Finite Iterated Prisoner's Dilemma Revisited: Belief Change and End-game Effect

Jiawei Li
University of Nottingham
Jubilee Campus, Wollaton Road
Nottingham, United Kingdom
jwl@cs.nott.ac.uk

Graham Kendall
University of Nottingham
Jubilee Campus, Wollaton Road
Nottingham, United Kingdom
gxk@cs.nott.ac.uk

ABSTRACT

We develop a novel Bayesian model for the finite Iterated Prisoner's Dilemma that takes into consideration belief change and end-game effect. According to this model, mutual defection is always the Nash equilibrium at any stage of the game, but it is not the only Nash equilibrium under some conditions. The conditions for mutual cooperation to be Nash equilibrium are deduced. It reveals that cooperation can be achieved if both players believe that their opponents are likely to cooperate not only at the current stage but also in future stages. End-game effect cannot be backward induced in repeated games with uncertainty. We illustrate this by analyzing the unexpected hanging paradox.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Economics; I.2.1 [**Applications and Expert Systems**]: Games

General Terms

Game theory, iterated Prisoner's Dilemma, end-game effect, belief change

1. INTRODUCTION

In a finite Iterated Prisoner's Dilemma (IPD), the only Nash equilibrium is mutual defection at each stage of the game. This is concluded by means of the so-called backward induction: both players will choose to defect at the last stage. Given mutual defection at the last stage, the optimal strategy at the second to last stage is to defect for each player. And so on, back to the first stage. Each rational player would play the strategy 'All Defect' (AllD). However, many experimental tests have revealed that cooperation between subjects at early stages of a finite IPD can be achieved if there are sufficient iterations [1,2,7,8,10]. Experimental tests show that the rate of cooperation at the beginning of a finite IPD is significantly high and it tends to be rare at the end of the game.

One of the many explanations of the deviation from theoretical prediction is incomplete information of the types of the players, that is, if one player is not absolutely certain about the opponent's payoff or one player assigns a probability that the other will play a conditional cooperating strategy, cooperation may be a rational choice [3,5,9]. For a review of IPD strategies (include recently appeared group strategies), see [4,6].

In a game of incomplete information, the strategies of players depend on their beliefs –estimations of some pieces of unknown information ex ante. Generally, a piece of information unavailable to one player is known by another (or some others). The players will therefore utilize the asymmetric information to maximize their payoffs. When the game is repeated many times, there is the possibility that some players change their beliefs by learning. In this paper, we follow the assumption of incomplete information and develop a new Bayesian model for finite IPD that takes into consideration belief change and end-game effect.

Copyright is held by the author/owner(s).

BQGT'10 May 14-16, 2010 Newport Beach, California USA
ACM 978-1-60558-919-0.

2. FINITE IPD

The general form of the prisoner's dilemma game is shown in Figure 1.

		COL	
		C	D
ROW	C	(*R*, *R*)	(*S*, *T*)
	D	(*T*, *S*)	(*P*, *P*)

Figure 1: General formulation of the Prisoner's Dilemma payoff matrix.

where R, S, T, and P denote, respectively, Reward for mutual cooperation, Suckers payoff, Temptation to defect, and Punishment for mutual defection, and $T > R > P > S$ and $2R > (T + S)$. The two constraints motivate each player to play non-cooperatively and prevent any incentive to alternate between cooperation and defection. In an IPD game, ROW and COL have to choose their strategies repeatedly, and have the option to retain a memory of the previous behaviors of both players.

Kreps et al explained the emergence of cooperation in finite IPD by means of incomplete information [5]. They suggested that the players might not be absolutely certain that their opponent would play AllD. When COL believes that there is a probability that ROW will play Tit-for-Tat (TFT), COL may choose to cooperate in order to achieve a higher payoff than that of mutual defection. If COL's belief is common knowledge, ROW knows that COL may choose to cooperate in response to ROW's strategy. Then, mutual cooperation may be achieved as a sequential equilibrium.

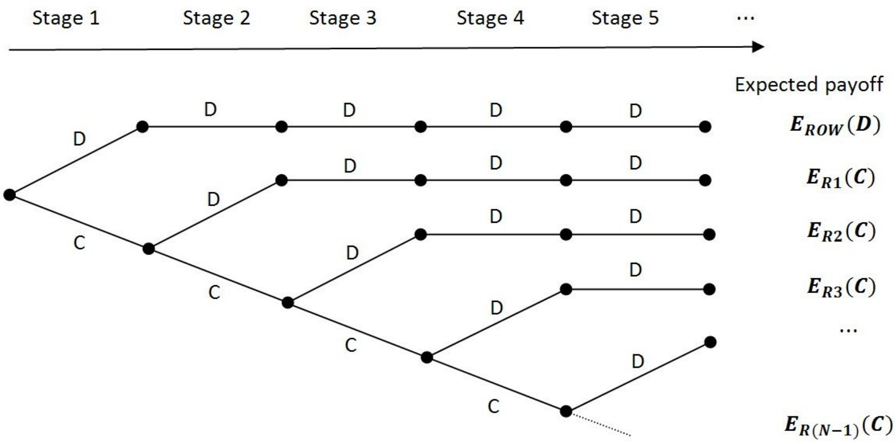

Figure 2: Strategy tree for ROW (COL) when choosing their strategy at stage one.

The model of [5] assumes that only one player's type is uncertain. With this model, the influence of end-game effect is described by a critical value such that, if the number of iterations remaining is more than this value, choosing to cooperate is strictly better. In the following section, we consider the situation where both players are not completely certain about the types of the other side.

3. A NEW BAYESIAN MODEL

Let ROW and COL denote the players in an n-stage IPD game. Both players are not certain that the other will play AllD or TFT strategy. Because of the end-game effect, a TFT player may possibly start to defect at any stage even if cooperation is achieved at the previous stage. We refer to the type of player ROW by $\{\delta_1, \delta_2, \cdots, \delta_n\}$ that denote the probabilities that, if ROW is a TFT player, ROW chooses to cooperate at the $1, \cdots, n$ stage. For example, δ_2 denotes the probability that ROW chooses to cooperate at stage two. Similarly, we define the type of player COL by $\{\theta_1, \theta_2, \cdots, \theta_n\}$ that denotes the probabilities that COL chooses to cooperate at each stage if COL is a TFT player. δ_i and θ_i ($i = 1, \cdots, n$) represent the beliefs of player COL and ROW respectively.

We note that δ_i and θ_i are changeable during the game. We have two assumptions about δ_i and θ_i.

A1: $1 \geq \delta_1 \geq \delta_2 \geq \cdots \geq \delta_n \geq 0$ and $1 \geq \theta_1 \geq \theta_2 \geq \cdots \geq \theta_n \geq 0$.

This reflects the end-game effect. The closer to the last stage it is, the less likely that cooperation will be achieved. Therefore, if we have $\delta_i = 0$, there must be $\delta_{i+1} = \cdots = \delta_n = 0$. It is the same for θ_i.

A2: If either player chooses to defect at stage i, there will be $\delta_{i+1} = 0$ and $\theta_{i+1} = 0$ after stage i.

Without loss of generality, assume ROW chooses to defect at stage i. Since ROW knows that COL must defect at stage $i+1$, there is $\theta_{i+1} = 0$. According to A1, there is also $\theta_{i+2} = \cdots = \theta_n = 0$, that is, ROW believes that COL will play AllD in the remaining stages. Then, ROW's optimal strategy is AllD and COL knows that ROW's optimal strategy is AllD. So there must be $\delta_{i+1} = 0$.

Under the assumptions A1 and A2, each player knows that a defection will trigger mutual defection in the remaining stages, so the finite IPD becomes a form of centipede game.

At the first stage, each player faces a strategy tree as shown in Figure 2.

Let $E^1_{ROW}(C)$ denote ROW's expected payoff if he chooses C at stage one. It is not easy to compute $E^1_{ROW}(C)$ because there are n branches below this node (see Fig. 2). We refer to the expected payoffs for different branches by $E_{R1}(C)$, $E_{R2}(C)$, $\cdots$, $E_{R(n-1)}(C)$. For example, $E_{R1}(C)$ denotes the expected payoff when ROW chooses C at the first stage and D in the remaining stages. We define,

$$E^1_{ROW}(C) = max\{E_{R1}(C), E_{R2}(C), \cdots, E_{R(n-1)}(C)\}$$

where

$E_{R1}(C) = \theta_1 R + (1-\theta_1)S + \theta_1(\theta_2 T + (1-\theta_2)P) + (1-\theta_1)P + (n-2)P$;

$E_{R2}(C) = \theta_1 R + (1-\theta_1)S + \theta_1(\theta_2 R + (1-\theta_2)S) + (1-\theta_1)P + \theta_1\theta_2(\theta_3 T + (1-\theta_3)P) + (1-\theta_1\theta_2)P + (n-3)P$;

...

$E_{R(n-1)}(C) = E_{R(n-2)}(C) - (\prod_{i=1}^{n-2}\theta_i)(\theta_{n-1}(T-R) + (1-\theta_{n-1})(P-S)) + (\prod_{i=1}^{n}\theta_i)(T-P)$;

Similarly, COL's expected payoff $E^1_{COL}(C)$ is defined as,

$$E^1_{COL}(C) = max\{E_{C1}(C), E_{C2}(C), \cdots, E_{C(n-1)}(C)\}$$

where

$E_{C1}(C) = \delta_1 R + (1-\delta_1)S + \delta_1(\delta_2 T + (1-\delta_2)P) + (1-\delta_1)P + (n-2)P$;

$E_{C2}(C) = \delta_1 R + (1-\delta_1)S + \delta_1(\delta_2 R + (1-\delta_2)S) + (1-\delta_1)P + \delta_1\delta_2(\delta_3 T + (1-\delta_3)P) + (1-\delta_1\delta_2)P + (n-3)P$;

...

$E_{C(n-1)}(C) = E_{C(n-2)}(C) - (\prod_{i=1}^{n-2}\delta_i)(\delta_{n-1}(T-R) + (1-\delta_{n-1})(P-S)) + (\prod_{i=1}^{n}\delta_i)(T-P)$;

Thus, at stage one, the payoff matrix for ROW and COL can be expressed by Figure 3. It is obvious that (D, D) is always a Nash equilibrium and (C, C) can be Nash equilibrium only if $E^1_{COL}(C) > \delta_1 T + (1-\delta_1)P + (n-1)P$ and $E^1_{ROW}(C) > \theta_1 T + (1-\theta_1)P + (n-1)P$.

		COL	
		C	D
Row	C	$(E^1_{ROW}(C), E^1_{COL}(C))$	$(\theta_1 S + (1-\theta_1)P + (N-1)P,$ $\delta_1 T + (1-\delta_1)P + (N-1)P)$
	D	$(\theta_1 T + (1-\theta_1)P + (N-1)P,$ $\delta_1 S + (1-\delta_1)P + (N-1)P)$	(NP, NP)

Figure 3: Payoff matrix for ROW and COL at stage one.

		COL	
		C	D
Row	C	$(E^i_{ROW}(C), E^i_{COL}(C))$	$(\theta_i S + (1-\theta_i)P + (N-i)P,$ $\delta_i T + (1-\delta_i)P + (N-i)P)$
	D	$(\theta_i T + (1-\theta_i)P + (N-i)P,$ $\delta_i S + (1-\delta_i)P + (N-i)P)$	$((N-i+1)P, (N-i+1)P)$

Figure 4: Payoff matrix for ROW and COL at stage i ($i = 1, \cdots, n$).

$$\begin{cases} E^1_{COL}(C) > \delta_1 T + (1-\delta_1)P + (n-1)P \\ E^1_{ROW}(C) > \theta_1 T + (1-\theta_1)P + (n-1)P \end{cases} \quad (1)$$

Both (C, C) and (D, D) are Nash equilibrium at stage one if (1) is satisfied. Therefore, it is possible to achieve mutual cooperation at the beginning of finite IPD. Whether or not cooperation can be achieved depends on the initial assignment of both players' beliefs.

The payoff matrix at any stage i can be expressed as Figure 4. $E^i_{ROW}(C)$ and $E^i_{COL}(C)$ denote the expected payoffs of ROW and COL in the stages from i to n if they choose C at stage i.

The condition for (C, C) to be a Nash equilibrium at stage $i$is,

$$\begin{cases} E^i_{COL}(C) > \delta_i T + (1-\delta_i)P + (n-i)P \\ E^i_{ROW}(C) > \theta_i T + (1-\theta_i)P + (n-i)P \end{cases} \quad (2)$$

where

$E^i_{ROW}(C) = max\{E_{Ri}(C), E_{R(i+1)}(C), \cdots, E_{R(n-1)}(C)\}$;

$E_{Ri}(C) = \theta_i R + (1-\theta_i)S + \theta_i(\theta_{i+1}T + (1-\theta_{i+1})P) + (1-\theta_i)P + (n-i-1)P$;

$E_{R(i+1)}(C) = E_{Ri}(C) - \theta_i(\theta_{i+1}(T-R) + (1-\theta_{i+1})(P-S)) + \theta_i\theta_{i+1}\theta_{i+2}(T-P)$;

$\cdots$

$E_{R(n-1)}(C) = E_{R(n-2)}(C) - (\prod_{k=i}^{n-2}\theta_k)(\theta_{n-1}(T-R) + (1-\theta_{n-1})(P-S)) + (\prod_{k=i}^{n}\theta_k)(T-P)$;

$E^i_{COL}(C) = max\{E_{Ci}(C), E_{C(i+1)}(C), \cdots, E_{C(n-1)}(C)\}$;

$E_{Ci}(C) = \delta_i R + (1-\delta_i)S + \delta_i(\delta_{i+1}T + (1-\delta_{i+1})P) + (1-\delta_i)P + (n-i-1)P$;

$E_{C(i+1)}(C) = E_{Ci}(C) - \delta_i(\delta_{i+1}(T-R) + (1-\delta_{i+1})(P-S)) + \delta_i\delta_{i+1}\delta_{i+2}(T-P)$;

$\cdots$

$E_{C(n-1)}(C) = E_{C(n-2)}(C) - (\prod_{k=i}^{n-2}\delta_k)(\delta_{n-1}(T-R) + (1-\delta_{n-1})(P-S)) + (\prod_{k=i}^{n}\delta_k)(T-P)$;

When (2) is satisfied, both (C, C) and (D, D) are Nash equilibrium at stage i; otherwise, (D, D) is the only Nash equilibrium. It is a necessary and sufficient condition. But its expression may be too complicated to be accurately computed, especially when the number of iteration is large. We now deduce a sufficient condition for mutual cooperation to be Nash equilibrium at stage i.

Let's replace $E^i_{ROW}(C)$ and $E^i_{COL}(C)$ by $E_{Ri}(C)$ and $E_{Ci}(C)$ respectively. According to (2), we have,

$$E_{Ci}(C) > \delta_i T + (1-\delta_i)P + (n-i)P$$

$$E_{Ri}(C) > \theta_i T + (1-\theta_i)P + (n-i)P$$

or,

$$\begin{cases} \delta_i(\delta_{i+1}T + (1-\delta i+1)P + R - T - S) > P - S \\ \theta_i(\theta_{i+1}T + (1-\theta i+1)P + R - T - S) > P - S \end{cases} \quad (3)$$

(3) is a sufficient condition for (C, C) to be Nash equilibrium, which denotes a depth-one induction, that is, if both players are likely to cooperate at both the current stage and the next stage, it is worth each player choosing to cooperate at the current stage. Mutual cooperation is possible at stage i if (3) is satisfied. For example, when T=5, R=3, P=1, and S=0, the condition for (C, C) to be a Nash equilibrium at stage i is,

$\delta_i(4\delta_{i+1} - 1) > 1$ and $\theta_i(4\theta_{i+1} - 1) > 1$.

Consider that $1 \geq \delta_i, \delta_{i+1} \geq 0$ and $1 \geq \theta_i, \theta_{i+1} \geq 0$, the necessary conditions for (3) to be hold are $\delta_i \geq \frac{1}{3}$, $\delta_{i+1} \geq \frac{1}{2}$, $\theta_i \geq \frac{1}{3}$, and $\theta_{i+1} \geq \frac{1}{2}$.

Although (3) is derived by assuming each player cooperates at the current stage and defects at the next stage, we cannot conclude that any player must choose to defect at the next stage because the players face different payoff matrix at different stages. When cooperation has been achieved at

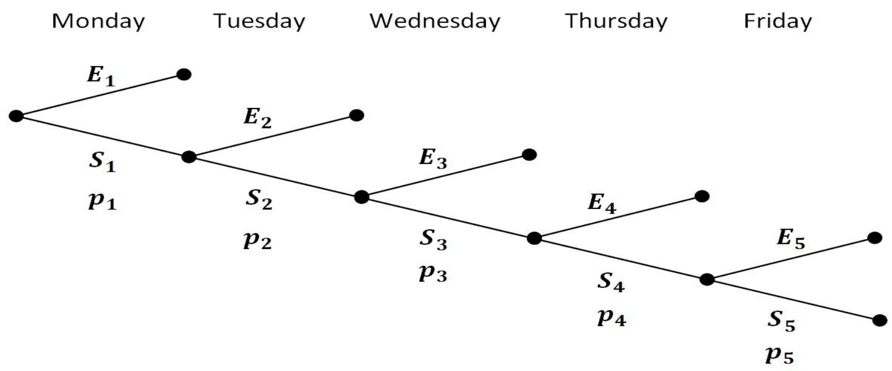

Figure 5: Possibility tree for the prisoner in the unexpected hanging paradox.

stage i, both players will face a new payoff matrix to determine their strategies at stage $i+1$, by which cooperation is still possible to be a Nash equilibrium.

Note that mutual cooperation at stage i is possible only if cooperation has been achieved in all the stages $1, \cdots, i-1$. This means that we cannot analyze strategies of players at stage i until the strategies at stage $i-1$ have been determined. This is important when analyzing the influence of the end-game effect on the strategies of the players.

4. END-GAME EFFECT

It is obvious that (D, D) is the only Nash equilibrium at the last stage. If we assume that $\theta_n = \delta_n = 0$ and it is independent of the strategies of both players in the previous stages, then (D, D) will be the only Nash equilibrium at the second to last stage. So the problem is, can we derive $\theta_{n-1} = \delta_{n-1} = 0$ from $\theta_n = \delta_n = 0$? In order to make it clear, let us first analyze a similar problem, the unexpected hanging paradox.

The paradox is as follows: A condemned prisoner is told that he will be hanged on one weekday in the following week but the execution will be a surprise to him. He will not know the day of the hanging until the executioner knocks on his cell door on that day. The prisoner concludes that he will not be hanged, by means of a backward induction. Firstly, he knows that he will not be executed on Friday. If he has not been hanged by Thursday, then the execution must be on Friday and it will not be a surprise. Secondly, he reasons that the execution cannot be on Thursday. If he has not been hanged by Wednesday, then Thursday must be the date of execution since Friday has been eliminated. By similar reasoning, Wednesday, Tuesday and Monday can also be eliminated. To his surprise, however, the execution is on Wednesday.

What is wrong with the prisoner's reasoning lies in the backward induction. Fig.5 shows the possibility tree for the prisoner to estimate the date of execution. He can be either Executed (E) or Safe (S) on each day of the week. Let $p_1, \cdots, p_5$ denote the probabilities that the prisoner is safe on Monday, Tuesday, $\cdots$ Friday respectively. Let's focus on the last node on Friday. If the prisoner has not been hanged by Thursday, then he must be hanged on Friday. So there should be $p_5 = 0$, which means that the branch of S_5 is unreachable and the branch of E_5 is reached with probability 1. The mistake in the prisoner's reasoning is to deduce $p_4 = 0$ from $p_5 = 0$. $p_4 = 0$ means that the branch of S_4 is unreachable so that both E_5 and S_5 branches become unreachable. Obviously, the prisoner's backward induction mistakenly cuts off the branch of E_5, E_4, $\cdots$, E_1.

The end-game effect leads to $p_5 = 0$ in this game and therefore, it is possible for the prisoner to be executed on any weekday except Friday. The unexpected hanging paradox reveals that backward induction should be used very carefully when there is uncertainty about the strategies or payoffs of the players in repeated games.

In the finite IPD, mutual defection at the last stage could not lead to $\theta_{n-1} = \delta_{n-1} = 0$ due to the same reason as the above example. The values of θ_i and δ_i cannot be determined by means of backward induction because they depend on the strategies of both players in the previous stages. Therefore, the end game effect only influences the strategies of both players at the last stage.

If the end-game effect only influences the strategies of both players at the last stage, why does the rate of cooperation in finite IPD experiments decrease as the game progresses? The reason is that the expected payoffs of both players will change over time. At the beginning of the game, the expected payoffs of both players are relatively high if mutual cooperation could be achieved. They decrease as the game is played. Noting that mutual defection is always Nash equilibrium at any stage, both players have more and more incentive to start defection as the game approaches the end. Consider a finite IPD with $T = 5$, $R = 3$, $P = 1$, and $S = 0$, for example. At the 20^{th} to last stage, the expected payoff at most is 57 for each player; while the expected payoff at most is 12 at the 5^{th} to last stage (see Fig.6). This explains why the rate of mutual cooperation decreases in finite IPD experiments.

5. CONCLUSIONS

A novel Bayesian model for finite IPD that takes into consideration changeable belief and end-game effect has been developed. When both players predict that their opponent is likely to choose to cooperate not only at the current stage but also in the future stages, both (C, C) and (D, D) could be Nash equilibrium. We deduce the conditions for mutual cooperation to be achieved. Specially, if the players only consider payoffs at the current stage and the following stage, a sufficient condition for cooperation is derived. At the last stage of finite IPD, (D, D) is the only Nash equilibrium be-

		COL	
		C	D
Row	C	**(57, 57)**	**(19,24)**
	D	**(24, 19)**	**(20, 20)**

(*a*) Expected payoffs at the 20th to last stage.

		COL	
		C	D
Row	C	**(12, 12)**	**(4, 9)**
	D	**(9, 4)**	**(5, 5)**

(*b*) Expected payoffs at the 5th to last stage.

Figure 6: Possibility tree for the prisoner in the unexpected hanging paradox.

cause of the end-game effect. The influence of end-game effect cannot be backward induced in repeated games with uncertainty.

This model can be extended to deal with infinite IPD and indefinite IPD. In infinite IPD, assumptions A1 and A2 are not suitable. So cooperation is possibly established after mutual defection and the expression of conditions for mutual cooperation will be more complicated. Condition (2) then becomes a sufficient condition for mutual cooperation to be Nash equilibrium. In indefinite IPD, there is a probability that the game ends at each stage. Let q_i denote the probability that the game ends after the i^{th} stage. A sufficient condition for (C, C) to be Nash equilibrium at the i^{th} stage is (4).

$$\begin{cases} \delta_i(\delta_{i+1}T + (1-\delta_{i+1})P + R - T - S) > P - S \\ \qquad\qquad + q_i\delta_i\delta_{i+1}(T-P) \\ \theta_i(\theta_{i+1}T + (1-\theta_{i+1})P + R - T - S) > P - S \\ \qquad\qquad + q_i\theta_i\theta_{i+1}(T-P) \end{cases} \tag{4}$$

Future work will address belief change in generalized repeated games of incomplete information.

6. ACKNOWLEDGMENTS

This study is supported by EPSRC (Engineering and Physical Sciences Research Council) grant (Ref: EP/E017975/1).

7. REFERENCES

[1] Andreoni J. and Miller J. (1993) Rational cooperation in the finite repeated prisoner's dilemma: experimental evidence, The Economic Journal, 103: 570-585.

[2] Cooper R. (1996) Cooperation without reputation: experimental evidence from prisoner's dilemma games, Games and Economic Behavior, 12(2: 187-218.

[3] Fudenberg D. and Maskin E. (1986) The Folk Theorem in repeated games with discounting and incomplete information, Econometrica, 54: 533-554.

[4] Yew Chong S., Humble J., Kendall G., Li J. and Yao X. (2007) Iterated Prisoner's Dilemma and Evolutionary Game Theory. Chapter 2 of The Iterated Prisoner's Dilemma: 20 Years On (Kendall G., Yao X. and Yew Chong S (eds)), World Scientific.

[5] Kreps D., Milgrom P., Roberts J., and Wilson R. (1982) Rational cooperation in the finite repeated prisoner's dilemma, Journal of Economic Theory, 27: 245-252.

[6] Li J. and Kendall G. (2009) A strategy with novel evolutionary features for iterated prisoner's dilemma, Evolutionary Computation, 17(2): 257-274.

[7] Normann H. and Wallace B. (2004) The impact of the termination rule on cooperation in a prisoner's dilemma experiment, Discussing papers in Economics, University of London.

[8] Rapoport A. and Chammah A. (1965) Prisoner's Dilemma. University of Michigan Press, Ann Arbor.

[9] Sarin R. (1999) Simple play in the prisoner's dilemma, Journal of Economic Behavior and Organization, 40(1):105-113.

[10] Dal Bo P. (2005) Cooperation under the shadow of the future: experimental Evidence from Infinitely Repeated Games, American Economic Review, 95: 1591-1604.