Stability of defection, optimisation pf strategy and the limits of memory in the Prisoner's Dilemma.

Nikoleta E. Glynatsi Vincent Knight

Abstract

In this manuscript we build upon a framework provided in 1989 to study best responses in the well known memory one strategies of the Iterated Prisoner's Dilemma. The aim of this work is to construct a compact way of identifying best responses of short memory strategies and to show their limitations in multi-opponent interactions. A number of theoretic results are presented.

1 Introduction

The Prisoner's Dilemma (PD) is a two player person game used in understanding the evolution of co-operative behaviour. Each player can choose between cooperation (C) and defection (D). The decisions are made simultaneously and independently. The normal form representation of the game is given by:

$$S_p = \begin{pmatrix} R & S \\ T & P \end{pmatrix} \quad S_q = \begin{pmatrix} R & T \\ S & P \end{pmatrix} \tag{1}$$

where S_p represents the utilities of the row player and S_q the utilities of the column player. The payoffs, (R, P, S, T) (the most common values used in the literature are $(3, 1, 0, 5)$ [4]), are constrained by equations (2) and (3). Constraint (2) ensures that defection dominates cooperation and constraint (3) ensures that there is a dilemma; the sum of the utilities for both players is better when both choose to cooperate.

$$T > R > P > S \tag{2}$$

$$2R > T + S \tag{3}$$

The PD is a one shot game, however it is commonly studied in a manner where the history of the interactions matter. The repeated form of the game is called the Iterated Prisoner's Dilemma (IPD) and in the 1980s following the work of [5, 6] it attracted the attention of the scientific community.

In [5] and [6], the first well known computer tournaments of the IPD were performed. A total of 13 and 63 strategies were submitted in computer code and competed against each other in a round robin tournament. All contestants competed against each other, a copy of themselves and a random strategy. The winner was decided on the average score a strategy achieved and not the total number of wins. How many turns of history that a strategy would use, the memory size, was left to the creator of the strategy to decide.

The winning strategy of both tournaments was a strategy called Tit for Tat. Tit for Tat is a strategy which starts by cooperating and then mimics the last move of it's opponent. This is a strategy which makes use of the previous move of the opponent only and a reactive strategy. In [16] a framework for studying such strategies was introduced. This was later used to introduce well known reactive strategies such as Generous Tit For Tat [17].

Reactive strategies are a subset of memory one strategies. Memory one strategies similarly are only concerned with the previous turn. However, they take into consideration both players' recent moves to decide on an action. Several successful memory one strategies are found in the literature, for example Pavlov [15].

A well known set of memory on strategies was introduced in [18], these were called zero determinant (ZD) strategies. The ZD strategies manage to force a linear relationship between the score of the strategy and the opponent. According to [18] the ZD strategies can dominate any evolutionary opponent in pairwise interactions by using a single slot of memory. Thus the usefulness of memory in the IPD was questioned.

The ZD strategies attracted a lot of attention. It was stated that "Press and Dyson have fundamentally changed the viewpoint on the Prisoner's Dilemma" [19]. In [19] a very similar tournament to Axelrod's tournament is run including ZD strategies and a new set of ZD strategies the Generous ZD. One specific advantage of memory one strategies is their mathematical tractability. As described in Section 2 they can be represented completely as an element of $\mathbb{R}^4$.

Even so, ZD and memory one strategies have also received criticism. In [14], 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 robust against multiple opponents.

The purpose of this work is to consider a given memory one strategy in a similar fashion to [18]. However whilst [18] found a way for a player to manipulate a given opponent, this work will consider a multidimensional optimisation approach to identify the best response to a group of opponents. In essence the aim is to produce a compact method of identifying the best memory one strategy against a given opponent.

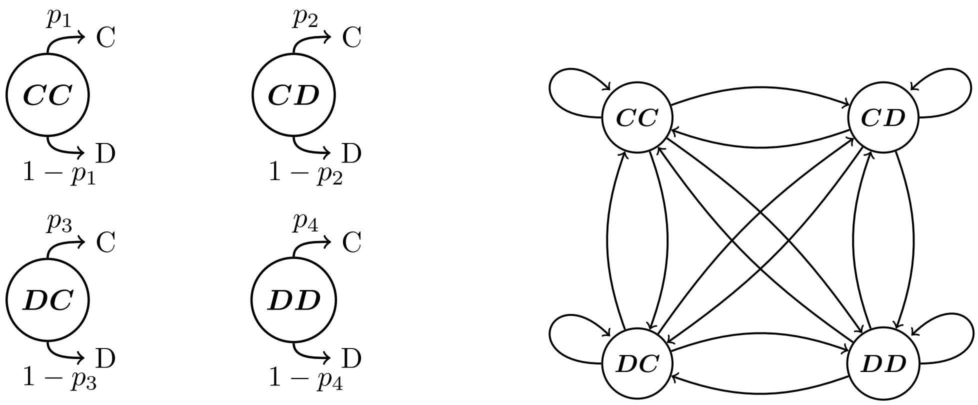
In the second part of this manuscript we explore the limitation of these best response strategies. This is achieved by comparing the performance of an optimal memory one strategy, for a given environment, with the performance of a more complex strategy that has a larger memory.

One particular benefit of this analysis is the identification of conditions for which defection is a best response. Thus, laos identifying environments for which cooperation can not occur.

2 The utility against memory one players

In [18] a framework is described to sufficient study the interactions of memory one strategies modelled as a stochastic process. In this manuscript it is stated that if a strategy is concerned with only the outcome of a single turn then there are four possible 'states' the strategy could be in. These are CC, CD, DC, CC. A memory one strategy is denoted by the probabilities of cooperating after each of these states, $p = (p_1, p_2, p_3, p_4) \in \mathbb{R}^4_{[0,1]}$. A diagrammatic representation of such strategy is given in Figure 1a.

Moreover, if two players are moving from state to state following a general memory one strategy this can be modelled as a Markov process. A diagrammatic representation of the Markov chain is shown in Figure 1b. The corresponding transition matrix M given by:



(a) Diagrammatic representation of a memory one strategy.

(b) Markov chain on a PD game.

$$M = \begin{bmatrix} p_1q_1 & p_1(-q_1+1) & q_1(-p_1+1) & (-p_1+1)(-q_1+1) \\ p_2q_3 & p_2(-q_3+1) & q_3(-p_2+1) & (-p_2+1)(-q_3+1) \\ p_3q_2 & p_3(-q_2+1) & q_2(-p_3+1) & (-p_3+1)(-q_2+1) \\ p_4q_4 & p_4(-q_4+1) & q_4(-p_4+1) & (-p_4+1)(-q_4+1) \end{bmatrix}. \tag{4}$$

The long run steady state probability v is the solution to $\bigtriangledown M = \bigtriangledown$ (given in the Appendix). Combining the stationary vector v with the payoff matrices of equation (1) allow us to retrieve the expected payoffs for each player. Thus, the utility for player p against q, denoted as $u_q(p)$, is defined by,

$$u_q(p) = v \times (R, P, S, T). \tag{5}$$

Here we present the first theoretical result which concerns the form of $u_q(p)$. That is that $u_q(p)$ is given by a ratio of two quadratic forms [12], as presented by Theorem 1.

Theorem 1 *The expected utility of a memory one strategy $p \in \mathbb{R}^4_{[0,1]}$ against a memory one opponent strategy $q \in \mathbb{R}^4_{[0,1]}$, denoted as $u_q(p)$, can be written as a ratio of two quadratic forms:*

$$u_q(p) = \frac{\frac{1}{2}pQp^T + cp + a}{\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a}}, \tag{6}$$

where $Q, \bar{Q} \in \mathbb{R}^{4\times 4}$ matrices defined by the transition probabilities of the opponent q_1, q_2, q_3, q_4 as follows:

$$\boldsymbol{Q} = \begin{bmatrix} 0 & -(q_1-q_3)(q_2-5q_4-1) & q_3(q_1-q_2) & -5q_3(q_1-q_4) \\ -(q_1-q_3)(q_2-5q_4-1) & 0 & (q_2-q_3)(q_1-3q_4-1) & (q_3-q_4)(5q_1-3q_2-2) \\ q_3(q_1-q_2) & (q_2-q_3)(q_1-3q_4-1) & 0 & 3q_3(q_2-q_4) \\ -5q_3(q_1-q_4) & (q_3-q_4)(5q_1-3q_2-2) & 3q_3(q_2-q_4) & 0 \end{bmatrix}, \tag{7}$$

$$\bar{Q} = \begin{bmatrix} 0 & -(q_1 - q_3)(q_2 - q_4 - 1) & (q_1 - q_2)(q_3 - q_4) & (q_1 - q_4)(q_2 - q_3 - 1) \\ -(q_1 - q_3)(q_2 - q_4 - 1) & 0 & (q_2 - q_3)(q_1 - q_4 - 1) & (q_1 - q_2)(q_3 - q_4) \\ (q_1 - q_2)(q_3 - q_4) & (q_2 - q_3)(q_1 - q_4 - 1) & 0 & -(q_2 - q_4)(q_1 - q_3 - 1) \\ (q_1 - q_4)(q_2 - q_3 - 1) & (q_1 - q_2)(q_3 - q_4) & -(q_2 - q_4)(q_1 - q_3 - 1) & 0 \end{bmatrix}. \tag{8}$$

c and $\bar{c} \in \mathbb{R}^{4\times1}$ defined by:

$$c = \begin{bmatrix} q_1(q_2 - 5q_4 - 1) \\ -(q_3 - 1)(q_2 - 5q_4 - 1) \\ -q_1q_2 + q_2q_3 + 3q_2q_4 + q_2 - q_3 \\ 5q_1q_4 - 3q_2q_4 - 5q_3q_4 + 5q_3 - 2q_4 \end{bmatrix}, \tag{9}$$

$$\bar{c} = \begin{bmatrix} q_1(q_2 - q_4 - 1) \\ -(q_3 - 1)(q_2 - q_4 - 1) \\ -q_1q_2 + q_2q_3 + q_2 - q_3 + q_4 \\ q_1q_4 - q_2 - q_3q_4 + q_3 - q_4 + 1 \end{bmatrix}. \tag{10}$$

and $a = -q_2 + 5q_4 + 1$ and $\bar{a} = -q_2 + q_4 + 1$.

Proof: The proof is given in Appendix.

Figure 2 indicates that the formulation of $u_q(p)$ as a quadratic ratio successfully captures the simulated behaviour. A data set offering further validation is available at.

The simulated utility, denoted as $U_q(p)$ has been calculated using [1]. It is an open research framework for the study of the IPD and is described in [13]. Note that when referring to $U_q(p)$ here onwards we mean the simulated utility calculated with [1].

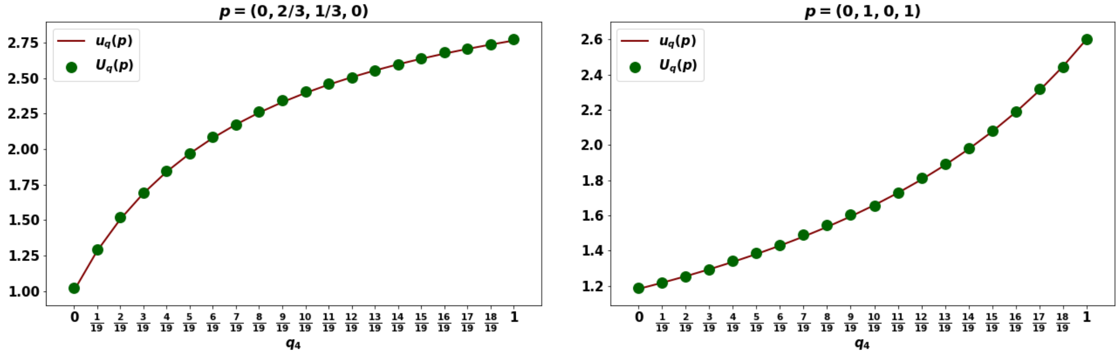


Figure 2: Differences between simulated and analytical results for $q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, q_4)$.

Moreover Theorem 1 can be expanded to multi opponent interactions. The IPD is commonly studied in tournaments where a strategy interacts with a number of opponents. There the payoff of a player is given by the average payoffs the player achieved. More specifically the expected utility of a memory one strategy against a N number of opponents is given by Theorem 2.

Theorem 2 *The expected utility of a memory one strategy $p \in \mathbb{R}^4_{[0,1]}$ against a group of opponents $q^{(1)}, q^{(2)}, \dots, q^{(N)}$, denoted as $\frac{1}{N}\sum\limits_{i=1}^{N} {u_q}^{(i)}(p)$ is given by:*

$$\frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p) = \frac{1}{N}\frac{\sum\limits_{i=1}^{N}(\frac{1}{2}pQ^{(i)}p^T + c^{(i)}p + a^{(i)}) \prod\limits_{\substack{j=1 \\ j \neq i}}^{N} (\frac{1}{2}p\bar{Q}^{(i)}p^T + \bar{c}^{(i)}p + \bar{a}^{(i)})}{\prod\limits_{i=1}^{N}(\frac{1}{2}p\bar{Q}^{(i)}p^T + \bar{c}^{(i)}p + \bar{a}^{(i)})}. \tag{11}$$

To validate the formulation of Theorem 2 we calculate the simulated and the theoretical payoffs of several memory one strategies against a set of 10 opponents. The opponents used are the memory one strategies for the tournament conducted in [19]. The names and a small explanation of the strategic rules is given in the Appendix A. Both values $\frac{1}{N}\sum u$ and $\frac{1}{N}\sum U$ appear to match, shown in Table 4, thus we conclude that the formulation of Theorem 2 is correct.

	p_1	p_2	p_3	p_4	$\frac{1}{10}\sum\limits_{i=1}^{10} u_q^{(i)}(p)$	$\frac{1}{10}\sum\limits_{i=1}^{10} U_q^{(i)}(p)$
0	0.0	$\frac{1}{3}$	$\frac{1}{3}$	1.0	2.158	2.166
1	0.0	0.0	$\frac{1}{3}$	1.0	2.165	2.173
2	0.0	$\frac{1}{3}$	1.0	1.0	2.149	2.157
3	0.0	$\frac{1}{3}$	$\frac{2}{3}$	1.0	2.139	2.149
4	0.0	0.0	0.0	$\frac{2}{3}$	2.191	2.200
5	0.0	$\frac{1}{3}$	1.0	$\frac{2}{3}$	2.157	2.167
6	0.0	0.0	$\frac{2}{3}$	1.0	2.156	2.166
7	0.0	0.0	$\frac{2}{3}$	$\frac{2}{3}$	2.145	2.156
8	0.0	$\frac{1}{3}$	0.0	$\frac{2}{3}$	2.199	2.211
9	0.0	0.0	1.0	$\frac{1}{3}$	2.186	2.198

Table 1: Results of memeory one strategies against the strategies in Table 4.

The analytical formulation of Theorem 2 will be used in the following sections to explore the best response to memory one strategies.

3 Best responses to memory one players

In the introduction a question was raised: which memory one strategy is the **best response** to a group of memory one strategies? This will be considered as an optimisation problem, where a memory one strategy p wants to optimise it's utility $u_q(p)$ against an opponent q. The decision variable is the vector p and the solitary constraint is that $p \in \mathbb{R}^4_{[0,1]}$. The optimisation problem is given by (12).

$$\begin{aligned} \max_p : \quad & u_q(p) = \frac{\frac{1}{2}pQp^T + c^T p + a}{\frac{1}{2}p\bar{Q}p^T + \bar{c}^T p + \bar{a}} \\ \text{such that} : \quad & p \in \mathbb{R}^4_{[0,1]}. \end{aligned} \tag{12}$$

This work is concerned with a fractional optimisation problem of quadratic forms. Initially, the convexity, whether or not $u_q(p)$ is concave [10], is checked (concave because is a maximisation problem).

To test the hypothesis that $u_q(p)$ is concave an empirical analysis was performed using computer code. It was shown that there exists at least one point for which the definition of concavity does not hold.

Several articles in fractional optimisation of quadratic forms that were non concave can be found [7, 8]. Though in these works both the numerator and denominator of the fractional problem were concave. In [3] it is stated that a quadratic form will be concave if and only if it's symmetric matrix is negative semi definite. In Appendix, it is proved that neither the numerator or the denominator of equation (6) are concave.

3.1 Best responses

The non concavity of $u(p)$ indicates multiple local optimal points. Thus we are not searching a single optimal point but a set of candidate optimal points. The aim is to introduce a compact way of constructing the candidate set. Once the set is defined the point that maximises $u(p)$ corresponds to the best response strategy.

The problem considered is a bounded because $p \in \mathbb{R}^4_{[0,1]}$. Because of this it is known that the candidate solutions will exist either at the boundaries of the feasible solution space, either within that space. The method of Lagrange Multipliers and Karush-Kuhn-Tucker conditions also agrees with this conclusion. The Karush-Kuhn-Tucker conditions are used because our constrains are inequalities. The proof is given in the Appendix.

Thus the candidate solution set is constructed as follows:

- any or all of p_1, p_2, p_3, p_4 are $\in \{0, 1\}$
- the rest or all of p_1, p_2, p_3, p_4 are given by the roots of $\frac{du}{dp}$.

The derivative $\frac{du}{dp}$ is given by,

$$\begin{aligned} \frac{du}{dp} = & \quad \frac{(\frac{1}{2}pQp^T + cp + a)'(\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a}) - (\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a})'(\frac{1}{2}pQp^T + cp + a)}{(\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a})^2} \\ = & \quad \frac{(pQ + c)(\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a}) - (p\bar{Q} + \bar{c})(\frac{1}{2}pQp^T + cp + a)}{(\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a})^2} \end{aligned} \tag{13}$$

For equation 13 to be zero, the numerator must fall to zero and the denominator can not nullified. Thus we conclude that the best response of a memory one strategy in match is given by Lemma 3.

Lemma 3 *The optimal behaviour of a memory one strategy player* (p^*) *against a given opponent* q *is given by:*

$$p^* = \text{argmax}(u_q(p)), \ p \in S_q,$$

where the set S_q *is defined as*

$$S_q = \{0, \bar{p}, 1\}^4$$

where the vector $\bar{p}$ *is the vector for which the following conditions are true:*

$$(\bar{p}Q + c)(\frac{1}{2}\bar{p}\bar{Q}\bar{p}^T + \bar{c}\bar{p} + \bar{a}) - (\bar{p}\bar{Q} + \bar{c})(\frac{1}{2}\bar{p}Q\bar{p}^T + c\bar{p} + a) = 0 \tag{14}$$

and

$$\frac{1}{2}\bar{p}\bar{Q}\bar{p}^T + \bar{c}\bar{p} + \bar{a} \neq 0 \tag{15}$$

Note that equation 14 is a $4-$ *polynomial system of* 4 *variables. Each polynomial corresponds to a partial derivative of* $u_q(p)$*.*

A question that arises immediately after capturing best responses of memory one strategies in pairwise interactions is: What is the optimal memory player against multiple opponents, in a tournament environment. Let us consider a collection of opponents: $\{q^{(1)}, q^{(2)}, \dots, q^{(N)}\}$, finding the optimal behaviour is captured as:

$$\begin{aligned} \max_p : & \frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p) \\ st : & \ p \in \mathbb{R}_{[0,1]} \end{aligned} \tag{16}$$

where,

$$\frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p) = \frac{1}{N} \frac{\sum\limits_{i=1}^{N} (\frac{1}{2}pQ^{(i)}p^T + c^{(i)}p + a^{(i)}) \prod\limits_{\substack{j=1 \\ j \neq i}}^{N} (\frac{1}{2}p\bar{Q}^{(i)}p^T + \bar{c}^{(i)}p + \bar{a}^{(i)})}{\prod\limits_{i=1}^{N} (\frac{1}{2}p\bar{Q}^{(i)}p^T + \bar{c}^{(i)}p + \bar{a}^{(i)})}. \tag{17}$$

Thus, we are optimising against the average utility over the set of opponents. Note that the best response can not be captured by optimising against the mean opponent. Thus,

$$\max_p \frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p) \neq \max_p u_{\frac{1}{N}\sum\limits_{i=1}^{N} q^{(i)}}(p). \tag{18}$$

A number of numerical experiments have been performed for cases where $p = (p, p, p, p)$ and $p = (p_1, p_2, p_1, p_2)$. This was done in order to compare the right hand side of equation (18) to the left. The fact that equation (18) holds is evident by Figure 3.

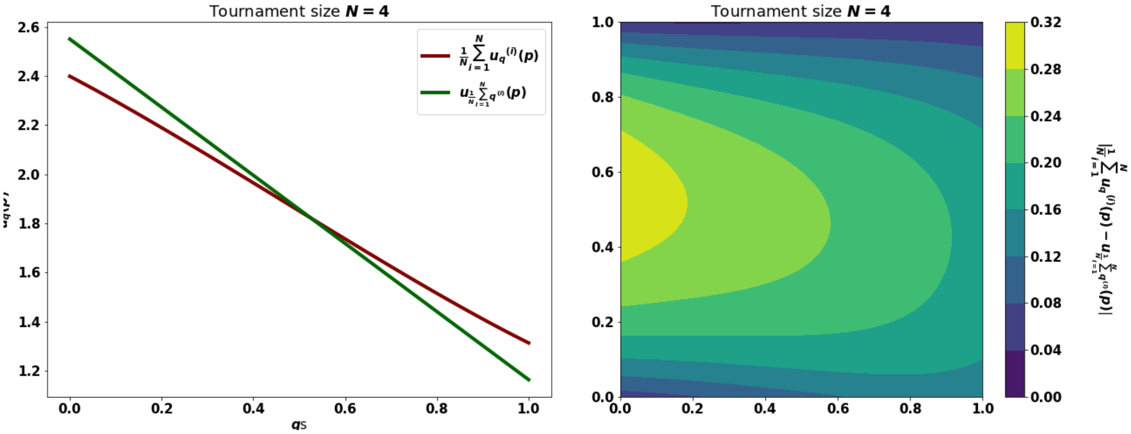


Figure 3: Numerical proof of equation 18 for $p = (p, p, p, p)$ and $p = (p_1, p_2, p_1, p_2)$.

For problem 16 a similar approach used for problem 12 is used. The candidate solutions set would be constructed by considering the bounds of the feasible space and the roots of the derivative $\frac{d}{dp}\frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p)$. The derivative is given by,

$$\frac{d}{dp}\frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p) =$$

$$= \frac{(\sum_{i=1}^{N} Q_N^{(i)'} \prod_{\substack{j=1 \\ j\neq i}}^{N} Q_D^{(i)} + \sum_{i=1}^{N} Q_D^{(i)'} \sum_{\substack{j=1 \\ j\neq i}}^{N} Q_N^{(i)} \prod_{\substack{l=1 \\ l\neq i \\ l\neq j}}^{N} Q_D^{(i)}) \times \prod_{i=1}^{N} Q_D^{(i)} - (\sum_{i=1}^{N} Q_D^{(i)'} \prod_{\substack{j=1 \\ j\neq i}}^{N} Q_D^{(i)}) \times (\sum_{i=1}^{N} Q_N^{(i)} \prod_{\substack{j=1 \\ j\neq i}}^{N} Q_D^{(i)})}{(\prod_{i=1}^{N} Q_D^{(i)})^2} \tag{19}$$

where,

$$\begin{aligned}
Q_N^{(i)} &= \frac{1}{2} p Q^{(i)} p^T + c^{(i)} p + a^{(i)}, \\
Q_N^{(i)'} &= p Q^{(i)} + c^{(i)}, \\
Q_D^{(i)} &= \frac{1}{2} p \bar{Q}^{(i)} p^T + \bar{c}^{(i)} p + \bar{a}^{(i)}, \\
Q_D^{(i)'} &= p \bar{Q}^{(i)} + \bar{c}^{(i)}.
\end{aligned}$$

Extracting the roots of equation's (19) numerator is not an easy task. This also applied to equation (14). Both equations are a system of 4 polynomials and the degree of the polynomials is gradually increasing every time an extra opponent is taken into account.

Because of that no further analytical consideration is given to problems 12 and 16. Instead both best responses of memory one strategies, pairwise and in multi interactions, will be solved using numerical methods. The methods and their results are discussed again in Section 4.3.

Though best responses can no longer be explored in an analytical way there are still many advantages by the formulation of Theorem 1. In the following subsections several theoretical results and exact ways of identifying best responses in constrained versions of problem 16 are presented. Moreover, the robustness of defection in specific interactions is investigated.

3.2 Purely random

The first constrained problem to be explored is that of the purely random strategies. Purely random strategies are a set of memory one strategies where the transition probabilities of each state are the same. The optimisation problem of (16) now has an extra constraint and is re written as,

$$\begin{aligned} \max_p : & \frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p) \\ \text{such that} : & \ 0 \leq p \leq 1 \\ & p_1 = p_2 = p_3 = p_4 = p. \end{aligned} \tag{20}$$

Due to the additional constrain, $\sum\limits_{i=1}^{N} {u_q}^{(i)}(p)$ is now a function of a single variable p and it can be handled analytically. To construct the set of candidate solutions a similar approach as the one described in previous sections is used. Thus:

- either p is $\in 0, 1$
- or p is given by the roots of $\frac{d\sum\limits_{i=1}^{N} {u_q}^{(i)}(p)}{dp}$.

The roots of the derivative are given by nullifying the numerator of the derivative. It has been proved, Appendix, that the degree of the numerator does not exceed $2N$ where N is the number of opponents. Thus there are maximum $2N$ possible roots in the feasible space. These results are summarized by Lemma 4.

Lemma 4 (Optimisation of purely random player in a tournament) *The optimal behaviour of a* ***purely random*** *player (p, p, p, p) in an $N-$memory one player tournament, $\{q_{(1)}, q_{(2)} \dots, q_{(N)}\}$ is given by:*

$$p^* = \text{argmax}(\sum_{i=1}^{N} {u_q}^{(i)}(p)), \ p \in S_{q(i)},$$

where the set S_q is defined as:

$$S_q = \{0, \lambda_i, 1\}, \ \textit{for } i \in [1, 2N].$$

Note that λ_i are the eigenvalues of the companion matrix corresponding to the numerator of

$$\frac{d}{dp}\sum_{i=1}^{N}{u_q}^{(i)}(p)$$

and λ_i are such that the denominator is not nullified.

The roots of the polynomial $\frac{d\sum\limits_{i=1}^{N}{u_q}^{(i)}(p)}{dp}$ are computed using the eigenvalues of the corresponding companion matrix. This is an algorithm resented and discussed in [9].

Furthermore, for the case of the purely random players two more theoretical results are discussed. These are the cases where the opponent has manage to make a random player indifferent and the case where a purely random player is better of playing a pure strategy.

There is importance in both results. Initially, being indifferent refers to our actions no having any effects on the match. Thus there is not optimal behaviour for player p.

Secondly, by a pure strategy we are referring to the $p = 0$ and $p = 1$. In this case it is know that p^* is $\in 0, 1$ without testing the roots of the derivative. The optimisation problem crumbles to a binary problem.

The results are given equivalently by Lemmas 5 and 6 and they are respective to the actions of the opponent. Figure 4 illustrates examples for both lemmas.

Lemma 5 *A given memory one player, (q_1, q_2, q_3, q_4), makes a* ***purely random*** *player, (p, p, p, p), indifferent if and only if, $-q_1+q_2+2q_3-2q_4 = 0$ and $(q_2-q_4-1)(q_1-2q_2-5q_3+7q_4+1)-(q_2-5q_4-1)(q_1-q_2-q_3+q_4) = 0$.*

Lemma 6 *Against a memory one player, (q_1, q_2, q_3, q_4), a* ***purely random*** *player would always play a pure strategy if and only if $(q_1q_4 - q_2q_3 + q_3 - q_4)(4q_1 - 3q_2 - 4q_3 + 3q_4 - 1) = 0$.*

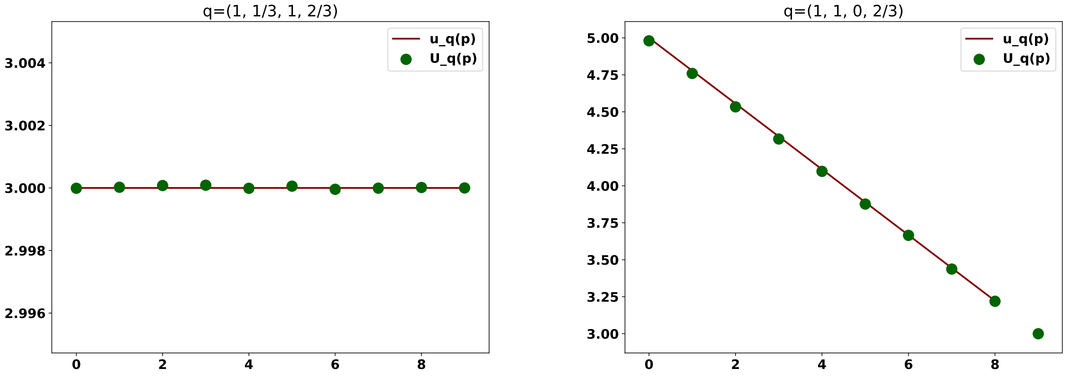


Figure 4: Proof of concept for Lemmas 5, 6.

3.3 Reactive Strategies

The next constrained case considered here is that of the reactive strategies. Reactive strategies are a set of memory one strategies where they only take into account the opponent's previous moves. As described in Section 1 Tit for Tat is a reactive strategy. The optimisation problem of (12) now has an extra constraint and is re written as,

$$\begin{aligned} \max_p &: u_q(p) \\ \text{such that} &: p_1 = p_3 \text{ and } p_2 = p_4 \\ & 0 \leq p_1, p_2 \leq 1. \end{aligned} \tag{21}$$

Reactive strategies allow us to study u_p as a function of two variables p_1, p_2. The candidate solution set can be constructed as follows:

- p_1, p_2 are $\in \{0, 1\}^2$
- the rest or all of p_1, p_2 are given by the roots of $\frac{du}{dp}$.

Note that now,

$$\frac{du}{dp} = 0$$

is a system of 2 polynomials over 2 variables. Each polynomial is equivalent to a partial derivative over p_1 and p_2. There are many methods that allow us to solve that analytical. In this work we will be using resultant theory to extract the roots from the partial derivatives.

The resultant is a symmetric function of the roots of the polynomials of a system and it can be expressed as a polynomial in the coefficients of the polynomials. The resultant will equal zero if and only if the system has at least one common root. Thus, the resultant becomes very useful in identifying whether common roots exist.

In this work the Sylvester's resultant [2] denoted as (R_S) is considered. The Sylvester's resultant is used to solve system of a single variable. However, for a system of two variables we solve over one variable and the second is kept as a coefficient. Thus we can find the roots of the equations and that is why the resultant is often refereed to as the eliminator.

Thus the best response of a reactive strategy is given in very similar approach as that described in Lemma 12. However now the partial derivatives can be solved using an exact method. Note that for pairwise interactions the maximum degree of the polynomials is equal to $2N$, however the degree increases as opponents are introduced.

3.4 Stability of defection

The final theoretical result explored is the stability of defection. Defection is known to be the dominant action in the PD and it can be proven to be the dominant strategy for the IPD for given environments. In this manuscript we try to provide a condition for when defection is the best response in the IPD.

This will be done by considering equation (13). Let equation (13) for $p_0 = (0, 0, 0, 0)$,

$$\frac{du}{dp_0} = \frac{c\bar{a} - \bar{c}a}{\bar{a}^2}. \tag{22}$$

The numerator $\bar{c}a - c\bar{a}$ is given by,

$$\begin{bmatrix} 0 \\ 0 \\ q_4 \left(4q_1q_2 - 3q_2^2 - 4q_2q_3 + 3q_2q_4 + 4q_3 - 5q_4 - 1\right) \\ -\left(q_2 - 1\right)\left(4q_1q_4 - 3q_2q_4 + q_2 - 4q_3q_4 + 4q_3 + 3q_4^2 - 6q_4 - 1\right) \end{bmatrix}$$

and the denominator $\bar{a}^2 = (-q_2 + q_4 + 1)^2$, which is always positive. In order for defection to be the best response the derivative must have a negative sign at the point p_0. That means that the utility is only decreasing after p_0. The point p_0 is by the edge of the feasible solution space thus points before that are not taken into account.

Because $\bar{a}^2$ is always positive the sign of the derivative is given by $\bar{c}a - c\bar{a}$. More specifically from equations,

$$q_4 \left(4q_1q_2 - 3q_2^2 - 4q_2q_3 + 3q_2q_4 + 4q_3 - 5q_4 - 1\right) \tag{23}$$

$$-\left(q_2 - 1\right)\left(4q_1q_4 - 3q_2q_4 + q_2 - 4q_3q_4 + 4q_3 + 3q_4^2 - 6q_4 - 1\right) \tag{24}$$

Both signs of the partial derivatives must be negative in order for the overall function to be decreasing, thus defection being the best response. The signs of equations (23) and (24) vary. There are cases that they have the same sign and cases that they do not, this is shown by numerical example summarized in Table 2.

					equation(23)	equation(24)
1	$q_1 = \frac{3}{10}$,	$q_2 = \frac{3}{20}$,	$q_3 = \frac{13}{20}$,	$q_4 = \frac{7}{100}$	+	+
2	$q_1 = \frac{11}{25}$,	$q_2 = \frac{3}{10}$,	$q_3 = \frac{9}{10}$,	$q_4 = \frac{1}{2}$	-	-
3	$q_1 = \frac{17}{20}$,	$q_2 = \frac{3}{4}$,	$q_3 = \frac{2}{5}$,	$q_4 = \frac{1}{4}$	-	+
4	$q_1 = \frac{13}{88}$,	$q_2 = \frac{21}{92}$,	$q_3 = \frac{21}{26}$,	$q_4 = \frac{20}{67}$	+	-

Table 2: Numerical examples of the derivative's sign.

For a tournament setting we substitute p_0 in equation (19):

$$\sum_{i=1}^{N} (c^{(i)T} \bar{a}^{(i)} - \bar{c}^{(i)T} a^{(i)}) \prod_{\substack{j=1 \\ j \neq i}}^{N} (\bar{a}^{(i)})^2 \tag{25}$$

The product term $\prod\limits_{\substack{j\,=\,1\\ j\,\neq\, i}}^{N} (\bar{a}^{(i)})^2$ is known to always be positive. However the sign of the sum term $\sum_{i=1}^{N}(c^{(i)T}\bar{a}^{(i)} - \bar{c}^{(i)T}a^{(i)})$ can vary based on the transition probabilities of the opponents, as discussed above. A condition that must hold in order for defection to be stable in a tournament is that the sum term must be negative. The results are summarised by Lemma **??**.

Lemma 7 *In a tournament of N players where $q^(i) = (q^(_1 i), q^(_2 i), q^(_3 i), q^(_4 i))$ defection is known to be a best response if the transition probabilities of the opponents satisfy the condition:*

$$\sum_{i=1}^{N}(c^{(i)T}\bar{a}^{(i)} - \bar{c}^{(i)T}a^{(i)}) <= 0. \tag{26}$$

Moreover lets us consider a constrained version of the problem once again. Lets us assume that in an pairwise interaction the opponent is a reactive player $q = (q_1, q_2, q_1, q_2)$. By substituting $q_3 = q_1$ and $q_4 = q_2$ equations (23) and (24) are now re written as follow,

$$\left[\begin{matrix}- q_{2} \left(4 q_{1} - 5 q_{2} - 1\right)\\\left(q_{2} - 1\right) \left(4 q_{1} - 5 q_{2} - 1\right)\end{matrix}\right]$$

The sign of both equations is now based on the same term,$(4q_1 - 5q_2 - 1)$, which is a term that can have both negative and positive values. This is shown by Figure 5. Following this the following result is retrieved,

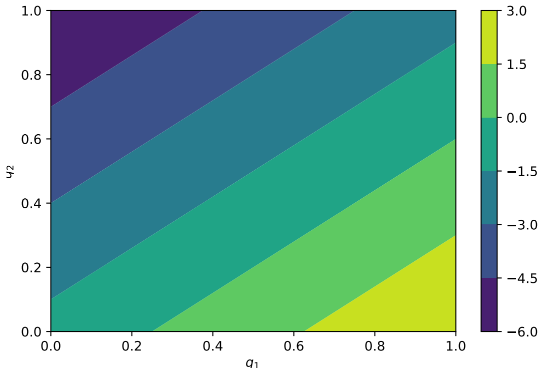


Figure 5: Sign of $(4q_1 - 5q_2 - 1)$.

Lemma 8 *Defection is the best responses of a memory one player p against a reactive player q if the transition probabilities of the opponent satisfy the condition:*

$$4q_1 - 5q_2 - 1 > 0 \tag{27}$$

4 Numerical Experiments

In this section several analytical results of Section **??** are validated using numerical experiments. Initially, numerical methods for best responses in the constrained versions of the problem are presented. We explore the case of purely random strategies and reactive strategies.

Moreover the answer to the questions discussed in Section 3.1 are answered via numerical algorithms. The question risen was identifying the best response memory one strategy.

4.1 Purely Random Strategies

Best responses of purely random strategies are given by Lemma 4, as presented in Section 3.2. Based on the results Algorithm 1 is constructed to carry out several numerical experiments for pairwise and multi agent interactions.

Algorithm 1 Best response algorithm for purely random strategies

```
procedure PURELY RANDOM SEARCH
    $N \leftarrow$ number of opponets
    $S_q \leftarrow \{0, 1\}$
    $u' \leftarrow \frac{d \sum\limits_{i=1}^{N} u}{d\bar{p}}$
    $\frac{u_N}{u_D} \leftarrow u'$
    $C(u_N) \leftarrow$ companion matrix of $u_N$
loop $i = 1$ to $2N$:
    $\lambda_i \leftarrow$ eigenvalue of $C(u_N)$
    if $u_D(\lambda_i) \neq 0$ then
        $S_q \cup \lambda_i$.
    goto loop.
    close;
    $p^* \leftarrow \text{argmax}(\sum\limits_{i=1}^{N} u_{q^{(i)}}(p)), p \in S_q$.
```

The results of pairwise interactions are given by Figure 6. There is it evident that the optimal behaviour has been captured by our search algorithm. Moreover, the algorithm is also validated for tournament interactions, as shown by Figure 7.

4.2 Reactive Strategies

Best responses of reactive strategies were discussed in Section 3.3. There it was stated that the field of resultant theory would be used to solve the partial derivatives of the utility. As a reminder, best responses in reactive build upon Lemma 3 however now condition (14) is a 2 polynomial system of 2 variables.

Based on the results Algorithm 2 is constructed and several examples are performed. Illustrated by Figure 8 are the results of these examples. The results suggest that the best response behaviour is captured by our

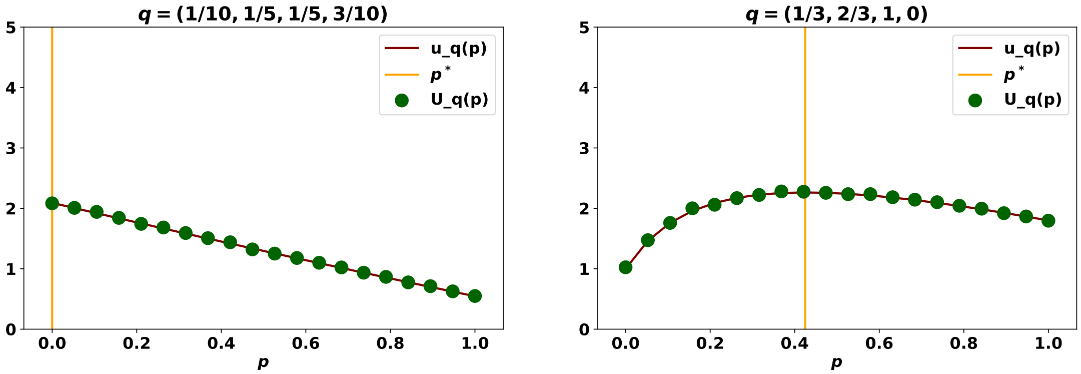


Figure 6: Numerical experiments for Algorithm 1 for $N = 1$.

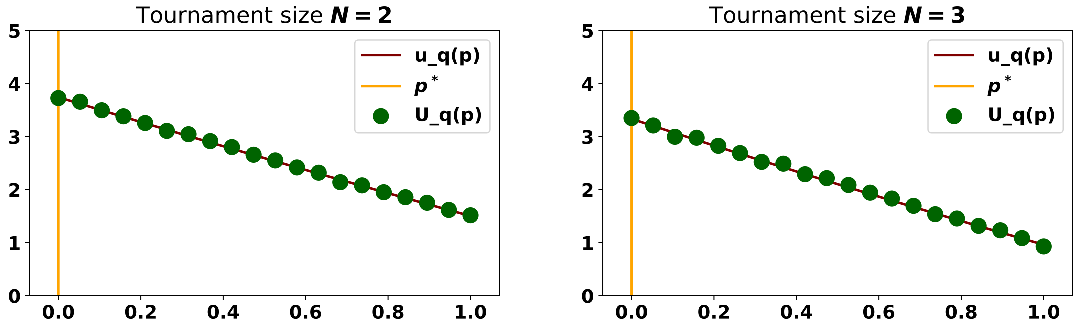


Figure 7: Numerical experiments for Algorithm 1 for $N > 1$.

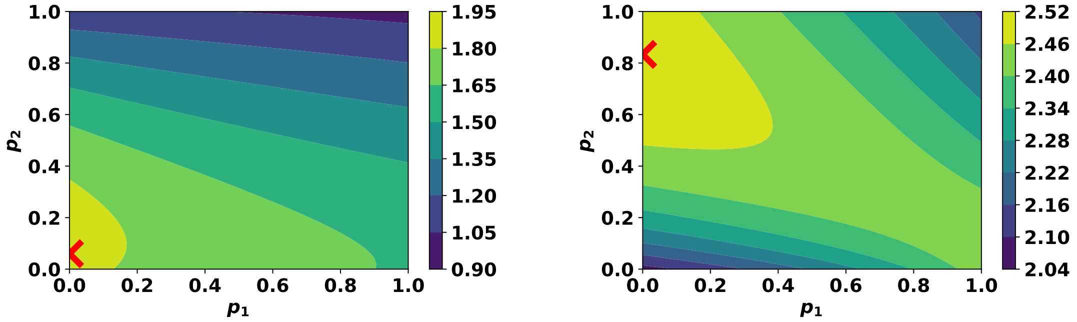


Figure 8: Numerical experiments for Algorithm 2 for $N = 1$.

algorithm.

Algorithm 2 Best response algorithm for reactive strategies

```
procedure REACTIVE SEARCH
    $N \leftarrow$ number of opponets
    $S_q \leftarrow \{0, 1\}^2$
    $u' \leftarrow \frac{d \sum_{i=1}^{N} u}{d\bar{p}}$
    $\frac{u_N}{u_D} \leftarrow u'$
    $S(u_N, p_2) \leftarrow$ Sylvester's matrix for$p_2$.Coefficients are polynomials of $p_1$
    $R_S(S) \leftarrow det(S)$
    roots$_{p_1} \leftarrow p_1$ for $det(M)_{p_1} = 0$
loop root in roots$_{p_1}$:
     system$(p_2) \leftarrow u_N(root)$
    root$_{p_2} \cup p_2$ for system$(p_2) = 0$
    if $u_D((\text{root}, \text{root}_{p_2})) \neq 0$ then
        $S_q \cup \{(\text{root}, \text{root}_{p_2})\}$
    goto loop.
    close;
    $p^* \leftarrow \text{argmax}(\sum_{i=1}^{N} u_{q^{(i)}}(p)), p \in S_q$.
```

4.3 Memory one strategies

As discussed in Section 3.1 best responses in memory one strategies will be captured using a numerical method. In this subsection we discuss the reasons for choosing the specific method and parameters.

Moreover, a parameter sweep has been performed and here we wil discuss the results.

Bayesian optimisation is ...

5 Limitation of memory

The second part of this paper focuses on the limitation of memory size. The aim is to show that memory one strategies suffer for multi agent interactions whilst they outperform any strategy in pairwise interactions. This is achieved by comparing the performance of an optimised short memory one strategies to a trained longer memory strategies in tournaments of size 3.

The complex strategy is trained using the same optimisation algorithm as the one described in Section 4.3, the bayesian optimisation. The trained strategy used is a strategy called Gambler. Gambler is introduced and discussed in [11].

5.1 Gambler

Several means of representing strategies have been used over the years for IPD strategies. In [11] several of those 'archetypes' are presented and used to train different successful strategies. One of the archetypes firstly introduced in that paper is the Gambler.

Gambler is based on a lookup table and encodes a probability of cooperating based on the opponent's first n_1 moves, the opponent's last m_1 moves, and the players last m_2 moves.

Several variants of Gambler have been trained for this work, a summary is given by Table. In essence Gambler can represent any generic strategy and that is why it has been chosen.

	n_1	m_1	m_2
1	1	1	2
2	2	2	0
3	2	2	1
4	2	2	2
5	4	4	4

Table 3: Variants of Gambler used.

5.2 Procedure

For a number of tournaments, where $N = 3$, using the results of Section, we find the optimal best memory one strategy and it's utility. Afterwards, the memory on strategy is removed from the tournament and it is replaced with a variant of Gambler which is then trained for that environment as well.

The performance of those two strategies is compared.

The results of a big parameter sweep are discussed in the following section.

5.3 Limitation Results

Results, results, results.

6 Discussion

A Appendix Tables

The memory one strategies used in the computer tournament described in [19] are given by Table 4.

	Name	Memory one representation	Explanation
1	Cooperator	$(1, 1, 1, 1)$	Always chooses C.
2	Defector	$(0, 0, 0, 0)$	Always chooses D.
3	Random	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	Randomly chooses between C and D with a probability of 0.5.
4	Tit for Tat	$(1, 0, 1, 0)$	Start with a C and then mimics the opponent's last move.
5	Grudger	$(1, 0, 0, 0)$	Starts by cooperating however will defect if at any point the opponent has defected.
6	Generous Tit for Tat	$(1, \frac{1}{3}, 1, \frac{1}{3})$	A more generous version of Tit for Tat.
7	Win Stay Lose Shift	$(1, 0, 0, 1)$	Starts with a C and then repeats it's previous move only if it was awarded with a payoff of R or T.
8	ZDGTFT2	$(1, \frac{1}{8}, 1, \frac{1}{4})$	A generous zero determinant strategy introduced in [19]
9	ZDExtort2	$(\frac{8}{9}, \frac{1}{2}, \frac{1}{3}, 0)$	An extortionate zero determinant strategy introduced in [19]
10	Hard Joss	$(\frac{9}{10}, 0, \frac{9}{10}, 0)$	Cooperates with probability $\frac{9}{10}$ when the opponent cooperates, otherwise emulates Tit for Tat.

Table 4: The memory one strategies from [19].

References

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

[2] Alkiviadis G. Akritas. *Sylvester's form of the Resultant and the Matrix-Triangularization Subresultant PRS Method*, pages 5–11. Springer New York, New York, NY, 1991.

[3] Howard Anton and Chris Rorres. *Elementary Linear Algebra: Applications Version.* Wiley, eleventh edition, 2014.

[4] R Axelrod and WD Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

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

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

[7] Amir Beck and Marc Teboulle. A convex optimization approach for minimizing the ratio of indefinite quadratic functions over an ellipsoid. *Mathematical Programming*, 118(1):13–35, 2009.

[8] Hongyan Cai, Yanfei Wang, and Tao Yi. An approach for minimizing a quadratically constrained fractional quadratic problem with application to the communications over wireless channels. *Optimization Methods and Software*, 29(2):310–320, 2014.

[9] Alan Edelman and H Murakami. Polynomial roots from companion matrix eigenvalues. *Mathematics of Computation*, 64(210):763–776, 1995.

[10] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products.* Elsevier/Academic Press, Amsterdam, seventh edition, 2007.

[11] Marc Harper, Vincent Knight, Martin Jones, Georgios Koutsovoulos, Nikoleta E. Glynatsi, and Owen Campbell. Reinforcement learning produces dominant strategies for the iterated prisoners dilemma. *PLOS ONE*, 12(12):1–33, 12 2017.

[12] Jeremy Kepner and John Gilbert. *Graph algorithms in the language of linear algebra.* SIAM, 2011.

[13] Vincent Knight, Owen Campbell, Marc Harper, Karol Langner, James Campbell, Thomas Campbell, Alex Carney, Martin Chorley, Cameron Davidson-Pilon, Kristian Glass, Tomáš Ehrlich, Martin Jones, Georgios Koutsovoulos, Holly Tibble, Müller Jochen, Geraint Palmer, Paul Slavin, Timothy Standen, Luis Visintini, and Karl Molden. An open reproducible framework for the study of the iterated prisoner's dilemma. 1(1), 2016.

[14] Christopher Lee, Marc Harper, and Dashiell Fryer. The art of war: Beyond memory-one strategies in population games. *PLOS ONE*, 10(3):1–16, 03 2015.

[15] Frederick A Matsen and Martin A Nowak. Win–stay, lose–shift in language learning from peers. *Proceedings of the National Academy of Sciences*, 101(52):18053–18057, 2004.

[16] Martin Nowak and Karl Sigmund. Game-dynamical aspects of the prisoner's dilemma. *Applied Mathematics and Computation*, 30(3):191–213, 1989.

[17] Martin Nowak and Karl Sigmund. The evolution of stochastic strategies in the prisoner's dilemma. *Acta Applicandae Mathematicae*, 20(3):247–265, 1990.

[18] William H. Press and Freeman J. Dyson. Iterated prisoners dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26):10409–10413, 2012.

[19] Alexander J. Stewart and Joshua B. Plotkin. Extortion and cooperation in the prisoners dilemma. *Proceedings of the National Academy of Sciences*, 109(26):10134–10135, 2012.