Iterated Prisoner's Dilemma

Peter Varshavsky

Version of June 26, 2014

Contents

Chapter 1

Cooperation

1.0.1 Introduction

There will be either one or two chapters with introduction and preliminaries defining the game and some of the concepts that are used here and giving some examples from biology and political science to motivate the problem.

1.0.2 Theorems

There are a number of theorems from Axelrod stated without proof. I could add proofs or just mention their statements in the body of text.

1.0.3 Draft of chapter

According to Robert Axelrod, there was a rich literature exploring the Prisoner's Dilemma by the late 1970s, but most of it was focused on analyzing situations that shared the game's structure in various social sciences rather than finding the best way to play the game [1, p.28].

Looking to understand how various strategies may perform against each other, he invited fourteen game theorists who specialized in disciplines of psychology, economics, political science, mathematics, and sociology to participate in a computer tournament. Each participant submitted a strategy (player) that described how to play the 200-move iterated Prisoner's Dilemma given the current position and history of prior interaction. In addition Axelrod added a player called RANDOM, which played a mixed strategy of $(0.5, 0.5)$ *(comment: change to (p_1, p_2, p_3, p_4) notation and define that notation in preliminaries)* on every move. Each of the fifteen strategies was set to play each other strategy and itself, and the tournament was repeated five times to get more stable estimates of scores for each pair of players.

The simplest submission, entered by Professor Anatol Rapoport, University of Toronto, went on to win the tournament. The strategy, called TIT FOR TAT (TFT) started by cooperating on the first move and on all other moves simply repeated its opponents previous move.

Axelrod circulated the results and solicited entries for the second round of the tournament, this time receiving sixty two entries from professionals and hobbyists. In the first tournament many strategies, knowing that the game only had 200 iterations attempted to use end-of-game tactics to boost their score or to avoid being taken advantage of in the final moves of the game. To avoid this end-of-game opportunism, Axelrod introduced the probability of repeat interaction of $w = 0.99654$, so with probability 0.00346 any iteration could be the last one of the game, and the median match length remained 200 moves. Once again Rapoport submitted TIT FOR TAT, and once again it won.

The unexpected success of TIT FOR TAT, which can never receive a score higher than its opponent in a single match, was subject of a series of papers and a book *The Evolution of Cooperation* by Axelrod in which he discusses robustness and possible real world examples of the strategy. He suggested that the following four characteristics made TIT FOR TAT successful.

1. *Don't be envious*
   TFT is not trying to beat its opponents, instead it is designed to do well for itself when playing against a wide range of strategies

2. *Don't be first to defect*
   The tournaments showed that strategies that were nice, that is that did not defect unless provoked did considerably better than strategies that used unprovoked defection

3. *Reciprocate cooperation and defection*
   A strategy that only reciprocates defection, like the grim trigger, gives up the possibility of letting bygones be bygones and returning to mutual cooperation. A strategy that only reciprocates cooperation is easily exploitable by unprovoked defections.

4. *Don't be too clever*
   Rules that constructed complex inferences about their opponents did not do very well likely because the advantages they received from correct inferences did not outweigh the losses due to incorrect ones

Many of these characteristics were shared by other strategies that did well in the tournaments. In fact submissions that tried to get ahead by using unprovoked defections generally did worse than the players that erred on a more forgiving side. In his analysis of the first tournament, Axelrod notes that TIT FOR TAT did not have to win. One strategy that would have done better is the more forgiving TIT FOR TWO TATS (TF2T), which only retaliates after two consecutive defections by its opponent. TF2T, however, is not better than TFT in any absolute way—it was submitted in the second tournament by John Maynard Smith and only finished in 24th place. The poor repeat performance of TF2T was due to it being exploited by strategies that tested how much defection they could get away with. It is a good illustration of the principle that there can not be a single best strategy independent of the population of other strategies: although these testing strategies spelled doom for TF2T, they themselves did not do well in the end [1, p.47].

To further study the robustness of TIT FOR TAT, Axelrod turned to the approach of evolutionary or ecological stability developed by John Maynard Smith in which the average payoff that a strategy receives in a given population determines its reproductive fitness. Each generation of strategies plays a round robin tournament similar to round two described above, and relative frequency of strategies in the next generation is determined by relative frequencies and average payoffs in the current generation. The chronology of this ecological computational experiment is interesting. In the early generations strategies that are too nice are exploited by meaner strategies, which leads to proliferation of mean strategies and near extinction of too forgiving ones. The mean strategies are usually not very good at playing against other mean strategies, and as their numbers increase while the prey becomes sparse, their fitness suffers, and they eventually lose prominence. TFT, which receives the payoff of $R = 3$ against itself and other nice strategies, and which does not allow exploitation by mean strategies, once again proved to be very robust in this ecological setting and came out on top in every simulation [1, p.53].

The idea that to remain ecologically or evolutionarily fit a strategy needs to perform well against opponents, but also against itself was first formalized by John Maynard Smith in his 1973 paper "The Logic of Animal Conflict" [6] *(comment: second time I bring up Maynard Smith, reads a little repetitive)*. He defined a strategy I to be an *evolutionarily stable strategy (ESS)* if a large population employing I could not be invaded by a mutant playing some other strategy J, or more precisely if $W(J, I)$ is the payoff to J in a contest with I, and $W(I, I)$ is the payoff to I playing against itself, then

$$\begin{cases} W(J,I) < W(I,I) \\ \text{ or} \\ W(J,I) = W(I,I) \text{ and } W(I,J) > W(J,J) \end{cases} \tag{1.1}$$

In his 1981 paper "The Emergence of Cooperation among Egoists" [2], Axelrod defined a slight modification of ESS restricting its inequality. He called a strategy *collectively stable (CS)* if $W(J, I) < W(I, I)$ *(comment: there's some confusion here – CS is technically not a restriction of ESS, need to review definitions)* for any mutant invading strategy J in a population of strategies I. This restriction allowed him to provide a characterization theorem for this class of strategies. The underlying idea for the characterization is that the incumbent can prevent the newcomer from invading if no matter what the newcomer does, the incumbent can keep his score sufficiently low. Axelrod introduces the definition: B has a *secure position* over A on move n if no matter what A does from move n onwards, $W(A|B) \leq W(B|B)$, assuming that B defects from move n onwards. If $V_n(A|B)$ denotes A's discounted cumulative score in the moves before move n, then we can say that B has a secure position over A on move n if

$$V_n(A|B) + w^{n-1}P/(1-w) \leq V(B|B)$$

or equivalently

$$V_n(A|B) \leq V(B|B) - w^{n-1}P/(1-w) \tag{1.2}$$

thus the characterization theorem, which is a version of the folk theorem of repeated games, "embodies the advice that if you want to employ a collectively stable strategy, you should only cooperate when you can afford an exploitation by the other side and still retain your secure position" [2, p.313].

Theorem 1.1. ***The characterization theorem.*** *B is a collectively stable strategy if and only if B defects on move n whenever the other player's cumulative score so far is too great, specifically when*

$$W_n(A|B) > V(B|B) - w^{n-1}(T + wP/(1-w)). \tag{1.3}$$

Proof. The converse implication uses induction. Assume B is a strategy that defects as required by 1.3. To have a secure position over A on move $n = 1$, B is required to defect on every move, that is B is the $ALLD$ strategy, and $V(A|ALLD) \leq V(ALLD|ALLD)$ for any strategy A, so B has a secure position over A on the first move.

If B has a secure position over A on move n, we need to show that it will maintain its secure position on move $n+1$. First, if B defects on move n, A gets at most P, so

$$V_{n+1}(A|B) \leq V_n(A|B) + w^{n-1}P$$

Using 1.2, we get

$$\begin{aligned} V_{n+1}(A|B) &\leq V(B|B) - w^{n-1}P/(1-w) + w^{n-1}P \\ &\leq V(B|B) - w^n P/(1-w) \end{aligned}$$

B is only allowed to cooperate on move n if B can afford to be exploited by A, that is

$$V_n(A|B) \leq V(B|B) - w^{n-1}(T + wP/(1-w)).$$

Since A can get at most T on the n'th move,

$$\begin{aligned} V_{n+1}(A|B) &\leq V(B|B) - w^{n-1}(T + wP/(1-w)) + w^{n-1}T \\ &= V(B|B) - w^n P/(1-w) \end{aligned}$$

thus if B follows the prescription of the theorem, it will preserve its secure position over A and will be collectively stable.

The forward implication is proved by contradiction. Suppose that B is collectively stable, and there is an A and an n such that B does not defect on move n when

$$V_n(A|B) > V(B|B) - w^{n-1}(T + wP/(1-w))$$

or, equivalently, when

$$V_n(A|B) + w^{n-1}(T + wP/(1-w)) > V(B|B). \tag{1.4}$$

Now define A' to be the same as A on the first $n-1$ moves, and from move n onward. Since B cooperates on move n, A' gets T, and at least P on the following moves. So

$$V(A'|B) \geq V_n(A|B) + w^{n-1}(T + wP/(1-w)).$$

combined with 1.4 this yields $V(A'|B) > V(B|B)$, so A' invades B, and hence B is not collectively stable. □

Axelrod proves a number of other results supporting his conviction that TIT FOR TAT is good at playing the game, but acknowledging that there cannot be a single best strategy.

Theorem 1.2. *If the discount parameter w is sufficiently high, there is no best strategy independent of the strategy used by the other player [2, Theorem 1].*

Theorem 1.3. *TIT FOR TAT is a collectively stable strategy if and only if*

$$w \geq \max\left(\frac{T-R}{T-P}, \frac{T-R}{R-S}\right)$$

or, alternatively, TIT FOR TAT is a collectively stable strategy if and only if it is invadeable neither by ALLD nor the strategy which alternates defection and cooperation [2, Theorem 2].

Theorem 1.4. *For a nice strategy to be collectively stable, it must be provoked by the first defection of the other player [2, Theorem 4]*

Theorem 1.5. *Any rule, B which may be the first to cooperate is collectively stable only when w is sufficiently large. [2, Theorem 5]*

Theorem 1.6. *ALLD is always collectively stable [2, Theorem 6]*

Clustering and emergence of cooperation

Axelrod's tournaments and Maynard Smith's ideas of evolutionary stability have given a rather positive answer on rationality of cooperation among self-interested individuals, but it seems at odds with the stability of ALLD (Theorem 1.8). If we can assume that at some point cooperation did not exist, then the stability of ALLD implies that cooperation would never emerge from individual mutations, since cooperating individuals would not be able to invade a population of defectors.

This pessimistic prediction changes if we allow cooperators to arrive in clusters to the population of defectors. Clusters of cooperation are likely to develop along close kin or individuals occupying the same or neighboring territories. Richard Dawkins, in his best seller *The Slefish Gene*[4], describes how such kin-based cooperation or altruism can be modeled and understood if we focus not on survival and procreation of individuals in a population, but of units of genetic information.

To model invasion by a cluster, we still assume that the majority of the population is using the strategy B, but now the newcomers A are arriving in sufficient numbers and sufficiently near each other so that the probability that A interacts with A is $p \in (0,1)$, and the probability that A interacts with B is $1-p$. Then the expected score for a newcomer is $pV(A|A) + (1-p)V(A|B)$, and for the incumbent $V(B|B)$. So the p-cluster of newcomers A will invade B if $pV(A|A) + (1-p)V(A|B) > V(B|B)$, or solving for p,

$$p > \frac{V(B|B) - V(A|B)}{V(A|A) - V(A|B)}.$$

We can now calculate the size of the cluster necessary for TIT FOR TAT to invade ALLD. Given our usual values of T, R, P, S, and $w = 0.9$, we get $p > 1/21$, that is if TIT FOR TAT newcomers have just a five percent chance of interactions with other TIT FOR TAT players, they can successfully invade a world of noncooperative players. As w increases, the minimum value of p necessary for TIT FOR TAT to invade is reduced even lower [2].

(Note: May leave the next two theorems out)

Theorem 1.7. *The strategies which can invade ALLD in a cluster with the smallest value of p are those which are maximally discriminating, such as TIT FOR TAT [2, Theorem 7]*

Theorem 1.8. *If a nice strategy cannot be invaded by a single individual, it cannot be invaded by any cluster of individuals either [2, Theorem 8]*

Comparing ESS, collective stability, and Nash equilibrium

So far this paper described several solution concepts all based on the idea that in an equilibrium it does not pay to unilaterally change behavior: Nash equilibrium, ESS, and collective stability. We denote the set of strategies that are in Nash equilibrium with themselves, also called a *symmetric Nash equilibrium,* Δ^{NE}, the set of ESS strategies Δ^{ESS}, and the set of collectively stable strategies Δ^{CS}. It follows immediately from definitions that $\Delta^{NE} = \Delta^{CS}$. The relationship between Δ^{ESS} and Δ^{CS} is more interesting. Our definition of ESS (Equation 1.1) is equivalent to

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

where $\beta^*(x) = \{y \in \Delta : u(y, x) \geq u(y', x) \ \ \forall y' \in \Delta\}$ is the set of best replies to a strategy x. In particular $\Delta^{ESS} \subset \Delta^{CS}$, so ESS is a more demanding requirement than collective stability. Axelrod motivated his relaxation of ESS to CS as a way to simplify the proofs of his propositions, and claimed that, except for the characterization theorem (Theorem 1.1), they hold for ESS strategies as well. However in a 1987 paper "No pure strategy is evolutionarily stable in the repeated Prisoner's Dilemma game" [3], R. Boyd and J.P. Lorberbaum showed that neither TFT nor any other strategy whose behavior on a given move is fully determined by the history of prior moves is evolutionarily stable. In his arguments Axelrod considers invasion of a pure strategy population by a single newcomer strategy, and he only considers ecologic dynamics, that is the dynamics of relative frequencies of a set of given strategies not altered by mutation. If the possibility of mutation is introduced, however, Boyd and Lorberbaum prove that collectively stable strategies can succumb to invasion by pairs of other strategies.

Theorem 1.9. *No strategy whose behavior during interaction t is uniquely determined by the history of the game up to that point is evolutionarily stable if*

$$w > \min\left[\frac{T-R}{T-P}, \frac{P-S}{R-S}\right]$$

Proof. ***(The proof is quoted almost verbatim from [3])*** Let S_e be a collectively stable strategy and let S_1 be a distinct strategy that behaves exactly the same way with S_e on each interaction with S_e as S_e does against itself. This implies that $V(S_1|S_e) = V(S_e|S_e) = V(S_e|S_1) = V(S_1|S_1)$. If a third strategy S_x exists in the population, S_e can be invaded by S_1 if $V(S_1|S_x) > V(S_e|S_x)$. Because S_e and S_1 are distinct, there must be some history h such that S_e and S_1 react differently to S_x for the first time on move t ***(technically it's not necessary for such history*** h ***to exist, but let's assume*** S_x ***is such that it makes*** S_1 ***react to it differently from*** S_e ***at some point*** t***).*** There are two possibilities. First, suppose S_e defects and S_1 cooperates on move t. Let S_2 be the strategy that generates the history h in response to both S_e and S_1, cooperates on move t and then defects forever in response to defection by S_e and cooperates forever in response to cooperation by S_1 on move t. S_e can be invaded by S_1 whenever

$$V(S_1|S_2) - V(S_e|S_2) \geq w^{t-1}(R-T) + w^t \frac{R-P}{1-w} > 0$$

or $w > (T-R)/(T-P)$. Next, let S_3 be a strategy which behaves exactly like S_2 for the first $t-1$ moves, defects on move t, and then defects forever in response to S_e's defection and cooperates forever in response to cooperation by S_1 on move t. In this case S_e can be invaded by S_1 whenever $w > (P-S)/(R-S)$. The second possibility is that S_e cooperates and that S_1 defects on move t. A similar argument shows that S_e can be invaded by S_1 for any value of w. □

Boyd and Lorberbaum note that this theorem agrees with Axelrod's insight that no strategy is strictly a best strategy regardless of the remaining population. They are careful to not imply that their theorem disproves the robustness of nice, provokable and forgiving strategies, but instead shows the existence of scenarios where such strategies, robust as they are, can still be invaded and displaced. They present the following scenario in which TIT FOR TAT is invaded by a combination of the more forgiving TIT FOR TWO TATS (TF2T), and SUSPICIOUS TIT FOR TAT (STFT) which defects on the first interaction and then plays TIT FOR TAT for the remainder of the game. In their example STFT is assumed to be maintained in the population by one-way mutation or phenotypic variation. Because $V(TF2T|TFT) = V(TFT|TFT)$ and since the forgiving $TF2T$ can induce $STFT$ to cooperate, while TFT and $STFT$ end up locked in a string of alternating defections, $V(TF2T|STFT) > V(TFT|STFT)$ when w is large enough, so TF2T can invade a predominantly TFT population if it also contains some STFT individuals.

Win-Stay, Lose-Shift

One of the main features that makes TFT so attractive for studying emergence of cooperation is its simplicity. It does not take a great leap of faith to imagine reciprocity-based behaviors emerge in human society or the animal kingdom. TFT does not require a long memory or sophisticated optimization calculations, it does not require that the individuals are from the same tribe or species, and that the payoffs are equal or even measured in the same units for the players involved. Whether or not it precisely described the real evolution of cooperation (which it probably did not) is not relevant. Its success lies in giving a simple and constructive proof that a simple system that starts as a population of defectors can develop into a cooperating system. But how unique is TFT in those regards?

In 1993 Martin Nowak and Karl Sigmund published a paper in Nature [8] that proposed that another simple and realistic rule called WIN-STAY, LOSE-SHIFT could outperform TIT FOR TAT. Dubbed PAVLOV for its reflex-like response to the payoffs, the rule repeats its previous move if it was rewarded by R or T points, but changes from C to D or from D to C if it was punished by receiving only S or P. The discovery of PAVLOV's dominance was accidental. Nowak and Sigmund were studying the performance of various memory one strategies (that is strategies that only use the previous move in determining their next move) in populations where errors were possible, that is there was a nonzero probability that TFT would mistakenly cooperate in response to a defection or vice versa. If we restrict our attention to memory one strategies, iterated prisoner's dilemma can be thought of as a Markov chain with states $\{CC, CD, DC, DD\}$, then a strategy can be described as a probability of that the player plays C given that the game is in any of the above states $\mathbf{p} = (p_1, p_2, p_3, p_4)$. Thus TFT is described by $\mathbf{p}^{TFT} = (1, 0, 1, 0)$, and PAVLOV corresponds to $\mathbf{p}^{PAVLOV} = (1, 0, 0, 1)$. Generous tit for tat GTFT with $\mathbf{p}^{GTFT} = (1, \gamma, 1, \gamma)$ is a variant of TFT that cooperates with some probability γ after the opponent's defection. Possibility of mistakes can be modeled by changing all unit probabilities in these representations to $1 - \epsilon$ and all null probabilities to ϵ, where ϵ is the probability of error.

When the possibility of errors is added to the model pairs of TFT run the risk of getting stuck in lengthy periods of alternating defection following a mistaken defection by one of the players. The generous morph GTFT has the advantage of being able to break these vicious cycles sooner than TFT, which increases its fitness in the new model. In fact Nowak and Sigmund were expecting a version of GTFT to come be the winner of their simulations, but in the very long run PAVLOV showed to be more fit, likely owing to its ability to take advantage of unconditional or very forgiving cooperators, while playing well against itself. If, due to his own mistaken defection, PAVLOV discovers that it can get away with defecting unilaterally, it will continue doing so until it either switches to C due to another error or the opponent's retaliation. Unlike a pair of TFTs, a pair of PAVLOV players do not get stuck in long loops of alternating defections following the mistake by one of the players—it only takes two moves to return to the state of mutual cooperation.

Nowak and Sigmund started their simulation with a population of random strategies $\mathbf{p}^{\text{RAND}} = (0.5, 0.5, 0.5, 0.5)$. To allow for errors they restricted $0.001 < p_i < 0.999$ for all i for all strategies. On average every 100 generations they introduced a randomly generated mutant strategy. The simulations were consistent with Axelrod's predictions that TFT mutants can invade predominantly ALLD populations, and with Boyd and Lorberbaum's findings that TFT can be invaded by mixtures of generous and suspicious newcomers. While the chronology of the simulations varied greatly, Nowak and Sigmund point to several interesting features that were consistent. Most of the time was spent in states of various equilibria lasting from tens of thousands to millions of generations. The transitions between equilibria were rare but fast, often taking just a few generations. Cooperations was observed in only 27.5% of the runs after $t = 10^4$ generations, but in 90% at $t = 10^7$. PAVLOV dominated 10% of the runs at $t = 10^4$ and 82.5% at $t = 10^7$. The periods spent in cooperating equilibria increased in length with time, but the threat of breakdown and reversion to defecting states never completely abated.

Conclusion of Part 1

This concludes the first part of the paper. It focused on emergence of cooperation mainly through the lens of Robert Axelrod tournaments. It compared the solution concepts of ESS by Maynard Smith and collective stability by Axelrod. Axelrod's approach was mainly ecological, meaning mutations were not allowed in the model. Once we allow mutations (which is the evolutionary approach), some of Axelrod's findings are no longer valid. His insight, however, can still be useful.

It is interesting that the strategies that do best in these tournaments are some of the simplest rules: TFT and WIN-STAY-LOSE-SHIFT. This simplicity makes it believable that similar processes could have actually been present in evolution of species, ecosystems, and behaviors.

Notes on plan for Part 2

The second body of work that I planned to use focuses on strategies that attempt to manipulate their opponents. The tentative plan for this part is to start with countervailing strategies, then go on to zero determinant strategies, explore what is the relationship between zero determinant and countervailing strategies, and then note that although zero determinant strategies are not ESS, they can perform interesting roles as catalysts of equilibrium shifts (like TFT was in Nowak and Sigmund's simulation). TFT and PAVLOV are both ZD. I will introduce evolutionary dynamics briefly only to be able to quote some results from papers that discuss stability of ZD strategies.

Bibliography

[1] R. Axelrod. *The evolution of cooperation*. Political Science / Science. Basic Books, 1984.

[2] Robert Axelrod. The emergence of cooperation among egoists. *The American Political Science Review*, 1981.

[3] Robert Boyd and Jeffrey P. Lorberbaum. No pure strategy is evolutionarily stable in the repeated Prisoner's Dilemma game. *Nature*, 327(6117):58–59, May 1987.

[4] R. Dawkins. *The Selfish Gene: 30th Anniversary Edition*. ISSR library. OUP Oxford, 2006.

[5] Menusch Khadjavi and Andreas Lange. Prisoners and their dilemma. *Journal of Economic Behavior and Organization*, 92(0):163 – 175, 2013.

[6] J. Maynard Smith and G. R. Price. The logic of animal conflict. *Nature*, 246(5427):15–18, 1973.

[7] M.A. Nowak. *Evolutionary Dynamics*. Harvard University Press, 2006.

[8] Martin Nowak and Karl Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364(6432):56–58, 1993.