Understanding responses to environments for the Prisoner's Dilemma; A machine learning approach

Nikoleta E. Glynatsi

Month 2020

Submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy.

School of Mathematics
Ysgol Mathemateg

Executive Summary

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse

platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetuer a, feugiat vitae, porttitor eu,

libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetuer. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Acknowledgements

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Contents

List of Figures

List of Tables

Summary

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Chapter 1

Introduction

Chapter 2

A systematic literature review of the Prisoner's Dilemma.

The Prisoner's Dilemma is a well known game used since the 1950's as a framework for studying the emergence of cooperation; a topic of continuing interest for mathematical, social, biological and ecological sciences. The iterated version of the game, the Iterated Prisoner's Dilemma, attracted attention in the 1980's after the publication of the "The Evolution of Cooperation" and has been a topic of pioneering research ever since. The aim of this paper is to provide a systematic literature review on Prisoner's Dilemma related research. This is achieved by reviewing selected pieces of work and partition the literature into five different sections with each reviewing a different aspect of research. The questions answered in this manuscript are (1) what are the research trends in the field (2) what are the already existing results within the field.

2.1 Introduction

Based on the Darwinian principle of survival of the fittest cooperative behaviour should not be favoured, however, cooperation is plentiful in nature. A paradigm of understanding the emergence of these behaviours is a particular two player non-cooperative game called the Prisoner's Dilemma (PD), originally described in **Flood1958**.

In the PD each player has two choices, to either be selfless and cooperate or to be selfish and defect. Each decision is made simultaneously and independently. The utility of each player is influenced by its own behaviour, and the behaviour of the opponent. Both players do better if they choose to cooperate than if both choose to defect. However, a player has the temptation to deviate as that player will receive a higher payoff than that of mutual cooperation. Players' payoffs are generally represented by (2.1). Both players receive a reward for mutual cooperation, R, and a payoff P for mutual defection. A player that defects while the other cooperates receives a payoff of T, whereas the

cooperator receives S. The dilemma exists due to constraints (2.2) and (2.3).

$$\begin{pmatrix} R & S \\ T & P \end{pmatrix} \tag{2.1}$$

$$T > R > P > S \tag{2.2}$$

$$2R > T + S \tag{2.3}$$

Another common representation of the payoff matrix is given by (2.4), where b is the benefit of the altruistic behaviour and c it's its cost (constraints (2.2) and (2.3) still hold).

$$\begin{pmatrix} b - c & c \\ b & 0 \end{pmatrix} \tag{2.4}$$

Constraints (2.2-2.3) guarantee that it never benefits a player to cooperate, indeed mutual defection is a Nash equilibrium. However, when the game is studied in a manner where prior outcome matters, defecting is no longer necessarily the dominant choice.

The repeated form of the game is called the Iterated Prisoner's Dilemma (IPD) and theoretical works have shown that cooperation can emerge once players interact repeatedly. Arguably, the most important of these works is Robert Axelrod's "The Evolution of Cooperation" **Axelrod1984**. In his book Axelrod reports on a series of computer tournaments he organised. In these tournaments academics from several fields were invited to design computer strategies to compete. Axelrod's work showed that greedy strategies did very poorly in the long run whereas altruistic strategies did better. "The Evolution of Cooperation" is considered a milestone in the field but it is not the only one. On the contrary, the PD has attracted attention ever since the game's origins.

This manuscript presents a qualitative description of selected pieces of work. These have been separated into five sections, each reviewing a different aspect of research. The topics reviewed at each section are the following:

- section 2.2, **Origins of the Prisoner's Dilemma**.

- section 2.3, **Axelrod's tournaments and intelligent design of strategies**.

- section 2.4, **Evolutionary dynamics**

- section 2.5, **Structured strategies and training**.

- section 2.6, **Software**.

The aim of this work is to provide a concrete summary of the existing literature on the PD. This is done to provide a review which will allow the research community to understand overall trends in the field, and already existing results.

2.2 Origins of the prisoner's dilemma

The origin of the PD goes back to the 1950s in early experiments conducted at RAND **Flood1958** to test the applicability of games described in **VonNeumann1944**. The game received its name later the same year. According to **Tucker1983**, Albert W. Tucker (the PhD supervisor of John Nash **Nash1951**), in an attempt to deliver the game with a story during a talk described the players as prisoners and the game has been known as the Prisoner's Dilemma ever since.

The early research on the IPD was limited. The only source of experimental results was through human subject research where pairs of participants simulated plays of the game. Human subject research had disadvantages. Humans could behave randomly and in several experiments both the size and the background of the individuals were different, thus comparing results of two or more studies became difficult.

The main aim of these early research experiments was to understand how conditions such as the gender of the participants **Evans1966**; **Lutzker1961**; **Mack1971**, the physical distance between the participants **Sensenig1972**, the effect of their opening moves **Tedeschi1968** and even how the experimenter, by varying the tone of their voice and facial expressions **Gallo1968**, could influence the outcomes and subsequently the emergence of cooperation. An early figure that sought out to understand several of these conditions was the mathematical psychologist Anatol Rapoport. The results of his work are summarised in **rapoport1965**.

Rapoport was also interested in conceptualising strategies that could promote international cooperation. Decades later he would submit the winning strategy (Tit for Tat) of the first computer tournament, run by Robert Axelrod. In the next section these tournaments, and several strategies that were designed by researchers, such as Rapoport, are introduced.

2.3 Axelrod's tournaments and intelligently designed strategies

As discussed in Section 2.2, before 1980 a great deal of research was done in the field, however, as described in **Axelrod2012**, the political scientist Robert Axelrod believed that there was no clear answer to the question of how to avoid conflict, or even how

an individual should play the game. Combining his interest in artificial intelligence
and political science Axelrod created a framework for exploring these questions using
computer tournaments. Axelrod asked researchers to design a strategy with the purpose
of wining an IPD tournament. This section covers Axelrod's original tournaments as
well as research that introduced new intelligently designed strategies.

Axelrod's tournaments made the study of cooperation of critical interest. As described
in **Rapoport2015**, "Axelrod's "new approach" has been extremely successful and im-
mensely influential in casting light on the conflict between an individual and the collec-
tive rationality reflected in the choices of a population whose members are unknown and
its size unspecified, thereby opening a new avenue of research". In a collaboration with
a colleague, Douglas Dion, Axelrod in **Axelrod1988** summarized a number of works
that were immediately inspired from the "Evolution of Cooperation", and **Jurisic2012**
gives a review of tournaments that have been conducted since the originals.

The first reported computer tournament took place in 1980 **Axelrod1980a**. A total
of 13 strategies were submitted, written in the programming languages Fortran or
Basic. Each competed in a 200 turn match against all 12 opponents, itself and a
player that played randomly (called **Random**). This type of tournament is referred
to as a round robin. The tournament was repeated 5 times to get a more stable
estimate of the scores for each pair of play. Each participant knew the exact number
of turns and had access to the full history of each match. Furthermore, Axelrod
performed a preliminary tournament and the results were known to the participants.
This preliminary tournament is mentioned in **Axelrod1980a** but no details were given.
The payoff values used for equation (2.1) were $R = 3, P = 1, T = 5$ and $S = 0$. These
values are commonly used in the literature and unless specified will be the values used
in the rest of the works described here.

The winner of the tournament was determined by the total average score and not by
the number of matches won. The strategy that was announced the winner was the
strategy submitted by Rapoport, **Tit For Tat**. The success of Tit for Tat came as a
surprise. It was not only the simplest submitted strategy, it would always cooperates on
the first round and then mimic the opponent's previous move, but it had also won the
tournament even though it could never beat any player it was interacting with.

In order to further test the results Axelrod performed a second tournament in 1980 **Axelrod1980b**.
The second tournament received much more attention and had a total of 62 entries.
The participants knew the results of the previous tournament and the rules were similar
with only a few alterations. The tournament was repeated 5 times and the length of
each match was not known to the participants. Axelrod intended to use a fixed proba-
bility (refereed to as 'shadow of the future' **Axelrod1988**) of the game ending on the
next move. However, 5 different number of turns were selected for each match 63, 77,

151, 308 and 401, such that the average length would be around 200 turns.

Nine of the original participants competed again in the second tournament. Two strategies that remained the same were Tit For Tat and **Grudger**. Grudger is a strategy that will cooperate as long as the opponent does not defect, submitted by James W. Friedman. The name Grudger was give to the strategy in **Li2014**, though the strategy goes by many names in the literature such as, Spite **Beaufils1997**, Grim Trigger **Banks1990** and Grim **Van2015**. New entries in the second tournament included **Tit for Two Tats** submitted by John Maynard Smith and **KPavlovC**. KPavlovC, is also known as Simpleton **rapoport1965**, introduced by Rapoport or just Pavlov **Nowak1993**. The strategy is based on the fundamental behavioural mechanism win-stay, lose-shift. Pavlov is heavily studied in the literature and similarly to Tit for Tat it is used in tournaments today and has had many variants trying to build upon it's success, for example **PavlovD** and **Adaptive Pavlov Li2007**.

Despite the larger size of the second tournament none of the new entries managed to outperform the simpler designed strategy. The winner was once again Tit for Tat. Axelrod deduced the following guidelines for a strategy to perform well:

- The strategy would start of by cooperating.

- It would forgive it's opponent after a defection.

- It would always be provoked by a defection no matter the history.

- It was simple.

The success of Tit for Tat, however, was not unquestionable. Several papers showed that stochastic uncertainties severely undercut the effectiveness of reciprocating strategies and such stochastic uncertainties have to be expected in real life situations **Milinski1987**. For example, in **Molander1985** it is proven that in an environment where **noise** (a probability that a player's move will be flipped) is introduced two strategies playing Tit for Tat receive the same average payoff as two Random players. Hammerstein, pointed out that if by mistake, one of two Tit for Tat players makes a wrong move, this locks the two opponents into a hopeless sequence of alternating defections and cooperations **Hammerstein1984**.

The poor performance of the strategy in noisy environments was also demonstrated in tournaments. In **Bendor1991**; **Donninger1986** round robin tournaments with noise were performed, and Tit For Tat did not win. The authors concluded that to overcome the noise more generous strategies than Tit For Tat were needed. They introduced the strategies **Nice and Forgiving** and **OmegaTFT** respectively.

A second type of stochastic uncertainty is misperception, where a player's action is made correctly but it is recorded incorrectly by the opponent. In **Wu1995**, a strategy

called **Contrite Tit for Tat** was introduced that was more successful than Tit for
Tat in such environments. The difference between the strategies was that Contrite Tit
for Tat was not so fast to retaliate against a defection.

Several works extended the reciprocity based approach which has led to new strategies.
For example Gradual **Beaufils1997** which was constructed to have the same qualities
as those of Tit for Tat except one, **Gradual** had a memory of the game since the
beginning of it. Gradual recorded the number of defections by the opponent and pun-
ished them with a growing number of defections. It would then enter a calming state in
which it would cooperates for two rounds. In a tournament of 12 strategies, including
both Tit for Tat and Pavlov, Gradual managed to outperformed them all. A strat-
egy with the same intuition as Gradual is **Adaptive Tit for Tat tzafestas-2000a**.
Adaptive Tit for Tat does not keep a permanent count of past defections, it maintains
a continually updated estimate of the opponent's behaviour, and uses this estimate to
condition its future actions. In the exact same tournament as in **Beaufils1997** with
now 13 strategies Adaptive Tit for Tat ranked first.

Another extension of strategies was that of teams of strategies **J.P.Delahaye1993Lp**;
J.P.Delahaye1995LIeP; **A.Rogers2007Ctpw** that collude to increase one mem-
ber's score. In 2004 Graham Kendall led the Anniversary Iterated Prisoner's Dilemma
Tournament with a total of 223 entries. In this tournament participants were allowed
to submit multiple strategies. A team from the University of Southampton submit-
ted a total of 60 strategies **A.Rogers2007Ctpw**. All these were strategies that had
been programmed with a recognition mechanism by default. Once the strategies recog-
nised one another, one would act as leader and the other as a follower. The follower
plays as a **Cooperator**, cooperates unconditionally and the leader would play as a
Defector gaining the highest achievable score. The followers would defect uncondi-
tionally against other strategies to lower their score and help the leader. The result
was that Southampton had the top three performers. Nick Jennings, who was part
of the team, said that "We developed ways of looking at the Prisoner's Dilemma in a
more realistic environment and we devised a way for computer agents to recognise and
collude with one another despite the noise. Our solution beats the standard Tit For
Tat strategy" **southampton˙blog**.

2.3.1 Memory one Strategies

A set of strategies that have received a lot of attention in the literature are **memory
one** strategies. In **nowak1989**, Nowak and Sigmund proposed a structure for studying
simple strategies that remembered only the previous turn, and moreover, only recorded
the move of the opponent. These are called **reactive** strategies and they can be
represented by using three parameters (y, p_1, p_2), where y is the probability to cooperate

in the first move, and p_1 and p_2 the conditional probabilities to cooperate, given that the opponent's last move was a cooperation or a defection. For example Tit For Tat is a reactive strategy and it can be written as $(1, 1, 0)$. Another reactive strategy well known in the literature is **Generous Tit for Tat Nowak1992**.

In **Nowak1990**, Nowak and Sigmund extended their work to include strategies which consider the entire history of the previous turn to make a decision. These are called **memory one** strategies. If only a single turn of the game is taken into account and depending on the simultaneous moves of the two players there are only four possible states that the players could be in. These are:

- Both players cooperated, denoted as CC.

- First player cooperated while the second one defected, denoted as CD.

- First player defected while the second one cooperated, denoted as DC.

- Both players defected, denoted as DD.

Thus a memory one strategy can be denoted by the probabilities of cooperating after each state and the probability of cooperating in the first round, (y, p_1, p_2, p_3, p_4). For example Pavlov's memory one representation is $(1, 1, 0, 0, 1)$.

Memory one strategies made an impact when a specific set of memory one strategies were introduced called **Zero-determinant** (ZD) **Press2012**. The American Mathematical Society's news section **hilbe2015** stated that "the world of game theory is currently on fire" and in **Stewart2012** it was stated that "Press and Dyson have fundamentally changed the viewpoint on the Prisoner's Dilemma". ZD are a set of extortionate strategies that can force a linear relationship between the long-run scores of both themselves and the opponent, therefore ensuring that the opponent will never do better than them.

Press and Dyson's suggested ZD strategies were the dominant family of strategies in the IPD. Moreover, they argued that memory is not beneficial. In **Adami2013**; **Knight2017**; **Hilbe2013**; **Hilbe2013b**; **hilbe2015**; **KnightHGC17**; **Knight2019**; **Lee2015**; **Stewart2012** the effectiveness of ZD strategies is questioned. In **Adami2013**, it was shown that ZD strategies are not evolutionary stable, and in **Stewart2012** a more generous set of ZDs, the **Generous ZD**, were shown to outperform the more extortionate ZDs. Finally, in **Knight2017**; **KnightHGC17**; **Knight2019**; **Lee2015**, the 'memory does not benefit a strategy' 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.

This section covered the original computer tournaments of Axelrod and the early success of Tit For Tat in these tournaments. Though Tit For Tat was considered to be the

most robust basic strategy, reciprocity was found to not be enough in environments
with uncertainties. There are at least two properties, that have been discussed in this
section, for coping with such uncertainties; generosity and contrition. Generosity is let-
ting a percentage of defections go unpunished, and contrition is lowering a strategy's
readiness to defect following an opponent's defection.

In the later part of this section a series of new strategies which were built on the basic
reciprocal approaches were presented, followed by the infamous memory one strategies,
the zero-determinant strategies. Though the ZDs can be proven to be robust in pairwise
interactions they were found to be lacking in evolutionary settings and in computer
tournaments. Evolutionary settings and the emergence of cooperation under natural
selection are covered in the next section.

2.4 Evolutionary dynamics

As yet, the emergence of cooperation has been discussed in the contexts of the one shot
PD game and the IPD round robin tournaments. In the PD it is proven that coopera-
tion will not emerge, furthermore, in a series of influential works Axelrod demonstrated
that reciprocal behaviour favours cooperation when individuals interact repeatedly.
But does natural selection favours cooperation? Understanding the conditions under
which natural selection can favour cooperative behaviour is important in understanding
social behaviour amongst intelligent agents **Boyd1987**.

Imagine a mixed population of cooperators and defectors where every time two individ-
uals meet they play a game of PD. In such population the average payoff for defectors
is always higher than cooperators. Under natural selection the frequency of defectors
will steadily increase until cooperators become extinct. Thus natural selection favours
defection in the PD (Figure 2.1). However, there are several mechanisms that allow
the emergence of cooperation in an evolutionary context which will be covered in this
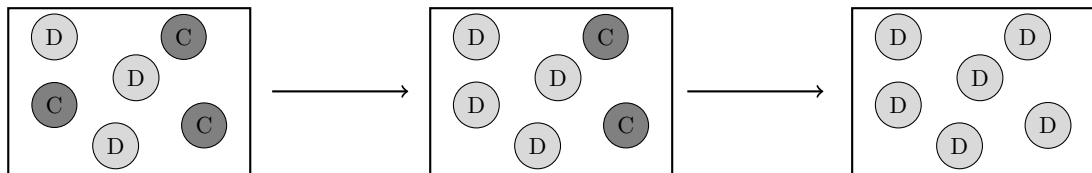section.

Figure 2.1: Natural selection favours defection in a mixed population of Cooperators
and Defectors.

In the later sections of **Axelrod1980b**, Axelrod discusses an ecological tournament
that he performed using the 62 strategies of the second tournament to understand the
reproductive success of Tit for Tat. In his ecological tournament the prevalence of each
type of strategy in each round was determined by that strategy's success in the previous

round. The competition in each round would become stronger as weaker performers were reduced and eliminated. The ecological simulation concluded with a handful of nice strategies dominating the population whilst exploitative strategies had died off as weaker strategies were becoming extinct. This new result led Axelrod to study the IPD in an evolutionary context based on several of the approaches established by the biologist John M. Smith **Smith1974**; **Smith1979**; **Smith1973**. John M. Smith was a fundamental figure in evolutionary game theory and a participant of Axelrod's second tournament. Axelrod and the biologist William Donald Hamilton wrote about the biological applications of the evolutionary dynamics of the IPD **Axelrod1984** and won the Newcomb-Cleveland prize of the American Association for the Advancement of Science.

In Axelrod's model **axelrod1981** pairs of individuals from a population played the IPD. The number of interactions between the pairs were not fixed, but there was a probability defined as the importance of the future of the game w, where $0 < w < 1$, that the pair would interact again. In **axelrod1981** it was shown that for a sufficient high w Tit For Tat strategies would become common and remain common because they were "collectively stable". Axelrod argued that collective stability implied evolutionary stability (ESS) and that when a collectively stable strategy is common in a population and individuals are paired randomly, no other rare strategy can invade. However, Boyd and Lorderbaum in **Boyd1987** proved that if w, the importance of the future of the game, is large enough then no pure strategy is ESS because it can always be invaded by any pair of other strategies. This was also independently proven in **Pudaite1987**.

All these conclusions were made in populations where the individuals could all interact with each other. In 1992, Nowak and May, considered a structured population where an individual's interactions were limited to its neighbours. More specifically, in **Nowak1992b** they explored how local interaction alone can facilitate population wide cooperation in a one shot PD game. The two deterministic strategies Defector and Cooperator, were placed onto a two dimensional square array where the individuals could interact only with the immediate neighbours. The number of immediate neighbours could be either, fourth, six or eight, as shown in Figure 2.2, where each node represents a player and the edges denote whether two players will interact. This topology is refereed to as spatial topology. Each cell of the lattice is occupied by a Cooperator or a Defector and at each generation step each cell owner interacts with its immediate neighbours. The score of each player is calculated as the sum of all the scores the player achieved at each generation. At the start of the next generation, each lattice cell is occupied by the player with the highest score among the previous owner and their immediate neighbours.

Local interactions proved that as long as small clusters of cooperators form, where they can benefit from interactions with other cooperators while avoiding interactions with defectors, global cooperation will continue. Thus, local interactions proved that even for the PD cooperation can emerge. Moreover in **Ohtsuki2006**, whilst using the payoff matrix (2.4), it was shown that cooperation will evolve in a structured population as long as the benefit to cost ratio b/c is higher than the number of neighbours. In **Perc2011**, graphs were a probability of rewiring ones connections was considered were studied. The rewire could be with any given node in the graphs and not just with immediate neighbours. Perc et al. concluded that "making new friends" may be an important activity for the successful evolution of cooperation, but also they must be selected carefully and one should keep their number limited.
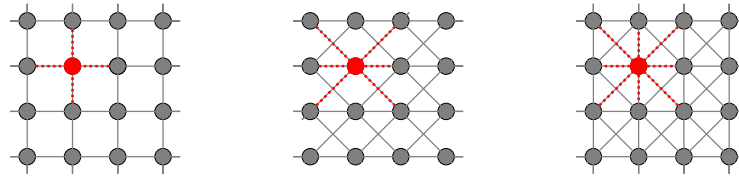

Figure 2.2: Spatial neighbourhoods

Another approach for increasing the likelihood of cooperation by increasing of assortative interactions among cooperative agents, include partner identification methods such as reputation **Janssen2006**; **Nowak1998**; **Suzuki2005**, communication tokens **Miller2002** and tags **Choi2006**; **Hales2000**; **Miller2002**; **Riolo2001**.

In this section evolutionary dynamics and the emergence of cooperation were reviewed. The following section focuses on strategy archetypes, training methods and strategies obtained from training.

2.5 Structured strategies and training

This section covers strategies that are different to that of intelligent design discussed in Section 2.3. These are strategies that have been **trained** using generic strategy archetypes. For example, in **Axelrod1987** Axelrod decided to explore deterministic strategies that took into account the last 3 turns of the game. As discussed in Section 2.3.1, for each turn there are 4 possible outcomes, CC, CD, DC, DD, thus for 3 turns there are a total of $4 \times 4 \times 4 = 64$ possible combinations. Therefore, the strategy can be defined by a series of 64 C's/D's, corresponding to each combination; this type of strategy is called a lookup table. This lookup table was then trained using a genetic algorithm **Koza1997**. During the training process random changes are made to a given lookup table. If the utility of the strategy has increased this change is kept,

otherwise not.

In 1996 John Miller considered finite state automata as an archetype **Miller1996**, more specifically, Moore machines **moore1956**. He used a genetic algorithm to train finite state machines in environments with noise. Miller's results showed that even a small difference in noise (from 1% to 3%) significantly changed the characteristics of the evolving strategies. The strategies he introduced were **Punish Twice**, **Punish Once for Two Tats** and **Punish Twice and Wait**. In **Ashlock2006b** finite state automata and genetic algorithms were also used to introduce new strategies. In a series of experiments where the size of the population varied, there were two strategies frequently developed by the training process and more over they were developed only after the evolution had gone on for many generations. These were **Fortess3** and **Fortess4**. Following Miller's work in 1996, the first structured strategies based on neural networks that had be trained using a genetic algorithm was introduced in **Harrald1996** by Harrald and Fogel. Harrald and Fogel considered a single layered neural network which had 6 inputs. These were the last 3 moves of the player and the opponent, similar to **Axelrod1987**. Neural networks have broadly been used to train IPD strategies since then with genetic algorithms **Ashlock2006a**; **Chong2005**; **Marks1999** and particle swarm optimization **Franken2005**.

In **Knight2017**; **KnightHGC17** both genetic algorithm and particle swarm optimization were used to introduce a series of structured strategies based on lookup tables, finite state machines, neural networks, hidden Markov models **eddy1996** and Gambler. Hidden Markov models, are a stochastic variant of a finite state machine and Gamblers are stochastic variants of lookup tables. The structured strategies that arised from the training were put up against a large number of strategies in (1) a Moran process, which is an evolutionary model of invasion and resistance across time during which high performing individuals are more likely to be replicated and (2) a round robin tournament. In a round robin tournament which was simulated using the software **axelrodproject** and the 200 strategies implemented within the software, the top spots were dominated by the trained strategies of all the archetypes. The top three strategies were **Evolved LookUp 2 2 2**, **Evolved HMM 5** and **Evolved FSM 16**.

In **KnightHGC17** it was demonstrated that these trained strategies would overtake the population in a Moran process. The strategies evolved an ability to recognise themselves by using a handshake. This recognition mechanism allowed the strategies to resist invasion by increasing the interactions between themselves, an approach similar to the one described in Section 2.4.

Throughout the different methods of training that have been discussed in this section, a spectrum of structured strategies can be found. Differentiating between strategies is not always straightforward. It is not obvious looking at a finite state diagram how a

machine will behave, and many different machines, or neural networks can represent
the same strategy. For example Figure 2.3 shows two finite automata and both are a
representation of Tit for Tat.

(a) Tit for Tat as a finite state machine (b) Tit for Tat as a finite state machine
with 1 state.   with 2 states.

Figure 2.3: Finite state machine representations of Tit for Tat. A machine consists of
transition arrows associated with the states. Each arrow is labelled with A/R where
A is the opponent's last action and R is the player's response. Finite state machines
consist of a set of internal states. In (a) Tit for Tat finite state machine consists of 1
state and in (b) of 2.

To allow for identification of similar strategies a method called fingerprinting was in-
troduced in **Ashlock2005**. The method of fingerprinting is a technique for generating
a functional signature for a strategy **Ashlock2008**. This is achieved by computing the
score of a strategy against a spectrum of opponents. The basic method is to play the
strategy against a probe strategy with varying noise parameters. In **Ashlock2005** Tit
for Tat is used as the probe strategy. In Figure 2.4 an example of Pavlov's fingerprint
is given. Fingerprinting has been studied in depth in **Ashlock2008**; **Ashlock2009**;
Ashlock2010; **Ashlock2006a**. Another type of fingerprinting is the transitive fin-
gerprint **axelrodproject**. The method represents the cooperation rate of a strategy
against a set of opponents over a number of turns. An example of a transitive finger-
print is given in Figure 2.5.

This section covered structured strategies and training methods. In the following
section software that has been developed with main aim simulating the IPD is pre-
sented.

2.6 Software

The research of the IPD heavily relies on software. This is to be expected as computer
tournaments have become the main means of simulating the interactions in an IPD
game. Many academic fields suffer from lack of source code availability and the IPD
is not an exception. Several of the tournaments that have been discussed so far were
generated using computer code, though not all of the source code is available. The
code for Axelrod's original tournament is known to be lost and moreover for the second
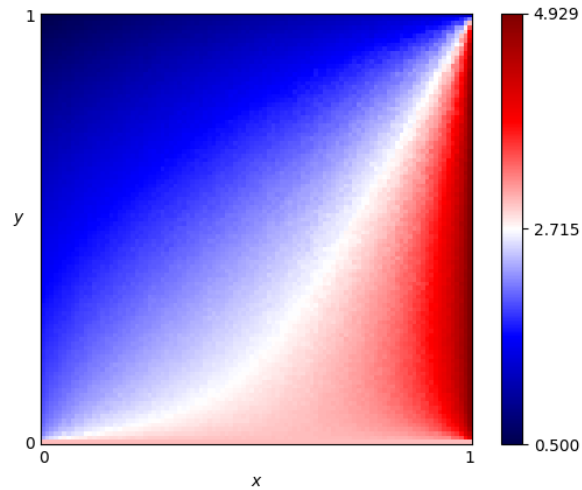
Figure 2.4: Pavlov fingerprinting with Tit for Tat used as the probe strategy. Figure
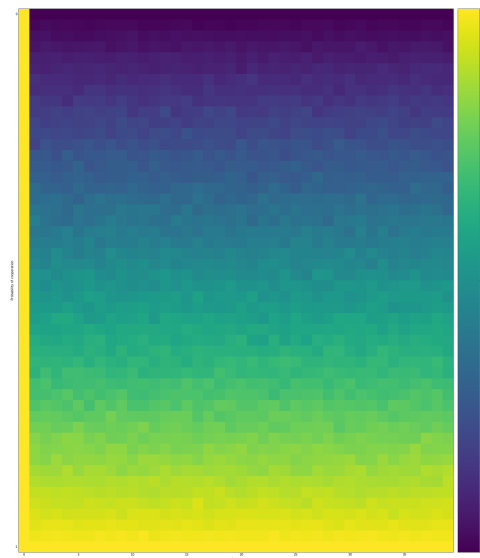was generated using **axelrodproject**.

Figure 2.5: Transitive fingerprint of Tit for Tat against a set of 50 random opponents.

tournament the only source code available is the code for the 62 strategies (found on
Axelrod's personal website **fortan˙code**).

Several projects, however, are open, available and have been used as research tools or
educational platforms over the years. Two research tools **prison**; **axelrodproject** and
two educational tools **pd˙trust**; **pd˙game** are briefly mentioned here. Both **prison**;
axelrodproject are open source projects. The "Game of Trust" **pd˙trust** is an on-
line, graphical user interface educational platform for learning the basics of game theory,
the IPD and the notion of strategies. It attracted a lot of attention due to being "well-
presented with scribble-y hand drawn characters" **trust˙blogb** and "a whole heap of
fun" **trust˙bloga**. Finally **pd˙game** is a personal project written in PHP. It is a
graphical user interface that offers a big collection of strategies and allows the user to
try several matches and tournament configurations.

PRISON **prison** is written in the programming language Java and a preliminary ver-
sion was launched on 1998. It was used by its authors in several publications, such
as **Beaufils1997**, which introduced Gradual, and **Beaufils1988**. The project includes
a good number of strategies from the literature but unfortunately the last update of
the project dates back to 2004. Axelrod-Python **axelrodproject** is a software used
by **Knight2017**; **KnightHGC17**; **Goodman2018**; **Wang2017**. It is written in the
programming language Python following best practice approaches **Aberdour2007**;
Benureau2018 and contains the largest collection of strategies, known to the author.
The strategy list of the project has been cited by publications **Anastassacos2018**;
Hayes2017; **Neumann2018**.

2.7 Conclusion

This manuscript presented a literature review on the Iterated Prisoner's Dilemma.
The opening sections focused on research trends and published works of the field,
followed by a presentation of research and educational software. More specifically,
Section 2.2 covered the early years of research. This was when simulating turns of
the game was only possible with human subject research. Following the early years,
the pioneering tournaments of Axelord were introduced in Section 2.3. Axelrod's work
offered the field an agent based game theoretic framework to study the IPD. In his
original papers he asked researchers to design strategies to test their performance with
the new framework. The winning strategy of both his tournaments was Tit for Tat.
The strategy however came with limitations which were explored by other researchers,
and new intelligently designed strategies were introduced in order to surpass Tit for
Tat with some contributions such as Pavlov and Gradual.

Soon researchers came to realise that strategies should not just do well in a tournament
setting but should also be evolutionary robust. Evolutionary dynamic methods were

applied to many works in the field, and factors under which cooperation emerges were explored, as described in Section 2.4. This was not done only for unstructured populations, where all strategies in the population can interact with each other, but also in population where interactions were limited to only strategies that were close to each other. In such topologies it was proven that even in the one shot game, cooperation can indeed emerge.

Evolutionary approaches can offer many insights in the study of the PD. In evolutionary settings strategies can learn to adapt and take over population by adjusting their actions; such algorithms can be applied so that evolutionarily robust strategies can emerge. Algorithms and structures used to train strategies in the literature were covered in Section 2.5. From these training methods several strategies are found, and to be able to differentiate between them fingerprinting was introduced. The research of best play and cooperation has been going on since the 1950s, and for simulating the game software has been developed along the way. This software has been briefly discussed in Section 2.6.

The study of the PD is still an ongoing field research where new variants and new structures of strategies are continuously being explored **Ohtsuki2018**. The game now serves as a model in a wide range of applications, for example in medicine and the study of cancer cells **archetti2018**; **Kaznatchee2017**, as well as in social situations and how they can be driven by rewards **Dridi2018**. New research is still ongoing for example in evolutionarily dynamics on graphs **Allen2017**; **hathcock2018**; **Liu2017**.