2013

Extortion and Evolution in the Iterated Prisoner's Dilemma

Michael J. Earnest
Harvey Mudd College

Extortion and Evolution in the Iterated Prisoner's Dilemma

Michael Earnest

Nicholas Pippenger, Advisor

Francis Su, Reader

Department of Mathematics

May, 2013

Abstract

The Prisoner's Dilemma is a two player game where playing rationally leads to a suboptimal outcome for both players. The game is simple to analyze, but when it is played repeatedly, complex dynamics emerge. Recent research has shown the existence of extortionate strategies, which allow one player to win at least as much as the other. When one player plays such a strategy, the other must either decide to take a low payoff, or accede to the extortion, where they earn higher payoff, but their opponent receives a larger share. We investigate what happens when one player uses this strategy against an "evolutionary" player, who makes small changes to her strategy over time to increase her score, and show that there are cases where such a player will not evolve towards the optimal strategy of giving in to extortion.

Contents

Chapter 1

Introduction

Game theory is a mathematical model for human interaction and decision making. Any interaction between people where each person's actions have an impact on the rest can be thought of as a game and modeled by this mathematical theory. A game of particular importance is the Prisoner's Dilemma.

1.1 The Prisoner's Dilemma

Before we can define the Iterated Prisoner's Dilemma, we first define what a two player game is. We denote the players X and Y. A *game* consists of both players' *strategies*, which are simply sets, and their *payouts*. If we denote the strategy sets by S_X and S_Y, then the *outcome* of a game is an element of $S_X \times S_Y$. The payouts, F_X and F_Y, are functions from $S_X \times S_Y$ to $\mathbb{R}$.

This is a very general definition of a game. There is a more restricted definition associated with the phrase "game theory," which the Prisoner's Dilemma is an example of. In such a game, each player has a finite number of *pure strategies*, P_X and P_Y. The pure outcomes are then elements of $P_X \times P_Y$, each of which yields a payout for X and Y. These payouts are usually represented in the form of a matrix, where the rows are labeled with X's pure strategies, the columns with Y's. Each entry contains a pair of numbers, the first being X's payout for that outcome, and the second being Y's.

However, we do not restrict X and Y to playing these pure strategies. Their entire strategy set consists of all possible probability distributions on their pure strategy sets. These are called *mixed strategies*. When X and Y

choose a mixed strategy, the pure outcome is then a random variable, and we say the payout for X and Y is the expected value of their payouts for this pure outcome.

Given a game, it is natural to ask how rational players will act. Here, rational means the player will choose strategies to maximize their payout. In order to simplify games, there is a way to see if certain strategies can be ignored, using the concept of dominance. Suppose $p_1, p_2 \in P_X$, and for all $p \in P_Y$, we have that $F_X(p_1, p) \geq F_Y(p_2, p)$, with strict inequality for some $p \in P_Y$. Then we say that p_1 *dominates* p_2. In this case, it would be irrational for X to choose p_2, since they can always perform at least as well, and sometimes better, with p_1.

We are now ready to define the Prisoner's Dilemma. Each player has two pure strategies, Cooperate and Defect. The payout matrix is as below:

	C	D
C	R, R	S, T
D	T, S	P, P

The numbers R, S, T, P can be anything satisfying these two conditions.

1. $T > R > P > S$
2. $2R > T + S$

This game is easy to solve. Since $T > R$ and $P > S$, Defection is the dominant strategy for both players, so they will both choose D and receive P. This is called a Dilemma because this outcome is suboptimal: they would be better off cooperating and receiving R. In fact, using the given inequalties, we can show that the outcome CC (the outcome where both players cooperate) is *Pareto optimal*, meaning that if any other outcome gives a player more that R, than it gives the other less than R. To see this, suppose X cooperates with probability p and Y with probaility q. Then a simple calculation shows that the expected *sum* of their scores will be

$$pq(2R) + [p(1-q) + (1-p)q](S+T) + (1-p)(1-q)(2P).$$

From the relations $R > P$ and $2R > S + T$, the above quantity is always at most $2R$. Since the outcome CC maximizes the sum of X and Y's score, it must also be Pareto optimal.

There are many cases where the Prisoner's Dilemma is a model for human interaction. When people work together, they can profit the most, but there is often a benefit to being dishonest and trying to gain more than one's

fair share at the other's expense. The previous analysis predicts that when humans act rationally in such situations, both players will defect, resulting in a suboptimal outcome. Since this result is very grim, and unrealistic in terms of how humans actually interact, a central question in the Prisoner's Dilemma is how to change the above game so that cooperative strategies become viable. A way to do this that seems possibly effective is to play the Prisoner's Dilemma many times against the same opponent. In the single case, their is no reason not to harm your opponent and gain the most for yourself, but in iterated play, it may be advantageous not to antogonize a player with whom you could benefit from cooperating in the future.

1.2 Past Formulations of the Iterated Prisoner's Dilemma

Robert Axelrod was one of the first to analyze the Prisoner's Dilemma under repeated play. In Axelrod (1984), he considers what happens when two players play the prisoner's dilemma an infinite number of times, but where the amount won from the i^{th} round is discounted by w^i for some $0 < w < 1$. This is equivalent to stipulating that players play together over and over, but after each round they stop playing with probability w. Each player's strategy can be thought of as a computer program: given an input of the outcomes of all the past rounds, a strategy outputs either the decision to cooperate or defect. There are infinitely many strategies, so analyzing this game is very difficult. In order to cope with this, Axelrod held a tournament, where people were invited to submit computer programs to play the Prisoner's Dilemma against each other repeatedly in a round-robin fashion, where every pair of programs played the Prisoner's Dilemma against each other a large number of times. There were two such tournaments, the first with 15 entries and the second with 63. In both tournaments, the most succesful strategy was Tit-for-Tat, which initially cooperates, and in subsequent rounds repeats its opponent's last move.

In general, Axelrod found that the stragies which performed well had the following properties:

1. Kindness (never defecting unless opponet does first),
2. Reciprocity (treat your oponent as they treat you),
3. Simplicity (described by a short computer program).

These results are in no way conclusive proof that Tit-for-Tat is the "best" strategy. Both of these tournaments used the values $R = 3, T = 5, S =$

$0, P = 1$, so we cannot conclude anything about the Prisoner's Dilemma in general. In addition, this experiment only shows Tit-for-Tat performs well when played against the particular makeup of competitors that was submitted into the tournament. However, these experiments still do illuminate many of the important properties of an effective strategy in the IPD, and Axelrod's work sparked further research into the subject.

The approach taken by Axelrod is not the only way to formulate the Iterated Prisoner's Dilemma. In Nowak and Sigmund (1990), the set of strategies each player can use is much more limited; their probability of cooperating on a given round is only a function of their opponents previous action. Using this limited set of strategies, Nowak was able to find exact formulas for the payoff of each player given their strategies. This is similar to the approach taken in this paper.

In this paper, we deal with a particular form of the IPD, where the each player bases their strategy for a particular round on the result of the previous round (not just their opponent's previous action). The exact definitions of this IPD formulation are in Chapter 2, along with the previous research done by Dyson and Press (2012). In Chapter 3, we examine classify the cases where the IPD is particularly easy to study. Finally, in Chapter 4, we examine the implications of the recenetly discovered *exortionate strategies* in the IPD.

Chapter 2

Background

Before defining the our version of the IPD, we need some results from the theory of Markov chains.

2.1 Markov Chains

Definition 2.1. *Let $X_0, X_1, X_2, \ldots$ be a sequence of random variables, which each take values in some finite set S. Then $\{X_n\}_{n=0}^{\infty}$ is a* Markov chain *if*

$$P(X_{n+1} = x | X_1 = x_1, \ldots, X_n = x_n) = P(X_{n+1} = x | X_n = x_n).$$

We say a Markov Chain is time homogenous *if $P(X_{n+1} = x | X_n = x')$ is independent of n.*

We will only consider time homogenous Markov processes. The elements of S will be referred to as *states*, and can be thought of as being indexed from 1 to s. The probability, given that a variable is in the i^{th} state, of the next variable being in the j^{th} state is denoted p_{ij}. We can arrange these probabilities in a matrix, M. This is useful, since, if we write the distribution of each X_k as a row vector v_k, whose i^{th} entry is $P(X_k = s_i)$, then one can show that

$$v_{k+1} = v_k M.$$

Applying this formula k times to the initial distribution v_0, we have that

$$v_k = v_0 M^k.$$

In the same way that the i, j entry of M determines the probability of transitioning from state i to state j, we can show that the i, j entry of M^k determines the probability of transitioning from state i to state j in exactly k steps. We denote this probability by $p_{ij}^{(k)}$. This leads to several definitions:

Definition 2.2. *If $p_{ij}^{(n)} > 0$ for some $n > 0$, then we say state j is* accessible *from state i. If both i and j are accessible from each other, then they communicate.*

The relation "i communicates with j" is an equivalence relation, which partitions S into equivalence classes called *communicating classes*.

Definition 2.3. *If a communicating class C has the property that no state outside of C is accessible from one in C, then we say that C is* ergodic. *Otherwise, it is* transient.

If a communicating class is a single state, then that state is absorbing.

A Markov chain always has at least one ergodic communicating class. This can be seen by forming a poset on the communicating classes, where $C_1 \geq C_2$ if there is a state in C_2 accessible from some state in C_1. This poset is finite, and must therefore have minimal elements, which are ergodic.

There is one more main property used to characterize Markov chains. It may be the case that when X_n is in some state, s_k, then it will only return to s_k after a number of steps later which is a multiple of some period, d. For instance, if the transition matrix M is

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and the state is initially in s_1, then it will only return to s_1 on even numbered steps. This motivates the below definition:

Definition 2.4. *The* period *of a state s_i is greatest common divisor of the set*

$$\{n | p_{ii}^{(n)} > 0\}.$$

If all states of a Markov chain have period 1, we say that the chain is aperiodic.

One of the central problems in Markov theory is determining what distribution the X_n converge to, if any. The X_n may not converge at all: this occurs when there are states with periods greater than 1, since this will cause the process to alternate between distributions instead of settling. However, it can be shown that the average of $X_0, \ldots, X_n$ will always approach some limit distribution. Specifically, we have

Proposition 2.1. *For any Markov chain, $\frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)}$ converges as $n \to \infty$.*

This is proved in (Tijms, 2003: p. 96). We then have that

Corollary 2.1. *For any initial distribution v_0,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} v_0 M^n$$

exists.

This is true since the j^{th} component of $\frac{1}{n}\sum_{k=0}^{n-1} vM^n$ is given by

$$\sum_{i=1}^{s} v_i \left(\frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)} \right),$$

which is a linear combination of series which converge. The above limit is called a *stationary distribution* of the Markov process, and it can be shown that when π is such a distribution, then

$$\pi M = \pi.$$

Such a distribution must exist, but it may not be unique. To see this, suppose that C_1 and C_2 are two distinct ergodic communicating classes. Any initial distribution which is only nonzero in C_1 will have this property for all future time steps, as will its limit. The same applies for C_2, so that there are two initial distributions which cannot converge to the same distribution. Since the property of having a unique stationary distribution is important, we give the following definition:

Definition 2.5. *If M is the transition matrix for a Markov process, and there is a π such that for all v_0, $\frac{1}{n}\sum_{k=1}^{n-1} v_0 M^k$ converges to π as $n \to \infty$, we say that the process is* well-behaved.

The previous paragraph showed that having only one ergodic class is necessary for being well-behaved. The next theorem proves that this is sufficient as well.

Theorem 2.1. *If a Markov chain with transition matrix M has a single ergodic communicating class, then*

1. *The matrices $\frac{1}{n}\sum_{k=0}^{n-1} M^k$ converge (componentwise) to a matrix A, whose rows are all the same probability vector π.*

2. *The Markov chain is well-behaved.*

This is also proved in (Tijms, 2003: p. 131-2).
We are now ready to define the Iterated Prisoner's Dilemma rigorously.

2.2 The Iterated Prisoner's Dilemma

As discussed, there are many ways to define the Iterated Prisoner's Dilemma. Our approach follows that of Dyson and Press (2012), and we use notation largely similar to this article. Each player chooses a strategy, which consists of their decision of how to act on the first round, and how to decide their decision for every subsequent round based on the outcomes of the previous rounds. To simplify the problem, we assume that each player has *finite memory*, meaning that their decision for a particular round is a function of the past m outcomes for some fixed number m. Press and Dyson's paper proved that if X has memory 1, then every higher memory strategy Y plays leads to the same outcome as Y playing a certain memory 1 strategy (Dyson and Press, 2012: p. 4). We then suppose that X has memory 1, so that without loss of generality, Y does as well.

Formally, each player chooses a vector $\mathbf{p} = (p_1, p_2, p_3, p_4)$, which are her probabilities of cooperating on a given round given the previous outcome was respectively CC, CD, DC or DD, where X's choice is the first letter and Y's is the second. Throughout this paper, we will always refer to the outcomes in this order. When we say the i^{th} outcome, we mean the i^{th} outcome in the list CC, CD, DC, DD. If X adopts the strategy $\mathbf{p} = (p_1, p_2, p_3, p_4)$ and Y plays $\mathbf{q} = (q_1, q_2, q_3, q_4)$, and we let random variable V_i be the outcome of the i^{th} round, then the sequence $\{V_i\}$ will be a markov chain with transition matrix M, where

$$M = \begin{bmatrix} p_1q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) \\ p_2q_2 & p_2(1-q_2) & (1-p_2)q_2 & (1-p_2)(1-q_2) \\ p_3q_3 & p_3(1-q_3) & (1-p_3)q_3 & (1-p_3)(1-q_3) \\ p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4) \end{bmatrix}$$

In order to define the payouts of X and Y, we look at stationary vector of this process. Let $S_X = (R, S, T, P)^T$ and $S_Y = (R, T, S, P)^T$ be column vectors which encode the payout for each player for each outcome. If π is a unique stationary vector of M, as shown in the previous section, the expected average payouts of each player tend to πS_X and πS_Y. We denote these payouts by s_X and s_Y. However, it may be the case that the V_i are not well-behaved, so that we must specify an initial distribution in order for these payouts to be well defined. A very general way to do this is for X and Y to choose probabilities p_0 and q_0 of cooperating on the first round, so that $v_0 = (p_0q_0, p_0(1-q_0), (1-p_0)q_0, (1-p_0)(1-q_0))$. Now, we have a well defined game: each players strategies consist of a probability vector

and a probability, and their payoffs are given by the below limits:

$$s_X = \left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} v_0 M^k \right), \qquad s_Y = \left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} v_0 M^k \right).$$

When the Markov process is well-behaved, the above formula will not depend on v_0. This is true for most strategies (for instance, when all of X and Y's probabilities are strictly between 0 and 1).

2.3 Previous Work

Suppose that X plays the strategy $\mathbf{p} = (p_1, p_2, p_3, p_4)$ and Y plays $\mathbf{q} = (q_1, q_2, q_3, q_4)$. In their paper, Press and Dyson gave a formula for the dot product of the stationary vector, $\mathbf{v}$, with any vector $\mathbf{f}$. Namely, if we define

$$D(\mathbf{p}, \mathbf{q}, \mathbf{f}) \equiv \det \begin{bmatrix} -1 + p_1 q_1 & -1 + p_1 & -1 + q_1 & f_1 \\ p_2 q_2 & -1 + p_2 & q_2 & f_2 \\ p_3 q_3 & p_3 & -1 + q_3 & f_3 \\ p_4 q_4 & p_4 & q_4 & f_4 \end{bmatrix},$$

then we have that

$$\mathbf{v} \cdot \mathbf{f} = \frac{D(\mathbf{p}, \mathbf{q}, f)}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})},$$

where $\mathbf{1} = (1, 1, 1, 1)$. Since the average payouts of each player, s_X and s_Y, are given by $v \cdot S_X$ and $v \cdot S_Y$, the average payouts of X and Y are given by

$$s_X = \frac{D(\mathbf{p}, \mathbf{q}, S_X)}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}, \qquad s_Y = \frac{D(\mathbf{p}, \mathbf{q}, S_Y)}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}.$$

Since $D(\mathbf{p}, \mathbf{q}, \mathbf{f})$ is linear in f, this also gives that

$$a s_X + b s_Y + c = \frac{D(\mathbf{p}, \mathbf{q}, a S_X + b S_Y + c\mathbf{1})}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}.$$

Notice that, in the formula for $D(\mathbf{p}, \mathbf{q}, \mathbf{f})$, the middle two columns are entirely under a single player's control. When it is possible for a player to choose a strategy such that their column is a multiple of $aS_X + bS_Y + c\mathbf{1}$, then $D(\mathbf{p}, \mathbf{q}, aS_X + bS_Y + c\mathbf{1})$ will be zero, so that $as_X + bs_Y + c = 0$. It is not possible for players to choose a strategy which makes this true for all choices of a, b and c. However, they did show that it is possible for X to force the relationship, for any $\chi > 1$,

$$s_X - P = \chi(s_Y - P)$$

This means that X's score, in excess of P, is a multiple of Y's score relative to P. They called such strategies *extortionate*, with extortion factor χ. Specifically, X can do this by choosing

$$\begin{bmatrix} 1-p_1 \\ 1-p_2 \\ p_3 \\ p_4 \end{bmatrix} = \phi \begin{bmatrix} (\chi-1)\frac{R-P}{P-S} \\ 1+\chi\frac{T-P}{P-S} \\ \chi+\frac{T-P}{P-S} \\ 0 \end{bmatrix}.$$

However, the range of ϕ which makes the above a valid probability vector is

$$0 \leq \phi \leq \frac{(P-S)}{(P-S)+\chi(T-P)}$$

When ϕ is chosen so that these inequalites are strict, then $0 < p_1, p_2, p_3 < 1$. There also exist extortionate strategies for Y, which are defined similarly.

When X plays an extortionate strategy, Press and Dyson showed that when Y cooperates fully (i.e. uses the strategy (1,1,1,1)), the payouts of both X and Y are maximized. In this case, X receives an amount more than R, and Y receives less that R. The only way for Y to win an equal amount as X is for Y to choose a strategy where $q_4 = 0$. Since X also will have $p_4 = 0$, the state DD, once entered, will never be left, so that both players will receive a payout of P, which is less than what she would have gotten if she had cooperated.

Extortionate strategies place X in a position which is at least as good as Y. This makes the strategy seem too good to be true. However, it is only beneficial to X if Y goes along with the extortion, so we must try to determine how a player will react to such a strategy. This is discussed in Chapter 4. Before that, we must examine the foundations on which Press and Dyson's results stand; namely, their formulas for s_X and s_Y.

Chapter 3

Uniqueness of the Stationary Distribution

The Markov process for the IPD will not always have a unique distribution. Suppose that both X and Y play Tit-for-Tat (repeating opponent's last move), meaning $\mathbf{p} = (1, 0, 1, 0)$ and $\mathbf{q} = (1, 1, 0, 0)$. Then the communicating classes are $\{CC\}$, $\{CD, DC\}$, and $\{DD\}$. These are all ergodic, meaning that it impossible to move from one class to the other. The Markov processes restricted to each of these classes each have a unique stationary distribution, and the limit distribution of the entire process will be some weighted average of these three distributions.

For instance, suppose that X initially cooperates, and Y initially cooperates with probability $2/3$, making the initial distribution $(2/3, 1/3, 0, 0)$. Then there will be a $2/3$ chance that the first outcome is CC, where it will be CC for all subsequent rounds, and a $1/3$ chance of CD, where play will alternate between CD and DC. In this case, the average of their distributions will approach $(\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, 0)$. In contrast, if they both initally defected, the distribution would approach $(0, 0, 0, 1)$.

In their paper, Press and Dyson gave a formula for the dot product of the stationary distribution with any vector, which was independent of the initial condition. We can see that this formula must be wrong in the cases where there are more than one ergodic class. In fact, when the values for Tit-for-Tat are substituted in the formula for s_X and s_Y, then the denominator, $D(\mathbf{p}, \mathbf{q}, \mathbf{1})$, becomes zero! This formula is important for many of Press and Dyson's results, as well as for our results in the next chapter. Thus, in this chapter, we characterize the choices of strategies for X and Y for which the formula does hold, which are given in Theorem 3.1.

3.1 When the IPD is not Well-Behaved

In order to simplify the calculations, we introduce an alternate way to write the players' strategy vector. Suppose that X plays a strategy $\mathbf{p} = (p_1, p_2, p_3, p_4)$, and Y plays $\mathbf{q} = (q_1, q_2, q_3, q_4)$. We define the vectors

$$\widetilde{\mathbf{p}} = (1 - p_1, 1 - p_2, p_3, p_4) \qquad \widetilde{\mathbf{q}} = (1 - q_1, q_2, 1 - q_3, q_4).$$

Whereas $\mathbf{p}$ gives probability of cooperating, $\widetilde{\mathbf{p}}$ gives the probability of *switching strategies*. For instance, if $\widetilde{p}_2 = 1/3$, then if the previous outcome was CD, meaning X just cooperated, there is a $1/3$ chance X will switch to D on the next round.

We break the gameplay chains which aren't well-behaved into several cases.

3.1.1 Multiple Absorbing States

Any Markov process with multiple absorbing states will not be well-behaved, since these states will be different ergodic classes. Notice that the i^{th} state is absorbing exactly when $\widetilde{p}_i = \widetilde{q}_i = 0$, since once state i is reached, neither player will switch strategies. Thus, having two states whose coordinates in $\widetilde{\mathbf{p}}$ and $\widetilde{\mathbf{q}}$ are zero is a sufficient condition for gameplay to not be well-behaved.

3.1.2 No Absorbing States

In order for the gameplay to not be well-behaved, there must be two ergodic communicating classes, and in this case they must be each of size two. There are 3 ways to partition $\{CC, CD, DC, DD\}$ into two sets of size two. In order for the communicating classes to be $\{CC, CD\}$ and $\{DC, DD\}$, it must be the case that X never changes her strategy, so that $\widetilde{\mathbf{p}} = (0, 0, 0, 0)$. Similarly, $\widetilde{\mathbf{q}} = (0, 0, 0, 0)$ yields gameplay with ergodic classes $\{CC, DC\}$ and $\{CD, DD\}$. Note when both $\widetilde{\mathbf{p}} = \widetilde{\mathbf{q}} = (0, 0, 0, 0)$, then all states are absorbing, which was covered in the last case. To get the remaining possibility, both players must switch strategies after every move, so that $\widetilde{\mathbf{p}} = \widetilde{\mathbf{q}} = (1, 1, 1, 1)$.

3.1.3 One Absorbing State

If exactly one state is absorbing, and there are two ergodic classes, then the other class must be of size 2 or 3. The next table shows all of the ways to

choose an $S \subseteq \{CC, CD, DC, DD\}$ of size 2 or 3, along with the necessary and sufficient conditions for S to be *closed* (meaning the probability of leaving S is zero). For ease of reading, we use the numbers 1 through 4 to refer to the states (in the order CC, CD, DC, DD), and omit the braces for around each set S.

S	is closed if...
$1,2$	$\widetilde{p}_1 = \widetilde{p}_2 = 0$
$3,4$	$\widetilde{p}_3 = \widetilde{p}_4 = 0$
$1,3$	$\widetilde{q}_1 = \widetilde{q}_3 = 0$
$2,4$	$\widetilde{q}_2 = \widetilde{q}_4 = 0$
$1,4$	$\widetilde{p}_1 = \widetilde{q}_1 = \widetilde{p}_4 = \widetilde{q}_4 = 1$
$2,3$	$\widetilde{p}_2 = \widetilde{q}_2 = \widetilde{p}_3 = \widetilde{q}_3 = 1$
$1,2,3$	$\widetilde{p}_1\widetilde{q}_1 = \widetilde{p}_2(1-\widetilde{q}_2) = (1-\widetilde{p}_3)\widetilde{q}_3 = 0$
$1,2,4$	$\widetilde{p}_1(1-\widetilde{q}_1) = \widetilde{p}_2\widetilde{q}_2 = (1-\widetilde{p}_4)\widetilde{q}_4 = 0$
$1,3,4$	$(1-\widetilde{p}_1)\widetilde{q}_1 = \widetilde{p}_3\widetilde{q}_3 = \widetilde{p}_4(1-\widetilde{q}_4) = 0$
$2,3,4$	$(1-\widetilde{p}_2)\widetilde{q}_2 = \widetilde{p}_3(1-\widetilde{q}_3) = \widetilde{p}_4\widetilde{q}_4 = 0$

For instance, the first row gives the conditions required for $\{CC, CD\}$ to be a closed set. In order for this to be true, it must mean that if X has just cooperated, she will cooperate again, so that she will not change strategies. This is equivalent to saying that $\widetilde{p}_1 = \widetilde{p}_2 = 0$, as shown in the second column. The other cases where $|S| = 2$ are derived similarly. In the $1,2,3$ row, where $S = \{CC, CD, DC\}$, the only way to leave S is to transition to DD. The probabilities of transitioning to DD from each state are $(1 - p_i)(1 - q_i)$, since p_i and q_i are the probabilities of each player cooperating in state i. For S to be closed, we must have each of the transition probabilities be zero. These three conditions are what is listed in the cell adjoining $1,2,3$, with each p_i and q_i converted to $\widetilde{p}_i$ and $\widetilde{q}_i$.

This table is useful for checking if the IPD is well-behaved, given that it has exactly one absorbing state. Suppose only state i is abosrobing. If none of the subsets of $\{CC, CD, DC, DD\} - \{i\}$ are closed, then they are not ergodic communicating classes, so there will not be two ergodic communicating classes. We know that none of the singleton subsets of $\{CC, CD, DC, DD\} - \{i\}$ are closed, since only i is absorbing, and the above table allows us to check if the subsets of size 2 or 3 are closed. If none are closed, the IPD will be well-behaved. If not, then some subset of the closed class will be a communicating class, so the IPD will not be behaved.

We have now exhausted all ways for the IPD to not be well-behaved, leading us to the main theorem of this chapter.

Theorem 3.1. *When X plays the vector $\widetilde{\mathbf{p}} = (\widetilde{p}_1, \widetilde{p}_2, \widetilde{p}_3, \widetilde{p}_4)$ and Y plays $\widetilde{\mathbf{q}} = (\widetilde{q}_1, \widetilde{q}_2, \widetilde{q}_3, \widetilde{q}_4)$, then their gameplay is well-behaved if and only if none of these conditions hold:*

(i) There are two distinct states i and j for which

$$\widetilde{p}_i = \widetilde{q}_i = \widetilde{p}_j = \widetilde{q}_j = 0.$$

(ii) Either $\widetilde{\mathbf{p}} = \mathbf{0}$ or $\widetilde{\mathbf{q}} = \mathbf{0}$.

(iii) Both $\widetilde{\mathbf{p}} = \widetilde{\mathbf{q}} = \mathbf{1}$.

(iv) There is a state i for which $\widetilde{p}_i = \widetilde{q}_i = 0$, and a row in the above table where $i \notin S$ and the listed conditions hold.

In these cases, the formula given by Dyson and Press holds. In all other cases, no formula for s_X and s_Y independent of their initial choices can exist.

3.2 Discussion

When we initially defined the IPD, we allowed for any choice of the vectors $\mathbf{p}$ and $\mathbf{q}$. This requires both players to also choose an initial probability of cooperating, because as we have just seen, there are several caes where the stationary distribution will depend on the initial. The IPD would be much cleaner if they did not have to specify this initial condition, since their strategies would be vectors of four numbers instead of five. There are several ways to modify the IPD in order to eliminated the need for p_0 and q_0. One way to is to fix p_0 and q_0; for example, we could assume that both players are "nice", and will initially cooperate. Though this works, it has a number of drawbacks. No choice of p_0 or q_0 seems natural, and making an arbitrary choice for the initial condition is unintuitive.

More importantly, when we do this, the payoffs will be a discontinuous function of the player's strategies. To see this, supose both players initially cooperate, and consider the case where they both play Tit-for-Tat. Then each of their payoffs will be R. However, if X, instead of playing $(1, 0, 1, 0)$, reduces p_1 by ε, for any $\varepsilon > 0$, then eventually X will defect, where the outcomes will then alternate between CD and DC. We see that X makes an arbitrarily small strategy change, and the payoffs for each player drop to $(S + T)/2$. The reason that continuity is important will become more apparent in the next section.

Another way to modify the IPD is to impose restrictions on the choices for $\mathbf{p}$ and $\mathbf{q}$, or equivalently, on $\widetilde{\mathbf{p}}$ and $\widetilde{\mathbf{q}}$. This allows us to ensure that the cases listed in the previous theorem never occur. For instance, if any $\mathbf{p}$ or $\mathbf{q}$ with two zero entries are made illegal, then cases (i) and (ii) will never occur. Similarly, we can avoid (iii) by eliminating $\mathbf{1}$ as an option for both players. In order to avoid case (iv), we can require that if any of $\widetilde{p}_i$ are zero, then all other entires must be striclty between 0 and 1, and simlarly for $\widetilde{\mathbf{q}}$. Though this eliminates the dependency of payoffs on the initial conditions, it has also eliminated several famous strategies, such as Always Cooperate, Always Defect, and Tit-for-Tat. Instead, players must play strategies which are close. For example, suppose that X and Y each does not trust the other to cooperate, so that they decide to defect as much as possible. It is illegal to defect all of the time, but they may choose to always cooperate with small probabilities δ and ε. When X plays $(\delta, \delta, \delta, \delta)$ and Y plays, $(\varepsilon, \varepsilon, \varepsilon, \varepsilon)$ one can find that Y's payoff is given by

$$s_Y = 1 + 4\delta - \varepsilon(1 + \delta)$$

for the conventional values $(R, T, S, P) = (3, 5, 0, 1)$. Notice that as Y decreases ε, her payout increases. In this situation, Y has no best strategy; she would always prefer a smaller value of ε. This is counter-intuitive, and slightly troubling.

In order to eliminate this strange behavior, we can impose more restrictions on the available strategies to X and Y. For some number $0 < \varepsilon < \frac{1}{2}$, if we only allow strategies where each p_i and q_i are in the interval $[\varepsilon, 1 - \varepsilon]$, then we avoid dependence on initial conditions, and avoid the previous situtation where Y is never satisfied with her strategy. This is exactly the approach taken by Bohning et al. (2002).

Chapter 4

Evolutionary Play

We might ask how a rational player would react when they know X is playing an extortionate strategy. If Y is rational, then she will play to maximize her gain, which means choosing the strategy $(1,1,1,1)$, and accepting the unfair payoff. This is because Y has only one chance to earn points. However, when the IPD is played many times (!), then there could be some advantage to being uncooperative. If Y played $(0,0,0,0)$, then X would receive P; if she does this repeatedly, there is a chance X could realize that the extortionate strategy wasn't effective, and instead switch to a more cooperative strategy.

In order to eliminate the need for analyzing the though processes of each player, we suppose that Y plays according to a simple rule. She intially plays a strategy $\mathbf{q}_0$ and makes small changes to this strategy to increase her score. We can then ask what strategy Y will approach when X plays an extortionate strategy. Even though Y's optimal strategy given that X plays an extortionate strategy is to cooperate ($\mathbf{q} = \mathbf{1}$), it may not be true that Y will always evolve towards this. Press and Dyson conjectured that for conventional payouts, i.e. $(R,T,S,P) = (3,5,0,1)$, play will always evolve towards cooperation.

To formalize this, suppose that X plays an extotrionate stategy with $\chi > 1$, where she choose ϕ so that $0 < p_1, p_2, p_3 < 0$. We wish to only consider cases where the stationary distribution is unique in order to simplify the analysis. Referring to Theorem 3.1, neither case (i) nor (iii) when X plays an extortionate strategy. However, we can see that case (iv) occurs when Y chooses a strategy of the form $\widetilde{\mathbf{q}} = (0, \widetilde{q}_2, 0, 0)$. In order to make s_Y a function of $\mathbf{q}$ alone, we allow Y to play all strategies except these. In this new domain, since s_Y is a rational function in q_1, q_2, q_3, q_4 whose denomi-

nator is never zero, both it and its derivative are Lipschitz continuous on this domain.

We want $\mathbf{q}(t)$ to change in the direction that increases Y's payout the most. Intuitively, the derivative of $\mathbf{q}(t)$ should be the gradient of s_Y, but there are cases where this would cause $q(t)$ to leave I^4, which would make it an invalid probability vector. For instance, along the wall where $q_4 = 0$, both players will receive a payout of P. Once Y is at this point, the gradient of s_Y will be zero in the first 3 coordinates (since in these directions, her payout is still P), and there are choices of R, T, S, P where it is negative in the fourth. In order to account for this, we define $F(\mathbf{q}) = (F_1(\mathbf{q}), F_2(\mathbf{q}), F_3(\mathbf{q}), F_4(\mathbf{q}))$ to be $\nabla s_Y(\mathbf{q}(t))$, except when $q_i(t) = 0$ and the i^{th} coordinate of ∇s_Y is negative, or $q_i(t) = 1$ and the i^{th} coordinate of ∇s_Y is positive. In these cases, we say that $F_i(\mathbf{q}(t)) = 0$.

Given an initial vector $\mathbf{q}_0$, we define Y's evolution of play to be a vector valued function $\mathbf{q}(t)$, for $t \in [0, \infty)$ which satisfies:

$$\mathbf{q}(0) = \mathbf{q}_0$$

$$\mathbf{q}'(\mathbf{t}) = \mathbf{F}(\mathbf{q}(t)).$$

when $\mathbf{q}(t) \in U$.

From the Picard-Lindelöf theorem, such a $\mathbf{q}$ exists and is unique as long as $\mathbf{q}(t)$ remains in the interior of I^4. Once a path hits the boundary of I^4, $\mathbf{q}(t)$ may fail to be continuous, as $F(t)$ may fail to be.

Our first result shows that there is a region near the origin in which all initial strategies will reach the situation described above.

Theorem 4.1. *When $2P > S + T$, there is an open region W such that when $\mathbf{q}_0 \in W$, the path $q(t)$ terminates on the boundary of $q_4 = 0$.*

Proof. To prove this, we first cite a result from Press and Dyson, which states that

$$\nabla s_Y\Big|_{\mathbf{q}=\mathbf{0}} = \mathbf{F}(\mathbf{0}) = \left(0, 0, 0, \frac{(T-S)(S+T-2P)}{(P-S)+\chi(T-P)}\right)$$

Since $2P > S + T$, the last coordinate is negative: call its value -2ε. Since $\mathbf{F}$ is continuous, there is a $\delta > 0$ such that when $0 < \|\mathbf{q}\| < \delta$, then $|F_i(\mathbf{q})| < \varepsilon$ when $i = 1, 2, 3$ and $F_4(\mathbf{q}) < -\varepsilon$. Here, we let $\|\mathbf{q}\|$ be the infinity norm, namely, the maximum of its coordinates.

Now, define W to be the region where $0 < q_4 < \delta/2$, and $|q_i - \delta/2| < \delta/2 - q_4$ for $i = 1, 2, 3$. We first show that any flow originating in W must leave W.

For any $\mathbf{q}_0 \in W$, assume that $\mathbf{q}(t) \in W$ when $t \leq \delta/(2\varepsilon)$. Consider the fourth coordinate of $q(t)$. This satisfies

$$\frac{dq_4(t)}{dt} = F_4(\mathbf{q}(t)).$$

Integrating, we have

$$q_4\left(\frac{\delta}{2\varepsilon}\right) = q_4(0) + \int_0^{\frac{\delta}{2\varepsilon}} F_4(\mathbf{q}(t))\, dt.$$

We have that $q_4(0) < \delta/2$, and since W is in the region where $\|\mathbf{q}\| < \delta$, that $F_4(\mathbf{q}(t)) < -\varepsilon$. Thus,

$$q_4\left(\frac{\delta}{2\varepsilon}\right) < \delta/2 + \int_0^{\frac{\delta}{2\varepsilon}} -\varepsilon\, dt = 0.$$

This shows $\mathbf{q}$ has exited W, contradicting our assumption.

Now, we show that $\mathbf{q}$ exits W where $q_4 = 0$. Since W is open, the set of times where $\mathbf{q}(t) \notin W$ is closed, and there is a minimum time t^* where $\mathbf{q}(t^*) \notin W$. This must either satisfy $|q_i(t^*) - \delta/2| \geq \delta/2 - q_4(t^*)$ or $q_4(t^*) \notin (0, \delta/2)$. However, for $i = 1, 2, 3$,

$$\begin{aligned}
\left|q_i(t^*) - \frac{\delta}{2}\right| &= \left|q_i(0) - \frac{\delta}{2} + \int_0^{t^*} F_i(\mathbf{q}(t))\, dt\right| \\
&\leq \left|q_i(0) - \frac{\delta}{2}\right| + \int_0^{t*} |F_i(\mathbf{q}(t))|\, dt \\
&< \left(\frac{\delta}{2} - q_4(0)\right) + \int_0^{t^*} \varepsilon\, dt \\
&< \frac{\delta}{2} - \left(q_4(0) + \int_0^{t^*} F_4(\mathbf{q}(t))\, dt\right) \\
&= \frac{\delta}{2} - q_4(t^*).
\end{aligned}$$

Thus, must be true that $q_4(t^*) \notin (0, \delta/2)$. Since q_4 is monotone decreasing, this implies $q_4(t^*) \leq 0$, so that $\mathbf{q}$ first exits W where $q_4 = 0$.

Once $q_4 = 0$, we have that $s_Y(\mathbf{q}) = P$, since this causes the state DD to be absorbing. This holds for all values of q_i, so the gradient of s_Y is zero in these coordinates. Since ∇s_Y is negative, we have that $\mathbf{F}$ is also zero in the fourth coordinate, so that $\mathbf{q}(t^*)$ is an equilibrium point. Thus, $\mathbf{q}(t) = \mathbf{q}(t^*)$ for all $t \geq t^*$.

□

Ultimately, this shows that W is a region where extortion is not an effective strategy for X. The hope, when X chooses an extortionate strategy, is for Y to cooperate, so that X can receive her maximum possible score. However, in this case, Y will evolve to choose a strategy where $q_4 = 0$, so that X actually recieves the lowest possible score.

The fact that $2P > S + T$ is a necessary condition for this proof suggests that when $2P < S + T$, initial strategies near the origin will evolve toward cooperation, where s_Y is maximized. This is the conjecture that Press and Dyson had. The case where $\mathbf{q}_0$ was near the origin was tractable since the formula for ∇s_Y at this point was relatively simple. In most places, the gradient is a complicated function which is difficult to even find useful bounds for. A good path for future research would be to try to prove the conjecture of Press and Dyson, and in the case where $2P > S + T$, to find the precise regions where Y evolves towards rejecting extortion.

Bibliography

Axelrod, Robert. 1984. *The Evolution of Cooperation*. Basic Books, Inc.

Bohning, Daryl, Jeffrey Lorberbaum, Ananda Shastri, and Lauren Sine. 2002. Are there really no evolutionarily stable strategies in the iterated prisoner's dilemma? *Journal of Theoretical Biology* 214:155–169.

Dyson, Freeman J, and William H. Press. 2012. Iterated prisoner's dilemma contains strategies that dominate every evolutionary opponent. *PNAS* URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3387070/.

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

Tijms, Hank C. 2003. *A First Course in Stochastic Models*. Wiley.