Partial Imitation Hinders Emergence of Cooperation in the Iterated Prisoner's Dilemma with Direct Reciprocity

Mathis Antony, Degang Wu and K. Y. Szeto

Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong
phszeto@ust.hk

Abstract. The evolutionary time scales for various strategies in the iterated Prisoner's Dilemma on a fully connected network are investigated for players with finite memory, using two different kinds of imitation rules: the (commonly used) traditional imitation rule where the entire meta-strategy of the role model is copied, and the partial imitation rule where only the observed subset of moves is copied. If the players can memorize the last round of the game, a sufficiently large random initial population eventually reaches a cooperative equilibrium, even in an environment with bounded rationality (noise) and high temptation. With the traditional imitation rule the time scale to cooperation increases polynomially with decreasing intensity of selection (or increasing noise), whereas partial imitation results in an exponential dependence. Populations with finite lifetimes are therefore unlikely to ever reach a cooperative state in this setting. Instead, numerical experiments show the emergence and long persistence of a phase characterized by the dominance of always defecting strategies.

1 Introduction

We use the Prisoner Dilemma [10] (PD) as an example of a two player game to study the impact of incomplete information in the imitation process. When two players play the PD game, each of them can choose to cooperate (C) or defect (D). Each player is awarded a payoff depending on his own and the opponent's move. Cooperation yields R (S) if the opponent cooperates (defects) and defection yields T (P) if the opponent cooperates (defects). R is the *Reward for cooperation*, S is the *Sucker's payoff*, T is the *Temptation to defect* and P is the *Punishment*. In the PD, $T > R > P > S$ and $2R > T + P$ to prevent collusion if the game is played repeatedly. The PD is a so called *non zero sum game* because one players loss does not equal his opponent's gain. By cooperating, both players win, by mutually defecting they both lose. In this paper we do not vary these payoff parameters but employ a set of commonly used values with high temptation: $T = 5$, $R = 3$, $P = 1$, $S = 0$. These values were also used in Axelrod's famous PD computer tournament [2]. For an excellent review of the PD literature, we refer the reader to [11].

The tragedy behind the PD briefly consists in the fact that the best strategy for a selfish individual (D) is the worst strategy for the society. The expectation of playing D is greater than the expectation of playing C (independent of the opponents strategy), but cooperating yields a higher total payoff. The state where no player has anything to gain by changing her own strategy (the so called *Nash Equilibrium*) occurs only when all players defect. Hence, if the players imitate the behaviour of the more successful players, defection will dominate if we do not provide any additional circumstances to encourage cooperative behaviour. Nowak et al. summarized *five rules for the emergence of cooperation* [5]: kin selection, direct reciprocity, indirect reciprocity, network reciprocity [6,12] and group selection. Network reciprocity has attracted an received particular attention recently in the light of co-evolutionary dynamics [8,9] where the network topology of the underlying interaction network evolves alongside the agent strategies.

The mechanism at work here is direct reciprocity. Players are given a memory or in other words the ability to remember a fixed number of recent outcomes of the *PD* games. Each player is then supplied with a set of answers to respond to every possible history [4,3]. We call this set of moves a Strategy[1].

Players will then imitate other players by adopting their strategies. But if the strategies are elaborate an imitation may be challenging. As an illustration, assume Alice and Bob are playing chess and Alice is winning every game. Even if Bob recalls all of Alice's moves he will not be able to imitate her strategy completely until a huge number of games have been played. Instead he may attempt to improve his own strategy by adapting elements of Alice's strategy exposed to him during previous games. Intuitively the more complex a strategy, the more difficult it should be for a player to imitate it. We incorporate this condition by means of an imitative behaviour we refer to as *partial Imitation Rule (pIR)*. According to this rule a player can only imitate based on her knowledge about the opponent's strategy gathered during the most recent encounter with this opponent. This is in contrast to what we call the *traditional Imitation Rule (tIR)* that allows players to imitate the complete strategy of their opponents. Numerical experiments are performed to analyse the impact of the adjustment to the imitation behaviour. We show that it leads to new phenomena that do not occur under *tIR*, such as a phase dominated by defecting strategies.

The rest of this paper is structured as follows: in section 2 the terms memory, strategy and the imitation rules are defined. We also provide details about our numerical experiments. Results are presented and discussed in section 3. Our conclusions follow in section 4.

2 Methods

In this section we explain the concepts of memory and strategies and define the partial imitation rule we previously introduced in [13,14].

[1] It is common to refer to *cooperate* and *defect* as strategies. A set of rules telling the player when to cooperate or defect is then called a *meta-strategy*. For convenience we refer to the former as "moves" and to the latter as "strategies" instead.

2.1 Memory and Strategies

A player who can remember the last n rounds of the PD game has a n-step memory. We denote the ensemble of n-step strategies as M_n. The total number of strategies in M_n is $|M_n|$. As a player with one-step memory we need to remember two moves, our move and the one of our opponent. There are four possible outcomes (DD,DC, CD and CC, where the first letter is the move of the first player and the second letter is the move of the second player) of the PD game. For our one-step memory we need to have a response (either C or D) to each of these four possible outcomes. Thus our strategy can be represented by a 4-bit string where every bit is a response to one outcome of the previous round of the game. We add a bit for the first move against an unknown opponent. A strategy in M_1 is denoted as $S_0|S_{DD}S_{DC}S_{CD}S_{CC}$ where S_0 is the first move and S_{DD}, S_{DC}, S_{CD} and S_{CC} are the moves that follow DD, DC, CD and CC histories respectively. Thus there are $|M_1| = 2^5 = 32$ possible strategies as there are two choices for each S_i, either C or D. Three famous strategies are Grim-Trigger (GT): $C|DDDC$, Tit-For-Tat (TFT): $C|DCDC$ and Pavlov or "win-stay-lose-shift": $C|CDDC$.

In this paper we focus entirely on M_1. For models with longer memory horizon see [4,3]. We assume players never play in contradiction to their strategy. Then, there are four always defecting strategies in M_1, namely $D|DDDD$, $D|DDDC$, $D|DDCD$ and $D|DDCC$. We refer to these strategies as *all-D* type strategies. For a given opponent, they all score the same. However when using the partial imitation rule described below they can produce different children strategies. The same applies to the four always cooperating strategies. A strategy is *nice* if the first move and all moves that follow mutual cooperation are C. In M_1 those are the strategies $C|XXXC$ where X may be either C or D. A *retaliating* strategy defects after its attempt to cooperate is met with defection. In M_1 retaliating strategies are $X|XXDX$. The only four *nice* and *retaliating* strategies in M_1 are therefore *TFT*, *GT*, *Pavlov* and $C|CCDC$.

2.2 partial Imitation Rule (pIR)

In general the *traditional Imitation Rule* (tIR) has the following simple character: a player i will imitate the strategy of player j, who is usually one of the players who interacts with i, with a certain probability given by a monotonically increasing smoothing function $g(\Delta U)$ where $\Delta U = U_j - U_i$ is the payoff difference between player i and j. For the rest of this paper we use the following smoothing function, which also introduces a temperature like noise factor K allowing for irrational choices by players. If player i has been selected to imitate player j then he will carry out the imitation with probability

$$P(i \text{ imitates } j) = g(\Delta U) = \frac{1}{1 + \exp\left(\frac{-\Delta U}{K}\right)} \tag{1}$$

We note that this probability has the form of Fermi distribution and is a step function at zero noise ($K = 0$). For low values of the noise factor K, player i

imitates player j very rationally. For high values of K however, player i imitates player j with a probability close to $1/2$ as for $K \gg 1$ we have $P \approx \frac{1}{2} + \frac{\Delta U}{4}\frac{1}{K}$. The imitation in this case is similar to a scenario with weak selection intensity with the addition of the constant $1/2$ which introduces noise in finite sized populations.

The traditional imitation rule implicitly makes a bold assumption in the case of memory agents: the imitating player is assumed to know the entire strategy of the role model. Depending on how the two players interacted, some of the role model's strategy may be unknown to the imitator. In order to strip the players from these "mind reading" abilities we use the *partial Imitation Rule* in which the imitator only adapts the parts of the role model's strategy which have been exposed during their interaction.

We illustrate the difference between the two imitation rules with the example of a $C|DDDD$-strategist, Alice, imitating a TFT-strategist, Bob. Figure 1 shows the transition graph for this encounter. As shown in the transition graph the CD state from Alice's perspective (or DC state from Bob's perspective) never occurs. Bob has therefore never used the S_{DC} move of his strategy in this encounter. According to pIR Alice cannot copy such hidden moves. Hence if the $C|DDDD$ player Alice imitates the TFT player Bob according to the partial Imitation Rule she will only imitate S_0, S_{DD}, S_{CD} and S_{CC}, hence becoming herself a GT- and not a TFT-player. If Alice imitates Bob using tIR she will copy the entire TFT strategy and become a TFT player herself. By using tIR we implicitly assume Alice has found a way to expose Bob's hidden response to the DC history.

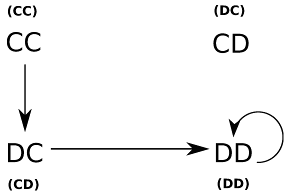

Fig. 1. Transition graph between the $C|DDDD$ player Alice and the TFT ($C|DCDC$) player Bob from Alice's point of view. Bob's point of view is described by the moves in parenthesis. The recurrent states is DD and the average recurrent state payoff is therefore P for both Alice and Bob.

2.3 Simulation

An important parameter in the iterated Prisoner's Dilemma with memory is the number f of rounds played during an encounter between two players i and j. If $f = 1$ we have a "one-shot" game and the agents can not make use of their memory. From our recent works in [13,14]we understand that the number f affects our results in a complex way. Here, we follow the approach employed in [3] for replicator dynamics: assuming that the number f is sufficiently large the average payoff per instance of the PD game played in a confrontation is well approximated by the average payoff from the recurrent states of the transition graph. In other words, we address the case $f \to \infty$ in our simulations by considering only payoff accumulated in recurrent states of the transition graph.

Let U_{ij} be the average recurrent state payoff obtained by an i strategist playing against a j strategist. If N is the total number of players and every player i plays against all other players and himself, his average payoff per encounter is

$$U_i = \frac{1}{N} \sum_{j=1}^{N} \boldsymbol{S_i^T U S_j} = \boldsymbol{S_i^T U} \left\langle \boldsymbol{S} \right\rangle \ , \quad i = 1, 2 \ldots |M_n| \tag{2}$$

Where $\boldsymbol{U}$ is the $|M_n| \times |M_n|$ real Matrix with coefficients $\boldsymbol{U_{ij}} = U_{ij}$. The vector $\boldsymbol{S_i^T} = (0 \ldots 0\,1\,0 \ldots 0)$ is a $|M_n|$ boolean vector where the m-th entry is equal to 1 only if player i is an m-strategist. $\langle \boldsymbol{S} \rangle$ is the $|M_n|$ dimensional real column-vector of the "average strategy" that can also be written as $\langle \boldsymbol{S} \rangle_m = \rho_m$ where $m = 1, 2, \ldots, |M_n|$ and ρ_m is the number density of m-strategists. Note that the summation of ρ_m over all m equals 1.

Initially every player is assigned a random strategy out of the 32 strategies in M_1. The system is then evolved with random sequential updating from time $t = 0$ until some final time t_f. We chose two players i and j and let them play against all opponents and themselves to accumulate an average payoff per encounter U_i and U_j. By using the reasoning above this can be achieved by randomly selecting two strategies i and j with probabilities equal to ρ_i and ρ_j respectively and evaluating U_i and U_j according to equation 2. Agent i then imitates agent j with a probability given by equation 1 according to pIR or tIR. If this imitation occurs we adjust ρ_i and ρ_k where k is the children strategy produced by the imitation process. The outline of this procedure is given in algorithm 1.

Algorithm 1 Outline of simulation procedure

```
poorest
  chose $N$ random strategies
  compute $\langle \boldsymbol{S} \rangle$
  for $t = 0$ to $t_f$ do
    for $n = 1$ to $N$ do
      pick random strategy $i$ and $j$ with probability $\rho_i$ and $\rho_j$ respectively
      compute $U_i$ and $U_j$
      $i$ imitates $j$ according to tIR or pIR with probability $g(\Delta U)$
      $\rho_i \leftarrow \rho_i - \frac{1}{N}$
      $\rho_k \leftarrow \rho_k + \frac{1}{N}$
    end for
  end for
```

A Monte Carlo sweep, generation or one time unit corresponds to N such updates. As a result of introducing noise, the strategy fractions during a typical simulation fluctuate considerably even if N is large. The fate of the entire population is then subject to the survival of a few key strategies, such as GT, in the early phase of evolution. As our primary interest is neither directed towards these special cases of evolution nor towards finite size effects we use a very large number of players. We have found that by choosing $N = 2.5 \cdot 10^7$ we can obtain reliable data for noise factors up to at least $K = 150$, which is sufficient for our observations. Based on our experiments we may state here that if the aforementioned extinctions of key strategies due to random fluctuations do not occur, the strategy fractions as a function of time for smaller population sizes are very similar to those we will observe below.

3 Results

In this section we first discuss a typical simulation at high noise in section 3.1 for illustrative purposes and to introduce the *all-D* phase. In section 3.2 we report and discuss the time scale of the evolution of cooperation in our model.

3.1 All-D Phase

The number density, concentration or fraction of a strategy i in a population of players is denoted as ρ_i. Figure 2 shows the number density of key strategies as a function of time for a typical pIR simulation with high noise factor (here $K = 100$). We notice that in a first phase the $D|DDCD$ and $D|DDCC$ fraction increase rapidly but die out soon thereafter to make room for the $D|DDDD$ and $D|DDDC$ strategy. These two strategies dominate for a long time but the $C|DDDD$ and GT fraction increase progressively up to a point where all four remaining strategies are about equally abundant. In figure 2 this occurs around $t = 9500$. The GT fraction then rises rapidly while the other strategies die out. Eventually we are left in an equilibrium state where all players cooperate by using the GT strategy.

We first give a intuitive explanation of these observations. When noise is high, the imitation probability is close to 1/2, thus, the imitation processes occur in many directions with only marginal drift towards imitation of better strategies. We denote the process of an A-strategist becoming a C-strategist by imitating a B-strategist as $A \xrightarrow{B} C$. If the C-strategist may turn back into an A-strategist by imitating the A-strategist we say that this imitation is directly reversible.

In an early tumultuous phase the majority of strategies die out rapidly. The famous and well scoring strategies TFT and *Pavlov* do not survive this phase either. Due to the nature of pIR there is a net drift away from these strategies at $K = 100$. In this early period of evolution, the players will obtain the highest payoff by exploiting the naive players in the initial random setup and adopt the *all-D* type strategies. As a result, most players go out of this early extinction phase as $D|DDCD$ and $D|DDCC$ defectors. The fate

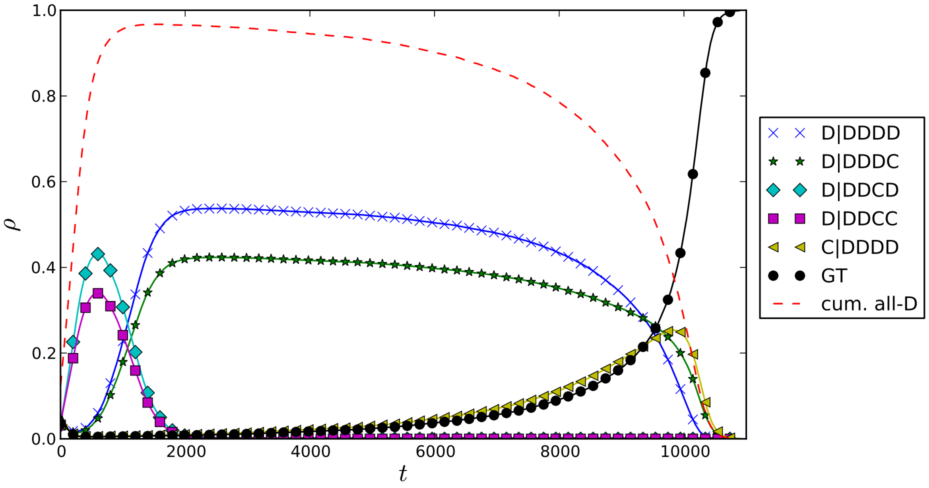

Fig. 2. Strategy fractions during a typical simulation at high noise, $K = 100$. The *cumulative all-D fraction* is given by $\rho_{\text{cum all-D}} = \rho_{D|DDDD} + \rho_{D|DDDC} + \rho_{D|DDCD} + \rho_{D|DDCC}$.

of these two strategies is governed by $D|DDCD \xrightarrow{GT \text{ or } C|DDDD} C|DDDD$ and $D|DDCC \xrightarrow{GT \text{ or } C|DDDD} GT$. Note that these imitation processes are not directly reversible under pIR. To illustrate this interesting phenomena, let's consider the example of a GT-strategist $(C|DDDC)$ imitating a $D|DDCC$-strategist. He will become a $D|DDDC$-strategist (and not a $D|DDCC$-strategist). This means that once a $D|DDCC$-strategist has imitated a GT-strategist, he may not turn back into a $D|DDCC$-strategist simply by imitating the GT strategy. The $D|DDCD$ and $D|DDCC$-strategists are gradually converted into $D|DDDD$ and $D|DDDC$ strategists by the interaction with GT and $C|DDDD$ players. The extinction of $D|DDCD$ and $D|DDCC$ is the inevitable consequence.

The main imitation processes during the following long phase of dominance of the $D|DDDD$ and $D|DDDC$ strategies are $D|DDDD \xrightarrow{GT \text{ or } C|DDDD} C|DDDD$, $D|DDDC \xrightarrow{GT \text{ or } C|DDDD} GT$ and $C|DDDD \xrightarrow{GT} GT$. As all these processes are directly reversible it takes a very long time for the players to drift towards the better scoring GT strategy. GT is the only strategy that scores higher than the other three remaining key strategies as GT players cooperate with GT players but defect against all the other remaining strategies.

3.2 Time Scale for Emergence of Cooperation

From extensive simulations we know that all populations eventually reach an equilibrium state[2] in which more than half the players use GT. We use this fact to analyse the time scale of our numerical experiments and define a quantity called the *GT First Passage Time* τ_{GT}, which is the time at which the population contains more GT players than defectors for the first time or in other words the *GT First Passage Time* is the lowest time t such that $\rho_{GT} > \rho_{\text{cum. all-D}}$. Where $\rho_{\text{cum. all-D}} = \rho_{D|DDDD} + \rho_{D|DDDC} + \rho_{D|DDCD} + \rho_{D|DDCC}$ is the *cumulative all-D fraction*. As mentioned before there is no *all-D phase* for *tIR*. Nevertheless we can use this definition for the *GT First Passage Time* τ_{GT} to compare the time scale to equilibrium of tIR and pIR populations. The first passage times τ_{GT} as function of K are shown in figure 3.

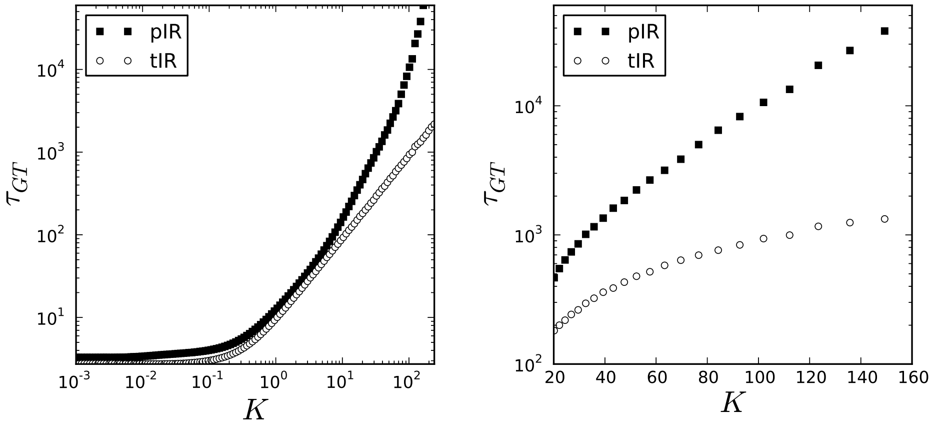

Fig. 3. First passage time τ_{GT} for $\rho_{GT} > \rho_{c.all-D}$ for both imitation rules. For illustrative purpose the data is shown in two figures, focusing on the higher end of the noise factor scale on right hand side. Note the logarithmic scale on both axes on the left and linear scale on the x-axis and logarithmic scale on the y-axis on the right.

We first examine the figure on the left hand side. For both imitation rules τ_{GT} increases monotonically with K. For $K < 0.1$ τ_{GT} is practically constant for *tIR* and increases only marginally with K for *pIR*. For *tIR* the growth is linear for $K \gtrsim 0.5$. Between $0.5 \lesssim K \lesssim 5$, τ_{GT} also exhibits a linear relationship with K for *pIR*. On the other hand by examining the figure on the right hand side we realize an exponential increase of τ_{GT} for *pIR* and $K > 40$. We find another

[2] Note that for tIR and pIR and all considered values of the noise factor K only *nice* strategies exist in the equilibrium state. In this state we have completely random drift among the surviving strategies for tIR. For pIR on the other hand there is no such drift as *nice* strategies do not change by imitating other nice strategies.

fundamental impact of our subtle adjustment to the imitation behaviour. The GT First Passage Time which can be considered as an indicator of the time the players need to become cooperators, scales very differently for the two imitation rules at high noise. The results also suggest, once more, that finite size effects are more important for higher noise factors, which imply stronger fluctuations due to the random nature of the imitation process.

The exact rate at which τ_{GT} increases with K depends on the values of the payoff parameters R, T, S and P. We note in passing that an exponential relationship for high values of the noise factor K is not unique to our choice of parameters. Notably, we also observe it for the entire range of T in the *weak Prisoner's Dilemma* [7] with $R = 1$, $1 < T < 2$ and $S = P = 0$.

In addition we observe that for *pIR* the time scale increases exponentially with K. These two phenomena coupled together put the survival of the cooperating strategies and the eventual emergence of a cooperative society at risk for finite sized populations. Furthermore, due to the exponential growth of τ_{GT} for *pIR* and because the lifetime of real populations are finite the cooperative equilibrium might never be reached. The partial imitation rule therefore introduces two new obstacles for the emergence of cooperation in the weak selection or high noise regime.

4 Conclusion

By performing numerical experiments on a large population of prisoners with finite memory playing the iterated Prisoner's Dilemma game on a fully connected network we have shown that a direct and fast route to cooperation may not exist unless the players are given the ability to copy unknown parts of their role models. Our adjustment to the imitative behaviour shows that incomplete information about opponent strategies has important consequences for the emergence of cooperation in such a society of prisoners. If the information about the wealth of opponents is vague (at high noise or in the weak selection regime) the majority of prisoners stick to defecting strategies for a very long time. As we have seen that in this environment the duration of this route to cooperation scales exponentially with the noise factor, we must question the significance of such a cooperative equilibrium not only for finite sized populations with finite life times.

The *partial Imitation Rule* also presents a new challenge for famous *PD* strategies. The *Grim Trigger* strategy has a fundamental advantage over other nice strategies such as *TFT* and *Pavlov*, despite the very similar performance of all these strategies: *GT* is the only strategy that is easy to imitate. This observation is the principle is sometimes referred to as *Occam's razor*. All other things being equal, the simplest strategy is the best strategy.

Many of the problems involving evolutionary games and memory could be reexamined under the new light of partial information. Although we have shown that considering partial rather than complete information has a strong impact already in the case of a one-step memory, we expect even more drastic effects in the case of longer memory. A natural extension of our work will therefore include more intelligent prisoners with longer memory. We expect that the simple GT strategy will not be as efficient anymore because once defecting it does not provide a way to reestablish cooperation.

Finally, finite size effects as well as the topology of the underlying network is an interesting topic for further investigation. In two dimensions [13,14], we also reported striking difference for the two imitation rules considered here. We may therefore extend our studies to other networks, such as *scale-free networks* which model the topology of real societies more accurately [1].

Acknowledgements. We thank the three anonymous referees for their valuable comments. K.Y. Szeto acknowledges the support of grant FSGRF13SC25.

References

1. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Reviews of Modern Physics **74**(1) (Jan 2002) 47–97
2. Axelrod, R.: The Evolution of Cooperation. Basic Books (1984)
3. Baek, S.K., Kim, B.J.: Intelligent tit-for-tat in the iterated prisoner's dilemma game. Physical Review E (Statistical, Nonlinear, and Soft Matter Physics) **78**(1) (2008) 011125
4. Lindgren, K., Nordahl, M.G.: Evolutionary dynamics of spatial games. Physica D Nonlinear Phenomena **75** (August 1994) 292–309
5. Nowak, M.A.: Five rules for the evolution of cooperation. Science **314**(5805) (December 2006) 1560–1563
6. Nowak, M.A., May, R.M.: The spatial dilemmas of evolution. Int. J. of Bifurcation and Chaos **3**(1) (1993) 35–78
7. Nowak, M.A., May, R.M.: Evolutionary games and spatial chaos. Nature **359**(6398) (Oct 1992) 826–829
8. Pacheco, J.M., Traulsen, A., Nowak, M.A.: Coevolution of strategy and structure in complex networks with dynamical linking. Phys. Rev. Lett. **97** (Dec 2006) 258103
9. Perc, M., Szolnoki, A.: Coevolutionary games–a mini review. Biosystems **99**(2) (2010) 109 – 125
10. Poundstone, W.: Prisoner's Dilemma: John Von Neumann, Game Theory and the Puzzle of the Bomb. Doubleday, New York, NY, USA (1992)
11. Szabo, G., Fath, G.: Evolutionary games on graphs. Physics Reports **446**(4-6) (July 2007) 97–216
12. Szabo, G., Vukov, J., Szolnoki, A.: Phase diagrams for an evolutionary prisoner's dilemma game on two-dimensional lattices. Phys. Rev. E **72**(4) (2005) 047107
13. Wu, D., Antony, M., Szeto, K.Y.: Evolution of grim trigger in prisoner dilemma game with partial imitation. In et al., C.D.C., ed.: EvoApplications 2010, Part I. Volume 6024 of Lecture Notes in Computer Science (LNCS). Springer-Verlag (Berlin), Barcelona, Spain (January 2010, Revised Papers 2010) 151–160
14. Wu, D., Antony, M., Szeto, K.Y.: Partial imitation rule in iterated prisoner dilemma game on a square lattice. In: NICSO. Volume 284 of Studies in Computational Intelligence. Springer (2010) 141–150