Iterated Prisoner's Dilemma

Peter Varshavsky

Version of May 29, 2014

Contents

Chapter 1

Abstract

One of the most famous models in game theory, the Prisoner's Dilemma, is an example of a game in which the Nash equilibrium solution is clearly not beneficial for either player. Regardless of the opponent's move, defection is always the dominant strategy in a single play of the game, and the resulting mutual defection costs both players the higher payoff they would have received from the unstable mutual cooperation play. If, however, the players have a nonzero chance of repeat interaction, turning the game into the Iterated Prisoner's Dilemma, defection may no longer be dominant, and the game allows for a rich field of strategies that utilize any available information about the opponent and vary their behavior between defection and cooperation (which can be viewed as punishment or reward) in the attempt to maximize the player's payoff or manipulate the opponent's payoff. Iterated Prisoner's Dilemma has been applied in ecology, sociology, evolutionary biology, international relations and behavior of firms in the market. This paper will review the history of Iterated Prisoner's Dilemma and its applications and will focus on the recent results by Press and Dyson (2012) showing the existence of a class of Zero Determinant strategies that allow a player to unilaterally determine her opponent's expected payoff, and to use extortion against some types of opponents. It will also examine the relationship between Zero Determinant strategies and the class of countervailing strategies proposed by J.P. Langlois and review the findings of Adami and Hintze, who show that the Zero Determinant strategies are not evolutionarily stable.

Chapter 2

Introduction

The question that inspired the body of research I focus on in this paper is simple to state: in a world of self-interested people (actors?), when does it make sense to cooperate? Real life examples abound. ***(Include examples: Price fixing. Athletes and steroid use. Governments and trade. Cold war.)***

The question extends beyond human behavior as we observe patterns of cooperation everywhere in nature from reciprocal food exchange in vampire bats (expand), to countless examples interspecie cooperation or symbiosis ***(Include examples in Axelrod, p90).***

These examples share a common structure. Each involves two participants making a choice to cooperate (fix prices, say no to steroids, remove trade restrictions, reduce nuclear weapons stocks, share blood with the less fortunate bat) or defect (undercut the competitor, use steroids, establish trade restrictions, increase weapons stockpiles, hoard the blood). Mutual cooperation yields a higher advantage to both parties than mutual defection, but if one individual cooperates, while the other defects, the defector gains a large advantage, while the cooperator is left a sucker.

This scenario was originally formalized in 1950 by Merrill Flood and Melvin Dresher at RAND corporation. Soon after Albert W. Tucker offered this illustrative parable that gave the game its name, the Prisoner's Dilemma.

Description of the prisoner dilemma with payoff table goes here.

At first look the proposition may seem pessimistic. When the game is played once by players who have no possibility of future interaction, the only rational move is to defect even though mutual cooperation would lead to a higher payoff for both players. To see this we assume the opponent's move is fixed and analyze our choices. ***Describe the solution. Motivate Nash Equilibrium.***

Definition 2.1. Nash equilibrium.

The game becomes more interesting—and less gloomy—with the possibility of repeat interaction, for it gives players a chance to reward cooperation and exact vengeance against defectors. When choosing a strategy, one must no longer just maximize the current payoff, but also consider the shadow of the future. Whereas in the singleton game the pure strategy space had only two points, the repeated game allows for great complexity and leads to some unexpected results that give insights about many problems in sociology, evolutionary biology, and political science.

The aim of this paper is to survey the history of the Iterated Prisoner's Dilemma focusing on four distinct approaches to its solution and their insights and consequences. In part one it will define and discuss the notion of evolutionary stability (ESS), due to John Maynard Smith, that paved the way for game theory applications to evolutionary biology and ecology. Then I will discuss the famous IPD tournaments hosted by Robert Axelrod and his conclusions about qualities shared by successful entrants. Part three will present an approach to modeling evolutionary systems using differential equations known as replicator dynamics. Finally part four will describe a class of Zero Determinant (ZD) strategies discovered by Press and Dyson in 2012 and discuss ZD strategies' evolutionary stability and role in mixed strategy populations.

Chapter 3

New organization of the paper

We may want to find the 'best' strategy, but then we need to develop some preliminaries in order to understand what it means for a strategy to be 'good'.

Proposition 1, (p 15) in Axelrod is a simple argument that with high enough discount parameter, there is no single best strategy independent of opponent's strategy. ***This uses discounting, so needs to be either introduced in Axelrod section or after discounting is introduced in preliminaries.***

3.1 Preliminaries

Definition 3.1. Game. Normal form game. Extensive form game.

Note that PD is not zero-sum, so the goal is not to beat the opponent but to do well for yourself, which is not the same thing.

Definition 3.2. Pure and mixed strategy.

Definition 3.3. Nash equilibrium.

Empty threats as motivation for subgame perfection.

Definition 3.4. Subgame Perfect Nash Equilibrium.

Definition 3.5. Generic game and equivalence. (This is used in Adami and Hintze, perhaps makes more sense to define it right before using in the ZD section of the paper).

3.2 Axelrod's Tournaments and TFT

this section's text is not the final text, just a roadmap

Axelrod held two rounds of tournaments. The first with 6? submissions, the second with 62. TFT won both tournaments. Axelrod analyzes the success of TFT. In particular he shows two other strategies that could have beat TFT in the first tournament. These 'better' strategies were among entrants in the second tournament, but did not do very well as they were thwarted by other new entrants. It's interesting to see that many strategies that did not end up doing well for themselves were important in creating a fitness landscape for other strategies, that is they had large effect on scores of other strategies, but did not have a strong effect on TFT - an argument for TFT's robustness.

Then Axelrod motivates the evolutionary approach with average payoffs of each tournament (generation) determining reproductive fitness. With the starting set of strategies of the second tournament, TFT prevails, that success is robust to changing the initial distribution of strategies. This provides the segue to ESS section that follows.

Discussion of *collective stability*

This section will motivate the discounted IPD model by showing that discounting is equivalent to a fixed probability of repeat encounter. Introduce TFT and describe its four characteristics that Axelrod concluded were key to its success ***(these are taken from Wikipedia, will find exact words in Axelrod later)***:

- Be nice: cooperate, never be the first to defect.
- Be provocable: return defection for defection, cooperation for cooperation.
- Don't be envious:: be fair with your partner.
- Don't be too clever: or, don't try to be tricky.

Discuss TFT and ESS (TFT is ESS if and only if the discount factor is sufficiently great). TFT does not earn more than opponent. TFT's success is in its ability to do well both against 'nice' and against 'bad' strategies. Discuss how TFT variants did not perform as well as TFT (JOSS - TFT with a probabilistic opportunistic defection vs. TFT example, p 39). Small group of TFT can invade AllD populations. Population can be split into different subpopulations that are TFT to members of the same subpopulation, but AllD to the other subpopulation.

Importance of being nice. JOSS v TFT showed us that opportunistic defection hurt the strategy's score in the tournament. Axelrod shows that TFTT (TIT FOR TWO TATS) would have actually won the tournament if entered. He notes that most submissions were trying to gain advantage from opportunistic defection, but the bigger advantage could be won by instead being more forgiving/nicer.

3.2.1 Examples

- 'Live and let live' cooperation in the trench warfare during WWI. Enabled by the static nature of the war when soldiers faced the same soldiers on the other side for long periods of time. Soldiers and officers ignored the orders of the high command and only fired warning/demonstration shots.
- There is a list of references to papers about arms race, oligopolistic competition and vote trading on (p. 28), and non-empirical papers on meaning of rationality (Luce and Raiffa), choices which affect other people (Schelling), Cooperation without enforcement (Taylor) (p. 29)

3.2.2 Other points and themes in Axelrod

The following quote perhaps better used or paraphrased in the introduction to motivate interest in IPD. "The model of the iterated Prisoner's Dilemma is much less restricted than it may at first appear. Not only can it apply to interactions between two bacteria or interactions between two primates, but it can also apply to the interactions between a colony of bacteria and, say, a primate serving as a host. There is no assumption that payoffs of the two sides are comparable. Provided that the payoffs to each side satisfy the inequalities that define the Prisoner's Dilemma, as given in chapter 1, the results of the analysis will be applicable" (p 95)

The following quote may be used in ESS section or Axelrod section "The chronological story that emerges from this analysis is the following. ALL D is the primeval state and is evolutionarily stable. But cooperation based on reciprocity can gain a foothold through two different mechanisms. First, there can be kinship between mutant strategies, giving the genes of the mutants some stake in each other's success, thereby altering the payoff of the interaction when viewed from the perspective of the gene rather than the individual. A second mechanism to overcome ego tap defection is for the mutant strategies to arrive in a cluster so that they provide a nontrivial proportion of the interactions each has, even if they are so few as to provide a negligible proportion of the interactions which the ALL D individuals have. Then the tournament approach described in chapter 2 demonstrates that once a variety of strategies is present, TIT FOR TAT is an extremely robust one. It does well in a wide range of circumstances and gradually displaces all other strategies in an ecological simulation that contains a great variety of more or less sophisticated decision rules... *Thus cooperation based on reciprocity can get started in a predominantly noncooperative world, can thrive in a variegated environment, and can defend itself once fully established* (emphasis added)." (p. 99)

Mechanisms for mutualism when players can't recognize each other.Continuous contact (hermit crab & sea-anemone, tree and its mycorrhizal fungi); constant meeting place (aquatic cleaner mutualizes in coastal and reef situations (Trivers 1971). (p 101). "Territory can serve this purpose.... Consistent with the theory, such male territorial birds show much more aggressive reactions when the song of an unfamiliar male rather than a neighbor is reproduced nearby" (p. 102)

Perhaps a curious side note. Prosopagnosia and the location of lesions that cause it: "This localization of cause, and specificity of effect, indicates that the recognition of individual faces has been an important enough task for a significant portion of the brain's resources to be devoted to it" (p. 102)

Another side note. There have been papers that applied IPD to model the increase in pathogen activity when the host is weakened (possibly by another pathogen), which could increase the chances that the first pathogen can transmit. (p. 103-104).

Another side note. Cooperation is not necessarily a socially desirable outcome. Examples: corruption is an act of cooperation between politicians and moneyed interests, price fixing is an act of cooperation between companies. (p. 18)

Another side note. "Examples of what is left out by this formal abstraction include the third parties, the problems of implementing a choice, and the uncertainty about what the other player actually did on the preceding move." (p. 19)

3.3 ESS

This section will contain motivation of ESS in terms of stability and resistance to invasion, formal definition using ϵ-neighborhoods. Based on Maynard Smith's article and book *The Logic of Animal Conflict* and *Evolution and the Theory of Games*.

Stability of AllD.

Segue to the next section: the ESS gives us both an intuitive framework and a formal topologic understanding of fitness of strategies. The proliferation of cheap computing power in the 1970s gave us another tool to study strategies – agent-based models.

Chapter 4

Preliminaries

Some preliminary definitions and general results are in order before we immerse ourselves in

4.1 Games and Solution concepts

This may be split up in subsection or just presented as a series of definitions with some motivation.

4.1.1 Games and Iterated Games

Normal and extensive form. Prisoner's dilemma. Repeated games and discounting. Continuous PD (if I discuss countervailing strategies).

Definition 4.1. Game.

Definition 4.2. Normal and extensive form games.

Example 4.1. Prisoner's dilemma, coordination game (or Stag Hunt; appears in Zero Determinant with Tags analysis in Adami and Hintze), Hawk-Dove (may exclude).

Definition 4.3. Generic game and equivalence. (Gintis, p.285)

"Suppose a normal form game is generic in the sense that no two payoffs for the same player are equal. Suppose $A = (a_{ij})$ and $B = (b_{ij})$ are the payoff matrices for Alice and Bob, so the payoff to Alice's strategy s_i against Bob's strategy t_j is a_{ij} for Alice and b_{ij} for Bob. We say that two generic 2×2 games with payoff matrices (A, B) and (C, D) are equivalent if", for all $i, j = 1, 2$

$$a_{ij} > a_{kl} \equiv c_{ij} > c_{kl}$$

and

$$b_{ij} > b_{kl} \equiv d_{ij} > d_{kl}$$

Proposition 4.1. *"Every generic 2×2 game is equivalent to either the prisoner's dilemma, the battle of the sexes, or the hawk-dove." (Gintis, p.286)*

Karl Sigmund in Calculus of Selfishness defines average payoff in addition to the geometric discounting with $\mathbb{P}(\text{another iteration}) = w < 1$ which is useful for the limiting case $w = 1$:

$$\lim_{n\to\infty} \frac{A(0) + \cdots + A(n)}{n+1}$$

where $A(n)$ is the payoff in round n.

4.1.2 Nash equilibrium

Definition 4.4. Nash equilibrium, strict Nash equilibrium.

Theorem 4.2. *Every symmetric game admits a symmetric Nash equilibrium. (Sigmund, Section 2.5)(May leave out)*

4.1.3 Subgame perfection

4.1.4 Markov perfection

If I discuss countervailing strategies.

4.1.5 ESS

May include criticisms of John Maynard Smith's ESS definition and discuss more modern alterations. Alternatively, the criticisms and discussion may be included in the discussion of papers.

Definition 4.5. $x \in \Delta$ is an evolutionary stable strategy (ESS) if for every strategy $y \neq x$ there exists some $\bar{\epsilon}_y \in (0,1)$ such that inequality

$$u[x, \epsilon y + (1-\epsilon)x] > u[y, \epsilon y + (1-\epsilon)x]$$

holds for all $\epsilon \in (0, \bar{\epsilon}_y)$.

Proposition 4.3.

$$\Delta^{ESS} = \{x \in \Delta^{NE} : u(y,y) < u(x,y) \;\; \forall y \in \beta^*(x), y \neq x\}$$

"Also if $(x,x) \in \Theta$ is a strict Nash equilibrium, then x is evolutionarily stable by default—then there are no alternative best replies. This observation has immediate implications concerning the connection between evolutionary stability and social efficiency: Evolutionary stability does not in general imply that average population fitness $u(x,x)$ is maximized." (Weibull, P38).

An example of potential social inefficiency of ESS is the one-shot prisoner's dilemma, whose only NE, and thus only ESS is to always defect.

There are games with no ESS, for example Rock Paper Scissors (Weibull, P.40)

May include results about ESS and trembling hand perfection (Weibull, P42). Evolutionary stability requires behavior that is not only "rational" and "coordinated" in the sense of Nash equilibrium but also "cautious."

Invasion barriers

In the "setting of finite games, evolutionary stability implies that $\bar{\epsilon}_y$ can be taken to be the same for all mutants; that is, an evolutionary stable strategy x has a uniform invasion barrier." (Weibull, P.43)

Proposition 4.4. *$x \in \Delta^{ESS}$ if and only if x has a uniform invasion barrier. (Weibull, P.43).*

4.1.6 Some strategies

TFT, Generous TFT, Pavlov (Win-Stay, Lose-Shift).

4.2 Replicator dynamics

This section will be based either on *Game Theory Evolving* (Gintis, H.), *Evolutionary Game Theory* (Weibull, J.W.), or *The Calculus of Selfishness* (Sigmund, K).

Definition 4.6. A replicator dynamic.

Definition 4.7. Fixed point.

Will possibly have some discussion of imitation dynamics, but for now it looks like imitation is not used in any of the papers I am discussing, so I will most likely leave out imitation.

Theorem 4.5. *In symmetric 2×2 games a population state is asymptotically stable in the replicator dynamics if and only if the corresponding mixed strategy is evolutionarily stable. (Weibull, p.75).*

Theorem 4.6. *Strongly dominated strategies do not survive in a replicator dynamic. (Theorem 12.3, Gintis, p.280). Stated without proof.*

Theorem 4.7. *Weakly dominated strategy cannot achieve unitary probability as $t \to \infty$ in a replicator dynamic. (Theorem 12.4, Gintis, p.281). Stated without proof.*

Theorem 4.8. *Weakly dominated strategy cannot achieve unitary probability as $t \to \infty$ in a replicator dynamic.*

Replicator dynamics and ESS

There is an example of a Nash equilibrium (Gintis, section 10.13) that "cannot be invaded by any pure strategy mutant but can be invaded by an appropriate mixed-strategy mutant. We can show that this Nash equilibrium is unstable under the replicator dynamic. This is why we insisted that the ESS concept be defined in terms of mixed- rather than pure-strategy mutants; an ESS is an asymptotically stable equilibrium only if the concept is so defined." (Gintis, p. 286)

4.3 Agent based modeling

I'm not yet sure I want to discuss this approach in the preliminaries. It seems pretty self-explanatory, so I may just refer to it when citing and discussing results.

4.4 Learning

A short description of learning. Press and Dyson talk about performance of extortionate ZD strategies against an opponent who learns by climbing up his payoff gradient.

Chapter 5

Results

I haven't yet decided if each of the papers I focus on will occupy its own section or if their results will be woven in a more continuous narrative. I may revisit Maynard Smith's ESS here.

5.1 Axelrod, tournaments, TFT

Axelrod's characterization of successful competitors (nice, retaliating, forgiving).

5.2 Press and Dyson: zero determinant strategies

This will be a summary of the paper's arguments along with the proofs of the theorems. The three results are:

Theorem 5.1. *There exists strategies that can unilaterally set the opponent's score or demand and get an extortionate share. Press and Dyson note that the ability to unilaterally set the opponent's score allows the ZD player to simulate an arbitrary fitness landscape for the evolutionary opponent. They also discuss what happens when extortionate ZD player plays against an evolutionary opponent.*

Theorem 5.2. *Shortest memory player sets the rules of the game.*

(Note from Adami & Hintze: this is true in a head-to-head competition, however in evolutionary setting longer memory may offer an advantage as the strategy would be able to recognize an opponent playing the same strategy, which would enable the long-memory player to conditionally cooperate, a conditionally cooperating ZD can be ESS).

Theorem 5.3. *ZD strategies succeed without Markov equilibrium. Press and Dyson used the game's stationary distribution to derive their ZD strategies. This result shows that the opponent cannot "'keep the game out of Markov equilibrium" or play "inside the Markov equilibration time scale."'*

There is another proof of existence of ZD strategies in (Hilbe, Nowak, Sigmund, 2013), but I will most likely use the Press and Dyson proof.

5.2.1 Stewart and Plotkin comments on Press and Dyson

Steward and Plotkin ran an Axelrod-style tournament using the usual set of strategies and two additional ZD strategies Extort-2 ($S_X - P = 2(S_Y - P)$) and Zero Determinant Generous Tit For Tat, ZDGTFT-2 ($S_X - R = 2(S_Y - R)$) and found that Extort-2 had the second number of head-to-head wins (the winningest strategy, of course, was the Pyrrhic AllD), while ZDGTFT-2 achieved the highest score.

5.3 Evolutionary stability of ZD

This section will be based on two papers. Their abstracts are.

5.3.1 Adami and Hintze, 2013

> "Here we show that ZD strategies are at most weakly dominant, are not evolutionarily stable, and will instead evolve into less coercive strategies. We show that ZD strategies with an informational advantage over other players that allows them to recognize each other can be evolutionarily stable (and able to exploit other players). However, such an advantage is bound to be short-lived as opposing strategies evolve to counteract the recognition".

Adami and Hintze define a new game in which players chose a role in $\{ZD, O\}$ (I think this is for ZD that are not extortionate), and the payoffs are the long-term average payoffs when playing an IPD with chosen roles. They then analyze this new game using replicator equations. In particular they look at two-strategy populations {AllD, ZD} and {Pavlov, ZD}. ZD loses to AllD in evolutionary contest. Pavlov, $q_{PAV} = (1, 0, 0, 1)$, cooperates with itself, is ESS, and loses to ZD in every direct competition, however Pavlov drives ZD to extinction (using replicator equations).

Then they acknowledge that the reformulation of the game may not perfectly reflect the dynamics that happen when the agents in the population play random one-shot games against each other. To address that, they created agent-based simulations and found that

5.3.2 Hilbe, Nowak, Sigmund, 2013

> "Here, were analyze the evolutionary performance of this new class of strategies. We show that in reasonably large populations, they can act as catalysts for the evolution of cooperation, similar to TFT, but that they are not the stable outcome of natural selection. In very small populations, however, extortioners hold their ground. Extortion strategies do particularly well in coevolutionary arms races between two distinct populations. Significantly, they benefit the population that evolves at the slower rate, an example of the so-called "Red King" effect. This may affect the evolution of interactions between host species and their endosymbionts".

5.4 Countervailing

Chapter 6

Other results and discussions

Some of these may make more sense to include in the above discussions than give them their own section. Some may just not be a good fit for this paper.

6.1 An empirical approach to PD

From Lave, L.B. 1962. There is evidence that people do not play the Nash Equilibrium even in iterated prisoner's dilemma with a fixed number of plays. In this setting . These results could be added to the solution concepts chapter.

6.2 Conclusion

From *The Selfish Gene, 3rd Ed, p.75*:

> This theoretical conclusion is not far from what actually happens in most wild animals. We have in a sense explained the 'gloved fist' aspect of animal aggression. Of core the details depend on the exact numbers of 'points' awarded for winning, being injured, wasting time, and so on. In elephant seals the prize for winning meat be near-monopoly rights over a large harem of females. The pay-off for winning must therefore be rated very high. Small wonder that fights are vicious and the probability of serious injury is also high. The cost of wasting time should presumably be regarded as small in comparison with the cost of being injured and the benefit of winning. For a small bird in a cold climate, on the other hand, the cost of wasting grime may be paramount. A great tit when feeding nestlings needs to catch an average of one prey per thirty seconds. Every second of daylight is precious. Even the comparatively short time wasted in a hawk/hawk fight should perhaps be regarded as more serious than the risk of injury to such a bird. Unfortunately we know too little at present to assign realistic numbers to the costs and benefits of various outcomes in nature*...

6.3 Notation

Notation from various sources. Once the content of the paper is clear, I will make sure notation is consistent.

Δ strategy space (Weibull)
Θ strategy profile (Weibull)