

Distribution of Strategies in a Spatial Multi-Agent Game

Frank Schweitzer
Chair of Systems Design, ETH Zurich
Kreuzplatz 5, 8032 Zurich, Switzerland
f Schweitzer@ethz.ch

Robert Mach
Chair of Systems Design, ETH Zurich
Kreuzplatz 5
8032 Zurich, Switzerland
rmach@ethz.ch

ABSTRACT

We investigate the adaptation of cooperating strategies in an iterated Prisoner's Dilemma (IPD) game. The deterministic IPD describes the interaction of N agents spatially distributed on a lattice, which are assumed to only interact with their four neighbors, hence, local configurations are of great importance. Particular interest is in the spatial-temporal distributions of agents playing different strategies, and their dependence on the number of consecutive encounters n_g during each generation. We show that above a critical n_g , there is no coexistence between agents playing different strategies, while below the critical n_g coexistence is found.

1. INTRODUCTION

When an agent chooses its strategy from a fixed set of possible ones, its decision has to take several factors into account: (i) *locality*, i.e. the subset of those agents, the respective agent most likely interacts with, (ii) *heterogeneity*, i.e. the possible strategies of these agents, (iii) *time horizon*, i.e. the number of foreseen interactions with these agents. The best strategy for a rational agent would be the one with the highest payoff for a given situation. To reach this goal, however, can be difficult for two reasons: *uncertainty*, i.e. particular realizations are usually subject to random disturbances and can therefore be known only with a certain probability, and *co-evolution*, i.e. the agent does not operate in a stationary environment, but in an ever-changing one, where its own action influences the decisions of other agents and vice-versa.

To reduce the risk of making the wrong decision, it often seems to be appropriate just to copy the successful strategies of others [16, 12]. Such an *imitation* behavior is widely found in biology, but also in cultural evolution. A similar kind of *local imitation* behavior will be used in this paper to explain the spatial dispersion of strategies in a multi-agent system. In order to address the problem systematically, we apply a well defined game-theoretic problem. Game theory embod-

ies many of the features mentioned above and thus has often been employed in the design of mechanisms and protocols for interaction in multi-agent systems [17, 21]. Many investigations on the relation between game theory and multiagent systems focus on mathematical investigations [3], and the learning and coordination of tasks [19, 6, 20, 18]. In contrast our paper mainly deals with the dynamics of strategy distribution.

We consider a system of N agents spatially distributed on a square lattice, so that each lattice cite is occupied by just one agent. Each agent i is characterized by two state variables, (i) its position r_i on the lattice, and (ii) a discrete variable θ_i , describing its possible actions, as specified in Sect. 3. Agents are assumed to directly interact only with their 4 nearest neighbors a number of n_g times. In order to describe the local interaction, we use the so-called *iterated Prisoner's Dilemma* (IPD) game – a paradigmatic example [1, 15] well established in evolutionary game theory with a broad range of applications in economics, political science, and biology.

In the simple Prisoner's Dilemma (PD) game, each agent i has two options to *act* in a given situation, to *cooperate* (C), or to *defect* (D). Playing with agent j , the outcome of this interaction depends on the action chosen by agent i , i.e. C or D , *without knowing* the action chosen by the other agent participating in a particular game. This outcome is described by a *payoff matrix*, which for the 2-person game, i.e. for the interaction of only two agents, has the following form:

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{array}{cc} R & S \\ T & P \end{array} \end{array} \quad (1)$$

In PD games, the payoffs have to fulfill the following two inequalities:

$$T > R > P > S \quad 2R > S + T \quad (2)$$

In this paper, we use the known standard values $T = 5$, $R = 3$, $P = 1$, $S = 0$. This means in a cooperating environment, a defector will get the highest payoff. From this, the abbreviations for the different payoffs become clear: T means (T)emptation payoff for defecting in a cooperative environment, S means (S)ucker' payoff for cooperating in a defecting environment, R means (R)eward payoff for cooperating in a likewise environment, and P means (P)unishment payoff for defecting in a likewise environment.

In any *one* round (or "one-shot") game, choosing action D is unbeatable, because it rewards the higher payoff for agent i whether the opponent chooses C or D . At the same

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AAMAS'06 May 8–12 2006, Hakodate, Hokkaido, Japan.
Copyright 2006 ACM 1-59593-303-4/06/0005 ...\$5.00.

time, the payoff for *both agents i and j* is maximized when both cooperate. But in a *consecutive* game played many times, both agents, by simply choosing D , would end up earning less than they would earn by collaborating. Thus, the number of games n_g two agents play together becomes important. For $n_g \geq 2$, this is called an *iterated Prisoner’s Dilemma* (IPD). It makes sense only if the agents can remember the previous choices of their opponents, i.e. if they have a memory of $n_m \leq n_g - 1$ steps. Then, they are able to develop different *strategies* based on their past experiences with their opponents, which is described in the following.

2. AGENT’S STRATEGIES

In this paper, we assume only a *one-step* memory of the agent, $n_m = 1$. Based on the known previous choice of its opponent, either C or D , agent i has then the choice between *eight* different strategies. Following a notation introduced by Nowak [14], these strategies are coded in a 3-bit binary string $[I_o I_c I_d]$ which always refers to *collaboration*. The first bit represents the *initial* choice of agent i : it is 1 if agent i *collaborates*, and 0 if it defects initially. The two other values refer always to the previous choice of agent j . I_c is set to 1 if agent i chooses to collaborate given that agent j has collaborated before and 0 otherwise. I_d is similarly set to 1 if agent i chooses to collaborate given that agent j has defected before and 0 otherwise. For the deterministic case discussed in this paper, the eight possible strategies ($s = 0, 1, \dots, 7$) are given in Tab. 1.

s	Strategy	Acronym	Bit String
0	suspicious defect	sD	000
1	suspicious anti-Tit-For-Tat	sATFT	001
2	suspicious Tit-For-Tat	sTFT	010
3	suspicious cooperate	sC	011
4	generous defect	gD	100
5	generous anti-Tit-For-Tat	gATFT	101
6	generous Tit-For-Tat	gTFT	110
7	generous cooperate	gC	111

Table 1: Possible agent’s strategies using a one-step memory.

Depending on the agent’s *first move*, we can distinguish between two different classes of strategies: (i) *suspicious* ($s = 0, 1, 2, 3$), i.e. the agent *initially defects*, and (ii) *generous* ($s = 4, 5, 6, 7$), i.e. the agent *initially cooperates*. Further, we note that four of the possible strategies do not pay attention on the opponent’s previous action, i.e. except for the first move, the agent continues to act in the same way, therefore the strategies sD, sC, gD, gC ($s = 0, 3, 4, 7$) can be also named *rigid* strategies.

The more interesting strategies are $s = 1, 2, 5, 6$. Strategy $s = 6$, known as (generous) “tit for tat” (TFT), means that agent i initially collaborates and continues to do so, given that agent j was also collaborative in the previous move. However if agent j was defective in the previous move, agent i chooses to be defective, too. This strategy was shown to be the most successful one in iterated Prisoners Dilemma games with *2 persons* [1]. Here, however, we are interested in

spatial interactions, where agents simultaneously encounter with 4 different neighbors.

Agents playing strategy gATFT ($s = 5$) initially also start with cooperation and then do the opposite of whatever the opponent did in the previous move, while agents playing strategy sATFT ($s = 1$) behave the same way, except for the first move where they defect. Agents playing strategy sTFT ($s = 2$) also start with defection, but then *imitate the previous move* of the opponent, as in gTFT. A closer look at the encounters reveals that sTFT and gTFT will exploit each other alternatively while gTFT will mutually cooperate. Also, sATFT and gATFT exploit each other alternatively, while sATFT will alternatively cooperate. This illustrates that the *first move* of a strategy can be vital to the outcome of the game. The number of interactions n_g is also a crucial parameter in this game, because, if n_g is even, gTFT and sTFT will gain the same, but in case of n_g being odd sTFT will gain more than gTFT.

Eventually, it can be argued that some of the strategies do not make sense from a human point of view. In particular sATFT or gATFT seem to be “lunatic” or “paranoid” strategies. Therefore, let us make two points clear: Such arguments basically reflect the intentions, not to say the preconceptions, of a human beholder. We try to avoid such arguments as much as possible. In our model, we consider a strategy space of $2^3 = 8$ possible strategies, and there is *no methodological reason* to exclude *a priori* some of these strategies. If they do not make sense, then this will be certainly shown by the evolutionary game theoretic approach used in our model. I.e., those strategies will disappear in no time, but not because of our private opinion, but because of a selection dynamics that has proven its usefulness in biological evolution. So, there is no reason to care too much about a few “paranoid” strategies. On the other hand, biological evolution has also shown that sometimes very unlikely strategies get a certain chance under specific (local?) conditions. In a complex system, it would be a priori not possible to predict the outcome of a particular evolutionary scenario, simply because of the path-dependence. We come back to this point in our conclusions, where we shortly discuss that in the case of eight strategies and $n_g = 2$ also unpredicted strategies survive.

3. SPATIAL INTERACTION

So far, we have explained the interaction of *two agents* with a one-step memory. This shall be put now into the perspective of a *spatial game* with *local* interaction among the agents. A spatially extended (non-iterative) PD game was first proposed by Axelrod [1]. Based on these investigations, Nowak and May simulated a spatial PD game on a cellular automaton and found a complex spatiotemporal dynamics [13, 14]. A recent mathematical analysis [2] revealed the critical conditions for the *spatial coexistence* of cooperators and defectors with either a majority of cooperators in large spatial domains, or a minority of cooperators in small (non-stationary) clusters.

In the following, we concentrate on the *iterated* PD game, where the number of encounters, n_g , plays an important role. We note that possible extensions of the IPD model have been investigated e.g. by Lindgren and Nordahl [7], who introduced players which act erroneously sometimes, allowing a complex evolution of strategies in an unbounded strategy space.

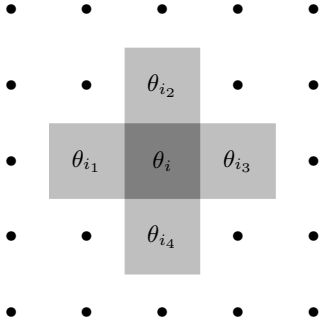


Figure 1: Local neighborhood of agent i . The nearest neighbors are labeled by a second index $j = 1, \dots, 4$. Note that $j = 0$ refers to the agent in the center.

In the spatial game, we have to consider *local configurations* of agents playing different strategies (see Fig. 1). As explained in the beginning, each agent i shall interact only with its *four nearest neighbors*. Let us define the size of a neighborhood by n (that also includes agent i), then the different neighbors of i are characterized by a second index $j = 1, \dots, n - 1$. The mutual interaction between them results in a n -person game, i.e. $n = 5$ agents interact simultaneously. In this paper, we use the common assumption that the 5-person game is decomposed into $(n - 1)$ 2-person games, that may occur *independently*, but *simultaneously* [7, 8, 15], a possible investigation of a “true” 5-person PD game is also given in [15].

We further specify the θ_i that characterize the possible actions of each agent as one of the strategies that could be played (Tab. 1), i.e. $\theta_i \in s = \{0, 1, \dots, 7\}$. The total number of agents playing strategy s in the neighborhood of agent i is given by:

$$k_i^s = \sum_{j=1}^{n-1} \delta_{s\theta_{i_j}} \quad (3)$$

where δ_{xy} means the Kronecker delta, which is 1 only for $x = y$ and zero otherwise. The vector $\mathbf{k}_i = \{k_i^0, k_i^1, k_i^2, \dots, k_i^7\}$ then describes the “occupation numbers” of the different strategies in the neighborhood of agent i playing strategy θ_i .

Agent i encounters with each of its four neighbors playing strategy θ_{i_j} in independent 2-person games from which it receives a payoff denoted by $a_{\theta_i\theta_{i_j}}$, which can be calculated with respect to the payoff matrix, eq. (1). The total payoff of an agent i after these independent games is then simply

$$a_i(\theta_i) = \sum_{j=1}^{n-1} a_{\theta_i\theta_{i_j}} = \sum_s a_{\theta_i s} \cdot k_i^s \quad (4)$$

We note again that the payoffs $a_{\theta_i s}$ also strongly depend on the number of encounters, n_g , for which explicit expressions have been derived. They are concluded in a 8×8 payoff matrix not printed here [11].

In order to introduce a *time scale*, we define a *generation* G to be the time in which each agent has interacted with its $n - 1$ nearest neighbors n_g times. During each generation, the strategy θ_i of an agent is not changed while it interacts with its neighbors simultaneously. But after a generation is

completed, θ_i can be changed based on a comparison of the payoffs received. I.e., payoff a_i is compared to the payoffs a_{i_j} of all neighboring agents, in order to find the maximum payoff within the local neighborhood during that generation, $\max\{a_i, a_{i_j}\}$. If agent i has received the highest payoff, then it will keep its θ_i , i.e. it will *continue* to play its strategy. But if one of its neighbors j has received the highest payoff, then agent i will adopt or *imitate*, respectively, the strategy of the respective agent. If

$$j^* = \arg \max_{j=0, \dots, n-1} a_{i_j} \quad (5)$$

defines the position of the agent that received the highest payoff in the neighborhood, the update rule of the game can be concluded as follows:

$$\theta_i(G + 1) = \theta_{i_{j^*}}(G) \quad (6)$$

We note that the evolution of the system described by eq. (6) is completely *deterministic*, results for stochastic CA have been discussed in [4].

The adaptation process leads to an evolution of the spatial distribution of strategies that will be investigated by means of computer simulations on a cellular automaton in the following section.

4. EVOLUTION OF SPATIAL PATTERNS OF 3 STRATEGIES

In order to illustrate the spatio-temporal evolution we have restricted the computer simulation here to only three strategies. The more complex (and less concise) case of eight strategies is discussed in detail in [11, 9, 10]. The three strategies were chosen as sD, sATFT and gTFT ($s = 0, 1, 6$) for the following reason. Strategy sD is known to be the winning strategy for the one-shot game, i.e. $n_g = 1$, while gTFT is known to be the most successful strategy for $n_g \geq 4$. We are interested in the *transition region*, $1 < n_g < 4$, thus we include those two strategies in our simulation and fixed n_g to values of 2 or 3. Agents playing sATFT are also added to the initial population, since they behave anti-cyclic, i.e. they defect when the opponent cooperated in the previous encounter and vice versa.

The apparent solution to describe the dynamics of the system by an dynamical system approach works only for the so-called mean-field case, which can be simulated by a random interaction. I.e., each agent interacts with four randomly chosen agents during each generation. In this case the dynamics can well be described by a selection equation of the Fisher-Eigen type [9]. The random interaction is also used as a reference case, to point out differences to the case of local interaction described in the following.

The simulation are carried out on a 100×100 lattice with periodic boundary conditions, in order to eliminate spatial artifacts at the edges. Initially, all agents are randomly assigned one of the three strategies. Defining the total fraction of agents playing strategy s at generation G as $f_s(G) = 1/N \sum_{i=1}^N \delta_{\theta_i s}$, $f_0(0) = f_1(0) = f_6(0) = 1/3$ holds for $G = 0$ (see also the first snapshot of Fig. 2).

Because each agent encounters with his 4 nearest neighbors n_g times during one generation, in each generation $(N/2 \times n_g \times 4)$ independent and simultaneous deterministic 2-person games occur. Fig. 2 shows snapshots of the spatio-temporal distribution of the three strategies for $n_g = 2$, while Fig. 3 shows snapshots with the same setting, but

for $n_g = 3$.

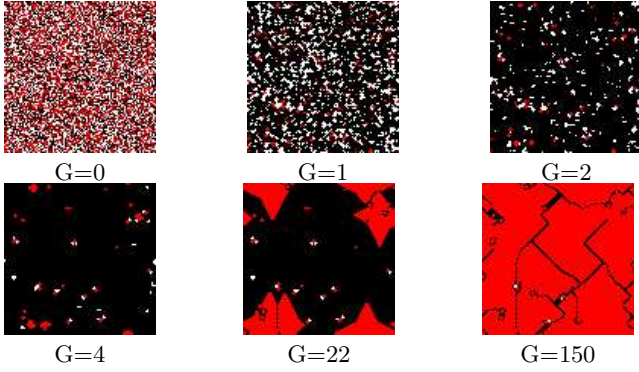


Figure 2: Spatial-temporal distribution of three strategies sD (black), sATFT (white), and gTFT (gray) on a 100×100 grid for $n_g = 2$.

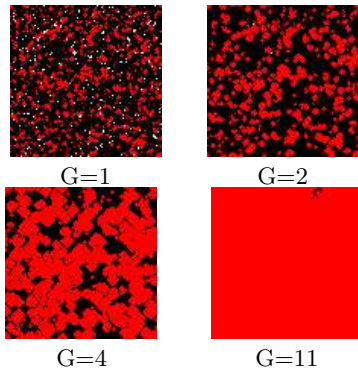


Figure 3: Spatial-temporal distribution of three strategies sD (black), sATFT (white), and gTFT (grey) on a 100×100 grid for $n_g = 3$. The comparison with Fig. 2 elucidates the influence of n_g .

For $n_g = 2$, we see from Fig. 2 that in the very beginning, i.e. in the first four generations, strategy sD grows very fast on the expense of sATFT and especially on gTFT. This can be also confirmed when looking at the global frequencies of each strategy (see left part of Fig. 4). Already for $G=4$, strategy sD is now the majority of the population – only a few agents playing gTFT and even fewer agents playing sATFT are left in some small clusters. Hence, for the next generation we would assume that the sD will take over the whole population. Interestingly, this is not the case. Instead, the global frequency of sD goes down while the frequency of gTFT starts to increase continuously until it reaches the majority. Only the frequency of sATFT stays at its very low value. On the spatial scale, this evolution is accompanied with a growth of domains of gTFT that are finally separated by only thin borders of agents playing sD (cf Fig. 2 for $G = 150$). The reasons for this kind of *crossover dynamics* will be explained later.

When increasing the number of encounters n_g from 2 to 3, we observe that the takeover of gTFT occurs much faster. Already for $G = 13$, it leads to a situation where *all agents* play gTFT, with no other strategy left. Hence, they will mutually cooperate. The fast takeover is only partly due to

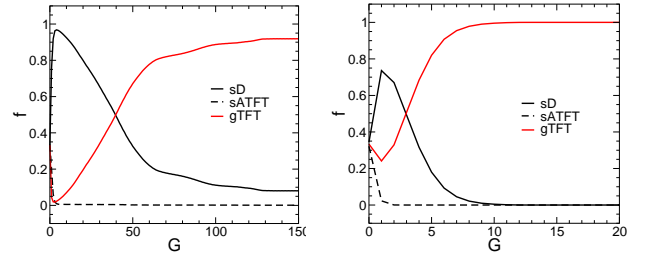


Figure 4: Global frequencies $f_s(G)$ of the three strategies for $n_g = 2$ (left) and $n_g = 3$ (right). For the spatial distribution, see Fig. 2 and Fig. 3, respectively.

the fact that the total number of encounters during one generation has increased. The main reason is that for $n_g = 2$ agents playing sATFT are able to *locally block* the spreading of strategy gTFT, while this is not the case for $n_g = 3$. This is because both of the n_g dependence of the agent's payoff and the local configuration of players: for $n_g = 2$, there is only one local configuration where strategy gTFT can invade sATFT, because of the higher payoff. After this invasion, however, the preconditions for further invasion have vanished. For $n_g = 3$, this situation is different in that there are more local configurations, where gTFT can invade sATFT. This in turn enables the further takeover. The crossover dynamics mentioned in conjunction with Fig. 4 can be explained in a similar manner. For $n_g = 2$ gTFT can not spread initially because of agents playing sATFT. Only sD is able to invade sATFT and gTFT, therefore its frequency increases. Once sATFT is removed, gTFT can spread [11].

5. GLOBAL PAYOFF DYNAMICS

The adaptation of strategies by the agents is governed by the ultimate goal of reaching a higher individual payoff. As we know from economic applications, however, the maximization of the private utility does not necessarily mean a maximization of the overall utility. So, it is of interest to investigate also the global payoff and the dynamics of the payoffs of the individual strategies [5].

The average payoff per agent \bar{a} is defined as:

$$\bar{a} = \frac{1}{N} \sum_{i=1}^N a_i(\theta_i) = \sum_s f_s(G) \cdot \bar{a}_s; \quad \bar{a}_s = \frac{\sum_i a_i(\theta_i) \delta_{\theta_i, s}}{\sum_i \delta_{\theta_i, s}} \quad (7)$$

where $f_s(G)$ is the total fraction of agents playing strategy s and \bar{a}_s is the average payoff *per strategy*, shown in Fig. 5 for the different strategies.

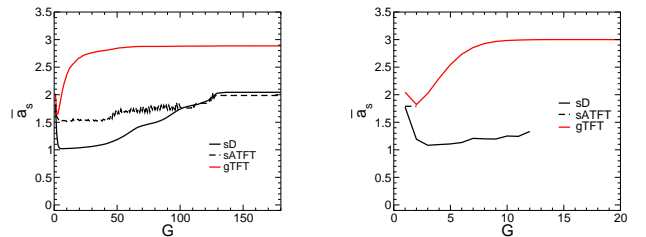


Figure 5: Average payoff per strategy, \bar{a}_s , eq. (7), vs. time for $n_g = 2$ (left) and $n_g = 3$ (right)

We note that the payoffs per strategy for the 2-person

games are always fixed dependent on n_g . However, the average payoff per strategy changes in the course of time mainly because the local configurations of agents playing different strategies change. For $n_g = 2$, we have the stable coexistence of all three strategies (cf Fig. 2 and Fig. 4left), while for $n_g = 3$ only strategy gTFT survives (cf Fig. 3 and Fig. 4right). Hence, in the latter case we find that the average payoff of gTFT reaches a higher value than for $n_g = 2$, while in Fig. 5 the corresponding curves for the other strategies simply end, if one of these strategies vanishes.

Eventually, the average global payoff is shown in Fig. 6 for different values of n_g . Obviously, the greater n_g , the faster the convergence towards a stationary global value, which is $\bar{a} = 3$ only in the “ideal case” of *complete cooperation*. As we have already noticed, for $n_g = 2$ there is a small number of defecting agents playing either sD or sATFT left, therefore the average global payoff is lower in this case.

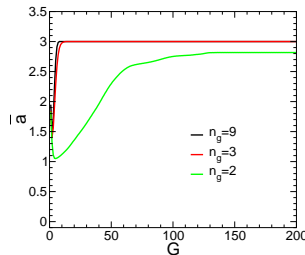


Figure 6: Average global payoff \bar{a} , eq. (7), vs. time for different values of n_g .

6. EXTENSIONS AND CONCLUSIONS

In this paper, we have investigated the spatial-temporal distributions of agents playing different strategies in an iterated Prisoner’s Dilemma game. Their interaction is restricted to the four nearest neighbors, hence, *local configurations* are of great influence. Particular importance was on the investigation of the number of consecutive encounters between any two agents, n_g . For the case of three strategies, we find that a critical value of n_g exists above which no coexistence between agents playing different strategies is observed. Hence, the most successful strategy, i.e. the one with the highest payoff in an iterated game, gTFT, is eventually adopted by all agents. This confirms the findings of Axelrod [1] also for the spatial case. Below the critical n_g , we find a coexistence between cooperating and defecting agents, where the cooperators are the clear majority (playing gTFT), whereas the defectors play two different strategies, either sD or sATFT. In both cases, we observe that the share of gTFT in the early evolution drastically decreases before it eventually invades the whole agent population.

We notice, however, that this picture holds only for a *random initial distribution* of strategies. It can be shown [11] that there are always specific initial distributions of strategies where gTFT fails to win. An interesting question is also under what conditions gTFT is not the most successful strategy any more. This is of particular interest if one considers the case where all *eight* strategies are present in the initial population. In [10] we have investigated this more complex case with $n_g = 2$. In contrast to the case of three strategies (gTFT, sD, sATFT) which all coexist for $n_g = 2$, we observe the coexistence of either *two* strategies

(gD, sTFT) or *three* strategies (sD, gD, gTFT) (given in order of frequency) in the final state. In particular, the most known strategy gTFT will usually become *extinct*, which is certainly different from the expected behavior.

A second point to be mentioned, we find for different runs with the same initial conditions different outcomes of the simulations. Hence, random deviations may lead the global dynamics to different attractors. Thus, *local* effects seem to be of great influence for the final outcome. This proves our point made at the end of Sect. 2, that path dependence plays an important role in the dynamics and the evolutionary game cannot be completely predicted. A more thorough analysis of the game three strategy game is given in [11], whereas [9, 10] concentrate on the eight strategy case.

Eventually, we note an important insight about spatial IPD [11]: Given a specific n_g , one can analytically deduce from the payoff matrix that the payoff of agent i playing strategy θ_i is always the same if it encounters with agent j playing particular strategies $\theta_j \in s$. These particular strategies can be grouped into certain classes, that yield the same payoff to agent i . For instance, for $n_g = 2$ it makes no difference for an agent playing strategy sD ($s = 0$) to play either against sATFT, sC, gD or gTFT ($s = 1, 3, 4, 6$), while for $n_g = 3$ the same is only true for sATFT and sC.

7. REFERENCES

- [1] R. Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.
- [2] L. Behera and F. Schweitzer. The invasion of cooperation in heterogeneous populations, I: Thresholds and attractor basins. *Journal of Mathematical Sociology*, (to be submitted)
- [3] P.E. Dunne M.J. Wooldridge. Preferences in qualitative coalitional games. In *Proc. Sixth Workshop on Game Theoretic and Decision Theoretic Agents (GTDT’04)*, pages 29–38. , New York, 2004.
- [4] R. Durrett. Stochastic spatial models. *SIAM Review*, 41(4):677–718, December 1999.
- [5] H. Gintis. *Game Theory Evolving*. Princeton University Press, June 2000.
- [6] B. Grosz and S. Kraus. Collaborative plans for complex group activities. *Artificial Intelligence*, 1996.
- [7] K. Lindgren and M. G. Nordahl. Evolutionary dynamics of spatial games. *Physica D*, 75:292–309, 1994.
- [8] K. Lindgren and J. Johansson. Coevolution of strategies in n-person prisoner’s dilemma. In J. Crutchfield and P. Schuster, editors, *Evolutionary Dynamics: Exploring the Interplay of Selection, Accident, Neutrality, and Function*, *SFI Studies in the Sciences of Complexity*, pages 341–360. Oxford University Press, May 2002.
- [9] R. Mach and F. Schweitzer. Adaptation of cooperation: Random vs. local interaction in a spatial agent model. *Journal of Artificial Societies and Social Simulation (JASSS)*, 2006. (to be submitted).
- [10] R. Mach and F. Schweitzer. Attractor structure in an IPD. *European Physical Journal*, 2006. (to be submitted).
- [11] R. Mach and F. Schweitzer. Food-web representation of a spatial IPD, 2006. (to be submitted).
- [12] A. Nowak, M. Kuś, J. Urbaniak, and Z. Tomasz.

- Simulating the coordination of individual economic decisions. *Physica A*, (297):613–630, 2000.
- [13] M. A. Nowak and R. M. May. Evolutionary games and spatial chaos. *Nature*, 359:826–829, October 1992.
- [14] M. A. Nowak and K. Sigmund. Tit for tat in heterogeneous populations. *Nature*, 355:250–253, January 1992.
- [15] F. Schweitzer, L. Behera, and H. Mühlenbein. Evolution of cooperation in a spatial prisoner’s dilemma. *Advances in Complex Systems*, 5(2):269–300, 2002a.
- [16] F. Schweitzer, J. Zimmermann, and H. Mühlenbein. Coordination of decisions in a spatial agent model. *Physica A*, 303(1-2):189–216, 2002b.
- [17] S. Sen. Reciprocity: A foundational principle for promoting cooperative behavior among self-interested agents. In Victor Lesser, editor, *Proceedings of the First International Conference on Multiagent Systems*, pages 322–329. MIT Press, 1996.
- [18] S. Sen, M. Sekaran, and J. Hale. Learning to coordinate without sharing information. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 426–431. Seattle, WA, 1994.
- [19] S. Sen and G. Weiss. Learning in multiagent systems. pages 259–298. MIT Press, Cambridge, MA, USA, 1999. ISBN 0-262-23203-0.
- [20] K. Tuyls, K. Verbeeck, and T. Lenaerts. A selection-mutation model for q-learning in multi-agent systems. In *AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 693–700. ACM Press, New York, NY, USA, 2003. ISBN 1-58113-683-8.
- [21] M. Wooldridge. *An Introduction to MultiAgent Systems*. Wiley, 2004.