Article

What You Gotta Know to Play Good in the Iterated Prisoner's Dilemma

Ethan Akin

Mathematics Department, The City College, 137 Street and Convent Avenue, New York City, NY 10031, USA; E-Mail: ethanakin@earthlink.net; Tel.: +1-(212)-650-5136

Abstract: For the iterated Prisoner's Dilemma there exist *good* strategies which solve the problem when we restrict attention to the long term average payoff. When used by both players, these assure the cooperative payoff for each of them. Neither player can benefit by moving unilaterally to any other strategy, *i.e.*, these provide Nash equilibria. In addition, if a player uses instead an alternative which decreases the opponent's payoff below the cooperative level, then his own payoff is decreased as well. Thus, if we limit attention to the long term payoff, these strategies effectively stabilize cooperative behavior. The existence of such strategies follows from the so-called Folk Theorem for supergames, and the proof constructs an explicit memory-one example, which has been labeled *Grim*. Here we describe all the memory-one good strategies for the non-symmetric version of the Prisoner's Dilemma. This is the natural object of study when the payoffs are in units of the separate players' utilities. We discuss the special advantages and problems associated with some specific good strategies.

Keywords: Prisoner's Dilemma; stable cooperative behavior; iterated play; Markov strategies; good strategies, individual utility

MSC classifications: 91A05, 91A20, 91A22, 60J20

1. Introduction

The *Prisoner's Dilemma* (hereafter PD) is a two person game which provides a simple model of a disturbing social phenomenon. It is a game in which each of the two players, X and Y, has a choice between two strategies, c (= *cooperation*) and d (= *defection*). The cooperative outcome cc leads to a

Pareto optimal pair of payoffs. On the other hand, defection *dominates* cooperation. This means that against any play, the opponent's best reply is defection. It follows that the dd outcome is the unique Nash equilibrium for the game. However, each player does worse at dd than at cc. The multiplayer version is Hardin's *tragedy of the commons* [1].

This collapse of cooperation can be controlled by repeated play. Given a game, the associated *supergame* is an infinite sequence of rounds of the original game. A strategy in the supergame allows each player to use as data the outcome of the previous rounds. A strategy is *memory-one* if, after the initial play, the player uses on each round just the outcome of the single preceding round. There are different ways of aggregating the payoffs on the individual rounds to obtain the payoff for the supergame. Following Press and Dyson [2] and some remarks of Aumann [3], we will use the limit of the averages of the payoffs, although there are some technical issues concerning the existence of this Cesaro limit. This has the effect of wiping out any advantages which one player obtained in the early rounds. The supergame we are considering is called the *Iterated Prisoner's Dilemma* (hereafter IPD).

The so-called *Folk Theorem* for supergames says that there are many Nash equilibria for supergames. The proof is sketched in Appendix 2 of [3] and in more detail in [4]. These are constructed using *trigger strategies*. In the IPD, this means that a player switches to permanent defection when the opponent has departed from some expected cooperative play. The memory-one strategy *Grim*, described later, is of this sort. Using such trigger strategies the players can stabilize the cooperative payoffs, but other Pareto optimal outcomes can be obtained as well.

For the IPD we will call a strategy *Nash type* if, when used by both players, the cooperative payoff is received and if, against it an opponent cannot do better than the cooperative payoff. We will call the strategy for X *good* if it is of Nash type and, in addition, if Y uses any strategy which yields the cooperative payoff for Y, then X receives the cooperative payoff for X as well. Grim and the well-known Tit-for-Tat strategies are both good.

The usual version of the PD is a symmetric game with the cooperative and defection payoffs the same for both players. In [5], the memory-one strategies of Nash type or good are completely described for the symmetric game. The evolutionary dynamics of such strategies is also considered. This means that the payoffs are measured in terms of fitness, and the dynamics is of selection between subpopulations using different strategies. In the language of [6], it is shown that good strategies are robust. That is, a population of good strategists cannot be invaded by a mutant which does better against them.

The symmetric version of PD is appropriate for this sort of evolutionary dynamics, but in classical game theory the payoffs have to be given in terms of utility functions, which measure the preferences of the players. When interpersonal comparison of utilities is excluded (does this dollar mean more to you than it does to me?), then the symmetric version is not appropriate. This is why the terms "cooperative payoff for X" and "cooperative payoff for Y" were used above. Here we extend the results of [5] to describe the memory-one strategies which are of Nash type or good in the general non-symmetric PD. The tricky bit, as we will see, is that the characterization of the strategies for X depends upon the payoff values for Y. Thus, in order to play good, X must estimate the utilities of the different outcomes for Y.

2. Good Strategies for the Iterated Prisoner's Dilemma

In the symmetric PD, each of the two players, X and Y, has a choice between the two strategies c and d. Thus, there are four outcomes which we list in the order: cc, cd, dc, dd, where, for example, cd is the outcome when X plays c and Y plays d. Each then receives a payoff given by the following 2×2 chart:

$X \backslash Y$	c	d
c	(R, R)	(S, T)
d	(T, S)	(P, P)

$$(2.1)$$

where the first entry of the pair is the payoff to X and the second is the payoff to Y.

Alternatively, we can describe the payoff vectors for each player.

$$\mathbf{S}_X \;=\; \begin{pmatrix} R \\ S \\ T \\ P \end{pmatrix}, \qquad \mathbf{S}_Y \;=\; \begin{pmatrix} R \\ T \\ S \\ P \end{pmatrix}. \tag{2.2}$$

Either player can use a *mixed* strategy by randomizing, adopting c with probability p_c and d with the complementary probability $1 - p_c$. The probability p_c lies between 0 and 1 with the extreme values corresponding to the pure strategies c and d.

The payoffs are assumed to satisfy

$$\begin{aligned} (i) \quad & T \;>\; R \;>\; P \;>\; S \\ (ii) \quad & 2R \;>\; T + S. \end{aligned} \tag{2.3}$$

The strategy c is *cooperation*. When both players cooperate they each receive the reward for cooperation (= R). The strategy d is *defection*. When both players defect they each receive the punishment for defection (= P). But if one player cooperates and the other does not then the defector receives the large temptation payoff (= T) while the hapless cooperator receives the very small sucker's payoff (= S). The condition $2R > T + S$ says that the reward for cooperation is larger than the players would receive from sharing equally the total payoff of a cd or dc outcome. Thus, the maximum total payoff occurs uniquely at cc. The cooperative outcome cc is clearly where the players "should" end up. If they could negotiate a binding agreement in advance of play, they would agree to play c and each receive R. However, the structure of the game is such that at the time of play, each chooses a strategy in ignorance of the other's choice.

Observe that, as described in the Introduction, strategy d *dominates* strategy c. This means that whatever Y's choice is, X receives a larger payoff by playing d than by using c. Hence, X chooses d and for exactly the same reason Y chooses d, and so they are driven to the dd outcome with payoff P for each. For helpful discussions see [7,8].

The search for a theoretical approach which will avert this depressing outcome has focused attention on *repeated play* in the IPD. X and Y play repeated rounds of the same game. For each round the players' choices are made independently, but each is aware of all of the previous outcomes. The hope is that the threat of future retaliation will rein in the temptation to defect in the current round.

Robert Axelrod devised a tournament in which submitted computer programs played against one another. Each program played a fixed, but unknown, number of plays against each of the competing programs, and the resulting payoffs were summed. The results are described and analyzed in his landmark book, [9]. The winning program, Tit-for-Tat, submitted by game theorist Anatol Rapaport, cooperates in the first round, then uses in each round the opponent's play in the just concluded round. A second tournament yielded the same winner. Axelrod extracted some interesting rules of thumb from Tit-for-Tat and applied these to some historical examples.

At around the same time, game theory was being introduced into biology by John Maynard-Smith, to study problems in the evolution of behavior. The books [10] and [11] provide good surveys of the early work. Tournament play for games, which has been widely explored since, exactly simulates the dynamics examined in this growing field of evolutionary game theory.

The choice of play for the first round is the *initial play*. A *strategy* is a choice of initial play together with what we will call a *plan*: A choice of play, after the first round, to respond to any possible past history of outcomes in the previous rounds.

Tit-for-Tat (hereafter TFT) is an example of a *memory-one plan* which bases its response entirely on outcome of the previous round. See, for example, [12] (Chapter 5). The TFT strategy is the TFT plan together with initial play c.

With the outcomes listed in order as cc, cd, dc, dd, a memory-one *plan vector* for X is a vector $\mathbf{p} = (p_1, p_2, p_3, p_4) = (p_{cc}, p_{cd}, p_{dc}, p_{dd})$ where p_z is the probability of playing c when the outcome z occurred in the previous round. If Y uses plan vector $\mathbf{q} = (q_1, q_2, q_3, q_4)$ then the Markov response is $(q_{cc}, q_{cd}, q_{dc}, q_{dd}) = (q_1, q_3, q_2, q_4)$ and the successive outcomes follow a Markov chain with transition matrix given by:

$$\mathbf{M} = \begin{pmatrix} p_1 q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) \\ p_2 q_3 & p_2(1-q_3) & (1-p_2)q_3 & (1-p_2)(1-q_3) \\ p_3 q_2 & p_3(1-q_2) & (1-p_3)q_2 & (1-p_3)(1-q_2) \\ p_4 q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4) \end{pmatrix}. \tag{2.4}$$

We use the switch in numbering from the Y plan vector $\mathbf{q}$ to the Y response vector, because switching the perspective of the players interchanges cd and dc. This way the "same" plan for X and for Y is given by the same plan vector. For example, the TFT plan vector for X and Y is given by $\mathbf{p} = \mathbf{q} = (1, 0, 1, 0)$ but the response vector for Y is $(1, 1, 0, 0)$. The plan vector *Repeat* is given by $\mathbf{p} = \mathbf{q} = (1, 1, 0, 0)$ but the response vector for Y is $(1, 0, 1, 0)$. The *Repeat* plan just continually repeats the player's previous play.

The two players' initial plays determine the initial distribution for the Markov chain:

$$\mathbf{v}^1 = (p_c^X p_c^Y, \ p_c^X(1 - p_c^Y), \ (1 - p_c^X)p_c^Y, \ (1 - p_c^X)(1 - p_C^Y)). \tag{2.5}$$

A memory-one strategy consists of a memory-one plan together with a pure or mixed initial play.

We will call a plan *agreeable* when it always responds to a cc with a play of c in the next round. The plan is *firm* when it always responds to a dd with a play of d in the next round. Thus, a memory-one plan vector $\mathbf{p}$ is agreeable when $p_1 = 1$ and is *firm* when $p_4 = 0$. TFT is both agreeable and firm, as is Repeat.

An *agreeable strategy* consists of an agreeable plan together with an initial play of c. If both players use agreeable strategies, then the outcome is fixed at cc. That is, both players receive the cooperative payoff at every play.

In the IPD we consider an infinite sequence of plays, yielding payoffs s_X^k, s_Y^k to the two players at round k. We will concern ourselves with the *long term average payoff* to each player. For X:

$$ s_X \quad = \quad Lim_{n \to \infty} \frac{1}{n} \Sigma_{k=1}^n s_X^k, \tag{2.6} $$

and similarly for Y.

If each player uses a (possibly mixed) initial play and then adopts a fixed memory-one plan, then the Markov chain leads to a probability distribution vector $\mathbf{v}^n$ on the outcomes at time n, with an average limiting distribution given by

$$ \mathbf{v} \quad = \quad Lim_{n \to \infty} \frac{1}{n} \Sigma_{k=1}^n \mathbf{v}^k. \tag{2.7} $$

The payoffs s_X and s_Y are just the expected values of the X and Y payoffs with respect to this distribution. That is,

$$ s_X \quad = \quad < \mathbf{v} \cdot \mathbf{S}_X >, \qquad s_Y \quad = \quad < \mathbf{v} \cdot \mathbf{S}_Y > . \tag{2.8} $$

In the general—not necessarily memory-one—case, the choice of strategies still determines a distribution vector $\mathbf{v}^n$ at each time n. The sequence of probability vectors $\{ \frac{1}{n} \Sigma_{k=1}^n \mathbf{v}^k \}$ need not converge in general. Any limit point $\mathbf{v}$ of the sequence is called a *limit distribution* associated with the choices of strategies. For each such limit distribution the associated long term payoff is given by (2.8). The name limit distribution comes from the following observation.

Proposition 2.1. *If X and Y use memory-one plans yielding the Markov matrix* $\mathbf{M}$*, then any limit distribution* $\mathbf{v}$ *for the play is an invariant distribution for* $\mathbf{M}$*. That is,* $\mathbf{v} \cdot \mathbf{M} = \mathbf{v}$*.*

Proof: $\mathbf{v}_i^{k+1}$ is the probability of outcome i at round $k + 1$. This is equal to the sum $\Sigma_{j=1}^4 \mathbf{v}_j^k \mathbf{M}_{ji}$ because $\mathbf{v}_j^k \mathbf{M}_{ji}$ is the probability of j after round k times the conditional probability of moving to i from j. It follows that

$$ (\frac{1}{n} \Sigma_{k=1}^n \mathbf{v}^k) \cdot \mathbf{M} \; - \; (\frac{1}{n} \Sigma_{k=1}^n \mathbf{v}^k) \quad = \quad \frac{1}{n}(\mathbf{v}^{n+1} - \mathbf{v}^1). \tag{2.9} $$

As $\mathbf{v}^{n+1} - \mathbf{v}^1$ has entries with absolute value at most 1, the limit of this expression is $\mathbf{0}$ as $n \to \infty$. Hence, every limit distribution satisfies $\mathbf{v} \cdot \mathbf{M} \; - \; \mathbf{v} \; = \; \mathbf{0}$.

$\square$

Corollary 2.2. *If X and Y use memory-one plan vectors* $\mathbf{p}$ *and* $\mathbf{q}$ *then* $\mathbf{e}_1 = (1, 0, 0, 0)$ *is an invariant distribution for* $\mathbf{M}$*, and so is a limit distribution, if and only if* $\mathbf{p}$ *and* $\mathbf{q}$ *are both agreeable.*

Proof: $\mathbf{e}_1$ satisfies $\mathbf{e}_1 \cdot \mathbf{M} = \mathbf{e}_1$ if and only if the first row of $\mathbf{M}$ is given by $(1, 0, 0, 0)$. Since $0 \leq p_1, q_1 \leq 1$, this is equivalent to $p_1 = q_1 = 1$.

$\square$

Because so much work had been done on this Markov model, the exciting new ideas of Press and Dyson [2] took people by surprise. They have inspired a number of responses, e.g., [13] and especially [14].

One achievement of the recent work inspired by Press and Dyson was a complete description in [5] of the *good strategies*, which, in a strong sense, are strategies which solve the IPD.

Definition 2.3. *A memory-one plan* $\mathbf{p}$ *for X is called good if it is agreeable and if for any strategy that Y chooses against it*

$$s_Y \geq R \qquad \Longrightarrow \qquad s_Y = R = s_X. \tag{2.10}$$

A good memory-one strategy is a good memory-one plan together with an initial play of c.

While good strategies were defined more generally in the Introduction, from now on we will use the term "good strategy" to refer to a good memory-one strategy.

Assume X uses a good strategy $\mathbf{p}$. If Y uses an agreeable strategy then the players receive the joint cooperative outcome. Furthermore, there is no strategy for Y which against $\mathbf{p}$ obtains more than the result of joint cooperation. In fact, if the effect of Y's strategy is to give X less than the cooperative outcome then Y receives less than the cooperative outcome as well.

The good strategies solve the Prisoner's Dilemma in the following sense. If X announces that she intends to use a good strategy, then Y cannot obtain any payoff better than the cooperative value. Furthermore, only joint cooperation yields the cooperative value for Y. The joint cooperative payoff is stabilized because Y has no incentive to behave any way but agreeably and a strong incentive to be agreeable. Furthermore, without an announcement, the statistics of the initial rounds can be used to estimate the entries of a memory-one strategy used by X. This would reveal that X is playing a good strategy. Then Y's best long term response is to begin to play good as well.

There are two caveats. The first is that we are only considering the long term payoff, ignoring any transient benefits from early defections. While this is worth investigating, our use of the long term payoff is just part of the structure of the game we are investigating.

The second is more interesting. It is possible for Y to choose a strategy against a good strategy $\mathbf{p}$, so that he does better than X, although both receive less than the cooperative payoff. For example, there is a large class of good strategies for X, called *complier strategies* in [14], against which it always happens that

$$R > s_X \qquad \Longrightarrow \qquad R > s_Y > s_X. \tag{2.11}$$

At this point we are confronted by a subtle change in viewpoint that was introduced by the evolutionary applications of game theory and their computer tournament models. In evolutionary game theory what matters is how a player is doing as compared with the competing players. Consider this with just two players and suppose they are currently considering strategies with the same payoff to each. From this comparative viewpoint, Y would reject a move to a strategy where he does better but which

causes X to do still better than he. That this sort of *altruism* is selected against is a major problem in the theory of evolution. Depending on what is going on in the rest of the population, a player may do better by giving up the joint cooperative payoff to force an opponent to do worse. In the language of evolutionary games this is called *spite*.

The good strategies do have good properties, described in [5], for the evolutionary game situation, but they need not eliminate all the alternatives from a population and may even be out-competed in certain circumstances.

However, in classical game theory X simply desires to obtain the highest absolute payoff. The payoffs to her opponent are irrelevant, except as data to predict Y's choice of strategy. It is the classical problem that we wish to consider here.

Recall that the payoffs must be measured by real desirability. The payoffs are often stated in money amounts or in years reduced from a prison sentence (the original "prisoner" version). But it is important to understand that the payoffs are really in units of *utility*. That is, the ordering in (2.3) is assumed to describe the order of desirability of the various outcomes to each player when the full ramifications of each outcome are taken into account. Thus, if X is induced to feel guilty at the dc outcome then the payoff to X of that outcome is reduced.

Adjusting the payoffs is the classic way of stabilizing cooperative behavior. Suppose prisoner X walks out of prison free after defecting, having consigned Y—who played c—to a 20 year sentence. Colleagues of Y might well do X some serious damage. Anticipation of such an event considerably reduces the desirability of the dc outcome for X, perhaps to well below R. If X and Y each have threatening friends then it is reasonable for each to expect that a prior agreement to play cc will stand and so they each receive R. However, in terms of utility this is no longer a Prisoner's Dilemma.

In the book which originated modern game theory [15], Von Neumann and Morgenstern developed an axiomatic theory of utility. What was needed was that the choices be made not only over fixed outcomes but also over *lotteries* which are probability distributions over a finite number of outcomes. They showed that, given the axioms, the utility function for an individual can be constructed so that the utility of such a lottery is exactly the expected value of the utilities of the outcomes. The utility function is uniquely defined up to positive affine transformation, *i.e.*, addition of a constant and multiplication by a positive constant. This allows us to make sense of such arithmetic relationships as inequality (ii) in (2.3).

This emphasis on utility raises a rather serious issue. The game described by (2.1) is a symmetric game. That is, reversing the outcomes for X and Y reverses their payoffs. This makes perfect sense if the payoffs are measured in some common unit like money, years in prison or evolutionary fitness (= relative growth rate of the subpopulation). It does not make sense for utility theory which excludes interpersonal comparison of utilities. In that context, the Prisoner's Dilemma should be represented as follows:

$X \backslash Y$	c	d
c	(R_1, R_2)	(S_1, T_2)
d	(T_1, S_2)	(P_1, P_2)

(2.12)

and for $p = 1, 2$:

$$
\begin{aligned}
(i) & \quad T_p > R_p > P_p > S_p \\
(ii) & \quad 2R_p > T_p + S_p.
\end{aligned}
\tag{2.13}
$$

The definitions of a good strategy and the related, slightly weaker, notion of a Nash type strategy are now given by

Definition 2.4. *A memory-one plan* **p** *for X is called* good *if it is agreeable and if for any initial play for X, any strategy chosen by Y, and any resulting limit distribution*

$$
s_Y \geq R_2 \quad \Longrightarrow \quad s_Y = R_2 \quad \text{and} \quad s_X = R_1.
\tag{2.14}
$$

The plan is called of Nash type *if it is agreeable and if for any initial play for X, any strategy chosen by Y, and any resulting limit distribution*

$$
s_Y \geq R_2 \quad \Longrightarrow \quad s_Y = R_2.
\tag{2.15}
$$

The name *Nash type* is used because if both players initially cooperate and use plans of Nash type then neither has a positive incentive to change strategy. That is, the pair of strategies provides a Nash equilibrium.

In the next section we will prove the following extension of the characterization in [5].

Theorem 2.5. *Let* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *be an agreeable plan vector for X, other than* Repeat. *That is,* $p_1 = 1$ *but* $\mathbf{p} \neq (1, 1, 0, 0)$.

The plan vector **p** *is of Nash type if and only if the following inequalities hold.*

$$
\frac{T_2 - R_2}{R_2 - S_2} \cdot p_3 \leq (1 - p_2) \quad \text{and} \quad \frac{T_2 - R_2}{R_2 - P_2} \cdot p_4 \leq (1 - p_2).
\tag{2.16}
$$

The plan vector **p** *is good if and only if both inequalities hold strictly (i.e., neither is an equation).*

Remarks: (a) Notice that $\frac{T_2 - R_2}{R_2 - S_2} < 1$ and so the first inequality always holds provided p_2 is sufficiently close to 0. There is no *a priori* bound on $\frac{T_2 - R_2}{R_2 - P_2}$.

(b) Just as for the symmetric game, see Corollary 1.6 of [5], the Nash type memory-one plans together with *Repeat* forms a closed, convex set whose interior in the set of agreeable memory-one plans is the set of good memory-one plans.

The strategy *Repeat* = $(1, 1, 0, 0)$ is an agreeable plan vector that is not of Nash type. If both players use *Repeat*, then the initial outcome repeats forever. If X initially cooperates and Y initially defects, then the initial outcome is cd and so $s_Y = T_2$ and $s_X = S_1$. This possibility shows that *Repeat* is not of Nash type.

An interesting aspect of this result is that the inequalities which are used by X to choose a good strategy depend upon the payoff values to player Y. This is understandable in that X's play is meant to constrain Y's response. However, it means that X has to perform some sort of estimate of Y's payoff values in order to choose what strategy to play.

3. The Good Strategy Characterization

Among other things, the Press–Dyson paper introduced a useful tool for the study of long term outcomes. For a memory-one plan $\mathbf{p}$ for X, we define the *Press–Dyson vector* $\tilde{\mathbf{p}} = \mathbf{p} - \mathbf{e}_{12}$, where $\mathbf{e}_{12} = (1, 1, 0, 0)$.

Considering its usefulness, the following result has a remarkably simple proof. In this form it occurs in [5], but I was inspired by some remarks of Sigmund, who referred to Hilbe, Nowak and Sigmund—Appendix A of [14].

Lemma 3.1. *Assume that X uses the memory-one plan vector* $\mathbf{p}$ *with Press–Dyson vector* $\tilde{\mathbf{p}}$. *If the opponent Y uses a strategy so that the play yields the sequence of distributions* $\{\mathbf{v}^n\}$, *then*

$$Lim_{n\to\infty}\frac{1}{n} \Sigma_{k=1}^n \ < \mathbf{v}^k \cdot \tilde{\mathbf{p}} > \ = \ 0,$$

$$and\ so\quad < \mathbf{v} \cdot \tilde{\mathbf{p}} > \ = \ v_1\tilde{p}_1 + v_2\tilde{p}_2 + v_3\tilde{p}_3 + v_4\tilde{p}_4 \ = \ 0 \tag{3.1}$$

for any associated limit distribution $\mathbf{v}$.

Proof: Let $v_{12}^k = \mathbf{v}_1^k + \mathbf{v}_2^k$, the probability that either cc or cd is the outcome in the kth round of play. That is, v_{12}^k, defined to be $< \mathbf{v}^k \cdot \mathbf{e}_{12} >$, is the probability that X played c in the kth round. On the other hand, since X is using the memory-one plan $\mathbf{p}$, p_i is the conditional probability that X plays c in the next round, given outcome i in the current round for $i = 1, 2, 3, 4$. Thus, $< \mathbf{v}^k \cdot \mathbf{p} >$ is the probability that X plays c in the $(k + 1)^{th}$ round, *i.e.*, it is v_{12}^{k+1}. Hence, $v_{12}^{k+1} - v_{12}^k \ = < \mathbf{v}^k \cdot \tilde{\mathbf{p}} >$. The sum telescopes to yield

$$v_{12}^{n+1} - v_{12}^1 \ = \ \Sigma_{k=1}^n \ < \mathbf{v}^k \cdot \tilde{\mathbf{p}} > . \tag{3.2}$$

As the left side has absolute value at most 1, the limit (3.1) follows. If a subsequence of the averages converges to $\mathbf{v}$, then $< \mathbf{v} \cdot \tilde{\mathbf{p}} > = \ 0$ by continuity of the dot product.
$\square$

To illustrate the use of this result, we examine the Tit-for-Tat plan vector: $TFT = (1, 0, 1, 0)$ and another plan vector which has been labeled in the literature $Grim = (1, 0, 0, 0)$, e.g., [16]. We consider mixtures of each with $Repeat = (1, 1, 0, 0)$. Notice that if $\mathbf{p} = (1, 1, 0, 0)$ then $\tilde{\mathbf{p}} = (0, 0, 0, 0)$.

Corollary 3.2. *Let* $1 \geq a > 0$.
(a) The plan vector $\mathbf{p} = aTFT + (1 - a)Repeat$ *is good.*
(b) The plan vector $\mathbf{p} = aGrim + (1 - a)Repeat$ *is good.*

Proof: (a) In this case, $\tilde{\mathbf{p}} = a(0, -1, 1, 0)$ and so (3.1) implies that $v_2 = v_3 = \frac{1}{2}(v_2 + v_3)$. Hence, $s_Y = v_1R_2 + (v_2 + v_3)\frac{1}{2}(T_2 + S_2) + v_4P_2$. So $s_Y < R_2$ unless $v_2 = v_3 = v_4 = 0$ and $v_1 = 1$. This implies $s_X = R_1$ and so $\mathbf{p}$ is good.

(b) Now $\tilde{\mathbf{p}} = a(0, -1, 0, 0)$ and so (3.1) implies that $v_2 = 0$. Thus, $s_Y = v_1R_2 + v_3S_2 + v_4P_2$ and this is less than R_2 unless $v_3 = v_4 = 0$ and $v_1 = 1$. Again this shows that $\mathbf{p}$ is good.
$\square$

Remark: We will call the plan vectors of (a) *TFT-like* and those of (b) *Grim-like*.

It will be convenient at times to normalize the payoffs. Recall that for each player the utility can be composed with a positive affine transformation. For player p (with $p = 1, 2$) we adjust the utility function by subtracting S_p and then dividing by $T_p - S_p$. The normalized game is given by:

$X \backslash Y$	c	d
c	(R_1, R_2)	$(0, 1)$
d	$(1, 0)$	(P_1, P_2)

(3.3)

with

$$
\begin{aligned}
1 \;>\; R_1 \;>\; P_1 \;>\; 0, \quad \text{and} \quad R_1 \;>\; \frac{1}{2}, \\
1 \;>\; R_2 \;>\; P_2 \;>\; 0, \quad \text{and} \quad R_2 \;>\; \frac{1}{2}.
\end{aligned}
$$

(3.4)

Lemma 3.3. *Let s_X and s_Y be the payoffs with respect to the distribution vector* **v**. *The following are equivalent.*

(i) $s_Y \;\geq\; R_2$ *and* $s_X \;\geq\; R_1$.
(ii) $\mathbf{v}_1 = 1$.
(iii) $\mathbf{v} \;=\; (1, 0, 0, 0)$.
(iv) $s_Y \;=\; R_2$ *and* $s_X \;=\; R_1$.

Proof: The conditions of (i) and (iv) are preserved by separate positive affine transformations of the utilities of each player and conditions (ii) and (iii) don't depend upon the payoffs at all. Hence, we may assume that the payoffs have been normalized as in (3.3).

Take the dot product of **v** with $\frac{1}{2}(\mathbf{S}_Y + \mathbf{S}_X) = \left(\frac{1}{2}(R_1 + R_2), \frac{1}{2}, \frac{1}{2}, \frac{1}{2}(P_1 + P_2) \right)$. Observe that $\frac{1}{2}(R_1 + R_2)$ is the unique maximum entry of the latter. It follows that

$$
\frac{1}{2}(s_Y + s_X) \;\leq\; \frac{1}{2}(R_1 + R_2),
$$

(3.5)

with equality if and only if $\mathbf{v} = (1, 0, 0, 0)$.

(i) $\Rightarrow$ (ii): From (i) and (3.5) we see that $\frac{1}{2}(s_Y + s_X) = \frac{1}{2}(R_1 + R_2)$ and so $\mathbf{v} = (1, 0, 0, 0)$ which implies (ii).

(ii) $\Leftrightarrow$ (iii): Obvious since **v** is a probability vector.

(iii) $\Rightarrow$ (iv) and (iv) $\Rightarrow$ (i): Obvious. $\square$

Remark: Notice that (3.5) is *not* independent of separate affine transformations on the two sets of payoffs. It requires the normalized form.

In Theorem 2.5, the ratios of the differences $\frac{T_2 - R_2}{R_2 - S_2}$ and $\frac{T_2 - R_2}{R_2 - P_2}$ are invariant with respect to positive affine transformation and so are independent of the choice of utility function. The definition of good strategy and Nash type are similarly invariant. Thus, we may again normalize to use the payoffs given by (3.3) with inequalities (3.4). After normalization, Theorem 2.5 becomes the following.

Theorem 3.4. *Let* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *be an agreeable plan vector for X, other than* $Repeat$. *That is,* $p_1 = 1$ *but* $\mathbf{p} \neq (1, 1, 0, 0)$.

The plan vector $\mathbf{p}$ *is of Nash type if and only if the following inequalities hold.*

$$\frac{1 - R_2}{R_2} \cdot p_3 \leq (1 - p_2) \qquad and \qquad \frac{1 - R_2}{R_2 - P_2} \cdot p_4 \leq (1 - p_2). \tag{3.6}$$

The plan vector $\mathbf{p}$ *is good if and only if both inequalities hold strictly.*

Proof: We first eliminate the possibility $p_2 = 1$. If $1 - p_2 = 0$, then the inequalities would yield $p_3 = p_4 = 0$ and so $\mathbf{p} = Repeat$, which we have excluded. On the other hand, if $p_2 = 1$, then $\mathbf{p} = (1, 1, p_3, p_4)$. If X initially plays c and against this Y plays $AllD = (0, 0, 0, 0)$ with initial play d, then fixation occurs at $\{cd\}$ with $s_Y = 1$ and $s_X = 0$. Hence, $\mathbf{p}$ is not of Nash type. Thus, if $p_2 = 1$, then neither is $\mathbf{p}$ of Nash type, nor do the inequalities hold for it. We now assume $1 - p_2 > 0$.

Observe that

$$
\begin{aligned}
s_Y - R_2 &= (v_1 R_2 + v_2 + v_4 P_2) - (v_1 R_2 + v_2 R_2 + v_3 R_2 + v_4 R_2) \\
&= v_2(1 - R_2) - v_3 R_2 - v_4(R_2 - P_2).
\end{aligned} \tag{3.7}
$$

Hence, multiplying by the positive quantity $(1 - p_2)$, we have

$$s_Y \;>=\; R_2 \quad \Longleftrightarrow \quad (1 - p_2)v_2(1 - R_2) \;>=\; v_3(1 - p_2)R_2 + v_4(1 - p_2)(R_2 - P_2), \tag{3.8}$$

where this notation means that the inequalities are equivalent and the equations are equivalent.

Since $\tilde{p}_1 = 0$, Equation (3.1) of Lemma 3.1 implies $v_2 \tilde{p}_2 + v_3 \tilde{p}_3 + v_4 \tilde{p}_4 = 0$ and so

$$(1 - p_2)v_2 = v_3 p_3 + v_4 p_4.$$

Substituting in the above inequality and collecting terms we get

$$
\begin{aligned}
s_Y \;>=\; R_2 \quad &\Longleftrightarrow \quad Av_3 \;>=\; Bv_4 \qquad \text{with} \\
A &= [p_3(1 - R_2) - (1 - p_2)R_2] \qquad \text{and} \\
B &= [(1 - p_2)(R_2 - P_2) - p_4(1 - R_2)].
\end{aligned} \tag{3.9}
$$

Observe that the inequalities of (3.6) are equivalent to $A \leq 0$ and $B \geq 0$. The proof is completed by using a sequence of little cases.

Case (i) $A = 0, B = 0$: In this case, $Av_3 = Bv_4$ holds for any strategy for Y. So for any Y strategy, $s_Y = R_2$ and $\mathbf{p}$ is of Nash type. If Y chooses any plan vector that is not agreeable, then by Corollary 2.2 $v_1 \neq 1$. From Lemma 3.3, $s_X < R_1$ and so $\mathbf{p}$ is not good.

Case (ii) $A < 0, B = 0$: The inequality $Av_3 \geq Bv_4$ holds if and only if $v_3 = 0$. If $v_3 = 0$, then $Av_3 = Bv_4$ and so $s_Y = R_2$. Thus, $\mathbf{p}$ is Nash.

Case (iia) $B \leq 0$, any A: Assume Y chooses a plan that is not agreeable and is such that $v_3 = 0$. For example, if Y plays $AllD = (0, 0, 0, 0)$ then after the first round dc never occurs. With such a Y choice, $Av_3 \geq Bv_4$ and so $s_Y \geq R_2$. By Corollary 2.2 again $v_1 \neq 1$ because the Y plan is not agreeable. Again, Lemma 3.3 implies $s_X < R_1$ and $\mathbf{p}$ is not good. Furthermore, $v_3 = 0, v_1 < 1, p_2 < 1$, and

$(1 - p_2)v_2 = v_4 p_4$ imply that $v_4 > 0$. So if $B < 0$, then $Av_3 > Bv_4$ and so $s_Y > R_2$. Thus, $\mathbf{p}$ is not Nash when $B < 0$.

Case (iii) $A = 0, B > 0$: The inequality $Av_3 \geq Bv_4$ holds if and only if $v_4 = 0$. If $v_4 = 0$, then $Av_3 = Bv_4$ and $s_Y = R_2$. Thus, $\mathbf{p}$ is Nash.

Case (iiia)$A \geq 0$, any B: Assume Y chooses a plan that is not agreeable and is such that $v_4 = 0$. For example, if Y plays $(0, 1, 1, 1)$ then after the first round dd never occurs. With such a Y choice, $Av_3 \geq Bv_4$ and so $s_Y \geq R_2$. As before, $v_1 \neq 1$ implies $s_X < R_1$ and $\mathbf{p}$ is not good. Furthermore, $v_4 = 0, v_1 < 1, p_2 < 1$, and $(1 - p_2)v_2 = v_3 p_3$ imply that $v_3 > 0$. So if $A > 0$, then $Av_3 > Bv_4$ and so $s_Y > R_2$. Hence, $\mathbf{p}$ is not Nash when $A > 0$.

Case (iv) $A < 0, B > 0$: The inequality $Av_3 \geq Bv_4$ implies $v_3, v_4 = 0$. So $(1 - p_2)v_2 = v_3 p_3 + v_4 p_4 = 0$. Since $p_2 < 1$, $v_2 = 0$. Hence, $v_1 = 1$. That is, $s_Y \geq R_2$ implies $s_Y = R_2$ and $s_X = R_1$ and so $\mathbf{p}$ is good.

□

Remark: In Case (i) of the proof, the payoff $s_Y = R_2$ is determined by $\mathbf{p}$ independent of the choice of strategy for Y. In general, strategies that fix the opponent's payoff in this way were described by Press and Dyson [2] and, earlier, by Boerlijst, Nowak and Sigmund [17], where they are called *equalizer strategies*. The agreeable equalizer strategies have $\tilde{\mathbf{p}} = a(0, -\frac{1 - R_2}{R_2}, 1, \frac{R_2 - P_2}{R_2})$ with $1 \geq a > 0$. In general, such a strategy for X, indeed any strategy with $A = 0$ or $B = 0$, requires precise knowledge of Y's utility and so is really unusable.

Christian Hilbe (personal communication) suggested a nice interpretation of the above results:

Corollary 3.5. *Let* $\mathbf{p}$ *be an agreeable plan vector for X with* $p_2 < 1$.

(a) If $\mathbf{p}$ *is good, then using any plan vector* $\mathbf{q}$ *for Y that is not agreeable forces Y to get a payoff* $s_Y < R_2$.

(b) If $\mathbf{p}$ *is not good, then by using at least one of the two plans* $\mathbf{q} = (0, 0, 0, 0)$ *or* $\mathbf{q} = (0, 1, 1, 1)$, *Y can certainly obtain a payoff* $s_Y \geq R_2$, *and force X to get a payoff* $s_X < R_1$.

(c) If $\mathbf{p}$ *is not Nash, then by using at least one of the two plans* $\mathbf{q} = (0, 0, 0, 0)$ *or* $\mathbf{q} = (0, 1, 1, 1)$, *Y can certainly obtain a payoff* $s_Y > R_2$, *and force X to get a payoff* $s_X < R_1$.

Proof: (a): If $\mathbf{p}$ is good, then $s_Y \geq R_2$ implies $s_Y = R_2$ and $s_X = R_1$, which requires $v = (1, 0, 0, 0)$. By Corollary 2.2, this requires that $\mathbf{q}$ as well as $\mathbf{p}$ be agreeable.

(b) and (c) follow from the analysis of cases in the above proof.

□

Remark: We saw above that if $p_2 = p_1 = 1$, then if Y uses $\mathbf{q} = (0, 0, 0, 0)$ with initial play d while X initially plays c, then fixation at cd occurs with $s_Y = 1$ and $s_X = 0$.

Suppose that for memory-one plan vectors $\mathbf{p}$ for X and $\mathbf{q}$ for Y, the associated Markov matrix M has a unique invariant distribution $\mathbf{v}$, *i.e.*, a unique probability vector $\mathbf{v}$ such that $\mathbf{v} \cdot \mathbf{M} = \mathbf{v}$. In that case, regardless of the initial plays, $\mathbf{v} = Lim_{n \to \infty} \frac{1}{n} \Sigma_{k=1}^{n} \mathbf{v}^k$ and so the long term payoffs

$s_X = < \mathbf{v} \cdot \mathbf{S}_X >, s_Y = < \mathbf{v} \cdot \mathbf{S}_Y >$ are independent of the initial plays. Furthermore, if an error occurs in the play, later play will move the averages of subsequent distribution sequence $\{\mathbf{v}^n\}$ back toward $\mathbf{v}$.

A nonempty subset C of the set of outcomes $\{cc, cd, dc, dd\}$ is a *closed set* for $\mathbf{M}$ if the probability of moving from a state in C to a state outside C is zero. This implies that the submatrix of $\mathbf{M}$ with rows and columns from C defines a Markov matrix on C. Hence, there is an invariant distribution $\mathbf{v}$ for $\mathbf{M}$ which has *support in* C. That is $\mathbf{v}_i = 0$ for $i \notin C$. If play leads to an outcome in C, then the resulting associated limit distribution has support in C. Conversely, if $\mathbf{v}$ is an invariant distribution then the *support* $\{i : \mathbf{v}_i > 0\}$ is closed.

A plan vector $\mathbf{p}$ is *firm* if $p_4 = 0$. We will call $\mathbf{p}$ *unforgiving* if $p_3 = p_4 = 0$ and *forgiving* otherwise. So $\mathbf{p}$ is forgiving exactly when $p_3 + p_4 > 0$. We will say that $\mathbf{p}$ *twists* if $p_2 = 0, p_3 = 1$. For example, $Grim$ and $Repeat$ are unforgiving. TFT is firm and twists.

Proposition 3.6. *For memory-one plan vectors* $\mathbf{p}$ *for X and* $\mathbf{q}$ *for Y, let* $\mathbf{M}$ *be the associated Markov matrix.*

(i) $\mathbf{p}$ *is unforgiving if and only if* $\{dc, dd\}$ *is closed.*
(ii) $\mathbf{q}$ *is unforgiving if and only if* $\{cd, dd\}$ *is closed.*
(iii) $\mathbf{p}$ *and* $\mathbf{q}$ *are both agreeable if and only if* $\{cc\}$ *is closed.*
(iv) $\mathbf{p}$ *and* $\mathbf{q}$ *are both firm if and only if* $\{dd\}$ *is closed.*
(v) *If* $\mathbf{p}$ *and* $\mathbf{q}$ *both twist then* $\{cd, dc\}$ *is closed.*

Proof: These are all easy to check. For example, $\mathbf{p}$ is unforgiving if and only if whenever X plays d at some round then she plays d at all subsequent rounds regardless of Y's play. Corollary 2.2 says that $(1, 0, 0, 0)$ is an invariant distribution if and only if $\mathbf{p}$ and $\mathbf{q}$ are both agreeable, proving (iii). If $\mathbf{p}$ and $\mathbf{q}$ both twist then cd and dc alternate forever, once either occurs.

□

Definition 3.7. *A memory-one plan vector* $\mathbf{p}$ *for* X *is called* stably good *if it is good and if, in addition,* $p_2, p_3, p_4 > 0$.

Theorem 3.8. *Assume X uses a stably good memory-one plan vector* $\mathbf{p}$. *If Y uses a memory-one plan vector* $\mathbf{q}$, *then* $\mathbf{e}_1 = (1, 0, 0, 0)$ *is the unique invariant distribution for* $\mathbf{M}$, *and so is the unique limit distribution regardless of initial play, if and only if* $\mathbf{q}$ *is agreeable and forgiving.*

Proof: First assume that $\mathbf{q}$ is agreeable and forgiving.

Theorem 2.5 implies that $p_2 < 1$. So, from cd there is a positive probability that X plays d. Since $p_i > 0$ for $i = 1, 2, 3, 4$ there is a positive probability that X plays c from any state.

Since $\mathbf{p}$ and $\mathbf{q}$ are agreeable, $\{cc\}$ is closed and so $\mathbf{e}_1 = (1, 0, 0, 0)$ is an invariant distribution.

We first show that cc is an element of any closed subset C.

Case i: Assume $q_3 > 0$. This implies that from cd there is a positive probability that Y plays c and so there is a positive probability that from cd, the next outcome is the closed state cc. Because $p_3, p_4 > 0$, from dc and dd the move to cy has positive probability either with $y = c$ or with $y = d$. If $y = c$ then the move is to cc. If $y = d$, then the move is to cd from which a move to cc occurs with positive probability. It follows that if C is closed then $cc \in C$.

Case ii: Assume $q_3 < 1$ and $q_4 > 0$. Since $p_4 > 0$, from dd there is a positive probability that both X and Y play c and so there is a positive probability that from dd the next outcome is the closed state cc. Since $p_2, q_3 < 1$, there is a positive probability that from cd the play moves to dd and then from there to cc. Because $p_3 > 0$, from dc the move to cy has positive probability either with $y = c$ or with $y = d$. If $y = c$ then the move is to cc. If $y = d$, then the move is to cd from which a move to dd and thence a move to cc occur with positive probability. It follows that if C is closed then $cc \in C$.

Now suppose that $\mathbf{v}$ is an invariant distribution for $\mathbf{M}$ with $v_1 < 1$. Since $\mathbf{e}_1$ is also an invariant distribution, the linear combination

$$\tilde{\mathbf{v}} \;=\; [1 - \mathbf{v}_1]^{-1}[\mathbf{v} - \mathbf{v}_1 \cdot \mathbf{e}_1] \;=\; [1 - \mathbf{v}_1]^{-1}(0, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4) \tag{3.10}$$

is an invariant distribution with support disjoint from $\{cc\}$. This support would be a closed set not containing $\{cc\}$. It follows that $\mathbf{e}_1$ is the only invariant distribution.

Conversely, if $\mathbf{q}$ is not agreeable then $\mathbf{e}_1$ is not an invariant distribution by Corollary 2.2. If $\mathbf{q}$ is unforgiving, then by Proposition 3.6 (ii) $C = \{cd, dd\}$ is closed and so there is an invariant distribution with support in C.

□

4. Discussion and Conclusions

Let us see how all this plays out in practice. As Einstein is supposed to have remarked: "In theory, theory and practice are the same thing, but in practice they are really not."

We are considering the classical version of the Iterated Prisoner's Dilemma. It is a model for a very large—but unknown—number of repeated plays with payoff the average of the payoffs in the individual rounds. The number is assumed large enough that any advantage to a player in a fairly long run of initial encounters is swamped by the averaging. Our players are assumed to be humans. Thus, we are not considering tournament play where the original strategy is locked in, nor the evolutionary variant where the strategy is a fixed phenotypic character. Instead, the possibility exists of changing strategy during the play. The players can change horses mid-stream, as it were.

First, we consider Nash-type strategies which are not good. For example, suppose that X is using an equalizer strategy. This assures Y his cooperative payoff, but he has no particular incentive to play so that X receives the cooperative payoff as well. Furthermore, since the Nash-type strategies occur at the boundary of the set of good strategies, the Computation Problem, considered below, is especially acute for such strategies. Hence, we will restrict attention to the good strategies.

If $\mathbf{p}$ is agreeable but unforgiving, then it is *Grim-like*, *i.e.*, a mixture of *Grim* and *Repeat* as considered in Corollary 3.2 (b). It is good provided that $p_2 < 1$, *i.e.*, $\mathbf{p} \neq Repeat$. In American parlance, this choice of strategy is a nuclear option, a kind of Mutually Assured Destruction. If X adopts this strategy and Y at any time fails to conform with a c play then X plays d from then on. Eventually, Y will have to play all d and so the players are locked into the dd punishment payoff unless X relents, at which point Y will have to recognize and respond to the change in X's behavior.

Similarly, if $\mathbf{p}$ is a good strategy which is also firm, for example, any TFT-*like* strategy, then there is a risk of ending in the dd closed set if Y plays firm as well. Errors or probes by Y could lead to

such a situation, which can be escaped only if Y plays c from a dd outcome and then plays c from the resulting dc outcome.

Instead, let us suppose that X adopts a stably good strategy, *i.e.*, initially plays c and then uses a fixed stably good plan.

At this point we introduce memory. While a memory-one strategy is a convenient device for responding, it is reasonable to assume that each player can keep track of a fairly long series of previous outcomes. Because the probabilities p_i are all positive, such a sequence of outcomes would allow Y to estimate **p** and so to detect the memory-one plan which X has adopted. The power of a good strategy is that, once it is recognized, Y's best response is to cooperate.

For example, in the symmetric game, Press and Dyson [2] describe certain strategies they call *extortionate*. This is the reverse of the complier strategies mentioned above. If Y is using an extortionate strategy and X plays so that s_X is larger than the punishment payoff P, then $s_Y > s_X$. The only way that X can increase her payoff is by increasing Y's to a still higher level. They believe that this forces X to accept the lesser payoff in order to obtain anything above P. I disagree. Instead, I believe that X should stick with her stably good strategy. After a while, each recognizes the opponent's memory-one plan. Y's threat to stay at the extortionate strategy is not credible, because by sticking to extortionate play Y locks himself, as well as X, below the cooperation payoff, while by switching he assures himself (and X as well) the cooperation payoff. Knowing this, X has no incentive to respond to Y's threat, while Y has every reason to respond to X's.

There remains the *Computation Problem*. As we have observed, the inequalities (2.16) which X uses to choose a good strategy depend upon the payoffs to Y. The values p_2 and p_4 should be chosen small enough that the inequality

$$\frac{T_2 - R_2}{R_2 - P_2} \cdot p_4 \; < \; (1 - p_2). \tag{4.1}$$

may be expected to hold. One can then ensure that

$$\frac{T_2 - R_2}{R_2 - S_2} \cdot p_3 \; < \; (1 - p_2) \tag{4.2}$$

by choosing p_3 arbitrarily in $(0, 1 - p_2]$. This suggests that X choose a positive ϵ small enough that it is very likely that

$$\frac{T_2 - P_2}{R_2 - P_2} \cdot \epsilon \; < \; 1, \tag{4.3}$$

or equivalently,

$$\epsilon \; < \; \frac{R_2 - P_2}{T_2 - P_2},$$

and then adopt

$$\mathbf{p}(\epsilon) \;\; = \;\; (1, \; \epsilon, \; 1 - \epsilon, \; \epsilon), \tag{4.4}$$

where ϵ becomes the—very small—probability of responding with c to the opponent's play of d. Of course, for $TFT = \mathbf{p}(0)$ the Computation Problem does not arise and such behavior is easy for an opponent to detect. The problem with it is, as mentioned above, the risk of landing in a closed set disjoint from $\{cc\}$.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Hardin, G. The tragedy of the commons. *Science* **1968**, *162*, 1243–1248.

2. Press, W.; Dyson, F. Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *PNAS* **2012**, *109*, 10409–10413.

3. Aumann, R. *Survey of Repeated Games*; Collected Papers; MIT Press: Cambridge, MA, USA, 2000; Volume 1, pp. 411–437.

4. Friedman, J.W. A non-cooperative equilibrium for supergames. *Rev. Econ. Stud.* **1971**, *30*, 1–12.

5. Akin, E. *Stable Cooperative Solutions for the Iterated Prisoner's Dilemma*; 2013, ArXiv-1211.0969v2. arXiv.org e-Print archive. Available online: http://arxiv.org/abs/1211.0969 (accessed on 4 April 2015).

6. Stewart, A.; Plotkin, J. Collapse of cooperation in evolving games. *PNAS* **2014**, *111*, 17558–17563.

7. Davis, M. *Game Theory: A Nontechnical Introduction*; Dover Publications: Mineola, NY, USA, 1983.

8. Straffin, P. *Game Theory and Strategy*; Mathematical Association of America: Washington, DC, USA, 1993.

9. Axelrod, R. *The Evolution of Cooperation*; Basic Books: New York, NY, USA, 1984.

10. Maynard-Smith, S.J. *Evolution and the Theory of Games*; Cambridge University Press: Cambridge, UK, 1982.

11. Sigmund, K. *Games of Life*; Oxford University Press: Oxford, UK, 1993.

12. Nowak, M. *Evolutionary Dynamics*; Harvard University Press: Cambridge, MA, USA, 2006.

13. Stewart, A.; Plotkin, J. Extortion and cooperation in the Prisoner's Dilemma. *PNAS* **2012**, *109*, 10134–10135.

14. Hilbe, C.; Nowak, M.; Sigmund, K. The evolution of extortion in iterated Prisoner's Dilemma games. *PNAS* **2013**, *110*, 6913–6918.

15. Von Neumann, J.; Morgenstern, O. *Theory of Games and Economic Behavior*; Princeton Univeristy Press: Princeton, NJ, USA, 1944.

16. Sigmund, K. *The Calculus of Selfishness*; Princeton University Press: Princeton, NJ, USA, 2010.

17. Boerlijst, M.; Nowak, M.; Sigmund, K. Equal pay for all prisoners. *Am. Math. Mon*, **1997**, *104*, 303–305.