

Iterated Prisoner's Dilemma

Peter Varshavsky

Version of July 8, 2014

Contents

Chapter 1

Manipulation

Much of the innovation in the history of IPD came from various refinements of the Nash equilibrium. Reframing the equilibrium in terms of ecology led to Axelrod's idea of collective stability and his analysis of the success of TIT FOR TAT. An evolutionary refinement of by Maynard Smith allowed the introduction of mutations and a stability robust to them. It exposed a weakness of TFT—it could be invaded by a pair of strategies. To account for possibility of error or miscommunication, Richard Selten proposed the trembling hand perfect equilibrium, and simulations showed that a simple strategy WSLs performed well in this error-prone environment. Not all successful strategies have the simple elegance of TFT and WSLs, however. In 2010 Iliopoulos, Hintze, and Adami showed that stability of strategies depended on the mutation rate, and analyzed a strategy they called GENERAL COOPERATOR $\mathbf{q}_{GC} = (0.935, 0.229, 0.266, 0.42)$, which was the evolutionary fixed point at low mutation rates [4].

In 2012 a new class of strategies was born out of yet another approach to solving the dilemma. If the IPD is restricted to memory-one strategies, which can only use the previous state of the game in each decision, it can be thought of as a Markov process with states CC, CD, DC, DD. Since the strategies can be described as vectors of probabilities the player will cooperate given each of the four states of the game

$$\begin{aligned}\mathbf{p} &= (p_{CC}, p_{CD}, p_{DC}, p_{DD}) \\ \mathbf{q} &= (q_{CC}, q_{DC}, q_{CD}, q_{DD})\end{aligned}$$

where $p_2 = \mathbb{P}(\text{player 1 cooperates} | \text{CD})$. Then the Markov transition probabilities are

$$\begin{bmatrix} p_1 q_1 & p_1(1 - q_1) & (1 - p_1)q_1 & (1 - p_1)(1 - q_1) \\ p_2 q_3 & p_2(1 - q_3) & (1 - p_2)q_3 & (1 - p_2)(1 - q_3) \\ p_3 q_2 & p_3(1 - q_2) & (1 - p_3)q_2 & (1 - p_3)(1 - q_2) \\ p_4 q_4 & p_4(1 - q_4) & (1 - p_4)q_4 & (1 - p_4)(1 - q_4) \end{bmatrix}.$$

If \mathbf{v} is the stationary distribution of the above chain, then we can define the long term average payoffs to be

$$\begin{aligned}S_X &= \mathbf{v} \cdot \mathbf{S}_X = \mathbf{v} \cdot (\text{R}, \text{S}, \text{T}, \text{P}) \\ S_Y &= \mathbf{v} \cdot \mathbf{S}_Y = \mathbf{v} \cdot (\text{R}, \text{T}, \text{S}, \text{P}).\end{aligned}$$

Press and Dyson's discovery was that the dot product of the stationary distribution \mathbf{v} with an arbitrary 4-vector \mathbf{f} is given by the following determinant

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

in which the second column is $\tilde{\mathbf{p}} \equiv (-1 + p_1, -1 + p_2, p_3, p_4)$ is entirely controlled by the first player, and the third column $\tilde{\mathbf{q}} \equiv (-1 + q_1, -p_3, -1 + q_2, q_4)$ is entirely controlled by the second player. This allows each player to unilaterally make the determinant vanish by setting their column to be a scalar multiple of \mathbf{f} . The opportunity for roguery arises when we consider $\mathbf{f} = \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1}$. Since $\mathbf{v} \cdot (\alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1}) = \alpha S_X + \beta S_Y + \gamma$, each player can unilaterally enforce a linear relation between long the players' long term average payoffs by setting determinant to zero. The strategies that enforce this linear relationship were christened *zero determinant (ZD)* strategies.

Press and Dyson focused on two types of ZD strategies. The first allows X to unilaterally set her opponent's score by setting $\alpha = 0$ and thus forcing $\beta S_Y + \gamma = 0$. Solving the equations

$$\begin{aligned} -1 + p_1 &= \beta R + \gamma \\ -1 + p_2 &= \beta T + \gamma \\ p_3 &= \beta S + \gamma \\ p_4 &= \beta P + \gamma \end{aligned}$$

to eliminate the parameters β and γ and expressing p_2 and p_3 as functions of p_1 and p_4 , they arrive at the strategy

$$\begin{aligned} p_2 &= \frac{p_1(T - P) - (1 + p_4)(T - R)}{R - P} \\ p_3 &= \frac{(1 - p_1)(P - S) + p_4(R - S)}{R - P} \end{aligned}$$

which sets Y's score to a weighted average of P and R

$$S_Y = \frac{(1 - p_1)P + p_4 R}{(1 - p_1) + p_4}.$$

This equation has feasible solutions when p_1 is close to but not equal to 1, and p_4 is close to but not equal 0.

Using a similar calculation Press and Dyson show that X cannot set her own score with $\tilde{\mathbf{p}} = \alpha \mathbf{S}_X + \gamma \mathbf{1}$, as

$$p_2 = \frac{(1 + p_4)(R - S) - p_1(P - S)}{R - P} \geq 1$$

except for $\mathbf{p} = (1, 1, 0, 0)$.

The second type of ZD strategies described by Press and Dyson is even more devilish. It allows X to demand and get an extortionate share of the payoff surplus over the mutual defection value P by setting

$$\tilde{\mathbf{p}} = \phi[(\mathbf{S}_X - P\mathbf{1}) - \chi(\mathbf{S}_Y - P\mathbf{1})]$$

which leads to

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

with feasible strategies existing for any χ and

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

. Computing the long-run average scores for X and Y with using Axelrod's values $(T, R, P, S) = (5, 3, 1, 0)$ Press and Dyson show

$$S_X = \frac{2 + 13\chi}{2 + 3\chi} > 3 = P, \text{ and } S_Y = \frac{12 + 3\chi}{2 + 3\chi} < 3 = P$$

for $\chi > 1$, while the limiting fair case $\chi = 1$ and $\phi = 1/5$ reduces to TIT FOR TAT with $\mathbf{p}^{\text{TFT}} = (1, 0, 1, 0)$.

When playing against an extortioner X, Y has two choices, the first is to try to maximize his own score, but that would feed into X's plot and increase X's score even more. The second is to refuse to be extorted by playing ALLD either out of spite or in hopes that X recognizes that extortion will not work and switches to a more equitable play. The key to understanding a match between two extortioners is in the $p_4 = q_4 = 0$. Once they reach a mutual defection, they will stay there indefinitely thus receiving a long term average of P . This should raise questions about evolutionary stability of extortionate ZD strategies. In the two years since Press and Dyson published their findings, a number of papers came out discussing how well can ZD strategies fare in ecological and evolutionary settings. Before moving on to those results, however, I should note two more theorems established by Press and Dyson.

The first shows that restricting our attention to memory-one strategies is not as limiting as it may seem. They prove that for any finite memory strategy there exists a memory-one strategy... (*comment: review*). In their 2012? paper Adami and Hintze [1] note that longer memory may provide evolutionary advantage to variants of ZD players who attempt to recognize each other by analyzing longer pattern of play and switch to mutual cooperation in order to increase their evolutionary fitness.

The second ancillary result dispels doubts that may arise about the steady state payoff $S_i = \mathbf{v} \cdot \mathbf{S}_i$, $i \in \{X, Y\}$. It may take hundreds of moves for the linear relationship

$$\alpha S_X + \beta S_Y + \gamma = 0$$

to be established, and one may wonder if there is any way the non-ZD player can prevent that relationship by changing their own play inside the equilibration time scale (*comment:*

find better wording). Press and Dyson answer that question negatively, thus showing that in a one-on-one play the non-ZD player cannot benefit from having longer memory or by keeping the game away from the Markov stationary distribution.

The strategies that we saw perform well in the previous chapter were mainly concerned with their own outcomes. The prescription to not be greedy and to reciprocate was shown to be robust, as was the simple idea of doing what works and not doing what doesn't work (Win-Stay Lose-Shift). In this chapter I will focus on approaches that attempt to manipulate the opponent's payoff. *(comment: This is not a good characterization of strategies in this chapter; some of them are extortionary, others aren't.)*

1.1 Results

IPD contains strategies that dominate any evolutionary opponent, Press, Dyson [5]
(April 19, 2012)

This paper introduced ZD strategies and talked about two types of ZDs. Extortioners (including the fair extortioner TFT), and strategies that unilaterally set opponent's score. The scenario that Press and Dyson describe is an infinite play between two opponents. If one of the opponents knows about ZD and the other attempts to optimize his score, then the ZD opponent can extort. However if both opponents play ZD, there is no surplus to share (this is easy to see since $p_4 = q_4 = 0$ for extortioners).

Stewart and Plotkin [6]
(June 26, 2012)

Stewart and Plotkin plug in two ZD strategies in Axelrod's first tournament in a short follow-up paper [6]. The strategies are Extort-2 with $S_X - P = 2(S_Y - P)$ that guarantees player X twice the share of payoffs above P compared with those received by Y. ZDGTFT-2 forces $S_X - R = 2(S_Y - R)$ offers Y a higher portion. ZDGTFT-2 won by average score. Extort-2 won the highest number of matches (except ALLD).

Evolutionary instability of Zero Determinant strategies demonstrates that winning isn't everything, Adami, Hintze [1]
(August 13, 2012 (v1)) (published in Nature Communications in 2013)

Adami and Hintze study evolutionary stability of two types of ZD strategies defined by Press and Dyson. The first type sets the opponent's score unilaterally by following

$$\mathbf{p} = \left(p_1, \frac{p_1(T - P) - (1 + p_4)(T - R)}{R - P}, \frac{(1 - p_1)(P - S) - p_4(R - S)}{R - P}, p_4 \right)$$

with p_1 close to but not equal 1, p_2 close to but not equal 0.

They define functions f, h, g to be the expected scores in a matches of ZD and some other strategy O against each other and themselves,

$$\begin{aligned} f(\mathbf{p}) &= \mathbb{E}(ZD, ZD) = \mathbb{E}(O, ZD) \\ h(\mathbf{q}) &= \mathbb{E}(O, O) \\ g(\mathbf{p}, \mathbf{q}) &= \mathbb{E}(ZD, O) \end{aligned}$$

and define a new game in which each player choses whether to play ZD or O with the matrix

$$\begin{pmatrix} f(\mathbf{p}) & g(\mathbf{p}, \mathbf{q}) \\ f(\mathbf{p}) & h(\mathbf{q}) \end{pmatrix}.$$

Can subtract a constant from each column (*comment: find reference*)

$$\begin{pmatrix} 0 & g(\mathbf{p}, \mathbf{q}) - h(\mathbf{q}) \\ 0 & 0 \end{pmatrix}$$

so if $g(\mathbf{p}, \mathbf{q}) - h(\mathbf{q}) > 0$ for every O, ZD is weak ESS (*comment: What is Weak ESS?*). A mixture of ZD and O cannot be mixed ESS because ESS enforces the same score on itself as it does on others.

They go on to compare ZD to PAVLOV and show that PAVLOV is the ESS (*comment: Didn't we decide that no pure strategy is ESS in Part 1? Could they be only looking at ESS given no mutations and a fixed starting finite set of strategies?*) They do this using the replicator equations

$$\dot{\pi}_i = \pi_i(w_i - \bar{w})$$

where π_i is the proportion of population using strategy i , w_i is the fitness of strategy i , and \bar{w} is the average fitness in the population. They show that independently of initial distribution of ZD and PAVLOV, over time PAVLOV takes over and ZD goes extinct. Agent-based simulations follow qualitatively identical trajectories (time scale may be different).

Another illustrative example is the performance of ZD against the “general cooperator (GC)” $\mathbf{q}_{GC} = (0.935, 0.229, 0.266, 0.42)$, which Iliopoulos, Adami, and Hintze showed to be an evolutionary fixed point at low mutation rates [4]. $E(Z, GC) = 2.125$ and $E(GC, GC) = 2.11$, so ZD is a weak ESS. How, then the GC is the evolutionary fixed point and not ZD? Adami and Hintze showed that if a mutation rate favoring GC is used in an agent based simulation with the ZD strategy, ZD evolves into GC. Thus ZD is not genetically or mutationally stable in addition to not being ESS.

Extortionate ZD strategies are not ESS due to their poor performance against other extortioners $p_4 = 0$ means that the game gets stuck in DD state following the first mutual defection. ZD_E strategies that can recognize other ZD_E s can be ESS by extorting others but cooperating among themselves. Of course the recognition mechanisms can lead to their opponents evolving methods to behave like ZD_E s and thus subvert their advantage. (*comment: Syrphidae flies are colored like wasps or bees*).

This possibility of self recognition suggests that short-memory players cannot set the rules of the game in an evolutionary setting. Press and Dyson’s claim that a long-but-finite memory strategy is equivalent to some memory-one strategy is correct, but in evolutionary

setting when strategies can use memory to recognize type of opponent, longer memory can be useful, and can lead to better stability if extortioners recognize each other and cooperate.

Other scenarios of useful or successful extortion are given in the following paper (Evolution of Extortion in IPD Games).

Evolution of extortion in IPD games, Hilbe, Nowak, Sigmund [2] (April 23, 2013)

Examine the contest between extortioners and four of the most important memory-one strategies. Show that extortion cannot be an outcome of evolution, but can catalyze the emergence of cooperation. Extortion strategies can only get a foothold if the population is very small. If IPD is played between members of two distinct populations, ZD strategies can emerge in the population that evolves more slowly. In particular, extortion strategies can allow host species to enslave their endosymbionts.

WSLS dominates extortioners.

From extortion to generosity, evolution in the IPD, Stewart, Plotkin [7] (July 25, 2013)

Extortion makes for a good headline, but there are ZD strategies that promote cooperation (*comment: Can use this as a way to bring the discussion back to Axelrod's arguments for TFT*)

The aftermath of discovery of ZD strategies was a volume of research into extortioners, while generous ZDs did not attract as much attention. It seems people always are trying to find clever ways to undermine their opponents in order to beat the game, these ways did not get them very far in Axelrod's tournaments, and once again not very far in using ZDs. Generous ZDs outperformed PAVLOV and TFT, while extortionate ZDs did not do well in long run.

This paper defines another version of evolutionary robustness for finite populations.

Adaptive Dynamics of Extortion and Compliance, Hilbe, Nowak, Traulsen [3] (November 1, 2013)

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.

Chapter 2

Countervailing

These are the payoffs for Langlois $\begin{pmatrix} 0 & -2 \\ 1 & -1 \end{pmatrix}$ and Axelrod $\begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}$ values of $\begin{pmatrix} R & S \\ T & P \end{pmatrix}$

$$\begin{aligned}
 & (1 - x_i, x_i) \begin{pmatrix} R & S \\ T & P \end{pmatrix} \begin{pmatrix} 1 - x_j \\ x_j \end{pmatrix} \\
 &= (1 - x_i)(1 - x_j)R + (1 - x_i)x_jS + x_i(1 - x_j)T + x_ix_jP \\
 &= (1 - x_i - x_j + x_ix_j)R + (x_j - x_ix_j)S + (x_i - x_ix_j)T + x_ix_jP \\
 &= R + (T - R)x_i + (S - R)x_j + (R - S - T + P)x_ix_j \\
 &= \begin{cases} x_i - 2x_j & \text{Langlois} \\ 3 + 2x_i - 3x_j - x_ix_j & \text{Axelrod} \end{cases}
 \end{aligned}$$

Question about discounting

In the derivation of Zero Determinant strategies, Press and Dyson use the transition probability matrix

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

for IPD with mixed strategies $\mathbf{p} = (p_1, p_2, p_3, p_4)$, $\mathbf{q} = (q_1, q_2, q_3, q_4)$ for players X, Y respectively. If \mathbf{v} is the stationary distribution of the above Markov chain, then they show that the dot product of \mathbf{v} with an arbitrary vector \mathbf{f} is given by the determinant

$$\mathbf{v} \cdot \mathbf{f} \equiv D(\mathbf{p}, \mathbf{q}, \mathbf{f}) = \det \begin{bmatrix} -1 + p_1q_1 & -1 + p_1 & -1 + q_1 & f_1 \\ p_2q_3 & -1 + p_2 & -q_3 & f_2 \\ p_3q_2 & p_3 & -1 + q_2 & f_3 \\ p_4q_4 & p_4 & q_4 & f_4 \end{bmatrix},$$

so if $\mathbf{S}_X = (R, S, T, P)$, then on average X 's payoff is $\mathbf{v} \cdot \mathbf{S}_X = D(\mathbf{p}, \mathbf{q}, \mathbf{S}_X)$. Press and Dyson use this average payoff for their definitions and proofs, so in their game there is no discounting.

Question: I'm not sure if this is a problem when comparing ZD strategies with countervailing strategies, which use a discount factor in their definition.

One way I could see doing this is making the game have five states: $(cc, cd, dc, dd, \text{sink})$, where sink is the end of game state associated with payoff of 0 for both players. I'm not sure what will happen to the ZD strategies if they are defined on this 5-state Markov chain.

The other way would be to say that if X 's average payoff is $S_X = \mathbf{v} \cdot \mathbf{S}_X$, then X 's discounted payoff should be $\frac{S_X}{1-w}$.

Is either of these approaches reasonable?

Are countervailing strategies ZD? Using your example.

The payoffs for S_X and S_Y are given by

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

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

so a linear (affine?) combination of payoffs is given by

$$\alpha S_X + \beta S_Y + \gamma = \frac{D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1})}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}$$

from the countervailing example with $\psi_j(x_i, x_j) = \frac{x_i + x_j}{2.7}$ yielding $\mathbf{p} = (0, \frac{1}{2.7}, \frac{1}{2.7}, \frac{2}{2.7})^T$

$$\begin{pmatrix} 0 \\ \frac{1}{2.7} \\ \frac{1}{2.7} \\ \frac{2}{2.7} \end{pmatrix} = \alpha \begin{pmatrix} 0 \\ -2 \\ 1 \\ -1 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \\ -2 \\ -1 \end{pmatrix} + \begin{pmatrix} \gamma \\ \gamma \\ \gamma \\ \gamma \end{pmatrix}$$

solving for α, β, γ , we get $\alpha = \beta = -\frac{1}{2.7}$, $\gamma = 0$, so the strategy \mathbf{p} is the ZD strategy that enforces the relationship $-\frac{1}{2.7}S_X - \frac{1}{2.7}S_Y = 0$ or simply $S_X + S_Y = 0 = R$.

Langlois' Countervailing example with Axelrod's R, S, T, P values

Taking $g_i(x_j) = -\gamma_i x_j$ does not work out. Instead here's an attempt with $g_i(x_j) = R - \gamma_i x_j = 3 - \gamma_i x_j$, this way it is consistent with *Langlois'* example, since $R = 0$ in Langlois' definition and $R = 3$ in Axelrod's definition.

Then we have

$$g_i(x_j) = 3 - \gamma_i x_j$$

solving the countervailing equation

$$3 - \gamma_i x_j = 3 + 2x_i - 3x_j - x_i x_j - \delta \gamma_i \psi_j$$

which yields

$$\psi_j(x_i, x_j) = \frac{2x_i + (\gamma_i - 3)x_j - x_i x_j}{\delta \gamma_i}$$

with the requirement that

$$\psi_j \in [0, 1]$$

setting $x_j = 0$ we get $\gamma_i > 0$, and setting $x_i = 0$ gives us $(\gamma_i - 3) > 0$ or $\gamma_i > 3$. To solve $2x_i + (\gamma_i - 3)x_j - x_i x_j \leq \delta \gamma_i$, we see that $\frac{\partial \text{LHS}}{\partial x_i} \geq 0$, so we can take $x_i = 1$, but $\frac{\partial \text{LHS}}{\partial x_j} \geq 0$ needs $\gamma_i \geq 4$. With that restriction, we can also take $x_j = 1$, and we get

$$\frac{2 + (\gamma_i - 3) - 1}{\delta \gamma_i} \leq 1$$

or

$$\frac{2}{1 - \delta} \geq \gamma_i.$$

Thus

$$\psi(x_i, x_j) \in [0, 1] \text{ whenever } 4 \leq \gamma_i \leq \frac{2}{1 - \delta}.$$

Taking $\delta = .9$, and $\gamma_i = 5$, we get

$$\psi_j(x_i, x_j) = \frac{2x_i + 2x_j - x_i x_j}{4.5}$$

Bibliography

- [1] Christoph Adami and Arend Hintze. Evolutionary instability of zero-determinant strategies demonstrates that winning is not everything, 2014.
- [2] Christian Hilbe, Martin A. Nowak, and Karl Sigmund. Evolution of extortion in iterated prisoner’s dilemma games. *Proceedings of the National Academy of Sciences*, 110(17):6913–6918, 2013.
- [3] Christian Hilbe, Martin A. Nowak, and Arne Traulsen. Adaptive dynamics of extortion and compliance. *PLoS ONE*, 8(11):e77886, 11 2013.
- [4] Dimitris Iliopoulos, Arend Hintze, and Christoph Adami. Critical dynamics in the evolution of stochastic strategies for the iterated prisoner’s dilemma. *PLoS Comput Biol*, 6(10):e1000948, 10 2010.
- [5] William H. Press and Freeman J. Dyson. Iterated prisoner’s dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26):10409–10413, 2012.
- [6] Alexander J. Stewart and Joshua B. Plotkin. Extortion and cooperation in the prisoner’s dilemma. *Proceedings of the National Academy of Sciences*, 109(26):10134–10135, 2012.
- [7] Alexander J. Stewart and Joshua B. Plotkin. From extortion to generosity, evolution in the iterated prisoner’s dilemma. *Proceedings of the National Academy of Sciences*, 2013.