Repeated Prisoner's Dilemma

Sandeep Pathak

***Abstract*—Situations involving cooperation and defect can be modelled using the model of Prisoner's Dilemma. The equilibrium outcome for such *one shot game* is that both players defect. The interesting and rich situations arise when the game is played iteratively. The idea is that two players may choose to cooperate by the threat of punishment by the other player in future for cheating in the present period. In this review, we look for possibilities of cooperation in repeated Prisoner's Dilemma.**

***Index Terms*—Prisoner's Dilemma, Repeated Games, Nash Equilibrium, Subgame perfect equilibrium**

I. Introduction

GAME theory is the study of interactions among intelligent decision makers (*players*). In a game, each player has a set of *strategies* from which he can choose his move or *action*. Due to interactions, the *payoff* or return of a player, not only depend on his action but also on all the players' actions. Thus, players choose strategies which will maximize their payoffs. The analysis of such interactions to find optimal or equilibrium outcomes lies at the heart of the subject. Game Theory has tremendous applications in political science, ethics, psychology, philosophy, economics, computer science and evolutionary biology ([1], [2], [3]).

One of the most well-known example of a game is that of a *Prisoner's Dilemma*. It can model various social situations such as Duopoly, Arms Race, Common Property, Tariff wars between countries, etc. It involves two players who both have two strategies – *"cooperation"* (C) and *"defection"* (D). The preferences over outcomes for player $i(\neq j)$ is in the following order:

- i chooses D, j chooses C
- i chooses C, j chooses C
- i chooses D, j chooses D
- i chooses C, j chooses D

So, we can see that irrespective of what his partner chooses, a player always has an incentive to choose D to C. Thus, this game has a unique steady state equilibrium solution or Nash equilibrium (to be defined later in the text) which is (D,D).

Sandeep Pathak is a graduate student at Materials Research Centre, Indian Institute of Science, Bangalore 560 012, India, e-mail: sandeep@mrc.iisc.ernet.in

In many social settings, same players may encounter each other more than once. At every point of time, each player can decide her action depending on the *history* (set of previous outcomes). Thus, a strategy of a player takes the form of a decision rule which is a function of previous histories. For example, a strategy can be as follows – start with C and continue to cooperate until the other player defects (chooses D) and choose D forever (or till the completion of the game) from the point when the other player defects (*grim trigger strategy*). A lot of research has gone into finding strategies which will ensure cooperation as an equilibrium outcome in repeated two person *Prisoner's Dilemma* ([4], [5], [6], [7], [8], [9], [10]).

An interesting application of this idea can be in the modelling of situation of a medieval trade fair. In such a fair, a transaction typically involved a transfer of goods in exchange for a promissory note to be paid at the next fair. It would be interesting to investigate – *how honesty can be ensured in such a traders' community in which each trader has an urge to cheat?* The problem can be modelled as N players interacting via *Prisoner's Dilemma* (two at a time) repeatedly. Milgrom *et. al* ([11]) have studied such a model and concluded that the threat of publicity of dishonest behavior (community enforcement) can ensure honest trade. Their model also illuminates the role of judiciary in attaining socially desirable outcome. They have gone ahead and studied the situation if the judge himself is dishonest and how the system can sustain such threats.

Repeated Prisoner's Dilemma can be useful in other situations as well. For example, it can be used to model situation of *Predator inspection by Stickleback fish*. Sticklebacks moves in pairs towards predators to collect information for the community. Moving forward is like cooperating and holding back is like defecting. A lot of work ([12], [13], [14], [15]) has been done in investigating the strategies used by two sticklebacks in such expedition.

Suppose two persons play repeated *Prisoner's Dilemma* with different strategies. The strategy which do well against other strategies is the dominant one and will evolve in a population at the cost of others[8]. This is known as *evolutionary competition.* This is analogous to biological evolution in which evolution process takes place by differential reproduction of the more successful individuals. Axelrod conducted tournaments in 1980s to find dominant strategy in such an evolutionary competition for repeated Prisoner's Dilemma problem.

The plan of this review is as follows. We start with an introduction to two types of games (strategic form games (section II.A) and extensive form (section II.B) games) in section II. Various solution concepts like Nash equilibrium and subgame perfect equilibrium are discussed in sections A.3 and B.4 respectively. Ideas of Repeated Games are discussed in section III. Specifically, the discussion is on how the ideas of strategies, outcomes etc. gets modified in the repeated game settings. Various strategies are analyzed for their possible candidature for Nash equilibria and subgame perfect equilibria. In section IV, a detailed overview of the current research in the field of Repeated Games is given. This review is concluded with the discussion of open problems in section V.

II. Types of Games

Broadly, games can be classified into two categories depending upon, whether the moves in the game are *sequential* or *simultaneous*:

A. Strategic form or Normal form game

A *strategic form game* is the simplest form of the game in which players are asked to choose their actions simultaneously without the knowledge about other players' actions.

A.1 Definition

A strategic form game, $\langle N, S_i, u_i \rangle$ involves:

- A set of players, N
- A strategy set for each player, $S_i, i = 1, \ldots, N$
- A payoff function for each player, $u_i : \times_{i=1}^{N} S_i \rightarrow \Re, i = 1, \ldots, N$

A.2 Representation – Matrix

Two player strategic form game can be represented with the help of a payoff matrix (See fig. 1). If there are two players, 1 and 2, then rows represents strategies of player 1, while columns represents a strategy for player 2. Here, a possible outcome, (s_1, s_2), where, $s_1 \in S_1, s_2 \in S_2$, is represented by any cell of the payoff matrix. Payoffs for each players are mentioned in each cell. Few common examples of strategic form games are Prisoner's Dilemma, BoS, Matching Pennies, Stag Hunt, Hawk-Dove, etc. (See any standard text e.g. [3] for more details)

	C	**D**
C	**3, 3**	**0, 4**
D	**4, 0**	**1, 1**

Fig. 1. A representative payoff matrix for Prisoner's Dilemma Game.

A.3 Nash Equilibrium

The steady state equilibrium outcome is known as *"Nash equilibrium"*. Nash equilibrium (NE) prescribes an outcome with the property that no player can do better by choosing a different strategy, given that other players adheres to their Nash equilibrium strategies.

Notation – $S_{-i} = \times_{j \neq i} S_j$

The strategy profile s^* in a strategic game is a Nash equilibrium if,

$$u_i(s^*) \geq u_i(s_i, s_{-i}^*) \; \forall s_i \in S_i, \; \forall i \tag{1}$$

The Nash equilibrium for Prisoner's Dilemma problem (See fig. 1) is given by (D, D). Some points should be noted regarding Nash equilibrium:

- Nash equilibrium need not be unique for a game.
- A Nash equilibrium represents a steady state solution which can be expected when an ensemble of people drawn from a population plays the game simultaneously, with each person playing the game only once.
- A Nash equilibrium need not be socially optimum solution. For example, in fig. 1, the socially optimum outcome is (C, C).
- Nash equilibrium for strategic form games can be calculated using *best response functions.*

B. Extensive form game

It is convenient to represent the game in *"extensive form"* when the moves are sequential – A player is assigned to start the game. The next assigned player, depending on the action of first mover, chooses his action from his action set. The game goes on in this fashion until the terminal point is reached. The examples of such games are tit-for-tat, chess, parlor games like poker etc.

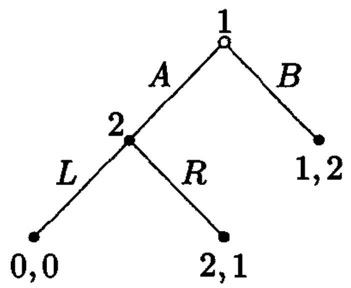

Fig. 2. An example of an extensive form game.

B.1 Representation – Tree

This game can be conveniently represented with the help of a tree diagram (fig. 2). The starting node is denoted by an open circle while the terminating node is represented by a filled circle. Each node is numbered with a player and the emanating branches are labelled by choices available to the player. The terminating node is labelled with the returns (payoffs) for the players involved. A feasible outcome in an extensive form game is any branch which starts at the starting node and ends in a terminating node.

B.2 Definition

An extensive form game, $\langle N, H, Z, P, u_i \rangle$, consists of

- A set of players, N
- A set of sequences, H with the following properties:
  - $\Phi \in H$
  - If $(a^k)_{k=1,\ldots,K} \in H$ and $L < K$, then $(a^k)_{k=1,\ldots,L} \in H$
  - If an infinite sequence $(a^k)_{k=1,\ldots} \in H$ satisfies $(a^k)_{k=1,\ldots,L} \in H$ for every positive integer L, then $(a^k)_{k=1,\ldots} \in H$

  Each member of H is called *history*. Each component of a history is an action taken by a player. A history $(a^k)_{k=1,\ldots,K}$ is called *terminal history* if it is infinite or there is no a^{K+1} such that $(a^k)_{k=1,\ldots,K+1} \in H$ i.e. no terminal history is a proper subhistory of any other sequence in H. That's why the terminology – terminal. Let's call the set of terminal histories as Z.
- a function (*player function*) that assigns a player to every non-terminal history, $P : H \backslash Z \rightarrow \mathcal{Z}$
- Payoff function over set of terminal histories for each player, $u_i : Z \rightarrow \mathcal{R}$, $i = 1, \ldots, N$

For example, consider fig. 2. Here,

- *Players*, $N = 1, 2$
- *Histories* – $H = \Phi, A, (A, L), (A, R), B$

	L	R
A	0,0	2,1
B	1,2	1,2

Fig. 3. Representation of extensive form game represented by fig. 2 in strategic form and its Nash equilibrium (marked with yellow colour).

- *Terminal histories* $Z = (A, L), (A, R), B$
- *Player function* $P(\Phi) = 1, P(A) = 2$
- *Payoff function*

$$\begin{aligned} u_1(A, L) &= 0, u_1(A, R) = 2, u_1(B) = 1 \\ u_2(A, L) &= 0, u_2(A, R) = 1, u_2(B) = 2 \end{aligned}$$

In extensive games, it is not required to specify the action set for each player explicitly. It can be deduced from the history set and the player function. The set of all actions available to the player who moves after history h is given by $Ac(h) = \{a : (h, a) \text{ is a history}\}$, For example, in fig. 2,

$$Ac(\Phi) = A, B, Ac(A) = L, R$$

An extensive game can be finite or infinite. It can be infinite in two ways:

- Length of longest terminal history is infinite – *Infinite horizon games or Games with infinite horizon*
- There are infinite number of terminal histories.

B.3 Strategies and Outcomes

The concept of strategy is very crucial in extensive form games.

- *Definition* – A strategy for player i in an extensive form game is a function that assigns to each history h after which it is player is turn to move, an action in $Ac(h)$.
- *For example*
  1. In fig. 2, player 1 moves only at the start of the game and thus have two strategies, namely, A and B. Player 2 also has two strategies, namely L and R. Thus, an extensive form game can be reduced to a strategic form game (See fig. 3) and we can calculate its Nash equilibrium using standard techniques. Nash equilibrium for this problem is (A, R) and (B, L).

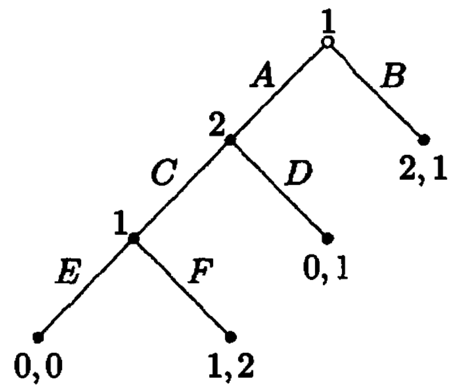

Fig. 4. An example of an extensive form game in which player 1 moves both before and after player 2.

	C	D
AE	0,0	0,1
AF	1,2	0,1
BE	2,1	2,1
BF	2,1	2,1

Fig. 5. Representation of extensive form game represented by fig. 4 in strategic form and its Nash equilibrium (marked with yellow colour).

2. In fig. 4, player 1 moves both before and after player 2. In each case, he has two actions, so he has four strategies, namely AE, AF, BE and BF. AE represents the strategy in which player 1 chooses A at the start of the game and history (A, C), he chooses E. Similarly, we can interpret nomenclature for his other strategies. One might wonder that how come action E or F possible after player 1 chooses B at the start, in which case, the history $(A.C)$ never occurs? Remember that our definition says that a strategy of any player i specifies an action for every history after which it is his turn to move, *even for histories that, if the strategy is followed, do not occur.* Player 2 in this game has two strategies – C and D. The Nash equilibrium are calculated after reducing the game to its strategic form (See fig. 5).

Outcomes The terminal history determined by a strategy profile is called as an outcome of the game.

B.4 Subgame perfect equilibrium

The strategic form game in fig. 3 has two Nash equilibrium (A, R) & (B, L). (A, R) gives more payoff to player 1 who has the first say. So it is natural that player 1 who is assumed to be rational will always prefer (A, R). Thus, (B, L) is some sort of a perturbed Nash equilibrium and is not robust. So the question arises – How to evaluate a robust NE ?

The answer to the above question is that we need to calculate Nash equilibrium for every subgame and not just the full game. This gives rise to the concept of *subgame perfect equilibrium.*

Subgame : Definition

Let $\tau = \langle N, H, Z, P, u_i \rangle$. For any non-terminal history h, we can define a subgame, $\tau(h)$ following the history h as following extensive game:

- *Players* – Players in τ
- *History Set* – The set of all sequences h' of actions such that (h, h') is a history
- *Player function* $P(h, h')$ is assigned to each proper subhistory h' of a terminal history.
- *Preferences* Each player prefers h' to h'' if he preferred (h, h') to (h, h'') in τ

In summary, a subgame of an extensive game is, simply, the game corresponding to a subtree of the tree corresponding to the full game.

A subgame perfect equilibrium ($s^* = (s_i^*, s_{-i}^*)$) induces Nash equilibrium in every subgame.

- A subgame perfect equilibrium is a Nash equilibrium.
- In no subgame, any player i can do better by choosing a strategy different from s_i^*, given that other players adheres to s_{-i}^*.
- A subgame perfect equilibrium need not be unique.
- *Existence* – Every finite extensive game has a subgame perfect equilibrium.
- *Calculation of Subgame perfect equilibrium* – A subgame perfect equilibrium for finite horizon games is calculated using *Backward Induction* technique. Let's define the *length of the subgame* to be the length of the longest history in the subgame. The process of backward induction is as follows – We start by finding optimum action for players who move in the last game (length 1). Then taking these actions as given, we find optimum actions of the players who move first in the subgames of length 2. The process continues until we get to the beginning of the game. The process is illustrated in figs. 6 and 7 for the games in figs. 2 and 4.
- It is known that in subgame perfect equilibrium for ticktacktoe and chess, either (a) one player has a winning strategy, or (b) both players have

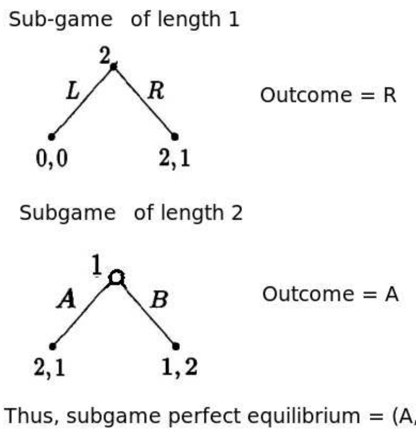

Fig. 6. Use of Back Induction to find subgame perfect equilibrium of extensive game shown in fig. 2

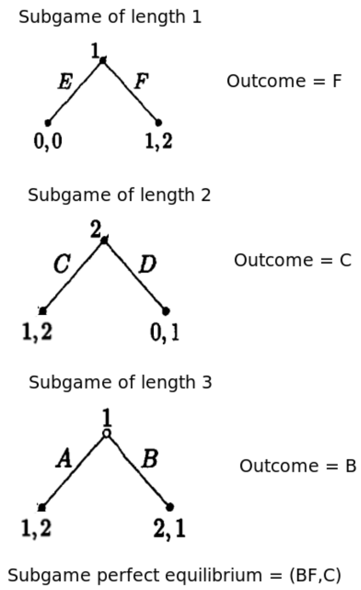

Fig. 7. Use of Back Induction to find subgame perfect equilibrium of extensive game shown in fig. 4

strategies which guarantees at worst a draw. In ticktacktoe (b) is true ([16]) though not much is known about chess. The empirical studies ([17], [18]) suggest that Black does not have a winning strategy but it has not been proved.

III. Repeated Prisoner's Dilemma

We saw that the *one shot* Prisoner's Dilemma, as it is called, has a unique Nash equilibrium which is that both player choose to defect. But when two players interact repeatedly, each player can condition his action at each point in time on the other player's previous actions and this may lead to cooperation. The main idea is that players may be forced to cooperate by the threat of future punishment from their partners for deterring from the socially favorable behavior.

A repeated Prisoner's Dilemma is an extensive game with simultaneous moves. A history is a sequence of action profiles in strategic from of Prisoner's Dilemma. For example – $((C,C),(C,C),(C,D),(D,D),(D,D),\ldots)$. After every non-terminal history, each player chooses an action (C or D) based on the history. A repeated Prisoner's Dilemma problem can be finite (T-period repeated Prisoner's Dilemma) or infinite, depending upon which the equilibrium outcomes vary.

A. Preferences : Discounting

Let's denote an outcome of a repeated game by sequence of outcomes of the strategic game, $(a^t)_{t=1,\ldots,T}$. Outcomes of repeated games are evaluated using discounted sum:

$$u_i(a^1)+\delta_i u_i(a^2)+\delta_i^2 u_i(a^3)+\cdots+\delta_i^{T-1}u_i(a^T) = \sum_{t=1}^{T}\delta_i^{t-1}u_i(a^t)$$

where $0<\delta_i<1$ is the discount factor which represents the patience level of player i. More the player i values his future, more is the value of δ_i. A very impatient player will have $\delta_i \approx 0$, while on the other, a very patient player will have $\delta_i \approx 1$. Let's assume that $\delta_i = \delta \ \forall i = 1,\ldots,N$, i.e. the patience levels of all the player are same.

Discounted Average: A player's preference over a sequence $(w^1, w^2, \ldots)$ of payoffs is represented by the discounted sum $V = \sum_{t=1}^{\infty}\delta^{t-1}w^t$. For the given sequence, we ask if there is a value c such that the player is indifferent between the given payoff sequence and $(c,c,\ldots)$? For that to happen,

$$c+c+\cdots = \frac{c}{1-\delta} = V \Rightarrow c = (1-\delta)V$$

Thus, the ***discounted average*** of any payoff sequence $(w^1, w^2, \ldots)$ for the discount factor, δ is $(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}w^t$. For any δ, and any number c, the discounted average of the constant sequence of payoffs $(c,c,\ldots)$ is equal to c.

B. Strategies

The concept of strategies has to be modified to be applied in the repeated setting. A strategy for a player i takes the form of a decision rule which depends on the previous history at any point in time. For example –

Fig. 8. Schematic representation of Grim trigger strategy

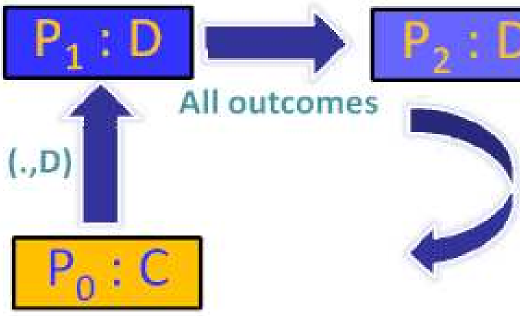

Fig. 9. Schematic representation of 1 period limited punishment strategy

- *Grim trigger strategy* Start with C, choose C as long as other player chooses C, if in any period other player chooses D, choose D in every subsequent period.
- *Tit-for-tat* Start with C, choose the same action as other players action in previous period.
- *Limited Punishment* Punish deviations for n periods and revert back to C
- *Delayed initiation* of punishment for defect
- *Win-stay, lose-shift (Pavlov)* Start with C, choose the same action again if the outcome was relatively good and switch action if it was not.

B.1 Strategies : Representations

Strategies can be represented, schematically, in form block diagrams as in figs. 8 and 9. For example, Grim trigger strategy can be thought of as a two state strategy Start with state P_0 in which C is chosen state changes to P_1 in which D is chosen, if the partner defects (See fig. 8).

C. Finitely repeated Prisoner's Dilemma

The discounted average for T-period repeated Prisoner's Dilemma is given by $(1-\delta)\sum_{t=1}^{T}\delta^{t-1}w^t$.

C.1 Nash Equilibrium

Suppose a player chooses D in each period, other player has no choice but to choose D in each period. Thus, this is a Nash equilibrium as no player can do better by deviating from this strategy, given that the other play sticks to D in all periods. In fact, every Nash equilibrium generates the same same outcome path, i.e. (D, D) in each period.

There is incentive to defect in the last step as it cannot be punished. Thus, (D, D) is the outcome in the last period in any Nash equilibrium. Suppose player 1 chooses C in the $(T-1)^{th}$ period. If player 2 chooses C, D is the better choice for player 1. If player 2 chooses D, D gives more payoff to player 1. Thus, (D, D) is the outcome in any NE in $(T-1)^{th}$ period. Using Backward Induction, every Nash equilibrium has outcome (D, D) in each period.

C.2 Subgame perfect equilibrium

Every subgame perfect equilibrium of an extensive game is a NE. Thus, every subgame perfect equilibrium of a finitely repeated PD generates the outcome (D,D) in every period.

D. Infinitely repeated Prisoner's Dilemma

D.1 Nash Equilibria

Clearly, the strategy pair in which both players chooses (D, D) is a Nash equilibrium of the infinitely repeated Prisoner's Dilemma problem. We look for other strategy pairs which yields cooperation as an equilibrium outcome of the infinitely repeated game:

- ***Grim Trigger Strategies (GTS)***
  Suppose that player 1 uses the GTS. If player 2 also uses the same strategy, then the outcome is (C, C) in every period and discounted average for both the players is 2.
  If player 2 defects in some period, player 1 uses D for every subsequent period. The equilibrium outcome of player 2 would be to choose D. Thus, the sequence of outcomes from the point of defection are $(C, D), (D, D), (D, D), \ldots$ and payoff sequence for player 2 is $(3, 1, 1, \ldots)$ whose discounted average is

$$(1-\delta)(3+\delta+\delta^2+\delta^3+\cdots)$$
$$= 3(1-\delta)+\delta$$

  So, for both players playing GTS to be a Nash equilibrium,

$$3(1-\delta)+\delta \leq 2$$

$$\boxed{\Rightarrow \delta \geq \tfrac{1}{2}}$$

- ***Limited k period punishment***
  Suppose player 1 uses limited k period punishment strategy, where k is an integer. If player

2 defects in some period, the payoff sequence is $(3, \underbrace{1, 1, \ldots, 1}_{k\ times})$ whose discounted average is

$$\begin{aligned}(1-\delta)(3+\delta+\delta^2+\delta^3+\cdots+\delta^k) \\ = 3(1-\delta)+\delta(1-\delta^k)\end{aligned}$$

If player 2 sticks to the strategy, then discounted average is

$$\begin{aligned}(1-\delta)\cdots 2\cdot(1+\delta+\delta^2+\delta^3+\cdots+\delta^k) \\ = 2(1-\delta^{k+1})\end{aligned}$$

For the strategy pair to be Nash equilibrium, So, for both players playing GTS to be a Nash equilibrium,

$$3(1-\delta)+\delta(1-\delta^k) \le 2(1-\delta^{k+1})$$

$$\boxed{\Rightarrow \delta^{k+1}-2\delta+1 \le 0}$$

- $k=1$, there is no solution satisfying above inequality. Thus, one period of punishment is not severe enough to discourage defection.
- $k=2$, $\delta \ge 0.62$
- $k=3$, $\delta \ge 0.55$

Thus, if players are patient enough, then short punishment can ensure the desirable outcome.

- ***Tit-for-tat(TFT)***
  Suppose player 1 sticks to TFT. If player 2 also sticks to TFT, then his discounted average is 2. If player 2 defects in some period, the player 1 uses D in the subsequent period. Player 2 has two options for the next period

1. He chooses D so that all subsequent outcomes are (D, D). Discounted average in this case is

$$\begin{aligned}d_1 = (1-\delta)(3+\delta+\delta^2+\delta^3+\cdots) \\ = 3(1-\delta)+\delta\end{aligned}$$

2. He chooses C and in the next period when player 1 chooses C, he chooses D. Thus, the outcome in such a case oscillates between (D, C) and (C, D). The discounted average in this case is

$$\begin{aligned}d_2 = (1-\delta)\cdot 3\cdot(1+\delta^2+\delta^4+\cdots) \\ = \frac{3}{1+\delta}\end{aligned}$$

For both players playing TFT to be Nash equilibrium,

$$d_1 \le 2 \;\;\&\;\; d_2 \le 2$$

$$\boxed{\Rightarrow \delta \ge \tfrac{1}{2}}$$

D.2 Nash equilibrium payoffs in an infinitely repeated Prisoner's Dilemma

Till now, we were concerned mainly in finding whether a given strategy pair is a Nash equilibrium for an infinitely repeated Prisoner's Dilemma or not. Now, let's ask the following question – Given a pair of numbers, is there a Nash equilibrium in which the discounted average payoff profile is equal to the given pair?

1. ***Feasible discounted average payoffs***
   Let's consider the simple case when all the player are very patient ($\delta \approx 1$). Consider an outcome in which the sequence $(C, C), (D, C), (C, D)$ is repeated. The discounted average payoff of a player in this case will be close to his average payoff in the sequence. So player 1's discounted average payoff is $\frac{1}{3}(2+3+0) = \frac{5}{3}$. Similarly, discounted average payoff player 2 is $\frac{5}{3}$ and thus discounted average payoff profile is $(\frac{5}{3}, \frac{5}{3})$. Thus, it is easy to see that in any outcome path, when $\delta \approx 1$, the discounted average profile is the weighted average of payoff profiles in the strategic game (here, $(2, 2), (3, 0), (0, 3), (1, 1)$) as these are the four possibilities in any point of time. This leads us to the concept of *Feasible payoff profiles.*
   Feasible payoff profiles – The set of feasible payoff profiles of a strategic game is the set of all convex combinations of payoff profiles in the game.
   If $\delta \approx 1$, the set of discounted average payoff profiles generated by the outcome paths of an infinitely repeated game is approximately equal to the set of feasible payoff profiles in the component strategic game.
2. ***Nash equilibrium discounted average payoffs***
   Clearly, by choosing D in every period, all players can, at least, obtain a payoff of $u_i(D, D)$. Thus, in any Nash equilibrium of an infinitely repeated Prisoner's Dilemma, each players' discounted average payoff is at least $u_i(D, D)$.
   Let (x_1, x_2) be a feasible pair of payoffs in a Prisoner's Dilemma with $x_i > u_i(D, D)$ for $i = 1, 2$. Then, by definition of feasibility, we can find a finite sequence of outcomes, $A = (a^1, a^2, \ldots, a^k)$ of the game for which player i's payoff approximates x_i for $i = 1, 2$ for $\delta \approx 1$. Thus, we can design a strategy pair in which A is repeated infinitely. This can be done analogous to grim trigger strategy formulation – play the game so that

the sequence A is maintained until the partner breaks it. On breaking of the sequence, play D. We can check that both players sticking to this strategy is the Nash equilibrium of the infinitely repeated Prisoner's Dilemma as $\delta \approx 1$.
There is a ***Nash's folk Theorem*** for arbitrary value of δ which we state without proof:

- For any value of δ , discounted average payoff is at least $u_i(D, D)$ in any Nash equilibrium of infinitely repeated Prisoner's Dilemma.
- Let $X = (x_1, x_2)$ be a feasible payoff pair s.t. $x_i > u_i(D, D)$, there exists $\bar{\delta}$ s.t. if $\delta > \bar{\delta}$, then infinitely repeated Prisoner's Dilemma has a Nash equilibrium in which discounted average payoff profile is X.
- Infinitely repeated Prisoner's Dilemma has a Nash equilibrium in which the discounted average payoff of each player is $u_i(D, D)$ for any value of δ.

D.3 Subgame perfect equilibrium

Subgame perfect equilibrium in the repeated game settings are calculated using *one deviation property*:
One Deviation Property – No player can increase his payoff by changing his action at the start of any subgame in which he is the first mover, given other players strategies and the rest of his own strategy. Clearly, if the one deviation property is satisfied, the outcome generates Nash equilibrium in every subgame and is, thus, a subgame perfect equilibrium.

The Nash equilibria in which both players chooses D after every history are subgame perfect equilibrium equilibrium also as if one player chooses D in every period, the other cannot deviate from D in any period. We look for other strategy pairs –

1. ***Grim Trigger Strategies (GTS)***
Grim trigger strategies do not satisfy one deviation property and is thus *not* a subgame perfect equilibrium.
Consider the subgame following the outcome (C, D). Let player 1 adheres to GTS. If player 2 adheres to GTS, then the outcome is (D, C) in first period and (D, D) in subsequent periods. Discounted average payoff in this case is
$$(1 - \delta)(0 + \delta + \delta^2 + \delta^3 + \cdots +) = \delta$$
If player 2 deviates from GTS, then the outcome is (D, D) in every period and thus, the discounted average payoff is 1. Thus, it is not optimal for player 2 to adhere to GTS. In fact, it can be shown that for $\delta > \frac{1}{2}$, it is not optimal for a player to punish deviation if the other player reverts back to C.

Fig. 10. A schematic block diagram showing the modified Grim Trigger strategies.

2. ***Modified Grim Trigger strategies (MGTS)***
If we modify GTS slightly, then we get a strategy which is the subgame perfect equilibrium of infinitely repeated Prisoner's Dilemma problem. The strategy prescribes to start with C and change to D for all outcomes except (C, C) (See fig. 10). Consider the subgame following empty history or a history containing (C, C). When both players follow the strategy, the outcome is (C, C) and thus, discounted average payoff is 2 for both the players. If player 1 deviates in the first period and chooses D but otherwise stick to the strategy, then outcome is $(D, C), (D, D), (D, D), \ldots$. The discounted average payoff in this case is
$$(1 - \delta)(3 + \delta + \delta^2 + \delta^3 + \cdots +) = 3 - 2\delta$$
which is less than 2 for $\delta \geq \frac{1}{2}$.
Consider any other subgame. The outcome is (D, D) in every period. Since (D,D) is a NE of PD, deviation is not optimal.
Thus, this modified Grim Trigger strategy is a subgame perfect equilibrium for $\delta \geq \frac{1}{2}$.
3. ***Limited Punishment*** –
This strategy pair is also not a subgame perfect equilibrium for the same reasons for which GTS pair is not a subgame perfect equilibrium.
4. ***Tit-for-tat (TFT)*** –
A players behavior in a subgame depends only on the last outcome. So, consider 4 cases –
   - *History ending in* (C, C) - This is subgame perfect equilibrium if $\delta \geq \frac{1}{2}$
   - *History ending in* (C, D) *(or* (D, C)*)*
     – If both the players adhere to TFT, the subsequent outcomes are $(D, C), (C, D), (D, C), (C, D), \ldots$ and the discounted average payoff for player 1 is $d_1^1 = 3/(1 + \delta)$.
     – If player 1 deviates in the first period but adheres to TFT later, the outcomes are

(C, C). Discounted average payoff is $d_1^2 = 2$ in this case.

– $d_1^1 \geq d_1^2 \Rightarrow \delta \leq \frac{1}{2}$, Similar analysis for player 2 yields $\delta \geq \frac{1}{2} \Rightarrow \delta = \frac{1}{2}$.

- *History ending in* (D, D) subgame perfect equilibrium if $\delta \leq \frac{1}{2}$.

Thus, strategy pair with both strategies as TFT is a subgame perfect equilibrium only when

$$\boxed{\delta = \frac{1}{2}}$$

D.4 Subgame perfect Folk Theorem for infinitely repeated Prisoner's Dilemma

There is a corresponding Folk Theorem for existence of subgame perfect equilibrium as well, which is stated without proof here –

- For any value of δ , discounted average payoff is at least $u_i(D, D)$ in any subgame perfect equilibrium of infinitely repeated Prisoner's Dilemma.
- Let $X = (x_1, x_2)$ be a feasible payoff pair s.t. $x_i > u_i(D, D)$, there exists $\bar{\delta}$ s.t. if $\delta > \bar{\delta}$, then infinitely repeated Prisoner's Dilemma has a subgame perfect equilibrium in which discounted average payoff profile is X.
- Infinitely repeated Prisoner's Dilemma has a subgame perfect equilibrium in which the discounted average payoff of each player is $u_i(D, D)$ for any value of δ.

IV. Recent Research Work

A. *Reciprocal Altruism among Sticklebacks*

A much discussed example of possible application of repeated Prisoner's Dilemma is that of predator inspection by sticklebacks (*Gasterosteus aculeatus*). Sticklebacks often moves in pairs for predator inspection. To obtain information about the predator, stickleback has to move closer to the predator. Moving forward is like cooperating, while holding back is like defecting.

Milinski ([12]) reported an experiment using a two compartment tank separated by a glass. A stickleback was placed in one compartment while a cichlid, which mimics the predator, was placed in the other compartment. A mirror was put in the stickleback compartment so that when stickleback moved forward, it got an impression that she had a partner mimicking her. In second condition, mirror was put at an angle s.t. when stickleback moved forward, the image retaliated. It was found that the fish had moved much closer to the cichlid in first condition. This meant that some kind of retaliatory strategy was used by Stickleback. They had proposed tit-for-tat strategy for this behavior. Later, they suggested that the strategy among sticklebacks can be closer to Pavlov than to tit-for-tat ([15]).

Similar kind of behavior was reported in guppier approaching a pumpkinseed fish ([13], [14]).

B. *Role of judiciary in revival of trade*

The motivation of this section comes from the social settings in the medieval world. The trade used to happen mainly at trade fairs in which people from distant locality used to come and trade. A transaction typically involved a transfer of goods in exchange for a promissory note to be paid at the next fairs. It might be possible that a trader once seen will never be seen again. There was no proper judiciary as well at that period. The situation is modeled by repeated Prisoner's Dilemma in which players are matched randomly in every period. Honesty is like cooperating and cheating is like defecting. Clearly each trader has an urge to cheat as he may never be caught. Then how can trust be ensured in such a society? This is the work done by Milgrom *et. al.* ([11]).

Milgrom *et. al.* have suggested two measures by which trust can be ensured in a trading community –

1. *Community enforcement or Word of mouth* – The idea is that traders can be stopped from cheating by the threat of spoiling of reputation. Thus, reputation system can serve honest trade. But this idea is valid only for small communities when the cost of sharing information about dishonest behavior is not too high. Strategy is called Adjusted TFT - Players are matched according to some matching rule in a given period. A player plays D if she played C in previous period and her opponent had played D on the previous period. Otherwise, player plays C.
2. *Judiciary* –
   If players are matched in such a way that they never have to face the same player again, then the outcome in any Nash equilibrium is to play cheat for all the players. Thus, a merchant Law or judiciary is required.
   Milgrom *et. al.* introduced an additional actor LM in N player repeated Prisoner's Dilemma problem, who serves both as a repository of information and as an adjugator of disputes. This game with additional actor is called as *LM system stage game.* The key features of this game are –
   (a) Players may query the LM about their current partner at utility cost $Q > 0$. Whatever is told

back becomes common knowledge.

(b) Two trader play Prisoner's Dilemma.

(c) Either player may appeal to LM at a personal cost $C > 0$, only if he has queried the LM

(d) On appeal, LM awards judgment, J, to the plaintiff if appeal is valid.

(e) If judgment is awarded, defendant may pay it at personal cost $f(J)$

(f) Any unpaid judgments are recorded by LM and becomes part of permanent record

They introduced a strategy called *LM System Strategy* ($LMSS$) –

- At substage (a), player queries LM only if he has no unpaid judgements
- At substage (b), if there is no query or it is established that either of the players has unpaid judgments, then both the players play D otherwise both play C
- At substage (c), if both parties queried at substage (a) and exactly one cheats at substage (b), then the victim appeals to the LM
- At substage (d), on valid appeal, judgment J is awarded.
- At substage (e), defendant pays judgment iff he has no other outstanding judgment.

They showed that if the costs of querying and appealing to LM is not too high, then $LMSS$ is the strategy that supports honest trade.

They also studied situation in which the judge may threaten to sully the reputations of the honest traders unless they pay bribe (Dishonest LM) and how system can survive such threats.

C. Axelrod's Tournaments

In 1980s, Axelrod organized two tournaments between strategies for repeated Prisoner's Dilemma. The entries were pitted against each other on a round-robin basis. Tit-for-tat was victorious (submitted by Anatol Rapoport) ([19], [20], [4]).

Using the entries in the tournament, Axelrod simulated an evolutionary environment in which the strategies that do well reproduce faster than the other strategies using techniques of Genetic algorithm ([8]). The process is analogous to biological evolution in which evolution process takes place by differential reproduction of the more successful individuals. Strategies were matched against each other and representatives of the one giving higher payoff were increased at each step after a large number of generations, TFT had the most representatives in the population.

Although Axelrod's results are very interesting but they are not very robust. A large community ([21], [22], [23], [24], [25]), believes that Tit-for-tat is not the strategy which dominates such an evolutionary competition. Also, Axelrod's simulation had not included mutation.

D. Other works

Akimov ([26]) applied similar approach in simulating evolutionary environment for N-person iterated Prisoner's Dilemma problem using automata theory exhibiting cooperation.

Altruism in finitely repeated Prisoner's Dilemma We saw that the defection is the dominant strategy in a finitely repeated PD. Andreoni ([27]) showed that if players have incomplete information, then cooperation in early rounds is rational (pretending to be altruistic) to develop a reputation for cooperation. Their work suggested that all models of altruism are combination of pure altruism, duty & reciprocal altruism. Also, they pointed out that the players need not be altruistic for cooperation. Cooperation is possible if the players believe that there are other altruistic players in the society.

THIS is a small attempt at introducing ideas of repeated games and a very brief account of current research in the field of repeated Prisoner's Dilemma has been provided here. The interesting readers may benefit from a detailed chronological bibliography available on Axelrod's Homepage (`http://www-personal.umich.edu/~axe/`).

V. Conclusions

Strategic and extensive form games and various solution concepts related to their equilibrium outcomes were reviewed.

We saw how these concept can be extended to be applicable to repeated games. Defection is the only equilibrium solution in strategic game of Prisoner's Dilemma, but in the repeated version, not only we found that cooperation is possible but there is a rich possibility of equilibrium involving combination of all the four outcomes of strategic PD game.

We saw two direct applications of repeated Prisoner's Dilemma in some detail - Predator inspection by Sticklebacks and role of judiciary in the revival of trade.

A. Suggestions for future work

Axelrods tournaments and simulations suggest that tit-for-tat is the dominant strategy in evolutionary competition involving Prisoner's Dilemma, while views differ on this point and it need further investigation.

It would also be interesting to find dominant strategy in N-player repeated Prisoner's Dilemma.

References

[1] Martin J. Osborne and Ariel Rubinstein, *A Course in Game Theory*, The MIT Press, 1st edition, 1994.

[2] Roger B. Myerson, *Analysis of Conflict*, Harvard University Press, 1997.

[3] Martin J. Osborne, *An Introduction of Game Theory*, Oxford University Press (Indian Edition), 1st edition, 2006.

[4] Robert Axelrod, *The Evolution of Cooperation*, NY:Basic Books, 1st edition, 1984.

[5] Robert Axelrod and Douglas Dion, "The further evolution of cooperation", *Science*, vol. 242, pp. 1385–1390, 1988.

[6] Wu Jianzhong and Robert Axelrod, "How to cope with noise in the iterated prisoner's dilemma", *Journal of Conflict Resolution*, vol. 39, pp. 183–189, 1995.

[7] Robert Axelrod, "On six advances in cooperation theory", *Analyse and Kritik*, vol. 22, pp. 130–151, 2000.

[8] Robert Axelrod, "The evolution of strategies in the iterated prisoner's dilemma", 2001, Available on the web at `http://www-personal.umich.edu/~axe/research/Evolving.pdf`.

[9] Rick L. Riolo, Michael D. Cohen, and Robert Axelrod, "Evolution of cooperation without reciprocity", *Nature*, vol. 414, pp. 441–443, 2001.

[10] Robert Axelrod, David E. Axelrod, and Kenneth J. Pienta, "Evolution of cooperation among tumor cells", *PNAS*, vol. 103, pp. 13474–13479, 2006.

[11] Paul R. Milgrom, Douglous C. North, and Berry R. Weingast, "The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs", *Economics and Politics*, vol. 2, pp. 1–23, 1990.

[12] Manfred Milinski, "Tit for tat in sticklebacks and the evolution of cooperation", *Nature*, vol. 325, pp. 433–435, 1987.

[13] L.A. Dugatkin, "Do guppies play tit for tat during predator inspection visits?", *Behavioral Ecology and Sociobiology*, vol. 23, pp. 395–399, 1988.

[14] L.A. Dugatkin, "Dynamics of the titfortat strategy during predator inspection in the guppy (poecilia reticulata)", *Behavioral Ecology and Sociobiology*, vol. 29, pp. 127–132, 1991.

[15] Manfred Milinski, "Cooperation wins and stays", *Nature*, vol. 364, pp. 12–13, 1993.

[16] Martin Gardner, *The Scientific American book of mathematical puzzles and diversions*, New York: Simon & Schuster, 1959.

[17] Ernst Zermelo, "über eine anwendung der mengenlehre auf die theorie des schachspiels", 1913, in *Proceedings of the fifth international congress of mathematicians (Volume II)*.

[18] Ulrich Schwalbe and Paul Walker, "Zermelo and the early history of the game theory", *Games and Economic Behavior*, vol. 34, pp. 123–137, 2001.

[19] Robert Axelrod, "Effective choice in the prisoners dilemma", *Journal of Conflict Resolution*, vol. 24, pp. 3–25, 1980.

[20] Robert Axelrod, "More effective choice in the prisoners dilemma", *Journal of Conflict Resolution*, vol. 24, pp. 379–403, 1980.

[21] Jack Hirshleifer and Juan Carlos Martinez Coll, "What strategies can support the evolutionary emergence of cooperation?", *Journal of Conflict Resolution*, vol. 32, pp. 367–398, 1988.

[22] Bruce G. Linster, "Evolutionary stability in the repeated prisoners' dilemma played by two-state moore machines", *Southern Economic Journal*, vol. 58, pp. 880–903, 1992.

[23] John H. Nachbar, "Evolution in the finitely repeated prisoner's dilemma", *Journal of Economic Behavior and Organization*, vol. 19, pp. 307–326, 1992.

[24] Martin A. Nowak and Karl Sigmund, "Tit for tat in heterogeneous populations", *Nature*, vol. 355, pp. 250–253, 1992.

[25] Bruce G. Linster, "Stochastic evolutionary dynamics in the repeated prisoners' dilemma", *Economic Inquiry*, vol. 32, pp. 342–357, 1994.

[26] Vladimir Akimov and Mikhail Soutchanski, "Automata simulation of n-person social dilemma games", *Journal of Conflict Resolution*, vol. 38, pp. 138–148, 1994.

[27] James Andreoni and John H. Miller, "Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence", *Economic Journal*, vol. 103, pp. 570–585, 1993.