

# One Dilemma – Different Points of View

Ivanna Ferdinandova

International Doctorate in Economic Analysis, Departament d'Economia i d'Història  
Econòmica, Universitat Autònoma de Barcelona

**Abstract.** We build a simulation model imitating the structure of human reasoning in order to study how people face a Repeated Prisoner's Dilemma game. The results are ranged starting from individual learning in which case the worst result-defection- is obtained, passing through a partial imitation, where individuals could end up in cooperation or defection, and reaching the other extreme of social learning, where mutual cooperation can be obtained. The influence of some particular strategies on the attainment of cooperation is also considered.

## 1 Introduction

Probably the most difficult problems are those that seem simple, and the most studied those faced by many people. A problem belonging to both groups is the so called Prisoner's Dilemma. The fable, which made it famous, is a story about two suspects in a crime who are put in a separated cells and told the following rules: if both of them confess, each will be sentenced to three years in prison, if only of them confesses, he will be free and the other one will be sent to jail for four months, if neither of them confesses they will be sent to prison only for one month. What makes the problem interesting is the conflict between individual interest to defect (D) and confess, compared to the collective interest of cooperating (C) with the opponent. If the game is played only once then both suspects have a common interest and they will defect. Solving the repeated problem becomes a challenge. On the one hand, the solution of the classical Game Theory to a Finitely Repeated Prisoner's Dilemma is defection in all periods, but on the other, experiments reveal a significant level of cooperation. This contradiction raises the question of whether there exist some conditions that promote cooperation.

In particular this work focuses on the learning process, and analyzes how different models of learning, can influence the final outcome.

We build an ACE model of the repeated game in which, players' learning is modeled using an explicit evolutionary process - genetic algorithms. The pioneer in simulating Repeated Prisoner's Dilemma (RPD) was Axelrod [3], who applied genetic algorithms to evolve RPD strategies against a constant environment. His work was continued by many economists and computer scientists' papers. Among them is the paper of Miller [7], who studies the evolution of cooperation starting from random population of strategies, which evolve in a changing environment. The authors mentioned above model learning at population level, i.e. each player

learns from the best individuals in the population, assumption that is difficult to justify in the context of PD. This model of learning implicitly means that a player imitates the other players strategies, even from players he has never played with, which is equivalent to removing the assumption of no communication and changing the essence of the dilemma. With this scenario cooperation is easily obtained.

Another possibility is to construct the other extreme of individual learning, where each individual learns only from his own experience and this becomes a reason for mutual defection.

The distinction between individual and social learning was studied in the work of Vriend [9], where he analyzed Cournot oligopoly and found that with social learning the results converge to the Walrasian output level, whereas with individual learning they converge to the Cournot - Nash outcome.

Our results move in the same directions as those of Vriend, i.e. social learning leads to socially optimal outcome, and individual to the egoistic one.

Having constructed these models we can notice that both assumptions are too extreme. People do not only learn from the others, neither do they learn only from proper experience. A scenario, which combines both assumptions, is more realistic. Here arises the problem of how to construct it. If we use the classical crossover for creating new strategies, half of the experiments end up defecting and the other half cooperating. But again we face a similar problem - how can a player imitate a strategy that he does not know. This problem can be avoided by using a different type of crossover, where each player is trying to understand and imitate to same extent (without copying the strategy) the behavior of his opponent (Vilà [8]). This reduces the number of cooperative outcomes and confirms that the type of crossover is not irrelevant to the outcome, in other words the way the information is exchanged is important.

Under this scenario the elements that determine the outcome are the type of strategies the players have at the beginning and how fast is the process of evolution. If the evolution process is fast it is more likely that players end up defecting, but if it is slow they will experiment with more strategies and could obtain cooperation.

To analyze the importance of players' strategies we introduce players who always play a given strategy. The effects they induce are different. A player who always either defects or cooperates will induce defection, but one who plays a variation of Tit-For-Tat (TFT) (cooperates when the opponent is observed to cooperate and defects when the opponent has defected) makes his opponent cooperate.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 is dedicated to the results and the last Section 4 summarizes the results and concludes.

## 2 Model

The purpose of our work is to distinguish between three models of learning - social, individual and imitation. We start by building a basic framework which can be easily adapted to one of the three scenarios.

Two individuals are about to play a RPD. When facing the problem for the first time, each of them is assumed to have  $K$  randomly generated strategies.<sup>1</sup>

Let  $S_i^k$  be the  $k$  strategy of an individual  $i$  ( $k = 1..K, i = 1,2$ ). For the purposes of our model we need an abstraction of the process by which the player implements this strategy. This role will be played by a machine called finite automata. A finite automata is a system which responds to discrete inputs (in our case the actions of the opponent) with discrete outputs (player's own actions). The type of finite automata adopted here is a Moore machine<sup>2</sup>.

Having constructed the sets of strategies for both players they are ready to play. Each player chooses one of his strategies to play  $r$  times PD. When the repeated game is over the strategy  $S_i^k$  receives payoff  $\Pi_i^k$  being the sum of the payoffs in each single game. We assume that the next time players meet they will experiment with another strategy.

Once the players have tried all their strategies, they analyze the results obtained. A strategy that has relatively high payoff will be kept in the memory as a good one, and used in the future, and the strategies that performed bad will be replaced with a combination of the existing strategies. A strategy that performed well in one trail is not necessarily good, it just appeared to be good against a given strategy of the opponent. For example, a strategy that always defects will be a very good choice if played against a strategy that always cooperates, the first one will have an average payoff of five and the second of zero. But the first strategy may be a bad choice against TFT.

The process explained above is evolutionary and will be modeled using Genetic Algorithms. In our context the population will be each player's strategies and the environment they play in - his opponent's strategies.

Using this framework and changing only the process of formation of the new strategies we are able to construct the three scenarios.

The first one resembles the model of Miller [7] and almost coincides with the classical framework in this problem. In Miller's model, each agent has only one strategy and the strategies of all agents evolve together i.e. there is only one population evolving. We use the same structure but our agents have more strategies. The strategies of both players evolve together, forming one population in which good strategies will be kept and bad replaced with a combination of the current strategies. This can be the case when players talk and discuss different plans. The result Miller obtained in the case of perfect information is that at the end all individuals have the same strategy and their behaviour converges to the cooperative result.

---

<sup>1</sup> A strategy is a plan of how to behave in all possible circumstances.

<sup>2</sup> For more detailed description of a Moore machine see Miller [7] or Vilà [8].

Since the assumptions of the social learning scenario contradict with the essence of the dilemma, another possibility is to construct the other extreme - individual learning, where each player learns only from his own experience. Players stay at different rooms and they do not try to interpret each other's behaviour, but just adapt. Technically this is a two population model based on Vilà [8], where a repeated discrete principal-agent game is analyzed. The difference between this model and the social learning is that now each player can use only his own strategies. The process of learning is internal and is based only on proper experience.

The two scenarios, as described above are good as benchmark models, but too extreme. A better option is the imitation scenario, which combines the assumptions of the previous two. Players are not allowed to communicate, but they do make an attempt, to understand each other. It starts like the individual learning scenario until the moment when each player has his strategies with the same structure. After this point each of them imitates the strategy of his opponent. If we use the classical method for creating new strategies (single cut crossover<sup>3</sup>) we are facing the same problem as in the social learning. In order to avoid it we use a different procedure for generating new strategies i.e. different type of crossover due to Vilà [8]. Our player can not observe the strategies of his opponent, but he observes the history of the repeated game, which includes both players actions. Let  $h_i$  be the history of player's  $i$  actions. Then assume that player 1 asks himself - what would I do if I were him and somebody played against me in the way I have played against him. Since he does not know the strategy of his opponent he can only imagine playing with his strategy, from his opponent's position a history of random length and comparing his action after this history with the real action of the opponent. If they do not coincide player one replaces his action with his partner's action.

All the variations of the original framework are summarized in the following table:

---

*The structure of the model:*

1. Generate K strategies for individual  $i$ ,  $i=1,2$ ;
2. Randomly match each strategy of one individual with one from the other;
3. Repeat until all strategies are played once. Each strategy receives payoff  $\Pi_i^k$   
 $i=1,2$ .  $k=1,..K$ ;
4. Form new strategies
  - 4.1 Social Learning (as in Miller [7])
  - 4.2. Individual Learning
    - a) include the best  $K/2$  strategies of each individual in his new group of strategies;

---

<sup>3</sup> With single cut crossover you combine two strategies by cutting them in two parts and interchanging the second parts.

b) select two strategies of an individual  $i$  to be parents. The probability of strategy  $S_i^k$  of being selected is:

$$P(S_i^k) = \frac{\Pi_i^k}{\sum_k \Pi_i^k} \tag{1}$$

c) create  $K/2$  new strategies for each individual using the crossover and then apply the mutation;

4.3. Imitation - steps a, b and c remain the same until both individuals have their strategies with the same structure. After that step b is changed with

b') one of the strategies selected belongs to the individual himself and the other is the best strategy of his opponent.

5. Repeat steps 1 ÷ 4  $R$  times, where  $R$  is the number of repetitions.

---

### 3 The Results

#### 3.1 Parameter Values

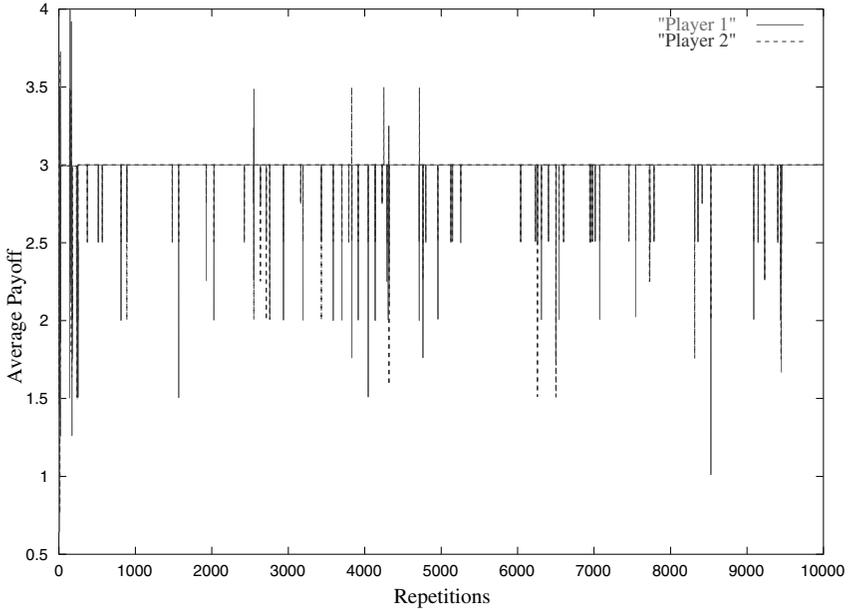
The simulations were performed under the following conditions:

Number of individuals	2	
Number of strategies of each individual	4	
Number of rounds	300	
Number of repetitions	10 000	
Probability of mutation	0.001	
Overlapping generations	1/2	(2)
Crossover type	partial imitation or single cut	
Length of bit string	60	
	C D	
Payoff structure	C 3 , 3 0 , 5	
	D 5 , 0 1 , 1	

The results are robust to changes in most of the parameters (number of strategies, length of bit string and payoff structure), but the choice of some of them requires some discussion. One of them is the number of strategies. Usually the size of the population chosen in similar models is higher, but having in mind that in our context the population consists of strategies, it is difficult to assume that people have 50 or 100 strategies. We have run simulations with 50 or 100 strategies, but the results obtained were not different. The number of rounds and the number of repetitions, were chosen to guarantee convergence. The probability of mutation determines, among other things, how adaptive the players are. The higher is the probability of mutation, the easier it will be for one player to adapt to the changes of the behavior of the other. But if it becomes too high they will modify their behavior too often. The value chosen is standard.

### 3.2 Results

Our results under the social learning scenario coincides with the one obtained in the literature i.e. cooperation. The outcome is independent of the type of crossover used. A typical evolution of the average payoffs of both players is depicted in Figure 1.

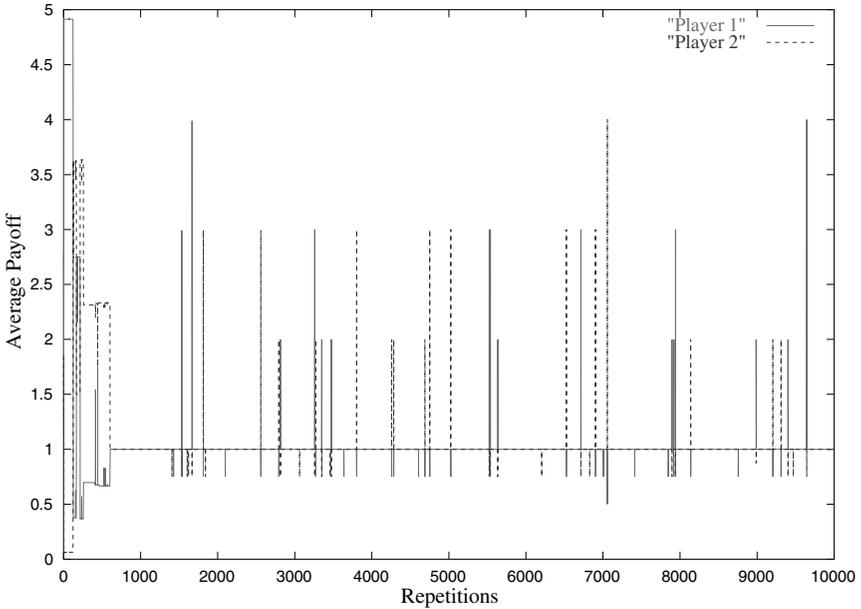


**Fig. 1.** In the case of *Social Learning* players cooperate and receive average payoff of three.

The same stability of the result is obtained also in the individual learning case but the outcome, as Figure 2 shows, is defection.

One possible explanation is that, when the evolution is common, the strategies that the players will have in the next period come out of the same process. Therefore it is very likely that they will be similar for both players and they will move together. Hence they have to choose between cooperate or not, and the first outcome is clearly better.

The imitation scenario has two possible results depending on the type of crossover assumed. If we use the traditional crossover (single cut) players can end up defecting or cooperating depending on how fast is the learning process and what type of strategy do they have before the imitation begins. This difference can be seen in the two graphs at Figure 3. Intuitively if players learns slowly the search field becomes bigger and the possibility of cooperation increases. But at the same time even if they learn fast but at least one of them uses a strategy



**Fig. 2.** In the scenario of *Individual Learning* the only possible result is defection, which gives both players an average payoff of one.

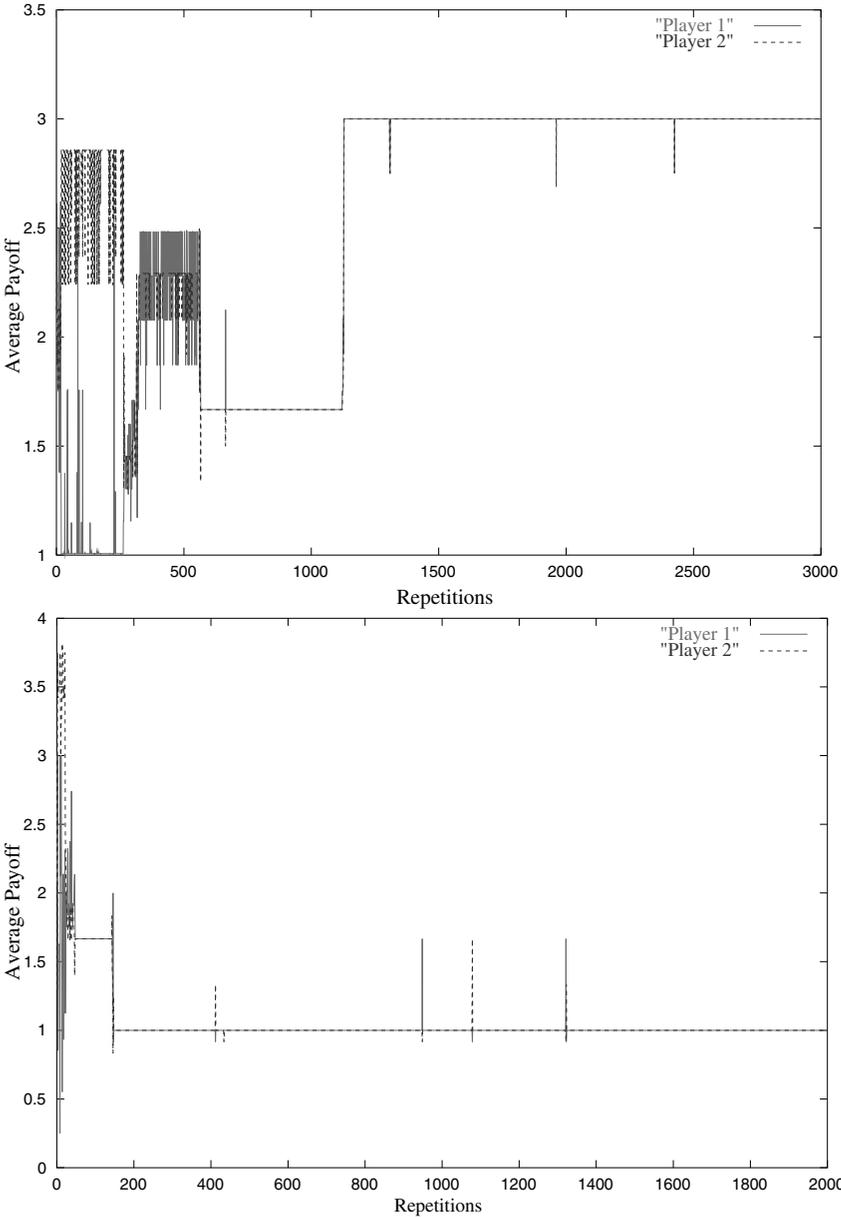
that induces cooperation this is enough to make cooperation the only possible outcome. In order to check the influence of different strategies we introduce a player who always plays the same strategy. The three possible players are a player who cooperates, player who defects and players who plays TFT. If the first two are introduced they induce defection, but the third one is able to "teach" his opponent to cooperate. One possible explanation for the experiments, which converge to defection is that it is very difficult for a strategy like TFT to appear in a random environment.

If instead of adopting the classical crossover we choose the one that does not violate the assumptions of no communication the only possible result as depicted in Figure 4 is defection.

## 4 Concluding remarks

The results of the three scenarios we have constructed suggest that one should be very careful when deciding which one to choose.

Special attention should be given to the assumption of no communication, which we found to be very influential. Intuitively if people cannot talk this lowers the confidence and decreases the probability of cooperation. Going back to the initial structure of the dilemma, if the two suspects can discuss, they will be able to find plans or strategies, which guarantee that neither will confess. Making



**Fig. 3.** *Imitation scenario* The left graph depicts a case in which the learning process is slow and the result is cooperation. Exactly the opposite can be seen at the right graph, where the players end up defecting

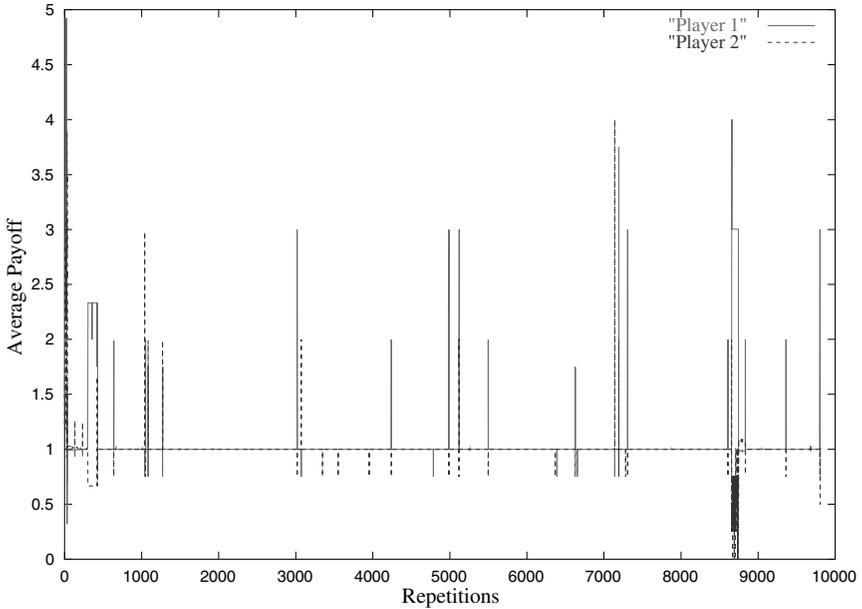


Fig. 4. *Imitation scenario with partial imitation crossover.* The result is defection.

the learning process personal, complicates it and independently on the possible attempts of reasoning about the other person's behavior, the cooperation fails. The imitation of the opponent's behavior leads to cooperation only if one of the agents plays TFT.

Summarizing, we have found two conditions, that promote cooperation. On the one hand, that is the possibility of communication and on the other the influence of a TFT player.

## References

1. Arthur, Brian W.: Complexity in Economic Theory: Inductive Reasoning and Bounded Rationality, American Economic Review, Vol.84,(1994) 406-411
2. Axelrod R.: The Evolution of Cooperation, Basic Books, New York (1984)
3. Axelrod R.: The complexity of cooperation: Agent based models of competition and collaboration, Princeton University Press (1997)
4. Goldberg David E.: Genetic Algorithms in Search, Optimization and Machine Learning, Addison Wesley (1989)
5. Holland J.H: Adaptation in Natural and Artificial Systems, University of Michigan Press (1975)
6. Holland J.H., Holyoak K. J., Nisbett R.E., Thagard P. R.: Induction: Process of Inference, Learning and Discovery, MIT Press (1989)

7. Miller, John H.: The Coevolution of automata in the repeated prisoner's dilemma, *Journal of Economic Behavior and Organization*, Vol.29, (1996) 87-112
8. Vilà Xavier: Adaptive Artificial Agents Play a Finitely Repeated Discrete Principal-Agent Game. *Lecture notes in Economics and Mathematical Systems*, Vol.456, Springer-Verlag, Berlin Heidelberg New York (1997) 437-56
9. Vriend, Nicolaas J.: An Illustration of the Essential Difference between Individual and Social Learning and Its Consequences for Computational Analyses, *Journal of Economic Dynamics and Control*, Vol.24, (2000) 1-19

### **Acknowledgments**

I am very grateful to Xavier Vilà. Without him this work would not exist. I would like also to thank Inés Macho Stadler and Andrés Romeu for their helpful comments. Special thanks to my friends and colleagues: Ariadna Dumitrescu, Patricia Trigales and Todora Radeva for their help and support.

This research was undertaken with support from the European Union's Phare ACE Programme 1997. The content of the publication is the sole responsibility of the author and it in no way represents the views of the Commission or its services.