** This strategy defects on the first round. If the opponent retaliates, it switches to TfT; otherwise, it alternates between cooperation and defection.
- **JOSS:** Similar to TfT, but with periodic defections to test the opponent's response.

2.3 Recipes for Success in Axelrod's Tournament

Axelrod identified several principles for successful strategies in the IPD:

- **Don't be envious:**

Avoid playing as if the game were zero-sum. Focus on maximizing your own payoff rather than minimizing your opponent's.

- **Be nice:**

Start by cooperating and reciprocate cooperation.

- **Retaliate appropriately:**

Punish defection immediately but use measured force. Avoid excessive retaliation.

- **Don't hold grudges:**

Always reciprocate cooperation immediately after it is offered.

These principles highlight the importance of conditional cooperation, retaliation, and forgiveness in fostering long-term cooperation and achieving higher payoffs in the IPD.

3. Methodology

3.1 Game Setup

We modelled the three-player Prisoner's Dilemma using a symmetric payoff matrix that preserves the strategic tension of the classical two-player game while accounting for ternary interactions. The payoffs for Player 1 (our agent) are defined as follows:

Player 1 (i)	Player (j)	Player 3 (k)	Payoff (Player 1)
0 (Cooperate)	0 (Cooperate)	0 (Cooperate)	CCC = 6
0 (Cooperate)	0 (Cooperate)	1 (Defect)	CCD = 3
0 (Cooperate)	1 (Defect)	0 (Cooperate)	CDC = 3
0 (Cooperate)	1 (Defect)	1 (Defect)	CDD = 0
1 (Defect)	0 (Cooperate)	0 (Cooperate)	DCC = 8
1 (Defect)	0 (Cooperate)	1 (Defect)	DCD = 5
1 (Defect)	1 (Defect)	0 (Cooperate)	DDC = 5
1 (Defect)	1 (Defect)	1 (Defect)	DDD = 2

Table 1: Payoff Matrix

Lowest
payoff

Highest
payoff

Ordering:

$$U(DCC) > U(CCC) > U(DDC) > U(CDC) > U(DDD) > U(CDD)$$

This payoff matrix illustrates the incentive structure of the game. Cooperating with others yields a moderate payoff when all parties cooperate (6), but cooperating when others defect results in the lowest payoff (0 or 3). Defecting, on the other hand, yields a higher payoff regardless of the other players' actions, with the highest payoff (8) occurring when both opponents cooperate. This creates a dilemma where individual rationality leads to defection, despite the potential for higher collective payoffs through cooperation.

3.2 Strategy Implementation

To analyse different strategic behaviours, we implemented various player strategies within the Java program. These strategies included:

1. **NicePlayer:** Always cooperates.
2. **NastyPlayer:** Always defects.
3. **RandomPlayer:** Randomly cooperates or defects with equal probability.
4. **TolerantPlayer:** Cooperates unless more than half of the opponents' actions have been defects.
5. **FreakyPlayer:** Randomly decides at the beginning of the game whether to always cooperate or always defect.
6. **T4TPlayer:** Mimics the last action of a randomly chosen opponent.

(Tit-for-Tat)

Each strategy was tested in multiple rounds of the game against other strategies to evaluate its effectiveness.

3.3 Tournament Simulation

The tournament was conducted by having each strategy play against every other strategy, including itself, in a series of matches. Each match consisted of 90 to 110 rounds, with the number of rounds determined randomly within this range. The scores for each match were calculated as the average of the payoffs from each round.

The following steps were followed to run the tournament:

1. **Player Creation:** Fresh copies of each player were created for each match.
2. **Rounds Calculation:** The number of rounds for each match was determined randomly between 90 and 110.
3. **Match Simulation:** Each match was simulated, and the scores for each player were calculated.
4. **Score Accumulation:** The scores from each match were added to the total scores for each player.
5. **Sorting and Results:** Players were sorted by their total scores, and the results were printed.

3.4 Evaluation Criteria

To optimize the performance of the `BalancedTitForTatPlayer`, we conducted a series of experiments varying the forgiveness threshold, which determines the maximum tolerated defection rate before the player retaliates. We tested thresholds of 0.15, 0.1, 0.05, and 0.06, running the tournament with each configuration. The results, presented in Table 2, demonstrate the impact of this parameter on the player's average score.

4. Results

4.1 Performance of Different Defection Rate Thresholds

Forgiveness Threshold	Average Score	Rank
0.15	209.1123	
0.10	213.8762	
0.05	215.4431	
0.06	217.7674	1st

Table 2: Performance of Different Defection Rate Thresholds

4.2 Analysis and Justification

- Thresholds above 0.06 (e.g., 0.1, 0.15) were too forgiving, allowing exploitative players to take advantage. The `BalancedTitForTatPlayer` cooperated too often, leading to lower overall payoffs.
- Thresholds below 0.06 (e.g., 0.05) were too aggressive, resulting in unnecessary defections. This led to lower cooperation levels and suboptimal scores.
- 0.06 provided the best balance between retaliation and cooperation. It ensured measured punishment against defectors while maintaining cooperation with trustworthy players.

4.3 Overall Tournament Performance

The tournament results for the various strategies, including the optimized `BalancedTitForTatPlayer` (with a 0.06 forgiveness threshold), are summarized in the table below. Each strategy was evaluated over five runs, and the average scores were calculated to determine their overall performance.

Player	Run 1	Run 2	Run 3	Run 4	Run 5	Averages
<code>BalancedTitForTatPlayer</code> :	219.6756	217.4991	209.0554	220.5134	222.0938	217.7674
<code>TriggerPlayer</code> :	214.1222	217.8737	209.3312	224.3493	215.2747	216.1902
<code>TolerantPlayer</code> :	210.8835	209.4095	202.7437	210.1637	202.3139	207.1028
<code>T4TPlayer</code> :	210.7355	204.2073	204.4157	206.2663	200.2558	205.1761
<code>NicePlayer</code> :	187.5589	197.5519	181.2373	189.7282	185.2840	188.2721
<code>FreakyPlayer</code> :	183.0234	175.2938	172.4321	180.8972	187.0608	179.7415
<code>NastyPlayer</code> :	169.2402	156.8480	162.8371	165.5685	150.3559	160.9699
<code>RandomPlayer</code> :	147.1756	146.5642	157.9835	164.1909	144.3847	152.0598

Table 3: Average Scores of All Strategies

4.4 Analysis of Results

The `BalancedTitForTatPlayer` with the optimized 0.06 forgiveness threshold achieved the highest average score in the tournament, demonstrating the effectiveness of a "measured force" approach. Strategies that prioritized cooperation and reciprocity, such as `TriggerPlayer`, `TolerantPlayer`, and `T4TPlayer`, also performed relatively well. In contrast, purely defective (`NastyPlayer`) and random (`RandomPlayer`) strategies resulted in the lowest scores.

`BalancedTitForTatPlayer` outperformed other strategies, including `TriggerPlayer` (216.1902), showing that a slightly forgiving but retaliatory approach is superior to always-triggering defections.

5. Discussion

The results indicate that strategies promoting cooperation, such as `BalancedTitForTatPlayer` and `TriggerPlayer`, generally performed better than purely defecting or random strategies. The `BalancedTitForTatPlayer`'s ability to balance cooperation, measured retaliation, and forgiveness led to the highest average payoffs. Through experimentation, it was found that a forgiveness threshold of 0.06 was optimal, highlighting the importance of adaptability and conditional cooperation in multi-agent interactions.

This success can be attributed to the well-balanced approach of the strategy, which follows four key principles: niceness, measured retaliation, controlled punishment, and forgiveness.

5.1 Key Rules in `BalancedTitForTatPlayer`

The `BalancedTitForTatPlayer` strategy follows several key rules to achieve its high performance:

1. **Be nice:**

The strategy always starts with cooperation to encourage mutual trust and maximize long-term rewards.

2. **Retaliate Appropriately:**

The strategy defects immediately if either opponent defects in the previous round.

3. **Use Measured Force:**

Instead of permanently defecting after a single betrayal, the strategy calculates the defection rates of both opponents. Defect if the opponent's defection rate exceeds the optimal 6% forgiveness threshold.

4. **Don't Hold Grudges:**

Once an opponent resumes cooperation, the strategy forgives past defections and immediately cooperates again. This prevents unnecessary cycles of mutual defection, ensuring that it does not hold grudges.

5.2 Why **BalancedTitForTat** Outperformed Other Strategies

Compared to other strategies, **BalancedTitForTat** demonstrated superior performance because:

- It avoided unnecessary defections, unlike **TriggerPlayer**, which switched to permanent defection too early.
- It was more forgiving than rigid **T4TPlayer**, preventing long chains of retaliation.
- It could defend itself against exploitative strategies, unlike **NicePlayer**, which was easily taken advantage of.

Overall, the success of **BalancedTitForTatPlayer** highlights the effectiveness of a measured, conditional cooperation strategy in the Iterated Three-Player Prisoner's Dilemma.

6. Conclusion

In this assignment, we explored strategic design in the Iterated Three-Player Prisoner's Dilemma through the implementation and evaluation of multiple agent strategies. Among these, the `BalancedTitForTatPlayer` emerged as the most effective, achieving the highest average score across tournament runs.

By adhering to well-established principles – being nice, retaliating appropriately, forgiving quickly, and avoiding envy – the strategy demonstrated that success in repeated games lies in conditional cooperation rather than pure aggression or blind trust. The fine-tuning of the forgiveness threshold to 0.06 proved crucial, striking the optimal balance between punishing defectors and maintaining cooperation with trustworthy opponents.

These findings reinforce the broader lesson from game theory and Axelrod's tournaments: robust cooperation can evolve and thrive when combined with measured and adaptive retaliation. The success of `BalancedTitForTatPlayer` highlights the importance of strategy tuning and theoretical grounding in designing intelligent agents for social dilemmas.

References

- [1] R. Axelrod, *The Evolution of Cooperation*, Basic Books, 1984.
- [2] D. Fudenberg and J. Tirole, *Game Theory*, MIT Press, 1991.
- [3] M. Nowak and K. Sigmund, “A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game”, *Nature* 364, vol. 37, pp. 56-58, 1993.
Available: <https://www.nature.com/articles/364056a0>
- [4] W. H. Press and F. J. Dyson, “Iterated Prisoner’s Dilemma contains strategies that dominate any evolutionary opponent”, *Proceedings of the National Academy of Sciences*, 109 (26), 10409-10413, 2012. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.1206569109>