Literature review paper for the iterated prisoner's dilemma.

Nikoleta E. Glynatsi

2016

1 Introduction

The emergence of cooperation is a topic of continuing and public interest for social, biological [21] and ecological sciences [22, 28, 36, 54]. The prisoner's dilemma is a popular game commonly used to represent situations of altruistic behaviour.

The prisoner's dilemma is a two players no-cooperative game where the decisions of the players are made simultaneously and independently. Both players can choose between cooperation (**C**) or defection (**D**).

The fitness of each player is influenced by its own behaviour, and the behaviour of the opponent. If both players choose to cooperate, both do better than if both defected. However, a player has the temptation to deviate. If one player were to defect while the other cooperates, the defector receives more than if both had cooperated. The reward for mutual cooperation is R units, for a mutual defection they receive P, and for cooperation-defection, the cooperator receives S where the defector receives T.

Thus, the game's payoff are defined by,

$$\begin{pmatrix} R & S \\ T & P \end{pmatrix}, \tag{1}$$

where, $T > R > P > S$ and $2R > T + S$. are the conditions for the dilemma to exist. Due to rational behaviour and the knowledge that an individual is tempted to defect, the game's equilibrium lies at a mutual defection and both players receive a payoff of P. Thus, the unbeatable strategy for the prisoner's dilemma is **D**.

Though the one shot game illustrate the conflict between individual and collective rationality, and how through mutual pursuit of self-interest players end up with a worse payodd than if they had behaved otherwise, greater insights can be achieved by studying the game in a manner where the prior outcomes matters. The repeated form is called the iterated prisoner's dilemma, and it will be discussed in Section 2 how it was proven to leave more room for cooperation to emerge.

The origin of the prisoner's dilemma go back to 1950 in early experiments conducted in RAND [17] to test the applicability of games described by [53]. In [17] they introduced a two player non-cooperative game but the story behind the game was given later the same year. A. W. Tucker, the advisor of the famous John Nash, in an attempt to delivery the game with a story during a talk formulated the story of the prisoners as it is known today [51].

Figure 1 illustrates a timeline generated using the open source library discussed in Section 3. It can be observed that the prisoner's dilemma has been under continuous research since it's origins. Several insights and applications have been offered in research by the game and several of these will be discussed in Section 2.

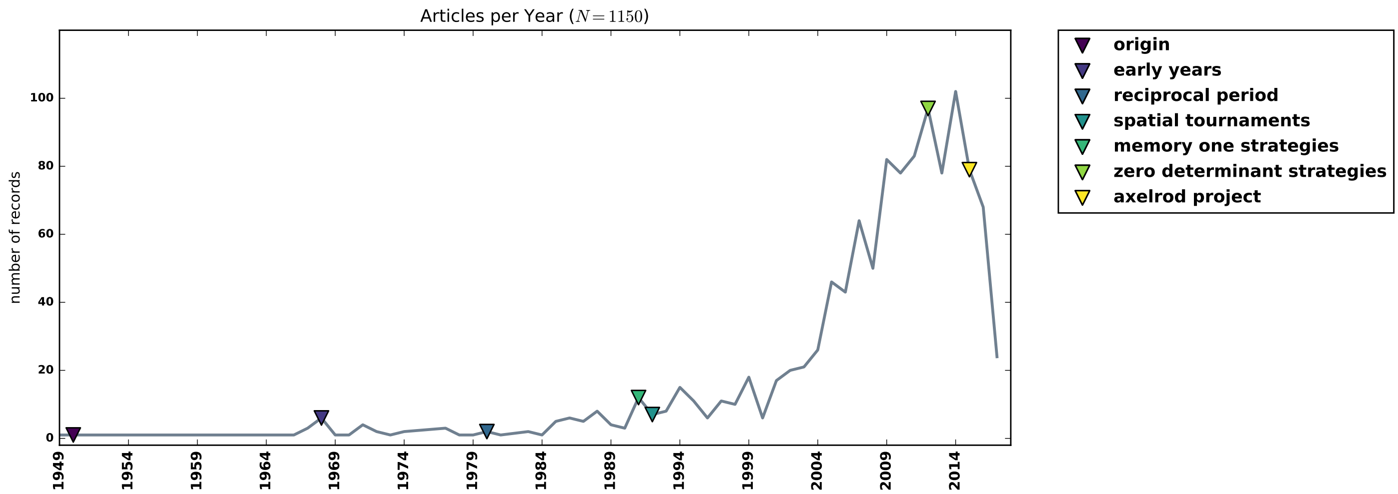

Figure 1: A timeline highlighting the milestones of the prisoner's dilemma.

2 Timeline

2.1 Early Years

The study of the prisoner's dilemma has attracted people from various fields across the years. An early figure within the field is Professor Anatol Rapoport, a mathematical psychologist, who's work focused on peacekeeping. In his early work [44] Rapoport conducted experiments using humans to simulate a play of the prisoner's dilemma. Experimental groups were not been used only by Rapoport but it was a common mean of studying the game [16, 19, 33, 34, 48] and are still being use to date.

These experiments explored the conditions under which altruist behaviour emerges in human societies. By analysing the play of a test subject researcher believed that they could identify an unbeatable strategy to play the game. Inspired by the work of Rapoport and the idea that AI was now being trained to play a game of chess the political scientist Robert Axelrod performed the first ever computer tournament, known to the author, of the iterated prisoner's dilemma [12, 10].

2.2 Reciprocal Period

2.2.1 Axelrod's Tournaments

In 1980 [7] a computer tournament of the iterated prisoner's dilemma took place with 14 participants. Each strategy played against all the 13 opponents, itself and the strategy called Random (a strategy that chooses between **C** and **D** randomly) a match of 200 turns. This topology is called round robin, illustrated in 2, and is the equivalent of a complete graph. The tournament was repeated 5 times to reduce variation in the results. Each participant knew the exact length of the matches and had access to full history of each match. The payoff values of (1) used by Axelrod were the following, $R = 3, P = 1, T = 5$ and $S = 0$. These values are commonly used in literature and assume that they are being used in the works referenced henceforward unless is stated otherwise.

The winner of the tournament was determined by the total average score and not by the number of matches wins. The strategy that was announced the winner was submitted by Rapoport and was called called Tit For Tat.

Tit for Tat, is a strategy that always cooperates on the first round and then mimics the opponent's previous move. The strategy is illustrated diagrammatically in Figure 3. To further test the robustness of the results a second tournament was performed later with a total of 63 strategies [8]. All the opponents knew the results of the previous tournament but this time the number of turns was not specified. Instead a probabilistic ending tournament was used. In a probabilistic

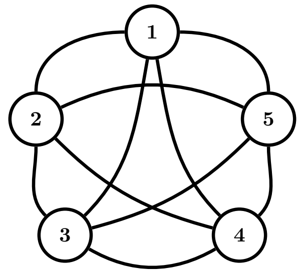

Figure 2: An example of a round robin tournament of 5 players. Each node represents a player and each edge a match.

ending tournament, each match has probability of ending after each turn. This is also refereed as 'shadow of the future' is some other works [11]. The winner of the second tournament was once again the strategy Tit for Tat.

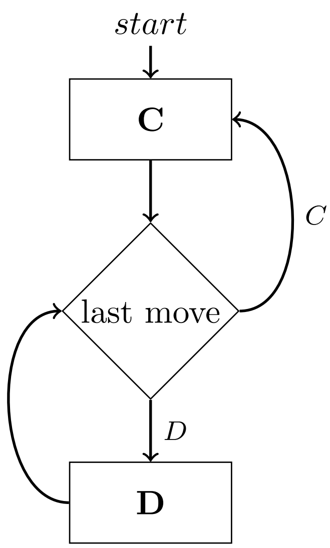

Figure 3: Diagrammatic representation of Tit for Tat.

Tit for Tat provided proof that reciprocity behaviour can allow cooperation to emerge in the iterated prisoner's dilemma game. According to Axelrod, the secrets behind the strategy's success have been 1) that it start of by cooperating 2) it would forgive it's opponent after a defection 3) after opponents identified that they were playing Tit for Tat choose to cooperate for the rest of the game.

The success of Tit For Tat was very soon known world wide and several researcher focused their work on the strategy ever since such as [21, 28, 36].

2.2.2 Stochastic Environments

But success often comes with criticism. Axelrod's tournaments assumed that each player has perfect information of the opponent's actions. In real life situations this is not always the case. Colleagues interactions often suffer from measures of uncertainty. In the original tournaments there was no possibility of mis implementation or misunderstanding. These stochastic variations are refereed to as **noise** and **mis perception**.Noise is the concept of flipping one's move based on a given probability. On the contrary, mis perception is the probability that the opponent's current move is flipped before being recorded. Noise will flip a player's action and it will be recorded correctly in the history where mis perception will not have an effect on the player move but it will be recorded wrong [26].

The performance of Tit for Tat was proven to suffer from such stochasticity in the tournament environment, especially against itself [4, 22, 38, 39, 49]. If two strategies playing Tit for Tat were to compete against each other in a noisy environment, the strategies get stuck in an unwanted cycle of cooperation followed by defection. In a non noisy environment the two strategies would have been cooperating for the entire match. An interesting result was introduced by [38]. If two players are both using the Tit for Tat strategy, both players would get the same average payoffs as two interacting Random players with $p = 0.5$. In his works [11] try to address the criticism by studying Tit for Tat in an evolutionary manner as well. It was shown that Tit For Tat does not perform as well in noisy and in environments with mis-perception, but there are variants of Tit for Tat that do. The work of [41] following a similar approach agreed with this result.

The pioneer work of computer tournaments in the iterated prisoner's dilemma sparked an interest within the field and several studies and approaches are discussed in the following sections.

2.3 Era of Strategies

Following the successful work of computer tournaments many researchers have sought to understand which strategies are dominant when playing the iterated prisoner's dilemma. These strategies vary from deterministic to more complex strategies. Strategies can reply on the history of the game, the length of the matches or choose to rely on none of above. The size of history a strategy takes into account is refereed to as memory size of the strategy.

In this section we will discuss several strategies of interest that suffered during the years.

2.3.1 Axelrod's Guests

In [12] a full list and an explanation of all the 13 attendants that competed in the first tournament is given. A list of names and the creators is shown in Table 1. On the contrary, not all 64 strategies of the second tournament are presented in much detail in [12]. The author mainly focuses on the high ranked participants, even so the name of the creators alongside the source code of each strategy written in Fortan is accessible in [1].

Author	Name
Tit For Tat	Anatol Rapoport
Tideman and Chieruzzi	T Nicolaus Tideman and Paula Chieruzz
Nydegger	Rudy Nydegger
Grofman	Bernard Grofman
Shubik	Martin Shubik
Stein and Rapoport	Stein and Anatol Rapoport
Grudger	James W Friedman
Davis	Morton Davis
Graaskamp	Jim Graaskamp
Downing	Leslie Downing
Feld	Scott Feld
Joss	Johann Joss
Tullock	Gordon Tullock
Unnamed	Strategy Unknown

Table 1: A list of all the strategies and creators of the original tournament.

2.3.2 Strategies of Interest

A growing number of strategic rules can be found in the literature of the iterated prisoner's dilemma. As discussed before, the strategies can vary from deterministic ones to more complex.

The two most common deterministic strategies used in various works are **Defector** and **Cooperator** introduced in [12]. Defector is a strategy that defects in each turn and on the other hand Cooperator always cooperates.

In [41], the space of re-active strategies was explored in a noisy environment. The strategy that was performing the best in that environment was the re-active strategy known as **Generous Tit for Tat**. A reactive strategy is a strategy that consider only the past move of the opponent, but they will be discussed later on in more detail. Generous Tit for Tat, attracted attention as it was a generous variant of the famous strategy Tit for Tat able to winstand noisy environments.

In 1993 [42], an interesting strategy with the tolerance of Generous Tit for Tat but the capability of resisting and invading an all-out cooperators population was introduced. The strategy is called **Pavlov**, and is based on the fundamental behavioural mechanism win-stay, lose-shift. The strategy starts off with a **C**, then Pavlov will repeat it's last move it was awarder with by R or T but will shift if punished by P or S.

Several other strategies were introduced as more generous versions and described as more dominant than Tit for Tat. These include, **Contrite Tit for Tat** [55] and **Adaptive Tit for Tat** [52]. **Tit for Two Tats** [11], which defects only if the other player defected on the two preceding moves). On the other hand, defector variants have also been studied [25]. **Anti Tit for Tat**, is a strategy that plays the opposite of the opponents previous move. Another limitation of the strategy was discussed in [49]. Tit for Tat was proven to hit a deadlock. Deadlock meaning a loop between cooperation and defection. **Omega Tit For Tat** was introduced and was a strategy capable of avoiding the deadlock [49].

Other strategies that made an impact have been **Gradual** [13] and **Handshake** [46] presented in 1997 and 1989 respectively. Gradual starts off by cooperating, then after the first defection of the other player, it defects one time and cooperates twice. After the second defection of the opponent, it defects two times and cooperates twice. After the n th defection it reacts with n consecutive defections and then two cooperations. Handshake is a strategy that starts with cooperation, defection. If the opponent plays in a similar way then it will cooperate forever, otherwise it will defect forever.

In 2011 [31] performed their own tournament where several interesting strategies made an appearance.

- **Periodic player CCD**, plays **C**, **C**, **D** periodically. Note that variations of a period player also make appearance in the article but will not be listed here.
- **Prober**, starts with the pattern **D**, **C**, **C** and then defects if the opponent has cooperated in the second and third move; otherwise, it play as Tit for Tat.
- **Reverse Pavlov**, a strategy that does the reverse of Pavlov.

In earlier work the same author introduced a strategy called **APavlov**, which stands for adaptive Pavlov [30]. The strategy attempts to classify the opponent as one of the following strategies, All Cooperator, All Defector, Pavlov, Random or **PavlovD**. PavlovD, is just Pavlov but it starts the game with a **D**. Once Adaptive Pavlov has classified the opponent plays to maximize it's payoff.

2.3.3 Memory One Strategies

Reactive strategies are a subset of memory one strategies introduced in 1989 [40]. Reactive strategies are denoted by the probabilities to cooperate after a **C** and a **D** of the opponent. Thus, a reactive strategy only considers the previous turn of the opponent. Strategies such as, Tit for Tat and Generous Tit for Tat are reactive.

Memory one strategies, are a set of strategies that consider only the last turn of the game to decide on the next action [41]. They are represented by the four conditional probabilities p_1, p_2, p_3 and p_4 to cooperate after CC, CD, DC and DD respectively (the four possible states a player can be in if only the last turn of the game was to be considered). Reactive strategies are just a constrained version where $p_1 = p_3$ and $p_2 = p_1$. The first action of the strategy (when the history does not exist yet) is assumed to be **C** unless is stated otherwise. For example, a reactive strategy called **Suspicious Tit for Tat**, studied in [39], has the same representation as Tit for Tat but plays **D** in the first round.

In [43], a new set of memory one strategies were introduced, called **zero determinant (ZD)** strategies. The ZD strategies, manage to force a linear relationship between the score of the strategy and the opponent. Press and Dyson, prove their concept of the ZD strategies and claim that a ZD strategy can outperform any given opponent.

The ZD strategies have attracted a lot of attention. It was stated that "Press and Dyson have fundamentally changed the viewpoint on the Prisoners Dilemma" [50]. In [50], a new tournament was performed including ZD strategies and a new set of ZD strategies the **Generous ZD**. Even so, ZD and memory one strategies have also received criticism. In [29], the 'memory of a strategy does not matter' statement was questioned. A set of more complex strategies, strategies that take in account the entire history set of the game, were trained and proven to be more stable than ZD strategies.

2.3.4 Complex Strategies

Complex strategies are defined here as a set of strategies that can make use of any information that can be provided during a match. The term complex can also be refereed to strategies that have been trained with evolutionary methods to be dominant. In [9], Axelrod used an evolutionary algorithm to identify a strategy that was equal to or better than Tit for Tat.

In [6] two new strategies are presented. These strategies have been trained using a finite state machine representation. They are called, **Fortress3** and **Fortress4**. Figure 4 illustrates their diagrammatic representation.

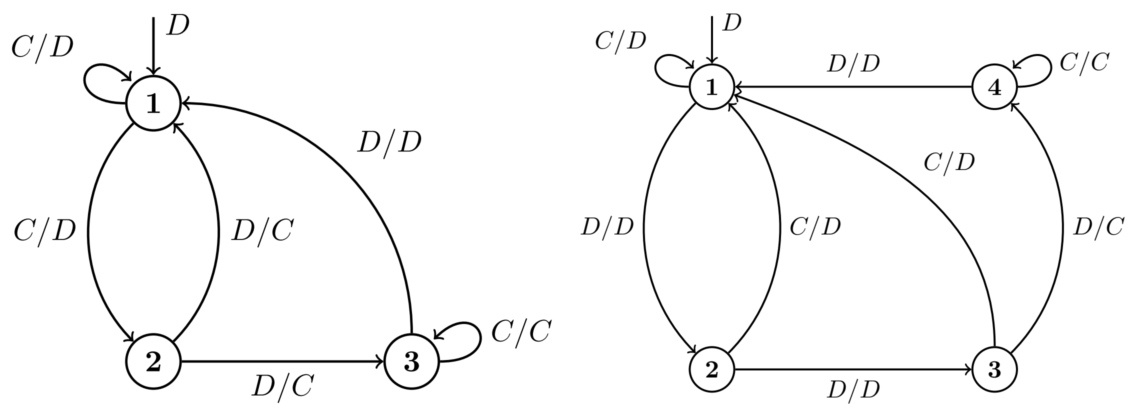

Figure 4: Representations of Fortress 3 and Fortress 4. Transition arrows are labelled O/P where O is the opponents last action and P is the players response. Note that the strategys first move, enters state 1, is defection for both strategies.

Finite state machines are commonly used to represent iterated prisoner's dilemma strategies [37, 47]. Other representation methods include lookup tables [9, 32] and artificial neural networks [24, 29].

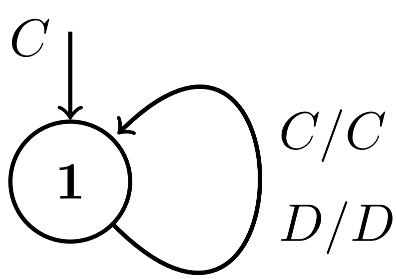

Figure 5: Finite state machine representation of Tit for Tat.

The work of [20] uses an evolutionary algorithm to train a finite state machine. The evolutionary algorithm used in both [9, 20] is called genetic algorithm. Other algorithms including particle swarm optimization have been used in research of the most dominant strategy [18].In [23], several of the representation methods have been used alongside evolutionary algorithms to represent as stated by the authors the best performing strategies for the iterated prisoners dilemma.

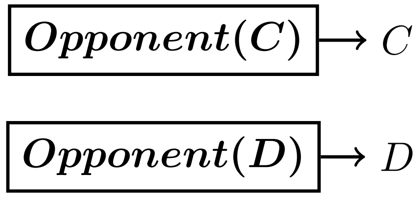

Figure 6: Look up table representation of Tit for Tat.

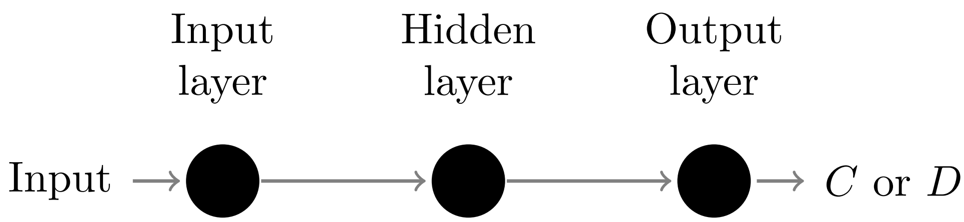

Figure 7: Neural network representation of Tit for Tat.

The different representation methods are refereed to as archetypes. In [23] the following combinations of archetypes and evolutionary algorithms have been used to introduce a number of successful new strategies:

- Lookup tables and genetic algorithm have been used to train a strategy called **LokkerUp**;
- Lookup tables and particle swarm optimization have introduced the strategy **PSOGambler**;
- Finally neural networks alongside genetic algorithm have trained a strategy called **Evolved HMM**.

2.4 Strategies Stability

In [12], the strategies set of second tournament was used to perform a ecological kind of tournament. The 63 strategies interact generation after generation to a round robin competition where their frequencies is proportional to their payoff in the previous round. Results showed that in a homogeneous population of Tit for Tat invasion by mutant strategies was not successful.

A strategy's dominance is tested through out the interactions and the performance of the strategy in a tournament against other strategies. Is the overall success of a strategy based only on it's performance in a round robin tournament, or should the strategy's performances be checked through other ways as well.

Following his initial tournaments Axelrod performed an 'ecological' tournament in 1981 [12]. The ability of strategies to be favoured under natural selection and their ability to withstand invasion from other strategies soon became an new measure of performance.

The ecological approach is based on the payoff matrix of the tournament. The highest performing strategies are adapted by lower scoring individuals within a fixed population. Over time a strategy takes over the population.

Axelrod and his reciprocal claims were put on the test once again. The results of [15] argued that no pure strategy is evolutionary stable in the iterated prisoner's dilemma. This was not proven analytically, instead a series of examples using strategies such as Tit for Tat, Suspicious Tit for Tat and All Defector where explored; a very constrained set of strategies.

The results were questioned by [35], they stated that much was still no fully explored and more research had to be put into the results. Another attempt to explore stability of strategies in the prisoner's dilemma was done in [14]. This time exploring the results in a noisy environment, but similarly a analytical prove was not given.

An extension to the natural selection was introduced in the 1992 [39], recommending a different type of topology. A population of two deterministic strategies, AllD and AllC, are placed on a a two dimensional square array where the

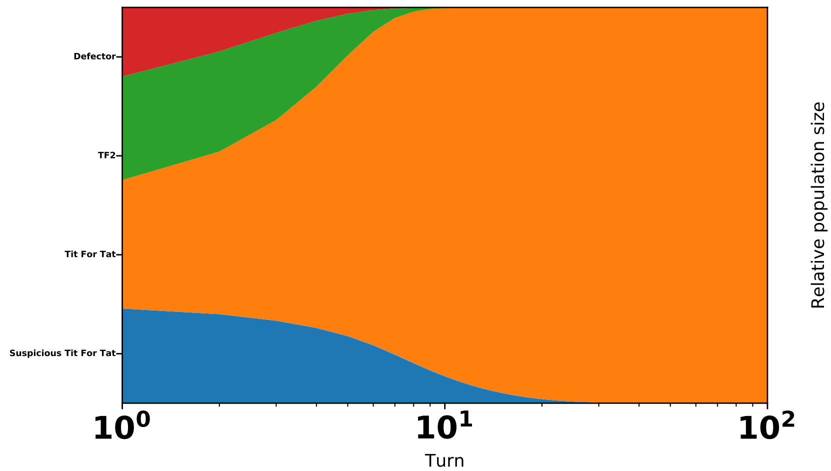

Figure 8: System evolving over time based on natural selection using [5].

individuals can now interact only with the immediate neighbours. The number of immediate neighbours can be either, fourth, six or eight. As shown in Figure 9. The authors claimed that the essential results remain true of all topologies; the results also hold whether self interactions are taken into account.

Thus each cell of the lattice is occupied by a **C** or a **D** and in each generation step each cell owner interacts with its immediate neighbour and play the game. The score of each player is the sum of the overall games the player competed in. At the start of the next generation, each lattice cell is occupied by the player with the highest score among the previous owner and the immediate neighbours. Nowak and all created this model where the model parameter has been the temptation payoff denoted as b. Thus $T = b$. For different values of the parameter b it was shown that cooperators and defectors can persist together indefinitely.

The new topology has been called every since, spatial topology in the literature.

Nowak and moran process.

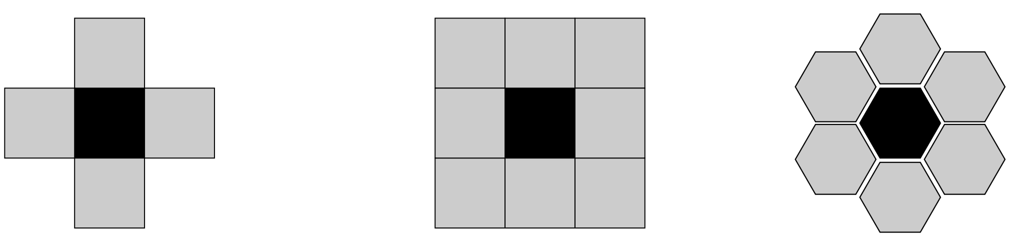

Figure 9: Spatial neighbor

2.5 Software

Due the nature of the research regarding the iterated prisoner's dilemma several software packages have been created in order to simulate the computer tournaments.

The earliest source code that can be found, to the authors knowledge, is that of Axelrod's second tournament [8]. The code has been written by Axelrod and several other contributors in the programming language Fortran [1]. The site specifically says that the course code of the initial tournament is not available.

Another piece of software includes a library called PRISON. PRISON is is written in the programming language Java and it has been used by it's authors in several publications. The project includes a good number of strategies from the literature but unfortunately the last update of the project dates back in 2004.

More recent projects include [2, 3]. Note that [2] is a user interface project that introduces several concept of the research. Such as the iterated game, computer tournaments and evolutionary dynamics. The list of strategies included is very limited. On the other hand, [3] offers a bigger collection of strategies. The users are also able to introduce the concept of noise in their tournaments. Even so, it remains a user interface project and makes it constrained for a research tool.

In 2015 an open source library, called the Axelrod project was introduced [5]. The project is written in the programming language Python, it is accessible and open source. To date the list of strategies implemented within the library exceed the 200. The project has been used in several publications including [23] and a paper describing it and it's capabilities was published in 2016 [27].

Software has a crucial role in research. Well written and maintained softwares allows the reproducibility of prior work and can accelerate findings withing the field. The field of the iterated prisoner's dilemma has suffered the consequences of poor research software. As stated above the source code of the initial computer tournament is not retrievable. Several of the strategies that competed in the tournament are not given a full explanation of how the decided on their next move.

2.6 Applications

2.6.1 Ecological Applications

The reciprocal period of the prisoner's dilemma spread the knowledge of the game not only worldwide but also across different scientific principles. The study of cooperation was once again a critical issue. The applications of the game soon found their way to ecological studies, for example [36] conducted an experiment using sticklebacks to test the robustness of the strategy Tit for Tat in the interactions of fish. Fish usually travel in pairs and monitor their hunters to gain information on the enemy. Other works that include applications to ecological settings have been those of [22, 54]. There the reciprocal food sharing between vampire bats was studied.

2.7 Biological Applications

2.7.1 not sure

In [45], the authors claim that they have managed to re-run the first tournament that Axelrod performed. They tried to push his work further by altering aspects such as, the format of the tournament, the objective and the population. One of the authors claimed to have been a contributor to the first tournaments, which would explain how it was managed to reproduce the tournament.

3 Analysis

References

[1] Complexity of cooperation web site. `http://www-personal.umich.edu/~axe/research/Software/CC/CC2.html`. Accessed: 2017-10-23.

[2] The evolution of trust. `http://ncase.me/trust/`. Accessed: 2017-10-23.

[3] The iterated prisoner's dilemma game. `http://selborne.nl/ipd/`. Accessed: 2017-10-23.

[4] When in doubt... cooperation in a noisy prisoner's dilemma. *The Journal of Conflict Resolution*, 35(4):691–719, 1991.

[5] The Axelrod project developers . Axelrod: ¡release title¿, April 2016.

[6] Wendy Ashlock and Daniel Ashlock. Changes in prisoners dilemma strategies over evolutionary time with different population sizes. pages 297–304, 2006.

[7] Robert Axelrod. Effective choice in the prisoner's dilemma. *The Journal of Conflict Resolution*, 24(1):3–25, 1980.

[8] Robert Axelrod. More effective choice in the prisoner's dilemma. *The Journal of Conflict Resolution*, 24(3):379–403, 1980.

[9] Robert Axelrod. The evolution of strategies in the iterated prisoner's dilemma. *Genetic Algorithms and Simulated Annealing*, pages 32–41, 1987.

[10] Robert Axelrod. Launching the evolution of cooperation. *Journal of Theoretical Biology*, 299(Supplement C):21 – 24, 2012. Evolution of Cooperation.

[11] Robert Axelrod and Douglas Dion. The further evolution of cooperation. *Science*, 242(4884):1385–1390, 1988.

[12] Robert Axelrod and William D. Hamilton. The evolution of cooperation, 1981.

[13] Bruno Beaufils, Jean paul Delahaye, and Philippe Mathieu. Our meeting with gradual: A good strategy for the iterated prisoner's dilemma. 1997.

[14] R Boyd. Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. *Journal of theoretical biology*, 136 1:47–56, 1989.

[15] R. Boyd and J. P. Lorberbaum. No pure strategy is evolutionarily stable in the repeated Prisoner's Dilemma game. *Nature*, 327:58–59, 1987.

[16] Gary W. Evans and Charles M. Crumbaugh. Payment schedule, sequence of choice, and cooperation in the prisoner's dilemma game. *Psychonomic Science*, 5(2):87–88, Feb 1966.

[17] Merrill M. Flood. Some experimental games. *Management Science*, 5(1):5–26, 1958.

[18] Nelis Franken and Andries Petrus Engelbrecht. Particle swarm optimization approaches to coevolve strategies for the iterated prisoner's dilemma. *IEEE Transactions on Evolutionary Computation*, 9(6):562–579, 2005.

[19] Philip S. Gallo and Irina Avery Dale. Experimenter bias in the prisoner's dilemma game. *Psychonomic Science*, 13(6):340–340, Jun 1968.

[20] Marco Gaudesi, Elio Piccolo, Giovanni Squillero, and Alberto Tonda. Exploiting evolutionary modeling to prevail in iterated prisoners dilemma tournaments. *IEEE Transactions on Computational Intelligence and AI in Games*, 8(3):288–300, 2016.

[21] Douglas R. Green. The evolution of forgiveness. *Nature Reviews Molecular Cell Biology*, 12:73, 2011.

[22] H. C. J. Godfray. The evolution of forgiveness. *Nature*, 355:206–207., 1992.

[23] Marc Harper, Vincent Knight, Martin Jones, Georgios Koutsovoulos, Nikoleta E. Glynatsi, and Owen Campbell. Reinforcement learning produces dominant strategies for the iterated prisoner's dil