Nice Invincible Strategy for the Average-Payoff IPD

Shiheng Wang,[1] **Fangzhen Lin**[2]

[2]HKUST-Xiaoi Robot Joint Lab

[1,2]Department of Computer Science and Engineering

[1,2]The Hong Kong University of Science and Technology

swangbv@cse.ust.hk, flin@cs.ust.hk

Abstract

The Iterated Prisoner's Dilemma (IPD) is a well-known benchmark for studying the long term behaviours of rational agents. Many well-known strategies have been studied, from the simple tit-for-tat (TFT) to more involved ones like zero determinant and extortionate strategies studied recently by Press and Dyson. In this paper, we consider what we call invincible strategies. These are ones that will never lose against any other strategy in terms of average payoff in the limit. We provide a simple characterization of this class of strategies, and show that invincible strategies can also be nice. We discuss its relationship with some important strategies and generalize our results to some typical repeated 2x2 games. It's known that experimentally, nice strategies like the TFT and extortionate ones can act as catalysts for the evolution of cooperation. Our experiments show that this is also the case for some invincible strategies that are neither nice nor extortionate.

Introduction

The Iterated Prisoner's Dilemma is a classic benchmark used to study rational agents' long term behavior. It involves two agents playing repeatedly the Prisoner's Dilemma (PD). In the PD, each player can choose between Cooperate (C) and Defect (D). If both choose C, they receive a payoff of R (rewards); If both choose D, they receive a payoff of P (penalty); If one chooses C and the other D, the defector receives a payoff of T (temptation to defect) and the cooperator receives a payoff of S (sucker's payoff). The assumptions are $T > R > P > S$ and $2R > T + S > 2P$. Table 1 gives a normal form representation of this game (Rapoport and Chammah 1965).

Table 1: Prisoner's Dilemma

	C	D
C	(R,R)	(S,T)
D	(T,S)	(P,P)

The profile (D, D) is the dominant Nash equilibrium of this game. But both players receive a higher payoff of R if they decide to cooperate, hence the dilemma. There is no controversy about what a rational agent should do when playing the PD. However, if the game is repeated indefinitely or infinitely, it is not clear which if any strategy is the best. In fact, it is easy to see that there is no one best strategy against all other possible ones (Axelrod and Hamilton 1981).

Researchers from diverse disciplines have used the IPD to study the emergence of cooperation among unrelated agents. Robert Axelrod (Axelrod and Hamilton 1981) was the first to run some computer tournaments on iterated prisoner's dilemma. Remarkably the simple Tit-For-Tat (TFT) strategy was the winner. In 2012 (Press and Dyson 2012) dramatically changed people's understanding of this game by deriving what they called *zero determinant* (ZD) strategies. Among them, of particular interests are what they called *extortionate* strategies that can enforce an extortionate linear relation between the players' scores. We will show that extortionate strategies are invincible, in the sense that no strategies can have a higher average payoff when they play against them for infinite rounds.

In this paper we define the class of invincible strategies, and show that they can be characterized by three simple conditions. Such invincible strategies can also be nice, which is never the first to defect. This result is then used to discuss its relationship with extortionate strategy (Press and Dyson 2012), Akin's good strategy (Akin 2016) and other related works (Hilbe, Traulsen, and Sigmund 2015; Hao, Li, and Zhou 2018).

The rest of this paper is organized as follows. We first introduce iterated prisoner's dilemma and briefly review some known results about it, especially the ones about memory-one strategies. We then formally define the class of invincible strategies, proving that they can be characterized by three simple conditions. We then relate them to Press and Dyson's extortionate and ZD strategies, as well as Akin's good strategies. We next consider generalizing our results to some other repeated 2x2 games, report some experimental results about invincible strategies and finally conclude the paper with some concluding remarks.

Table 2: Outcome and Strategy

Outcome*	CC	CD	DC	DD
Stationary Distribution	v_1	v_2	v_3	v_4
Strategy of $\mathbf{p}$	p_1	p_2	p_3	p_4
Strategy of $\mathbf{q}$	q_1	q_3	q_2	q_4

*Outcome is defined from $\mathbf{p}$'s perspective

Iterated Prisoner's Dilemma

The IPD is the repeated PD given in Table 1 under the constraint that $T > R > P > S$ and $2R > T + S > 2P$ so that cooperation pays off in the long run. An often used example of the PD that satisfies these constraints is $(R, S, T, P) = (3, 0, 5, 1)$.

In a repeated game like the IPD, a player's strategy is a function from histories of interactions to actions. Often one restricts strategies to some specific forms, such as Turing machines (Chen and Tang 2015; Knoblauch 1994; Megiddo and Wigderson 1986), finite automata (Rubinstein 1986; Ben-Porath 1990; Gilboa 1988; Zuo and Tang 2015), ones with limited memories (Hauert and Schuster 1997; Lindgren 1992; Chen et al. 2017), and other forms of bounded rationality (e.g. (Osborne and Rubinstein 1994; Shoham and Leyton-Brown 2008)).

For the IPD, (Press and Dyson 2012) proved that the player with the shortest memory sets the rule of the game, and that the co-player cannot evade memory-one strategy by changing his own strategy on a short timescale, even arbitrarily on every move of the game. Thus one needs only consider *memory-one* mixed strategies, which also win against more complex and human-like strategies, such as intention recognition (Anh, Moniz Pereira, and Santos 2011), theory of mind (Devaine, Hollard, and Daunizeau 2014) and neural-network based strategies (McNally, Brown, and Jackson 2012).

A memory-one (mixed) strategy decides with certain probabilty what action to do based on the outcome of the previous round. Thus it can be defined by the probabilities p_{CC}, p_{CD}, p_{DC}, and p_{DD} of playing C when the previous outcomes are CC, CD, DC, and DD, respectively. In the following, we write X's (player 1's) strategy $\mathbf{p}$ as a tuple in the following order:

$$\mathbf{p} = (p_1, p_2, p_3, p_4) = (p_{CC}, p_{CD}, p_{DC}, p_{DD}),$$

and Y's strategy as a tuple in the following order:

$$\mathbf{q} = (q_1, q_2, q_3, q_4) = (q_{CC}, q_{DC}, q_{CD}, q_{DD}).$$

Notice that the orders for X and Y are the same when they are viewed from the player's *own* perspective.

A probability distribution $\mathbf{v}$ on the set of outcomes is a non-negative vector $\mathbf{v} = (v_1, v_2, v_3, v_4) = (v_{CC}, v_{CD}, v_{DC}, v_{DD})$ with unit sum: $v_1 + v_2 + v_3 + v_4 = 1$. Given an initial distribution, the probability distribution after the r-th iteration is noted by $\mathbf{v}^r$. The relationship between outcome and strategies is shown in Table 2.

An effective way to study memory-one strategies is to view their interactions as Markov chains. Our following pre-

sentation follows mostly after (Press and Dyson 2012) and (Akin 2016).

If X uses initial probability p_0 (for playing C) and strategy $\mathbf{p} = (p_1, p_2, p_3, p_4)$, Y uses initial probability q_0 and memory-one strategy $\mathbf{q} = (q_1, q_2, q_3, q_4)$, then the probability distribution of the first iteration is $\mathbf{v}^1 = (p_0q_0, p_0(1 - q_0), (1 - p_0)q_0, (1 - p_0)(1 - q_0))$ and the successive outcomes follow a Markov chain with transition matrix given by:

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

Each entry of $\mathbf{M}$ represents the probability of transition between different states, and $\mathbf{v}^r\mathbf{M} = \mathbf{v}^{r+1}$. If $\mathbf{v}\mathbf{M} = \mathbf{v}$, then we say that $\mathbf{v}$ is stationary. Following (Akin 2016), we call M convergent when there is a unique stationary distribution vector for M. Although the sequence of $\mathbf{v}^i (i = 1, 2, ...)$ may cycle through several states and thus not converge, the sequence of the Cesaro averages $\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{v}^i\}$ of the outcome distributions always converges to some stationary distribution $\mathbf{v}$ (Akin 2016). Thus this limit of average always exists given the game is infinitely repeated. That is, if $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbf{v}^k = \mathbf{v}$, then $\mathbf{v}\mathbf{M} = \mathbf{v}$.

Press and Dyson (Press and Dyson 2012) showed that if $\mathbf{v}$ is a stationary vector of the Markov matrix $\mathbf{M}$, then its dot product with any vector $\mathbf{f} = (f_1, f_2, f_3, f_4)$ is proportional(noted as $\equiv$) to $D(\mathbf{p}, \mathbf{q}, \mathbf{f})$, which can be written as:

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

A normalization converts this proportion to equivalent, and the payoffs s_X and s_Y of X and Y, respectively, in $\mathbf{v}$ can be computed as

$$s_X = \frac{\mathbf{v} \cdot \mathbf{S_X}}{\mathbf{v} \cdot \mathbf{1}} = \frac{D(\mathbf{p}, \mathbf{q}, \mathbf{S_X})}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}$$
$$s_Y = \frac{\mathbf{v} \cdot \mathbf{S_Y}}{\mathbf{v} \cdot \mathbf{1}} = \frac{D(\mathbf{p}, \mathbf{q}, \mathbf{S_Y})}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})} \quad (2)$$

where $\mathbf{S_X} = (R, S, T, P)$ is the X's payoff matrix and $\mathbf{S_Y} = (R, T, S, P)$ the Y's.

Since $\mathbf{f}$ can be any four-vector, we shall use equation (1) to calculate the value of v_2 and v_3 in the next section.

Invincible Strategy

Stationary Distribution

Since only outcome CD and DC make a difference to players' payoffs, what we really care about is the relation between v_2 and v_3, which can be calculated according to Eq. (1). The normalization is similar to Eq. (2), which guarantees $v_1 + v_2 + v_3 + v_4 = 1$.

Let $\mathbf{f}_2 = (0,1,0,0), \mathbf{f}_3 = (0,0,1,0)$, then,

$$v_2 = \frac{\mathbf{v} \cdot \mathbf{f}_2}{\mathbf{v} \cdot \mathbf{1}} = \frac{D(\mathbf{p},\mathbf{q},\mathbf{f}_2)}{D(\mathbf{p},\mathbf{q},\mathbf{1})}$$

$$v_3 = \frac{\mathbf{v} \cdot \mathbf{f}_3}{\mathbf{v} \cdot \mathbf{1}} = \frac{D(\mathbf{p},\mathbf{q},\mathbf{f}_3)}{D(\mathbf{p},\mathbf{q},\mathbf{1})} \quad (3)$$

For simplicity, let $D = D(\mathbf{p},\mathbf{q},\mathbf{1})$, $D_2 = D(\mathbf{p},\mathbf{q},\mathbf{f}_2)$, $D_3 = D(\mathbf{p},\mathbf{q},\mathbf{f}_3)$. Then, $v_2 = D_2/D, v_3 = D_3/D$.

If $D = 0$, there may be more than one stationary distributions, depending on the initial distribution. For example, when both players take strategy $Repeat$, that is, $\mathbf{p} = \mathbf{q} = (1,1,0,0)$, then $D = 0$ and the vector of stationary distribution $\mathbf{v}$ is exactly the same as initial distribution $\mathbf{v}^1$. Such special cases will be discussed separately as the calculation is different from above.

Theorem 1. *Assume $p_1,...,p_4,q_1,...,q_4 \in [0,1]$, then for $D = D(\mathbf{p},\mathbf{q},\mathbf{1})$, $D \leq 0$*

Proof. $\forall z \in \{p_1,...,p_4,q_1,...,q_4\}$, $\partial^2 D/\partial^2 z = 0$. So, when all variables in $\{p_1,...,p_4,q_1,...,q_4\}$ except z are fixed, $\partial D/\partial z = C$, where C is a constant independent of z. Thus, when other variables are fixed, $D(z)$ is a monotonic function. Thus we can get all extrema of D by letting

$$z = 0 \quad or \quad z = 1, \quad \forall z \in \{p_1,...,p_4,q_1,...,q_4\}$$

Since all of $2^8 = 256$ extrema are less or equal to zero, we can conclude that $D \leq 0$. $\square$

Invincible Strategy

Definition 1 (Invincible Strategy). *Player X plays a memory-one strategy $\mathbf{p}$ against player Y who plays memory-one strategy $\mathbf{q}$. Player X and Y get the average payoff of s_X and s_Y respectively. A memory-one strategy $\mathbf{p}$ is invincible if against any other memory-one strategy $\mathbf{q}$, for any initial distribution v_0, the players' payoffs satisfy $s_X \geq s_Y$.*

According to our definition, a memory-one strategy $\mathbf{p}$ is invincible if it will not lose against any other memory-one strategy. Notice that according to (Press and Dyson 2012), this implies that $\mathbf{p}$ will not lose against any other strategy, memory-one or not, as having longer memory or playing more complex strategy will not help.

In Table 1, both players share the same payoff on action profile CC and DD, therefore only profile CD and DC matter when we compare payoffs, hence we have the following lemma.

Lemma 1. *Assume s_X and s_Y exist according to (2), and that $\mathbf{v} = (v_1,v_2,v_3,v_4)$ is the unique stationary distribution. Under $T > S$, $s_X \geq s_Y \iff v_2 \leq v_3$.*

Proof.

$$\begin{aligned} s_X \geq s_Y &\iff \mathbf{v} \cdot \mathbf{S_X} \geq \mathbf{v} \cdot \mathbf{S_Y} \\ &\iff (T-S) * (v_3 - v_2) \geq 0 \\ &\iff v_2 \leq v_3 \end{aligned}$$

$\square$

We need Akin's lemma as well as a similar new lemma in the proof of our theorems.

Lemma 2 (Akin's Lemma). *Assume that X uses the strategy $\mathbf{p} = (p_1,p_2,p_3,p_4)$, we call $\tilde{\mathbf{p}} = (p_1 - 1, p_2 - 1, p_3, p_4)$ the Press-Dyson vector of $\mathbf{p}$. If the opponent Y uses a strategy pattern that yields a sequence of distributions $\{\mathbf{v}^n\}$, then*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{v}^k \cdot \tilde{\mathbf{p}} = 0$$

Or shortly,

$$\mathbf{v} \cdot \tilde{\mathbf{p}} = v_1(p_1 - 1) + v_2(p_2 - 1) + v_3 p_3 + v_4 p_4 = 0. \quad (4)$$

Lemma 3. *Assume that X uses the strategy $\mathbf{p} = (p_1,p_2,p_3,p_4)$ and Y uses the strategy $\mathbf{q} = (q_1,q_2,q_3,q_4)$. Let $\mathbf{u} = (p_1 q_1 - 1, p_2 q_3, p_3 q_2, p_4 q_4)$, $\mathbf{v}$ is the stationary vector, then*

$$\mathbf{v} \cdot \mathbf{u} = v_1 u_1 + v_2 u_2 + v_3 u_3 + v_4 u_4 = 0.$$

Proof. Let $\mathbf{f} = \mathbf{u} = (p_1 q_1 - 1, p_2 q_3, p_3 q_2, p_4 q_4)$, then according to Eq. (1),

$$\mathbf{v} \cdot \mathbf{u} \equiv D(\mathbf{p},\mathbf{q},\mathbf{u}) = 0 \quad (5)$$

Eq. (5) is zero because the first and fourth column in this determinant are the same. $\square$

Theorem 2. *If $\mathbf{p} = (p_1,p_2,p_3,p_4)$ is invincible, then*

$$p_2 + p_3 \leq 1 \quad (6)$$

$$p_4 = 0 \quad (7)$$

$$p_2 \neq 1 \quad (8)$$

Proof. Suppose $\mathbf{p}$ is invincible.

Assume $p_2 = 1$. Consider when $\mathbf{p}$ plays against *Always Defect* and assume the first iteration is $\{CD\}$. The unique stationary distribution is $(0,1,0,0)$. So for this game $\mathbf{p}$ has lower score, a contradiction with our assumption that $\mathbf{p}$ is invincible. This shows $p_2 \neq 1$.

Suppose $p_4 > 0$. Consider again when $\mathbf{p}$ plays against *Always Defect*. From $\mathbf{p}$'s perspective, the possible outcomes are either $\{DD, CD\}$ or $\{CD\}$. Either case, strategy $\mathbf{p}$ loses. Thus $p_4 = 0$ for $\mathbf{p}$ to be invincible.

Thus we can assume $\mathbf{p} = (p_1,p_2,p_3,0)$. Now consider it plays against $\mathbf{q} = (0,0,0,1)$. Apply Lemma 2 to $\mathbf{p}$, we get,

$$(p_1 - 1)v_1 + (p_2 - 1)v_2 + p_3 v_3 = 0 \quad (9)$$

Apply Lemma 3 to both $\mathbf{p}$ and $\mathbf{q}$, we have,

$$-v_1 = 0 \quad (10)$$

From Eq. (9) and (10), we get, $(1 - p_2)v_2 = p_3 v_3$.
Since $\mathbf{p}$ is invincible, by Lemma 1, $v_2 \leq v_3$.
Thus $1 - p_2 \geq p_3$, hence $p_2 + p_3 \leq 1$. $\square$

Theorem 3. *If conditions (6), (7) and (8) hold, then $\mathbf{p} = (p_1,p_2,p_3,p_4)$ is invincible.*

Proof. Suppose agent X and Y take strategy $\mathbf{p}$ and $\mathbf{q}$ respectively, $D = D(\mathbf{p}, \mathbf{q}, \mathbf{1})$.

$$Let \quad \mathcal{L} = (1 - p_2 - p_3)(1 - p_1 q_1)$$
$$+ (1 - p_1)p_3 q_2 + (1 - p_1)p_2 q_3$$

Given that $p_1, ..., p_4, q_1, ..., q_3 \in [0, 1]$, and that we assume $p_2 + p_3 \leq 1$ in equation (6), we have

$$\forall q_1, ..., q_3 \in [0, 1], \quad \mathcal{L} \geq 0$$

Because we assume $p_4 = 0$ and $0 \leq q_4 \leq 1$, notice that

$$q_4 \cdot \mathcal{L} = D_2 - D_3$$

We have

$$\forall \mathbf{q}, \quad D_2 \geq D_3$$

(1)Assume $D \neq 0$, according to Theorem 1, $D < 0$, we have

$$\forall \mathbf{q}, \frac{D_2}{D} \leq \frac{D_3}{D}$$

It's the same as, $\forall \mathbf{q}, v_2 \leq v_3$. According to Lemma 1, $\forall \mathbf{q}, s_X \geq s_Y$.

Therefore, strategy $\mathbf{p}$ is invincible.

(2)When $D = 0$, there are more than one terminal sets, and the stationary distribution depends on the initial one, which occurs when some variables in $(p_1, ..., p_4, q_1, ..., q_4)$ equal to 0 or 1. It can be proved by computer solvers that if $\mathbf{q} \neq (1, 1, 0, 0)$ and $0 < p_i < 1$, then $D \neq 0$.

The following discussion is from $\mathbf{p}$'s perspective, i.e. $\{CD\}$ refers to the state where $\mathbf{p}$ plays C and $\mathbf{q}$ plays D.

Case 1. $\mathbf{p}=(0,0,0,0)$. Obviously, no matter what $\mathbf{q}$ is, *Always Defect* is invincible.

Case 2. $\mathbf{p}=(0,0,1,0)$. Because $\{CD\}$ can only appear exactly after $\{DC\}$ except for this first iteration, this strategy is invincible if the game is played for infinite rounds.

Case 3. $\mathbf{p}=(0,1,0,0)$. This strategy is **NOT** invincible since $\{CD\}$ can be a stationary distribution.

Case 4. $\mathbf{p}=(1,0,0,0)$. *Trigger Strategy* is invincible since $\{CD\}$ can only appear at most once in game history and we assume the game is played for infinite rounds.

Case 5. $\mathbf{p}=(1,0,1,0)$. *Tit for Tat* is invincible since $\{CD\}$ can only appear exactly after $\{DC\}$, except for the first time after $\{CC\}$ or the first iteration, which doesn't matter when the game is played for infinite rounds.

Case 6. $\mathbf{p}=(1,1,0,0)$. Obvious, *Repeat* is **NOT** invincible if it plays C in the first iteration.

Case 7. $\mathbf{q}=(1,1,0,0)$. *Repeat* is the only situation where $\mathbf{q}$ can unilaterally set $D = 0$ when $\forall z \in \{p_1, p_2, p_3, p_4\}, z \neq 0$ and $z \neq 1$.

All edge cases can be avoided when $p_2 \neq 1$. In case 1,2,4,5, in the infinite sequence of game history, there is one more $\{CD\}$ than $\{DC\}$. When the game is played for infinite rounds, losing one shot game doesn't make any difference on the average payoff. As for case 7, suppose $\mathbf{q}$ is *Always Defect*, $\{DD\}$ is the only absorbing state while $\{CD\}$ is only a transient state, thus strategy $\mathbf{p}$ is still invincible. $\square$

Theorem 2 and 3 lead to the following conclusion, our main technical result of the paper:

Theorem 4. *A strategy $\boldsymbol{p} = (p_1, ..., p_4)$ is invincible iff the three conditions (6), (7) and (8) hold.*

Invincible strategies can also be nice. A strategy is nice if it is never the first to defect. There is no constraint on p_1 in Theorem 4, so an invincible strategy is nice if it cooperates in the initial iteration and its $p_1 = 1$, together with conditions (6), (7) and (8).

Invincible strategies are largely robust against noise due to fuzzy mind or trembling-hand. As p_4 is fixed to 0, there is no excuse of noise on p_4. However, as there is no constraint on p_1, it's totally robust against noise. As for $p_2 + p_3 \leq 1$ and $p_2 \neq 1$, they are generally robust against noises under the constraints. Some other strategies, like generous Tit-for-Tat ($\mathbf{p} = (1, 1/3, 1, 1/3)$) in (Nowak and Sigmund 1992), may be exploited by their opponents although they're robust against errors.

Invincible strategies account for a large proportion of all strategies. Half of firm strategies ($p_4 = 0$) are invincible since the hyper plane $p_2 + p_3 \leq 1$ bisects the 3D cube $(p_1, p_2, p_3) \in [0, 1]$ when $p_4 = 0$. Well-known strategies that are invincible include *Tit-for-Tat* $(1, 0, 1, 0)$ which equalizes the payoff of both players, and *Always Defect* which never allow itself to be taken advantage of.

We now show the properties of invincible strategies and discuss its relationship with some well-studied strategies.

Press and Dyson's Extortionate Strategies

Press and Dyson's extortionate strategies are special zero determinant strategies (ZDS). According to Press and Dyson (Press and Dyson 2012), a strategy $\mathbf{p} = (p_1, ..., p_4)$ is a ZDS if for some α, β and γ, $\tilde{\mathbf{p}} = \alpha \mathbf{S_X} + \beta \mathbf{S_Y} + \gamma \mathbf{1}$ where $\tilde{\mathbf{p}} = (p_1 - 1, p_2 - 1, p_3, p_4)$.

A strategy $\mathbf{p} = (p_1, ..., p_4)$ is extortionate if for some ϕ and $\chi \geq 1$,

$$p_1 = 1 - \phi(\chi - 1)\frac{R - P}{P - S}, \quad p_2 = 1 - \phi(1 + \chi\frac{T - P}{P - S})$$

$$p_3 = \phi(\chi + \frac{T - P}{P - S}), \quad p_4 = 0 \quad (11)$$

χ is called the extortion factor because of the following equation which follows from (11):

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

For instance, $\mathbf{p} = (11/13, 1/2, 7/26, 0)$ is an extortionate strategy with the extortion factor $\chi = 3$. There are extortionate strategies for any χ and sufficiently small ϕ. According to (Press and Dyson 2012) the allowed range of ϕ is

$$0 < \phi \leq \frac{(P - S)}{(P - S) + \chi(T - P)} \quad (12)$$

As Press and Dyson showed, for given (R, S, T, P), one can use the constraint $\tilde{\mathbf{p}} = \alpha \mathbf{S_X} + \beta \mathbf{S_Y} + \gamma \mathbf{1}$ to deduce necessary conditions of zero determinant strategy. For example, if R=3, S=0,T=5, and P=1, then a zero determinant strategy must satisfy the following equation:

$$3 * p_1 - 2 * p_2 - 2 * p_3 + p_4 - 1 = 0 \quad (13)$$

Theorem 5. *All extortionate strategies are invincible strategies.*

Proof. According to equation set (11), extortionate strategies satisfy $p_2 \neq 1$ and $p_4 = 0$.

$$p_2 + p_3 = 1 + \phi(1-\chi)(\frac{T+S-2P}{P-S})$$

Recall in the setting of the IPD, $T+S > 2P$ and $P > S$. In extortionate strategies, $\chi \geq 1, \phi > 0$. Then we have $p_2 + p_3 \leq 1$. Therefore, all extortionate strategies are invincible. □

Extortionate strategies are also zero determinant. However, not all zero determinant strategies are invincible. For example, $\mathbf{p} = (3/7, 0, 5/7, 2/7)$ is a zero determinant strategy that sets the co-player's score to a fixed value, but itself can receive a lower payoff.

While we have shown that all extortionate strategies are invincible, not all invincible strategies are extortionate. For example, $\mathbf{p}=(0.5, 0.2, 0.7, 0)$ is invincible, but does not satisfy equation (13). In fact, this strategy is neither zero determinant nor extortionate.

Payoff Control

Theorem 6. *Suppose player X and Y play the IPD. If player X takes an invincible strategy, no matter what strategy player Y takes, both players' payoffs satisfy,*

$$s_X \geq P, s_Y \leq R. \tag{14}$$

Proof. This follows from the two constraints $s_X \geq s_Y$ and $2P \leq s_X + s_Y \leq 2R$. The former because X's strategy is invincible. The latter because of the constraints of the IPD. □

Theorem 6 guarantees that invincible strategies are as good as *Always Defect* in the worst case, and that mutual cooperation is the best outcome for the other player. Thus *Always Cooperate* is a best response to nice invincible strategies.

Payoff control is also discussed in (Hao, Li, and Zhou 2018), but the authors only give an example of extortionate strategy, which has already been defined in (Press and Dyson 2012). The simple characterization of invincible strategies is not pointed out. Without considering special edge cases, payoff control may not be robust against every deterministic strategy. For example, $p_1 = 1, p_4 = 0, p_2 + p_3 = 1$ may not guarantee $s_X = s_Y$, because if $p_2 = 1, p_3 = 0$, when the co-player plays D after initial outcome CD, the game converges to outcome CD. It's the worst case to the payoff controller.

Akin's Good Strategies

The invincible strategies include extortionate ones. But they also include nice ones like the TFT, which is also a *good* one according to Akin (Akin 2016). While not all invincible strategies are Akin's good strategies, all *nice* ones are. We first review Akin's definition of good strategies.

Definition 2 (Akin's Good Strategies (Akin 2016)). *X's strategy $\mathbf{p}$ is nice if $p_1 = 1$. It is good if it is nice and for any strategy chosen by Y, we have that*

$$\text{if } S_Y \geq R \text{ then } S_Y = S_X = R.$$

Notice that $p_1 = 1$ means that this strategy always cooperate when the previous outcome is CC. We now show that nice and invincible strategies are good.

Theorem 7. *If X's strategy is nice and invincible, then it is good.*

Proof. By Theorem 6, $s_Y \leq R$. Suppose $S_Y \geq R$, then $S_Y = R$. Since X is invincible, $S_X \geq S_Y = R$. Because $s_X + s_Y \leq \max\{2R, T+S, 2T\} = 2R$, we get $s_X = R$. □

However, not all good strategies are invincible. This follows from (Akin 2016).

Competitive Strategies

Invincible strategies are also studied in (Hilbe, Traulsen, and Sigmund 2015), where it's called competitive strategy. However, the concept is defined for the discounted-payoff IPD and all calculations depend on a discount factor $\delta < 1$. Its Proposition 2 conducts an explicit calculation of payoffs, showing that,

$$s_X - s_Y = -\frac{(T-S)((1-\delta)p_0 + \delta p_{DD})}{1 + \delta(p_{DD} - p_{CD})} \tag{15}$$

And their competitive strategy is defined as,

$$p_0 = p_{DD} = 0, \quad \delta(p_{CD} + p_{DC}) \leq 1. \tag{16}$$

We cannot simply let $\delta = 1$ to generalize this conclusion to average-payoff IPD, because the denominator in (15) may vanish to zero. Moreover, such competitive strategy always defects in the first iteration, hindering itself from being nice. While playing with cooperative strategies like *Trigger*($\mathbf{q} = (1,0,0,0)$) or even *Tit-for-Tat*($\mathbf{q} = (1,0,1,0)$), the game can immediately converge to outcome DD, which may not be what we want.

Other Solution Concepts

Also related is work that introduces alternative solution concepts (e.g. (Halpern and Pass 2012; Capraro et al. 2013)) to capture desirable behaviours of agents in repeated games. In particular, Capraro *et al.* (2013) proposed the concept of iterated cooperative equilibria (ICE) to explain human long term cooperative behaviours. An interesting question is whether certain invincible strategies are ICE or whether ICE are all nice and invincible. We leave this as future work.

Generalization

Repeated 2×2 Game

We have presented invincible strategies in the context of IPD, by far the most studied repeated game. Invincible strategies actually exist in other repeated 2×2 normal-form games as well.

A general 2×2 normal-form game a tuple $(\mathcal{N}, \mathcal{A}, u)$, where

Table 3: Payoff Matrix

(a) PD-like Games

	B_1	B_2
A_1	a, a	$b, b + \epsilon$
A_2	$c + \epsilon, c$	d, d

(b) SUG

	Accept	Deny
a	(a,b)	(0,0)
b	(b,a)	(0,0)

note: in table (a), $\epsilon > 0$; in table (b), $a > b > 0$.

- $\mathcal{N} = \{X, Y\}$ is a pair of players.
- $\mathcal{A} = A \times B$, where $A = \{A_1, A_2\}$ is a set of actions available to player X, while $B = \{B_1, B_2\}$ is that to player Y. Each vector $o = (a, b) \in \mathcal{A}$ is called an action profile(or outcome).
- $u = (u_X, u_Y)$, where $u_i : \mathcal{A} \mapsto \mathbb{R}$ is a real-valued utility(or payoff) function for player i.

The game is played for infinite rounds, and similar to the IPD, we consider only memory-one strategies. For X (the row player), her memory-one strategy is defined in terms of the probability of playing action A_1, and for Y (the column player), that of playing B_1. Again similar to the IPD, the strategies are represented by four-element vectors, in the following orders for X's strategy $\mathbf{p}$ and Y's strategy $\mathbf{q}$:

$$\mathbf{p} = (p_1, p_2, p_3, p_4) = (p_{A_1B_1}, p_{A_1B_2}, p_{A_2B_1}, p_{A_2B_2}),$$
$$\mathbf{q} = (q_1, q_2, q_3, q_4) = (q_{A_1B_1}, q_{A_2B_1}, q_{A_1B_2}, q_{A_2B_2}).$$

Outcome $\mathbf{o}$ and stationary distribution $\mathbf{v}$ can be defined similarly,

$$\mathbf{v} = (v_1, v_2, v_3, v_4) = (v_{A_1B_1}, v_{A_1B_2}, v_{A_2B_1}, v_{A_2B_2})$$
$$\mathbf{o} = (o_1, o_2, o_3, o_4) = (A_1B_1, A_1B_2, A_2B_1, A_2B_2)$$

The average payoffs of both players are calculated by,

$$s_X = \sum_{i=1}^{4} v_i * u_X(o_i), \quad s_Y = \sum_{i=1}^{4} v_i * u_Y(o_i) \quad (17)$$

Player X's strategy $\mathbf{p}$ is invincible iff $\forall \mathbf{q}$, $s_X \geq s_Y$. Similarly, player Y's strategy $\mathbf{q}$ is invincible iff $\forall \mathbf{p}$, $s_Y \geq s_X$. As the average payoffs depend on utility functions, it is not possible to have a general theorem that characterizes invincible strategies in all repeated 2x2 games. Below we consider some of them.

PD-like Games

Table 4: Example of Games

(a) Game of Chicken

	Chick	Dare
Chick	6,6	2,7
Dare	7,2	0,0

(b) Battle of the Sexes

	Opera	Game
Game	0,0	2,3
Opera	3,2	0,0

Consider a 2x2 game with a payoff matrix of the form in Table 3a. Games of this form includes the PD, the Game of Chicken (Table 4a), and the Battle of the Sexes (Table 4b).

Tacit collusion in repeated second-price auctions (Skrzypacz and Hopenhayn 2004) can also be modeled into such game.

For repeated versions of these games, invincible strategies for both players can be characterized in the same way as for the IPD, with similar proofs.

Simplified Ultimatum Game

Our second class of repeated 2x2 games for which we can characterize the invincible strategies is a simplified Ultimatum Game(SUG) (Thaler 1988), given in Table 3b. Player X makes a division of $a + b$ dollars ($a > b > 0$). She can either keep a or b dollars to herself and give the rest to Y. Player Y can either accept or decline the offer. If she accepts, both receive the money according to this division; otherwise none of them get any money. The game is repeated for infinite rounds and we characterize invincible strategies for X.

For the repeated simplified Ultimatum Game, we can show that $s_X \geq s_Y$ if and only if $v_1 \geq v_3$. The proof is similar to that of Lemma 1. Intuitively, $v_1 \geq v_3$ can be achieved by playing "more" action A_1 than A_2. However, the necessary and sufficient condition turns out to be more complex.

Following equation (3), let $\mathbf{f}_1 = (1, 0, 0, 0)$, $\mathbf{f}_3 = (0, 0, 1, 0)$.

$$v_1 = \frac{\mathbf{v} \cdot \mathbf{f}_1}{\mathbf{v} \cdot \mathbf{1}} = \frac{D(\mathbf{p}, \mathbf{q}, \mathbf{f}_1)}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}, \quad v_3 = \frac{\mathbf{v} \cdot \mathbf{f}_3}{\mathbf{v} \cdot \mathbf{1}} = \frac{D(\mathbf{p}, \mathbf{q}, \mathbf{f}_3)}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}$$

Let $g(\mathbf{p}, \mathbf{q}) = D(\mathbf{p}, \mathbf{q}, \mathbf{f}_3) - D(\mathbf{p}, \mathbf{q}, \mathbf{f}_1)$, from theorem 1, we know that $D(\mathbf{p}, \mathbf{q}, \mathbf{1}) \leq 0$. First we assume $D \neq 0$, then

$$v_1 \geq v_3 \Leftrightarrow D(\mathbf{p}, \mathbf{q}, \mathbf{f}_1) \leq D(\mathbf{p}, \mathbf{q}, \mathbf{f}_3) \Leftrightarrow g(\mathbf{p}, \mathbf{q}) \geq 0$$

Similar to Theorem 1, we can prove that $g(\mathbf{p}, \mathbf{q})$ is monotonic w.r.t. $q_i, i = 1, ..., 4$, so $g(\mathbf{p}, \mathbf{q})$ reaches its extreme value when $q_i = 0$ or $q_i = 1$. Then $\forall \mathbf{q}, g(\mathbf{p}, \mathbf{q}) \geq 0$ is

Table 5: Extreme Values

ID	Extreme Value	$g(\mathbf{p}, \mathbf{q}) \geq 0$
g_0	0	0
g_1	$(1 - p_2)(2 * p_4 - 1)$	$p_4 \geq 0.5$
g_2	$p_4 * (2 * p_2 - 1)$	$p_2 \geq 0.5$
g_3	$p_1 * p_2 - p_1 * p_4 + p_2 * p_3$ $-p_3 * p_4 + 2 * p_4 - 1$	
g_4	0	0
g_5	$(1 - p_2) * (p_3 + p_4 - 1)$	$p_3 + p_4 \geq 1$
g_6	$p_4 * (p_2 + p_3 - 1)$	$p_2 + p_3 \geq 1$
g_7	$p_1 * p_2 - p_1 * p_4$ $+p_3 + p_4 - 1$	
g_8	0	0
g_9	$(1 - p_2) * (p_1 + p_4 - 1)$	$p_1 + p_4 \geq 1$
g_{10}	$p_4 * (p_1 + p_2 - 1)$	$p_1 + p_2 \geq 1$
g_{11}	$p_1 + p_2 * p_3$ $-p_3 * p_4 + p_4 - 1$	
g_{12}	0	0
g_{13}	$(1 - p_2) * (p_1 + p_3 - 1)$	$p_1 + p_3 \geq 1$
g_{14}	$p_4 * (p_1 + p_3 - 1)$	$p_1 + p_3 \geq 1$
g_{15}	$p_1 + p_3 - 1$	$p_1 + p_3 \geq 1$

(a) WSLS, Defector
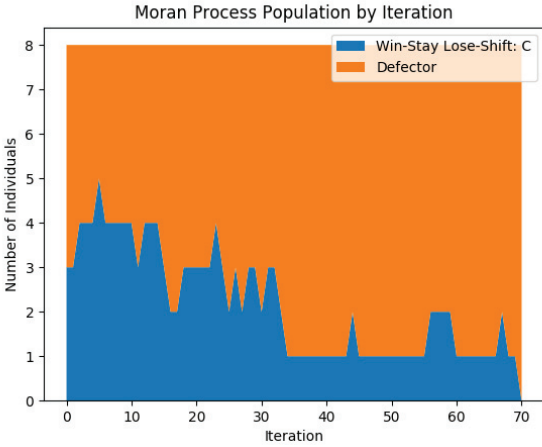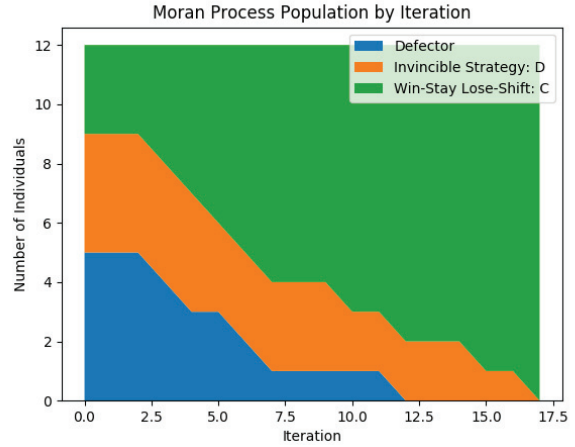
(b) WSLS, Defector, Invincible

Figure 1: Evolutionary Experiment

equivalent to all these extrema are no less than zero. We calculate all extrema with *SymPy*, a Python library for symbolic mathematics (Meurer et al. 2017). Table 5 displays all these extrema together with the if and only if condition that $g(\mathbf{p}, \mathbf{q}) \geq 0$.

Although g_3, g_7, g_{11} look complex, they can be derived from other inequations, namely when all other inequations hold, these three will hold automatically. For instance,

$$g_3 = (1 - p_1) * p4 + (1 - p_3) * p_4 + (p_1 + p_3) * p_2 - 1$$
$$\geq 0.5 * (1 - p_1) + 0.5 * (1 - p_3) + 0.5 * (p_1 + p_3) - 1$$
$$= 0$$

To avoid the case that $D(\mathbf{p}, \mathbf{q}, \mathbf{1}) = 0$ and the Markov chain is trapped in v_3, let $p_3 \neq 0$. Then we can conclude X's invincible strategies for repeated simplified Ultimatum Game.

Theorem 8. *In repeated simplified Ultimatum Game in table 3b, X's memory-one strategy* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *is invincible iff*

$$p_2 \geq 0.5, p_4 \geq 0.5, p_3 \neq 0$$
$$\forall i \neq j \in \{1, 2, 3, 4\}, p_i + p_j \geq 1$$

Evolutionary Experiment

Since Press and Dyson's work, there have been several experiments about ZD and extortionate strategies. One was by Stewart and Plotkin (Stewart and Plotkin 2012) who ran a Axelrod-style tournaments that include a few extortionate strategies. One of the notable results of Stewart and Plotkin's tournament is that the extortionate strategy named Extort-2 won the second most head-to-head matches. We now know this is not really surprising given that extortionate strategies are invincible. Actually, no invincible strategy will loose a head-to-head match if the game is repeated for sufficient number of rounds.

Another was by Hilbe *et al.* (Hilbe, Nowak, and Sigmund 2013) who ran an experiment to analyze the evolutionary performance of extortionate strategies. They concluded that extortionate strategies can act as catalysts for the evolution of cooperation but that they themselves are not the stable outcome of natural selection. We rerun their experiment and replace extortionate strategy with the invincible strategy (0.9,0.7,0.2,0), and the result turns out to be similar. Notice that (0.9,0.7,0.2,0) is just invincible. It is not even a zero determinant strategy.

Figure 1 shows some representative runs. The Axelrod python library (Knight et al. 2016) make it easy to run experiments about the IPD. Here are four experiments similar to some of those in (Hilbe, Nowak, and Sigmund 2013) but with extortionate strategies replaced by invincible ones.

The evolutionary behavior of invincible strategies turn out to be similar to that of extortionate ones.
(a) The population begins with cooperative strategy *Win-Stay-Lose-Shift* and defective strategy *Always Defect*. After 70 iterations, defective strategies dominates this population.
(b) After adding agents of the invincible strategy (0.9, 0.7, 0.2, 0), defectors are firstly eliminated, followed by invincible strategies. Cooperative strategy becomes the stable outcome. Thus, some invincible strategies can act as catalyst of cooperation, but they are not a stable outcome of evolution.

Conclusion

Inspired by our initial observation that no strategies can defeat an extortionate strategy, we went on to study the class of all such strategies that will never lose a head-to-head match and call them invincible.

Invincible strategies are interesting for a variety of reasons. We list four below. Firstly, they have a very clear and intuitive definitions - never lose a match. Secondly, they are surprisingly simple to characterize - our main technical result in this paper is that they are captured by three simple conditions (6), (7) and (8). Thirdly, they are closely related

with some other well-studied strategies such as Press and Dyson's extortionate strategies and Akin's good strategies. Finally, we have seen from our experiments that some invincible strategies that are neither extortionate nor cooperative like the TFT can also act as catalyst for cooperation.

Repeated games are notoriously difficult to analyse, partly because of the large space of complex strategies. The IPD is by far the most studied repeated game and by identifying an interesting new class of strategies with well-defined properties, we have contributed to the work on repeated game in general and the IPD in particular.

Acknowledgments

References

Akin, E. 2016. The iterated prisoner's dilemma: good strategies and their dynamics. *Ergodic Theory, Advances in Dynamical Systems* 77–107.

Anh, H. T.; Moniz Pereira, L.; and Santos, F. C. 2011. Intention recognition promotes the emergence of cooperation. *Adaptive Behavior* 19(4):264–279.

Axelrod, R., and Hamilton, W. D. 1981. The evolution of cooperation. *science* 211(4489):1390–1396.

Ben-Porath, E. 1990. The complexity of computing a best response automaton in repeated games with mixed strategies. *Games and Economic Behavior* 2(1):1–12.

Capraro, V.; Venanzi, M.; Polukarov, M.; and Jennings, N. R. 2013. Cooperative equilibria in iterated social dilemmas. In *International Symposium on Algorithmic Game Theory*, 146–158. Springer.

Chen, L., and Tang, P. 2015. Bounded rationality of restricted turing machines. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, 1673–1674. International Foundation for Autonomous Agents and Multiagent Systems.

Chen, L.; Lin, F.; Tang, P.; Wang, K.; Wang, R.; and Wang, S. 2017. K-memory strategies in repeated games. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 1493–1498. International Foundation for Autonomous Agents and Multiagent Systems.

Devaine, M.; Hollard, G.; and Daunizeau, J. 2014. Theory of mind: did evolution fool us? *PloS One* 9(2):e87619.

Gilboa, I. 1988. The complexity of computing best-response automata in repeated games. *Journal of economic theory* 45(2):342–352.

Halpern, J. Y., and Pass, R. 2012. Iterated regret minimization: A new solution concept. *Games and Economic Behavior* 74(1):184–207.

Hao, D.; Li, K.; and Zhou, T. 2018. Payoff control in the iterated prisoner's dilemma. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 296–302. AAAI Press.

Hauert, C., and Schuster, H. G. 1997. Effects of increasing the number of players and memory size in the iterated prisoner's dilemma: a numerical approach. *Proceedings of the Royal Society of London B: Biological Sciences* 264(1381):513–519.

Hilbe, C.; Nowak, M. A.; and Sigmund, K. 2013. Evolution of extortion in iterated prisoner's dilemma games. *Proceedings of the National Academy of Sciences* 110(17):6913–6918.

Hilbe, C.; Traulsen, A.; and Sigmund, K. 2015. Partners or rivals? strategies for the iterated prisoner's dilemma. *Games and economic behavior* 92:41–52.

Knight, V.; Campbell, O.; Harper, M.; Langner, K.; Campbell, J.; Campbell, T.; Carney, A.; Chorley, M.; Davidson-Pilon, C.; Glass, K.; et al. 2016. An open reproducible framework for the study of the iterated prisoner's dilemma. *arXiv preprint arXiv:1604.00896*.

Knoblauch, V. 1994. Computable strategies for repeated prisoner's dilemma. *Games and Economic Behavior* 7(3):381–389.

Lindgren, K. 1992. Evolutionary phenomena in simple dynamics. In *Artificial life II*, 295–312.

McNally, L.; Brown, S. P.; and Jackson, A. L. 2012. Cooperation and the evolution of intelligence. *Proc. R. Soc. B* rspb20120206.

Megiddo, N., and Wigderson, A. 1986. On play by means of computing machines: preliminary version. In *Proceedings of the 1986 Conference on Theoretical aspects of reasoning about knowledge*, 259–274. Morgan Kaufmann Publishers Inc.

Meurer, A.; Smith, C. P.; Paprocki, M.; Čertík, O.; Kirpichev, S. B.; Rocklin, M.; Kumar, A.; Ivanov, S.; Moore, J. K.; Singh, S.; Rathnayake, T.; Vig, S.; Granger, B. E.; Muller, R. P.; Bonazzi, F.; Gupta, H.; Vats, S.; Johansson, F.; Pedregosa, F.; Curry, M. J.; Terrel, A. R.; Roučka, v.; Saboo, A.; Fernando, I.; Kulal, S.; Cimrman, R.; and Scopatz, A. 2017. Sympy: symbolic computing in python. *PeerJ Computer Science* 3:e103.

Nowak, M., and Sigmund, K. 1992. Tit-for-tat in heterogeneous populations. *Nature* 355:250–253.

Osborne, M. J., and Rubinstein, A. 1994. *A course in game theory*. MIT press.

Press, W. H., and Dyson, F. J. 2012. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences* 109(26):10409–10413.

Rapoport, A., and Chammah, A. M. 1965. *Prisoner's dilemma: A study in conflict and cooperation*, volume 165. University of Michigan press.

Rubinstein, A. 1986. Finite automata play the repeated prisoner's dilemma. *Journal of economic theory* 39(1):83–96.

Shoham, Y., and Leyton-Brown, K. 2008. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

Skrzypacz, A., and Hopenhayn, H. 2004. Tacit collusion in repeated auctions. *Journal of Economic Theory* 114(1):153 – 169.

Stewart, A. J., and Plotkin, J. B. 2012. Extortion and cooperation in the prisoner's dilemma. *Proceedings of the National Academy of Sciences* 109(26):10134–10135.

Thaler, R. H. 1988. Anomalies: The ultimatum game. *The Journal of Economic Perspectives* 2(4):195–206.

Zuo, S., and Tang, P. 2015. Optimal machine strategies to commit to in two-person repeated games. In *AAAI*, 1071–1078.