SONDERFORSCHUNGSBEREICH 303
„Information und die
Koordination wirtschaftlicher Aktivitäten“

Projektbereich B
Discussion Paper No. B-330

Spatial Evolution of Automata in the Prisoners' Dilemma

Oliver Kirchkamp[1]

September 1995

---

[1] University of Bonn, Wirtschaftstheorie III, Adenauerallee 24–26, D-53113 Bonn, email kirchkamp@witch.econ3.uni-bonn.de, http://witch.econ3.uni-bonn.de.

An electronic version of the paper is available at http://witch.econ3.uni-bonn.de/∼oliver/spatEvol.shtml

Support by the Deutsche Forschungsgemeinschaft, SFB 303, is gratefully acknowledged. I thank George Mailath, Karl Schlag, Avner Shaked, Bryan Routledge and several seminar participants for comments.

RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT BONN

D-53113 Bonn, Lennéstraße 37

Abstract

The paper applies the idea of evolution to a spatial model. We assume that prisoners' dilemmas are played repeatedly within neighborhoods and that evolutionary pressure works locally.

Discriminatory behavior of players is introduced representing strategies as small automata, identical for a player but possibly in different states against different neighbors. Extensive simulations show that success among surviving strategies may differ and that in contrast to undiscriminative behavior cooperation persists even in a stochastic environment.

Keywords: Evolutionary Game Theory, Networks, Prisoners' Dilemma, Overlapping Generations. JEL-Code: C63, C73, D62, D83, R12, R13.

Contents

1 Introduction

Recently evolutionary models have become popular among game theorists and economists. The (evolutionary) idea of successful strategies or behavioral patterns driving out less successful ones was already a common assumption in social sciences for a long time.

Consider e.g. firms that produce a homogeneous product in a larger country. Firms may occupy different locations and compete for customers mainly with neighboring firms. Finding *optimal* strategies for such a firm can be a complex problem which firms could be either unwilling or unable to solve. Instead strategies could be the result of some *evolutionary* or *learning* process (In the following we will use the expressions *evolution* and *learning* synonymously. What we have in mind in both cases is always some learning process[1]).

Evolutionary game theory in its classical form formalized such an approach assuming large populations where each agent is *equally likely* to interact with *any other* agent in the population (see e.g. Maynard Smith and Price (1973) for static concepts and Taylor and Jonker (1978) and Zeeman (1981) for a dynamic model). These approaches have their merits for their simplicity, but they implicitly assume *global* interaction and *global* evolution, which are rarely found in reality. A more realistic model of evolution of social or economic strategies should allow for individuals that interact more often with neighbors or colleagues than with some randomly chosen individual.

Returning to the above spatially distributed firms, let us first notice that often firms do not *interact* globally. Changes in the price set by one firm may only influence the demand of nearby firms. Therefore we should introduce *local interaction* into such an evolutionary model.

But not only interaction of players is a local phenomenon. The same applies to the *evolution* of players' behavior. Learning of successful strategies is less likely to occur if a successful agent is far away, and is more likely if the shining example is proximate. Note that *interaction* and *evolution* are two different features that have to be interpreted separately if we change from a global to a local model.

Introducing local interaction and local evolution makes standard models more complex and difficult to solve.

Analytical solutions of such a model are often restricted to simple topologies (societies that live on a line with only small variations in neighborhoods, see e.g. Eshel, Samuelson and Shaked (1994)). It is often difficult to decide which properties of these models depend on their particular structure.

An alternative possibility to gain insight into spatial models are simulations (see e.g. Axelrod (1984, p. 158ff) and May and Nowak (1992, 1993)). An advantage of simulated models is that we can model more complex and more realistic environments. A problem is that understanding properties of these models may again be difficult.

In this paper we will study a simulation model of local interaction and local evolution which is based on Axelrod's (1984, p. 158ff) model. Within an extended and modified framework we analyze the effect of more complex strategies and the impact of synchronous

---

[1] We think it is sensible to use the term *evolution* both for behavioral changes of *individuals* and *populations*. *Learning* will be used only for behavioral changes on the *individual* level. While *learning* can mean both *conscious* and *unconscious* behavioral changes we will always use it as unconscious learning.

versus asynchronous timing of an evolutionary process.

An issue that we will address several times will be the one of *timing of evolutionary dynamics*. This point becomes of particular interest with local models where the same set of players interacts repeatedly. Repeated interaction gives rise to repeated game strategies, strategies where players condition their behavior on past experience. In such a setting timing of evolutionary steps and timing of interaction might interfere. In *classical* evolutionary dynamics[2] theory bases on the paradigm of animals that are randomly matched for a single interaction and never meet again. For this kind of random interaction it is appropriate to assume players to have no memory. Thus standard evolutionary dynamics can separate the evolutionary dynamics from the repeated game dynamics. With local interaction we have to analyze a model that combines these two dynamics and takes care of their mutual influence. Even after an evolutionary step where some strategies did replicate still interactions that took place *before* this evolutionary step might influence the behavior of some players in the current period.

In this paper we study a model where a population is represented as a cellular automaton. The advantage of the cellular automaton model is that it might come closer to what we observe in reality. All players are connected through a chain of neighborhoods but these connections can be very diverse. The disadvantage of the cellular automaton model is that due to the diversity of connections among players this structure it is substantially harder to analyze. Therefore cellular automata are often analyzed using simulations.

In the recent literature cellular automata are used frequently to model population behavior. Naturally there is more than one way to model population behavior with a cellular automaton. Some authors (e.g. Sakoda (1971) and Schelling (1971)) take players' states as fixed and introduce dynamics of the cellular automaton through movements of players. Others (Axelrod (1984, p. 158ff) and Nowak and May (1992)) take players' positions as fixed but allow players to change their states. Further there are models where both players are allowed to move and to change their state (see Hegselmann (1994)).

Sakoda (1971) presents a ‘checkerboard model of social interaction’. He studies a cellular automaton where cells can be either empty or occupied by players of one of two types. Types have different attitudes towards each other and players have randomly the possibility to make small steps in order to move to an empty position where attitudes towards their neighbors improve. Sakoda then considers different combinations of attitudes which he also links to different situations in reality and explains why groups mix or segregate in certain patterns. He views his model as a “breakthrough in the wall separating psychological concepts from sociological ones” (Sakoda 1971, p. 119).

Schelling (1971) studies similarly a model where two types of players live on a line or, as in Sakoda's model, on a checkerboard. Players of each type prefer to live in a neighborhood which consists mainly of their own type. Randomly they get the opportunity to move to more convenient place. In this framework Schelling studies various initial configurations and explains why segregations appears simply via unorganized individual behavior without any collective enforcement or economic need.

Sakoda and Schelling both analyze a model where *states* of players remain constant. The dynamics comes in through *traveling* of players. John Conway invented in 1970 the

---

[2] See Maynard Smith and Price (1973), Taylor and Jonker (1978) and Zeeman (1981).

game of live, a cellular automaton, where players do not have the opportunity to travel but may *change their states*. Each player can be in one of two states, alive or dead, and changes its state according to the state of its eight neighbors. A living player remains alive only if it has two or three living neighbors, otherwise it dies of loneliness or overpopulation. A dead player becomes a living player only if it has exactly four neighbors. This cellular automaton produces, starting from various initial configurations, an enormous number of different patterns of behavior, some of them stable, cycling, moving into a certain direction and possibly generating new structures up to chaotic behavior.

Wolfram (1984) could classify *some* simple cellular automata but most of them seem to be too complex to have analytically predictable properties.

Axelrod (1984, p. 158ff) analyzed a cellular automaton where all cells are occupied by players that are equipped with different strategies. Players achieve each period payoffs of a tournament against all their neighbors respectively. Between periods players copy successful strategies from their neighborhood. This process gives rise to complex patterns of different strategies. Axelrod finds that most of the surviving strategies are strategies that are also successful with global evolution. Nevertheless some of the locally surviving strategies show only intermediate success in the Axelrod's global model. These are strategies that do good against themselves but are only moderately successful against others. They can be successful in the spatial model because they typically form clusters where they only play against themselves.

While Axelrod studies a population with a large number of possible strategies, Nowak and May (1992) analyze a population playing a prisoner's dilemma where players can behave either cooperatively or defectively. They study various initial configurations and analyze the patterns of behavior that evolve with synchronous interaction.

A cellular automata model where both players change their state and their position in the network has been proposed by Hegselmann (1994). He studies a model where payoffs of the prisoners' dilemma to be played depends on the 'risk-class' of the opponents. Players may not change their risk-class, but they may choose the strategy and at least sometimes their location. He finds convergence both to clusters of players that belong to a similar risk-class and cooperation among members of the same class.

In the following we will study a model where players are not allowed to travel but where they change their states following some simple evolutionary mechanism. Thus our model can be compared with the one of Nowak and May and the one of Axelrod.

In contrast to Nowak and May's model we allow for different actions against different opponents if histories against these opponents differ. We introduce discriminatory behavior of players in representing strategies of the repeated game as small automata that are identical for a player, but possibly in different states against different neighbors.

While introducing discriminative behavior might seem similar to the repeated game strategies in Axelrod's model there are some important differences. In Axelrod's model players' behavior previous to a learning step influences players' behavior after the learning step only through the learning process and *not* due to the repeated game strategy. In Axelrod's model repeated game strategies are *only* important *between* learning steps. It is thus possible to interpret Axelrod's evolutionary process as one with a particular *synchronization of learning and memory* where after each learning event all players completely forget their experiences from the previous round. We think that these synchronizations are rarely present in reality and argue below that properties of such a synchronized model differ substantially from an asynchronous model. Another possible interpretation of Axelrod's model would be to represent the choice of a repeated game strategy for a single tournament as the choice of a stage game strategy for a coordination game. Thus one major difference to Axelrod's model is that we assume neighbors of learning players to preserve their memory.

A further substantial difference to the models of Axelrod and May and Nowak is that not everybody learns at the same time. While the *synchronization of learning and interaction* simplifies the analysis a lot, it might rarely appear in reality. We argue below in detail that again properties of a model with such a synchronization differ substantially from those of a model which is asynchronous. We therefore concentrate mainly on a model where both interaction and learning are independent stochastic events. That introducing stochastic interaction and evolution might matter has recently also been mentioned by Glance and Huberman (1993) who argue that cooperation might be extinguished by introduction of stochastic behavior. While we completely agree with the spirit of their argument we, nevertheless, try to point out that *if* stochastic behavior is introduced, cooperation does not break down necessarily. It is the way *how* stochastic behavior is introduced that determines persistence or breakdown of cooperation.

The evolution of discriminative behavior in prisoners' dilemmas has also been studied recently by Ashlock, Stanley and Tesfatsion (1994) who allow players to 'refuse to play' with undesired opponents. Their model presupposes no spatial structure ex ante—the structure is determined endogenously. The setting that we analyze in the following differs in that the spatial structure is determined ex ante such that 'refusal' is impossible. We introduce a different kind of 'variety' in the space of repeated game strategies allowing for all strategies that can be represented as small automata.

While the cellular automaton model of a population and the introduction of discriminatory behavior incorporates some real life flavor, it is on the other hand difficult to analyze. We have therefore tried to replace analytical beauty by extensive simulations. We carried out about 60 000 simulations on tori ranging from 80×80 up to 160×160 and continuing from 1000 to 20 000 periods. It turns out that most of the results vary only slightly and in an intuitive way with the parametrization of the model. Thus, results can be regarded as robust.

We will present the model in section 2. Section 3.1 defines convergence, section 3.2 describes the space of games that we consider, section 3.3 discusses the representation of our results, section 3.4 connects our results to the findings of May and Nowak (1992, 1993) investigating several models with simple strategies that are more similar to their model, section 3.5 extends the model to more complex (two-state) strategies, section 3.6 extends the analysis to coordination games and sections 3.7 and 3.8 discuss robustness of the results. Section 4 finally draws some conclusions.

2 The Model

2.1 Spatial Structure

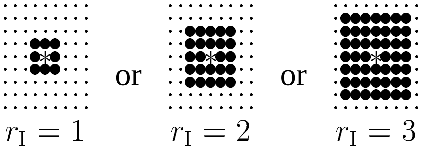
Players are assumed to live on a rectangular grid, e.g. a huge checkerboard, or on a long line. Each cell of the grid is occupied by one player. To avoid boundary effects, the edges of the grid (or the endpoints of the line) are pasted together. Thus, the neighborhood of all players has the same structure. The resulting network is now a torus (or a circle).

Thus a location may represent a geographical position and interaction only with geographical neighbors, location also can be interpreted as producing a certain differentiated product and interaction only with producers that manufacture a similar product. In the context of a model with overlapping generations one dimension of location can represent time where interaction takes place only with the next one or two generations.

Most simulations were carried out on a square of size 80×80. The results presented in sections 3.7 and 3.8 show that for sufficiently large networks neither the exact size of the network nor the dimension matters significantly.

2.2 Neighborhoods

In contrast to models of global evolution and interaction we will consider players who interact only with their neighbors and who only learn from their neighbors. Below we sketch some neighborhoods characterized by a different 'interaction radius' r_{I}. The possible interaction partners of a player $*$ are marked with $\bullet$.

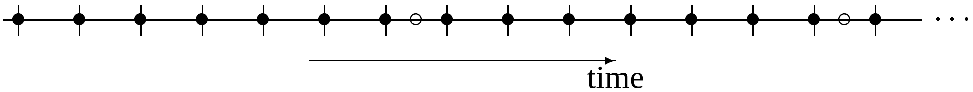


Each player has an 'interaction neighborhood' of radius r_{I} which determines the set of player i's possible opponents N_{I}^i. As will be explained below a player need not interact in a given period with *all* members of N_{I}^i.

In the same way we construct a 'learning neighborhood' N_{L}^i with a similar structure, but with a possibly different radius r_{L}. In this paper we will assume always that whenever a player learns he has information on *all* members of N_{L}^i (see Kirchkamp and Schlag (1995) for a simplified model where a learning player learns in a given period only from *one* member of N_{L}^i).

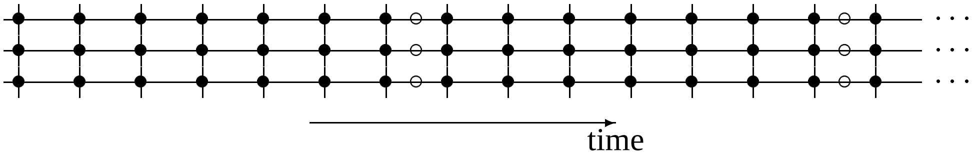
2.3 The Role of Time

When players interact and learn in the above described neighborhoods we will assume their behavior to be *asynchronous*. This is possibly not the standard way to model evolution of repeated game strategies. Let us consider the following example: We want to model a population where neighboring players interact about once a day and change their strategy about once a week. We describe interaction of a player with his top and bottom neighbor by ┼ and learning by ○. Two weeks in the life of a member of this population could be represented as follows.



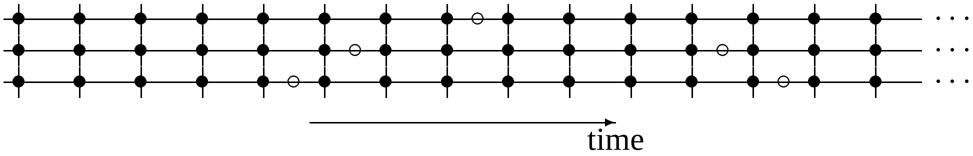
(1)

For simplicity we could model this behavior assuming that all possible interactions take place daily while learning and change of strategies happens only on Sundays. The following diagram shows a part of the life of three neighbored players that interact each with their top and bottom neighbor and learn after seven interactions:



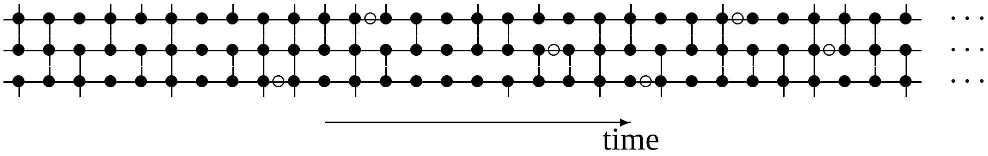
(2)

We will see in sections 3.4 and 3.5 that this simplification can influence outcomes considerably. To introduce what we call 'asynchronous learning' we could assume that not everybody learns on Sundays, but that learning for each player is a random event that is equally distributed over the whole week. Interactions still take place synchronously each day at noon. Success or failure of strategies is observable in the afternoon, so that each day some players could learn in the evening. The group of learning players would be different each day. Here again, a small subset of the population:



(3)

This model still requires players to interact synchronously (at noon) and we will see that this apparently harmless simplification affects the results. If we look closer at such a population we might find that on some days a player might interact twice with his neighbor while on other days he might not interact at all. Given that he might interact twice on a single day makes it necessary to split one day into at least two periods. Like the learning events discussed above also interaction could now occur stochastically in the morning or in the afternoon. If we denote interaction with the top player by ┴ and by ┬ for the bottom player respectively then the following sequence might be possible:



(4)

Evolution of repeated game strategies is often analyzed in a framework that is similar to diagram 2. When we discuss evolution of repeated game strategies here, we prefer an environment like the one represented in diagram 4.

To be precise we will assume that each period for each possible interaction a random draw decides (typically with probability $p_{\mathrm{I}} = 1/2$) whether this interaction takes place.

Thus, each period a player will at time t not play against all his neighbors N_{I}^i but only against a subset $N_{\mathrm{I}}^{i,t}$ which has on average half the size of N_{I}^i. Each period t the composition of his opponents will be different. We have used an interaction probability $p_{\mathrm{I}} = 1/2$ most of the time because it is small enough to avoid synchronization and not too small that simulations become too slow.

Since we also want the timing of *learning* to be stochastic, we will assume that strategies have something like a stochastic 'lifetime' t_{L} that in our simulations is typically distributed equally between 20 and 28 periods. Once the 'lifetime' expires players 'learn'. They possibly change to a new strategy and get a new 'lifetime' which is again a random number between 20 and 28. We mostly used a 'lifetime' in this range because it is large enough to give even complex strategies an opportunity to unfold. Thus, learning is a rare event as compared with interaction, but still learning occurs often enough to keep the simulations fast.

2.4 The Stage-Game

The above setting could be applied to all two player games. We will concentrate here on symmetric 2×2 games and in particular to the case of the prisoners' dilemma. Strategies will be named C and D.

Notice that all the dynamics of population behavior that will be discussed below are invariant to transformations of payoffs like adding constants or multiplying with a positive number, therefore we can represent the space of *all* symmetric prisoners' dilemmas with the game

		Player II	
		C	D
Player I	C	g g	1 h
	D	h 1	0 0

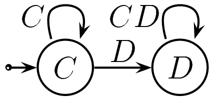
(5)

where $0 < g < 1$, $h < 0$ and $g > \frac{1}{2} - \frac{1}{2}h$.

2.5 Repeated-Game Strategies

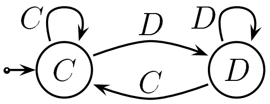
Nowak and May (1992, 1993) consider a model of spatial evolution of undiscriminative strategies. A player could either always play C against all his neighbors or always play D. A player having e.g. one neighbor playing always C and another neighbor playing D might, however, be tempted not to use the same strategy of the stage game against both neighbors. He might want to discriminate among his two neighbors.

To model discriminative behavior we here assume that players use repeated game strategies that can be represented as particular automata that are also called Moore machines. An example for such an automaton is 'grim':



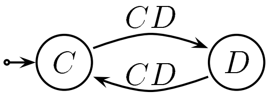
Grim is a strategy that has two 'states'. The stage game strategy that is actually used by grim in either of the two states is shown within the two circles. In the left state grim plays C, in the right state it plays D. The little arrow to the left of grim indicates that grim starts always with its left state, hence begins a game always with C. The notion of automata allows us to describe how a strategy reacts on the opponents behavior. Possible actions of an opponent are written next to an arrow that leads from the current state to the next state. Thus, if grim plays currently C while the opponent plays D then grim will follow the D-arrow and move to the right state. Being in the right state means that grim will from now on play D as well. If on the other hand a C-playing grim meets an opponent who plays C as well, then grim takes the arrow that is labeled C. This leads back to the left state and grim will play C next time. We further see that once grim is in its second state then there is only one arrow that is taken. Regardless whether the opponent plays C or D this arrow leads always back to the second state. To summarize, grim will always start friendly and play C. Grim will remain there as long as the opponent plays C as well. As soon as the opponent plays a single D grim will switch to its second state and play D forever. Thus, the behavior of grim is particularly unforgiving.

Another common strategy is tit-for-tat:



Tit-for-tat does exactly what the opponent did in the last period. If the opponent played D the tit-for-tat replies with D next period, if the opponent was nice and played C then tomorrow tit-for-tat will play C as well

The following automaton, which we call blinker, may be particularly stupid, but since we will meet it again in section 3.5 we will explain its behavior:



Blinker starts always playing C and, regardless of what the opponent does, continues with D, and then plays C in the next period again.

Of course, it is possible to construct automata with more than two states. In the remainder of this paper we will focus on populations where only strategies with less than three states are present. We make this restriction mainly to limit the number of possible automata (there are only 26 automata with less than three states). We argue that this is not a severe restriction since many interesting strategies (like grim, tit-for-tat, tat-for-tit etc.) are already present in this set. We have done some simulations with more complex automata and found that our results do not change. We argue that evolution of strategies applies only to contexts where players do not calculate in a particularly clever way the optimal strategy for a game, but instead are guided by a simple learning process. Then modeling players' strategies to be less sophisticated is only consistent. We further do not claim that *all* automata with less than three states are sensible strategies, but we expect that in a sensible model odd strategies are eliminated through evolution.

Let us now check what happens if a grim plays against a blinker. In the first period both will start with their initial state and, thus, both will play C. Grim will follow the C-arrow that leads back to the left state. Thus, grim will still play C in the next period. Blinker on the other hand is now in its second state and will play D. Observing this, grim will now follow the D-arrow and switch to the second state and thus play D in the next period. Blinker meanwhile switched back to C. From now on grim will always play D, while blinker switches constantly between C and D. The sequence of actions is then as follows:

Period:	1	2	3	4	5	6	7	8	9	$\cdots$
Grim's action:	C	C	D	D	D	D	D	D	D	$\cdots$
Blinker's action:	C	D	C	D	C	D	C	D	C	$\cdots$

Let us assume on the other hand two grims that play against each other. Both will start to play C and will never have any reason to switch to their second state. Thus, the pattern of actions will be the following:

Period:	1	2	3	4	5	6	7	8	9	$\cdots$
1st Grim's action:	C	C	C	C	C	C	C	C	C	$\cdots$
2nd Grim's action:	C	C	C	C	C	C	C	C	C	$\cdots$

So far we have only considered interaction of no more than two strategies. In the networks that we will discuss below a single player will have several neighbors. For each of his neighbors he has a copy of his preferred automaton. While all automata of one player are identical for all of his neighbors they can be in different states, thus allowing for distinguishing behavior.

To give an example, imagine three players that might occupy three floors of a house. All of them play a prisoners' dilemma against their immediate neighbor. The second floor player interacts both with the third floor player and with the first floor player while the latter do not interact with each other. Now assume both the second and the third floor plays grim while the first floor plays blinker. Following the same considerations as above we have the following interactions:

Period:	1	2	3	4	5	6	7	8	9	$\cdots$
3rd's behavior vs. 2nd:	C	C	C	C	C	C	C	C	C	$\cdots$
2nd's behavior vs. 3rd:	C	C	C	C	C	C	C	C	C	$\cdots$
2nd's behavior vs. 1st:	C	C	D	D	D	D	D	D	D	$\cdots$
1st's behavior vs. 2nd:	C	D	C	D	C	D	C	D	C	$\cdots$

This gives already an example for discriminating behavior of the second floor player. Since we combine interaction and evolution it might happen that at some stage the first floor player learns to use a different strategy. Let us assume this happens in period 15 where he becomes a grim. The sequence of actions is then as follows:

Period:	$\cdots$	12	13	14	15	16	17	18	19	$\cdots$
3rd's behavior vs. 2nd:	$\cdots$	C	C	C	C	C	C	C	C	$\cdots$
2nd's behavior vs. 3rd:	$\cdots$	C	C	C	C	C	C	C	C	$\cdots$
2nd's behavior vs. 1st:	$\cdots$	D	D	D	D	D	D	D	D	$\cdots$
1st's behavior vs. 2nd:	$\cdots$	D	C	D	C	D	D	D	D	$\cdots$

The new grim starts thus cooperatively in period 15, but soon finds out that his opponent from the second floor already defects and therefore switches to his second state as well. Thus, we see two pairs of grim here. The top pair cooperates all the time while the bottom pair defects from now on. This behavior occurs due to asynchronous learning. If players were to learn synchronously in one and the same period and afterwards everybody would start in the first state such phenomena would be excluded.

2.6 Relevant history

When players learn, one source of information they will use will be average payoffs of their neighbors. But which payoffs and interactions are the ones that count for determining the average payoff. We will denote the set of periods that player i considers as relevant or that he can access in period t with $E^{i,t}$. We will consider two extreme cases. A player could regard only today's payoffs and interactions as relevant. We will call this 'short memory'.

$$E^{i,t} := \{t\} \tag{6}$$

The other extreme case that we consider assumes that *all* the payoffs and interactions that a player experienced while using the same strategy without interruption are relevant for his learning decision. Thus, if a player used strategy A for the last 24 periods, then all payoffs and interactions achieved during these 24 periods count for the average payoff of this player. We will call this 'long memory'. Denote player i's strategy at time t with $x^{i,t}$ then we formulate 'long memory'

$$E^{i,t} := \left\{t' | \forall_{\tau \geq t'} : x^{i,\tau} = x^{i,t}\right\} . \tag{7}$$

Long memory is particularly reasonable in the context of the repeated game strategies that we analyze. We have seen in section 2.5 that the usage of repeated game strategies may lead to patterns of changing payoffs. E.g. the blinker that plays against a grim alternates in a prisoners' dilemma between low payoffs in odd periods and high payoffs in even periods. Observing only current period's payoff may lead to serious misperceptions of a strategy's performance. Averaging over several periods avoids this problem.

We sum up the total number of player i's interactions with his current strategy during the relevant history $E^{i,t}$ as

$$n_{\mathrm{e}}^{i,t} := \sum_{\tau \in E^{i,t}} |N_{\mathrm{I}}^{i,\tau}| . \tag{8}$$

We further sum up player i's payoff during $E^{i,t}$ as

$$\pi_{\mathrm{e}}^{i,t} := \sum_{\tau \in E^{i,t}} \pi^{i,\tau} . \tag{9}$$

Below we further need a definition of 'users' of a strategy s at time t in the neighborhood of player i:

$$U_s^{i,t} := \left\{j | j \in N_{\mathrm{L}}^i \wedge x^{j,t} = s\right\} \tag{10}$$

2.7 Update of Strategies

In the following we will describe the behavior of players who are not fully rational, who are not able or not willing to analyze games, and who do not try to predict their opponents' behavior. Nevertheless, players' behavior will have some structure since it is guided by imitation of successful strategies. We argue that in many situations players simply copy good examples without knowing why the example was so successful and without spending too much effort in checking whether this strategy might be as promising for the copying player himself.

We further assume that such an imitating player has incomplete information about the total population. All he can observe are strategies and payoffs in his neighborhood N_{L}^{i}. We may imagine that, when in period t a player learns he asks his neighbors $j \in N_{\mathrm{L}}^{i}$ for their strategy $x^{j,t}$ and their respective payoff $\pi^{j,t}$. Let us consider the following population:

player:		$i-4$	$i-3$	$i-2$	$i-1$	i	$i+1$	$i+2$	$i+3$	$i+4$	
strategy:	$\ldots$	$\boxed{A}$	$\boxed{A}$	$\boxed{A}$	$\boxed{B}$	$\boxed{C}$	$\boxed{B}$	$\boxed{A}$	$\boxed{C}$	$\boxed{D}$	$\ldots$
average payoff per interaction:		12	4	5	6	4	0	3	–	3	
number of interactions:		5	4	5	2	4	4	4	0	2	

$$\underbrace{i-3 \quad i-2 \quad i-1 \quad i \quad i+1 \quad i+2 \quad i+3}_{\text{player } i\text{'s learning neighborhood } N_{\mathrm{L}}^{i}}$$

Assume that player i learns and that he can see three players to the left and three players to the right. Thus, he does not realize that farther to the left there is an A with a high average payoff of 12. Nor can he see that there even exists a strategy D. He only observes three As with payoffs 3, 4 and 5, two Bs with payoffs 0 and 6 and two Cs, one of them with payoff 4 (he himself) and one of them with no interactions at all.

How can a player evaluate this information? In the following we will study two possible learning rules:

2.7.1 Copy best player

A learning player could simply look around in the neighborhood which he observes and determine the *player* with the highest average payoff. In our example he will find that the highest payoff (6) is achieved by a B.

A learning player that uses the rule 'copy best player' will adopt the strategy of the most successful player, which is in our example a B. Of course, it could well be that there is more than a single player who has the maximal payoff. Then let players use the following tie breaking rule. Define the set of most successful players as

$$M^{i,t} := \operatorname*{arg\,max}_{j \in N_{\mathrm{L}}^{i}} \left(\frac{\pi_{\mathrm{e}}^{j,t}}{n_{\mathrm{e}}^{i,t}} \right) . \tag{11}$$

Then the probability to choose strategy s in period $t+1$ is determined as

$$P(x^{i,t+1}=s) := \begin{cases} 1 & \text{if } x^{i,t} \in \{x^{j,t} | j \in M^{i,t}\} \text{ and } s = x^{i,t} \\ 0 & \text{if } x^{i,t} \in \{x^{j,t} | j \in M^{i,t}\} \text{ and } s \neq x^{i,t} \\ \dfrac{\sum\limits_{j \in M^{i,t} \wedge x^{j,t}=s} n_{\mathrm{e}}^{j,t}}{\sum\limits_{j \in M^{i,t}} n_{\mathrm{e}}^{j,t}} & \text{otherwise} \end{cases} . \tag{12}$$

Thus the player that is to be copied is chosen randomly with probabilities that are proportional to the number of interactions the respective best players had. In the special case where the player's own strategy is among the best strategies, we assume that the player prefers to keep his own strategy.

2.7.2 Copy best strategy

A learning player could also look at the *average payoffs* of strategy s at time t in the neighborhood of player i which we denote with $f_s^{i,t}$:

$$f_s^{i,t} := \begin{cases} \dfrac{\sum\limits_{j \in U_s^{i,t}} \pi_{\mathrm{e}}^{j,t}}{\sum\limits_{j \in U_s^{i,t}} n_{\mathrm{e}}^{j,t}} & \text{if } \sum\limits_{j \in U_s^{i,t}} n_{\mathrm{e}}^{j,t} > 0 \\ -\infty & \text{otherwise} \end{cases} \tag{13}$$

If a strategy is not used in a neighborhood we define its fitness to be $-\infty$ to make sure that it will be never selected by an evolutionary process.

In our example the learning player would find out that A has an average payoff of 5, B has an average payoff of 2, and C has an average payoff of 4.

A learning player that uses the rule 'copy best strategy' switches to the *strategy* with the highest average payoff, thus in our example he will become an A. Again there could be more than one strategy with maximal payoff. Then we use the following tie breaking rule. Define the set of most successful strategies as

$$N^{i,t} := \arg\max_s \left(f_s^{i,t}\right) . \tag{14}$$

As discussed above for the 'copy best player' rule, two strategies could achieve exactly the same average payoff. The probability that player i uses strategy s in the next period is then

$$P(x^{i,t+1}=s) := \begin{cases} 1 & \text{if } x^{i,t} \in N^{i,t} \text{ and } s = x^{i,t} \\ 0 & \text{if } x^{i,t} \in N^{i,t} \text{ and } s \neq x^{i,t} \\ \dfrac{\sum\limits_{j \in U_s^{i,t}} n_{\mathrm{e}}^{j,t}}{\sum\limits_{\sigma \in N^{i,t}} \sum\limits_{j \in U_\sigma^{i,t}} n_{\mathrm{e}}^{j,t}} & \text{otherwise} \end{cases} . \tag{15}$$

If a player already uses one of the best strategies he prefers his own. Otherwise he adopts one of the best strategies randomly with probabilities proportional to the number of interactions the users of the respective strategies had.

2.7.3 Symmetry of learning rules

Notice that both learning rules described above are *symmetric* in the sense that a player puts the same weight on his own experience (payoff) and on the experience of the observed players. Kirchkamp and Schlag (1995) find that once we let evolution select a player's learning rule, these rules turn out to be *asymmetric* and put more weight on their own payoff. We nevertheless think that symmetric rules have some appeal for their simplicity.

2.7.4 Learning repeated game strategies

If players learn *stage game strategies* they obviously have to use this stage game strategy from the next period on. If players on the other hand learn *repeated game strategies* that are represented as automata we have to explain which state of the automaton players use when they start using it. We find it reasonable to assume that players start with the initial state of the automaton against all their neighbors when they learned a new automaton. Further if players had the opportunity to learn, but do not change to a different automaton we assume that they continue to use the same automata in whatever states they are against their different neighbors.

2.8 Initial State

Throughout the paper we assume that the network is initialized randomly. I.e. first proportions of the available repeated game strategies are determined randomly (following an equal distribution over the simplex of relative frequencies) and then for each location in the network an initial repeated game strategy is selected according to the defined frequency of strategies. Thus, all simulations start from very different initial configurations. If nevertheless results are structured (as they are) they can be viewed as particularly robust.

3 Results

3.1 Convergence

Before we talk about results of simulations, we have to check whether our simulations converge.

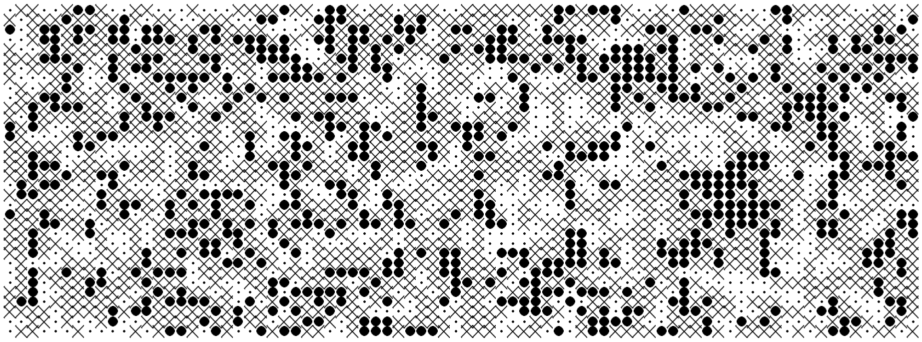
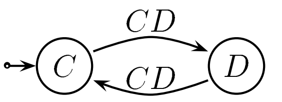
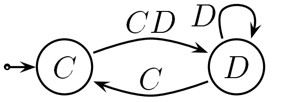
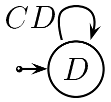


Figure 1: Part of an example net after 2000 periods.

For illustration let us consider a very simple simulation, where the game

		Player II	
		C	D
Player I	C	0.3 0.3	1 -0.7
	D	-0.7 1	0 0

is played on a torus of size 80×80 with neighborhood radius $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, deterministic interaction $p_{\mathrm{I}} = 1$ and stochastic timing of evolution $t_{\mathrm{L}} \in \{10 \dots 14\}$. The grid is initialized randomly. For simplicity of graphic representation we take only the following three automata that are denoted with the following three symbols:

Automata	Symbols
	×
	·
	●

A typical state of a grid after 2000 generations is displayed in figure 1. Here players are still permanently changing their strategy. Thus, at least for this example, even after 2000 periods the population has not converged to a state where each individual's behavior remains constant. Nevertheless the system shows some stability: Proportions of automata and proportions of actions remain more or less constant even if a single player never uses one and the same automaton forever.

Figure 2 on the next page shows a typical pattern of convergence. Here we have monitored the frequency of the combination of strategies CD (and DC) over time. While after a period of stabilization a more or less constant value for the proportion of CD is

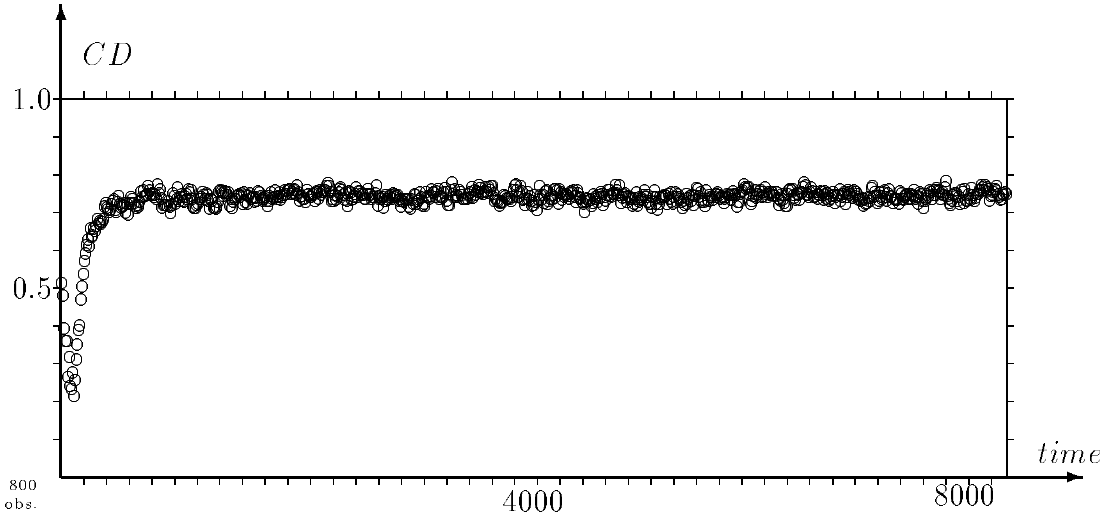


Figure 2: Convergence: Two-State-Strategies, $p_{\mathrm{I}} = \frac{1}{2}$, $t_{\mathrm{L}} \in \{20, \ldots, 28\}$, copy-best-strategy, $E^{i,t} := \{t\}$, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, grid=$80 \times 80$, $g = 0.3$, $h = -0.7$,

achieved there is no asymptotic convergence. The observed value oscillates within a small radius around some attractor. This behavior persists even during very long simulations[3]. We claim that for a stochastic process that describes e.g. the proportion of a certain type of automaton or the proportion of a combination of strategies these oscillations are relatively small.

We will observe later that proportions of automata and proportions of combinations of strategies do not depend on the initial configuration of the network if the network starts from a sufficiently random initial configuration.

In the following we will, therefore, not look at the exact state of the network (because this is confusing as figure 1 on the preceding page shows), but we will look only at relative proportions of automata.

To focus on the essentials we will concentrate mainly on the proportion of combinations of stage game strategies CC, CD, DC and DD. Thus, we do not know exactly, *why* players play a certain combination of stage game strategies, (which automaton a player uses) but we know at least *what* happens.

3.2 The Space of Considered Stage-Games

In this paper we only look at symmetric 2×2 games. Evolutionary dynamics as given by the learning rules ‘copy best player’ and ‘copy best strategy’ (see sections 2.7.1 and 2.7.2) will not change if we multiply all payoffs of a game with a positive constant or if we add a constant to all payoffs of the game. But then it is sufficient to analyze only a small subset of all possible 2×2 games. It turns out that all possible symmetric 2×2 games can be derived from the games given in figure 3 on the next page. These latter games are described by only two parameters g and h, thus, they are easily represented in a plane.

---

[3] Most of our simulations continued for 1000 to 2000 periods. We compared these results with 2000 simulations that continued for 20 000 periods and that lead exactly to the same result.

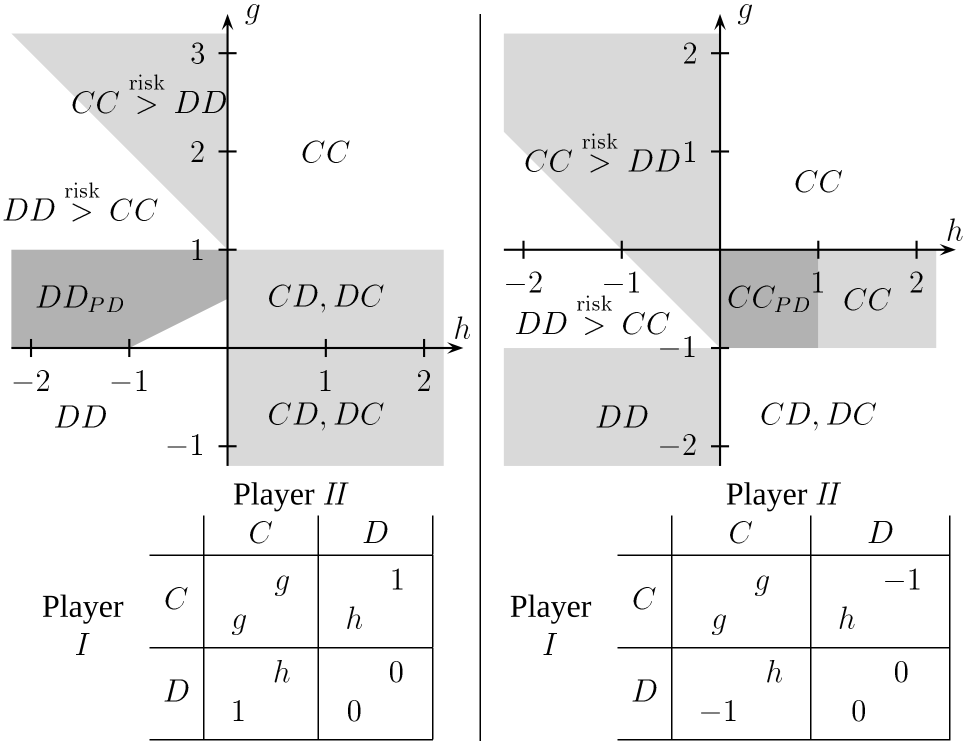


Figure 3: The space of considered stage-games.

Figure 3 shows for both types of games the regions of different equilibria. CC and DD denote regions where we have only one equilibrium that is not a prisoners' dilemma, CC_{PD} and DD_{PD} denote prisoners' dilemmas with one equilibrium, CD, DC is a region where CD and DC are symmetric equilibria, and $CC \overset{\text{risk}}{>} DD$ and $DD \overset{\text{risk}}{>} CC$ denote regions with two equilibria in pure strategies where one risk dominates the other.

Since all prisoners' dilemmas are (by symmetry) contained in the left hand game, we will start by analyzing the DD_{PD}-section of the left game.

3.3 Representation of the Results

To explain the representation of the results, take for example figure 4 on page 18. The figure shows the results of 800 different simulations. A computer program chose 800 times randomly different combinations of g and h. Taking the structure of game 5 on page 7 as given, specific values of g and h define a specific game. g and h were chosen so that most of these games were prisoners' dilemmas ($0 \leq g \leq 1$ and $-2.5 \leq h \leq 0$). With such a game a simulation is started and runs for 2000 periods. Following the arguments given in section 3.1 after 2000 periods we can expect that population statistics like proportions of stage game strategies or automata have converged.

Figure 4 shows three such statistics which are represented as circles in each of the three diagrams (one circle for each of the 800 simulations). The *position* of the circle indicates the payoffs g and h and, thus, the respective game. The *size* of the circle is in the first diagram proportional to the number of observed combinations of strategies CC, in the second proportional to CD (and DC respectively) and in the third proportional to DD. If the frequency of the respective combination of strategies is zero, no circle is plotted.

3.4 Simple (one-state) strategies

Before turning to two-state strategies we will first analyze a simple model (similar to May and Nowak's (1992, 1993)) where only two simple (one-state) strategies 'always C' and 'always D' (in the following denoted with 'C' and 'D') are allowed.

Why can cooperation persist with local evolutionary dynamics, whereas with global dynamics it necessarily dies out? Certainly a *single* C-playing automaton cannot survive if it is surrounded and exploited by Ds. Imagine on the other hand a *cluster* of Cs surrounded by Ds. Here Ds that are located close to the cluster of Cs can observe that the Cs receive a fairly good payoff, because they cooperate with each other. So a D might learn that C is a successful strategy and, thus, become a C. This explains why Cs do not die out necessarily.

Also a C that is situated close to the borderline between C and D is likely to change to a D. His payoff from interaction with a C becomes low if gains from cooperation g are low. Further his D-playing neighbors have a fairly attractive payoff because they are able to exploit at least one C. If gains from cooperation are sufficiently low the average success of D close to the border between C and D is higher than the average success of C. Therefore, we can expect survival of Cs only for games where cooperation is not too costly. The size of the set of games where cooperation survives depends on the dynamics of the particular evolutionary process.

Nowak and May (1992, 1993) consider a spatial model where only Cs and Ds play a prisoners' dilemma. They assumed the learning rule 'copy best player' (see section 2.7.1 on page 11), synchronous interaction ($p_{\mathrm{I}} = 1$), and synchronous evolution ($t_{\mathrm{L}} = 1$). One of their main results is that for certain initial configurations and for certain payoffs cooperation may indeed persist in the prisoners' dilemma.

Do we also observe cooperation for other initial configurations and does cooperation persist in other games even when it is more costly? The upper left part of figure 4 on the following page shows a small dark area which indicates the small range of payoffs where most simulations lead to mutual cooperation in May and Nowak's model. It is no surprise that this small area is close to $g = 1$ and $h = 0$, i.e. close to a range of payoffs where cooperation does not cost too much.

We also see that *some* simulations lead to mutual cooperation for smaller values of h. Closer inspection shows that this was only the case for certain initial configurations, since most of the simulations in this area lead to a mutually defecting population (see the bottom left part of figure 4).

Observation 1 *Local evolutionary dynamics with only simple (one-state) strategies explain deviation from the Nash-Equilibrium solution only to a small extent.*

Glance and Huberman (1993) questioned May and Nowak's model arguing that due to the deterministic dynamics the network might run into cycles that are unstable against small perturbations. They proposed random sequential interaction to model more realistic timing. In a particular setting, where May and Nowak found cooperation, random

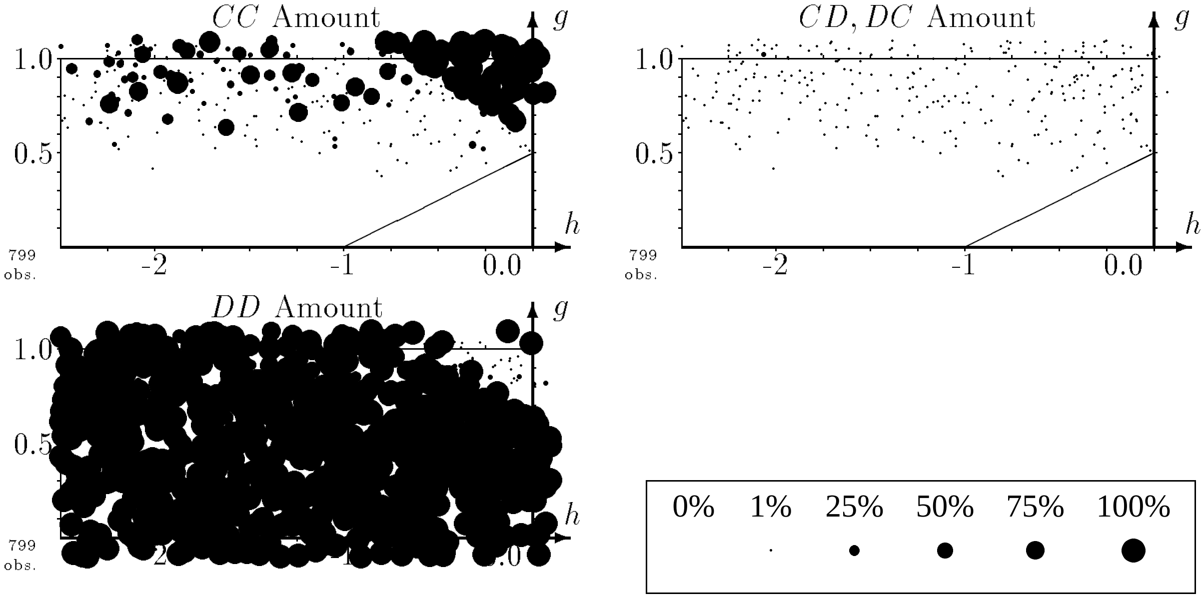


Figure 4: One-State-Strategies, $p_{\mathrm{I}} = 1$, $t_{\mathrm{L}} = 1$, copy-best-player, $E^{i,t} := \{t\}$, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, grid=$80 \times 80$.

sequential interaction lead to general defection. We agree with their finding that for a given initial state results might be altered. We have analyzed the influence of stochastic behavior in the framework that was described in section 2.3 on page 5. The precise way to introduce stochastic behavior differs from Glance and Huberman's model but, as we observe in figure 5 on the following page, we can replicate their result. To save space we display only the amount of mutual cooperation for a variety of models. The top left diagram in figure 5 displays the amount of cooperation in May and Nowak's model. We have seen this diagram already in figure 4. The top right diagram in figure 5 displays the same model, but now with stochastic learning and interaction. Players do not interact deterministically each period with all their neighbors, but carry out each interaction with probability $1/2$. Further learning is not performed deterministically each period, but learning is a stochastic event that happens each $20 \ldots 28$ periods. We see that for most of the prisoners' dilemmas where we found cooperation in the deterministic model on the left, cooperation disappeared in the stochastic model on the right. Thus, Glance and Huberman's criticism holds not only for a particular game and initial configuration, but also for a lot of games and initial configurations.

So far we use only *short memory* (as described in section 2.6) as a base for learning decisions. Matters change if players use *long memory*. With long memory introduction of stochastic behavior has almost no influence on cooperation. In the right model in the second line of figure 5 we find about the same amount of mutual cooperation than in the original deterministic model as used by May and Nowak. The deterministic model with short memory turns out to be a good approximation of the perhaps more realistic stochastic model with long memory. Nowak and May's model has its merits for its simplicity since it is much easier to program than any stochastic model which not only relies on short memory. But if it is overstretched results become at least confusing as the point raised by

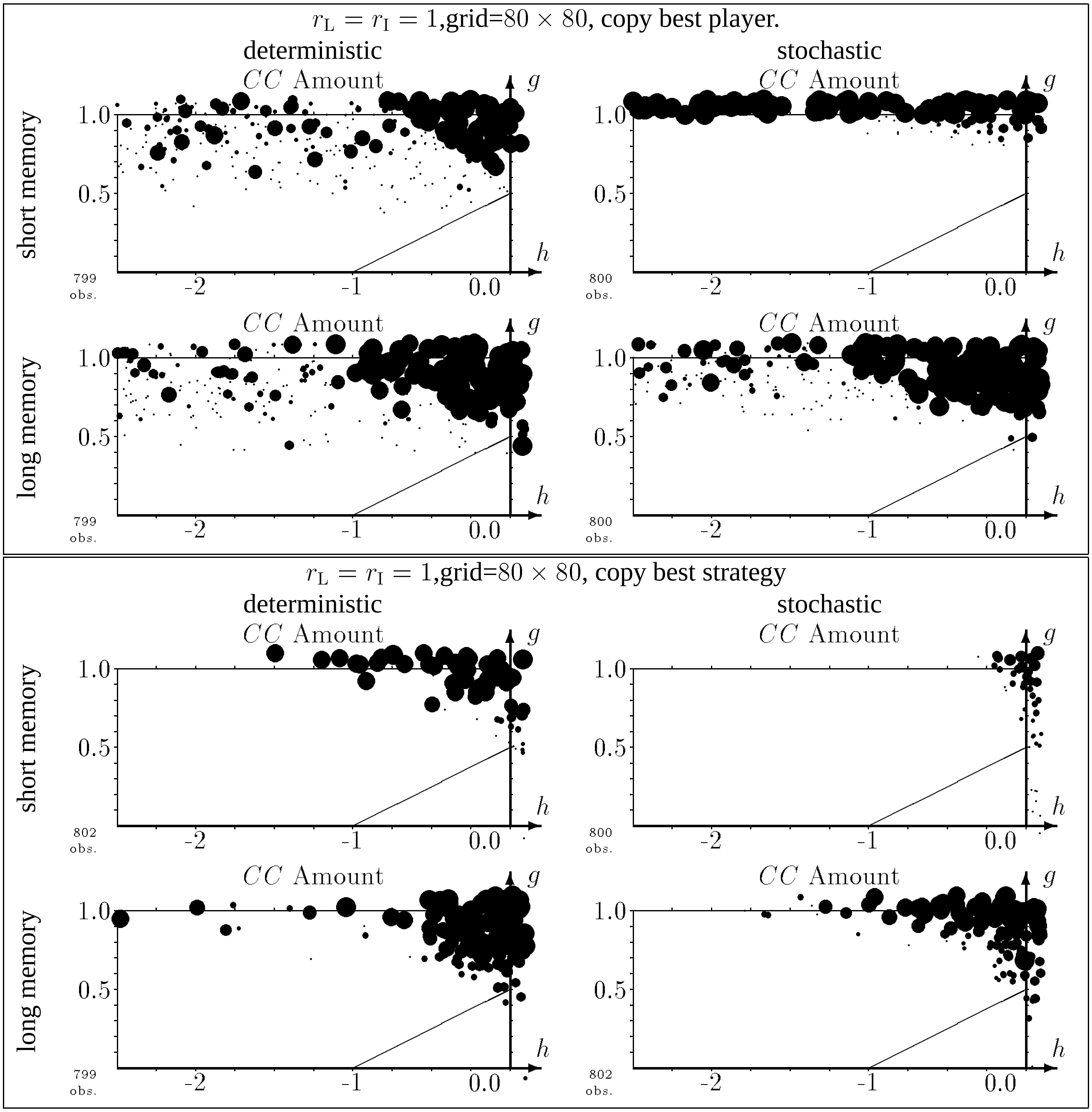


Figure 5: However modeled, one-state-strategies explain cooperation only to a limited degree.

Glance and Huberman shows.

To facilitate comparison with May and Nowak's model we have carried out the above discussion assuming that players use the learning rule 'copy best player'. In the lower half of figure 5 we show the results for the 'copy best strategy' learning rule. It turns out that changing the learning rule has little impact on the amount of mutual cooperation.

Observation 2 *With simple (one-state) strategies introduction of stochastic behavior may change the amount of cooperation depending on the length of players' memory.*

What is not challenged here is the fact that for *the same* initial configuration introduction of stochastic behavior affects the result of a simulation. Nevertheless, depending on the length of players' memory the amount of initial configurations that lead to cooperation in a certain game can be about the same with and without stochastic behavior.

3.5 Introducing discriminatory behavior

The above discussed simple strategies forced players to treat their neighbors without distinction. Either they had to play C against all of them, or they played D. In the following we try to capture discriminatory behavior, assuming that players' repeated game strategies can be represented as small (two-state) automata as described in section 2.5. Each player uses only one automaton, but this automaton can be in a different state against different neighbors. Results change substantially now.

As mentioned above we assume now that players use 'copy best strategy' as defined in section 2.7.2 as a learning rule. Most of the results we find also hold for 'copy best player' (see section 2.7.1).

3.5.1 Deterministic interaction

In this section we treat the subject of deterministic behavior mainly as a modelling issue. In the discussion preceding observation 2 we already tried to motivate that deterministic (or synchronous) interaction is not what we find in reality. Nevertheless we have seen that with simple (one-state) strategies switching from deterministic to stochastic interaction had not much influence on the results. Matters change with more complex (two-state) strategies.

More cooperation with complex automata: Let us first have a look at figure 6 on the next page where we display simulation results for a population that may use automata with less than three states as repeated game strategies.

Observation 3 *Two-state automata lead to more cooperation than simple (one-state) automata. If* $g > \frac{2}{3}$ *predominantly only the pair of strategies* CC *is played.*

This observation is explained easily. For one-state automata (see section 3.4) we have already used the image of a cluster of Cs (always cooperating) surrounded by Ds (always defecting) and, thus, motivated cooperation. With decreasing gains from cooperation g the situation of a cluster of Cs becomes quickly uncomfortable because Cs at the border of

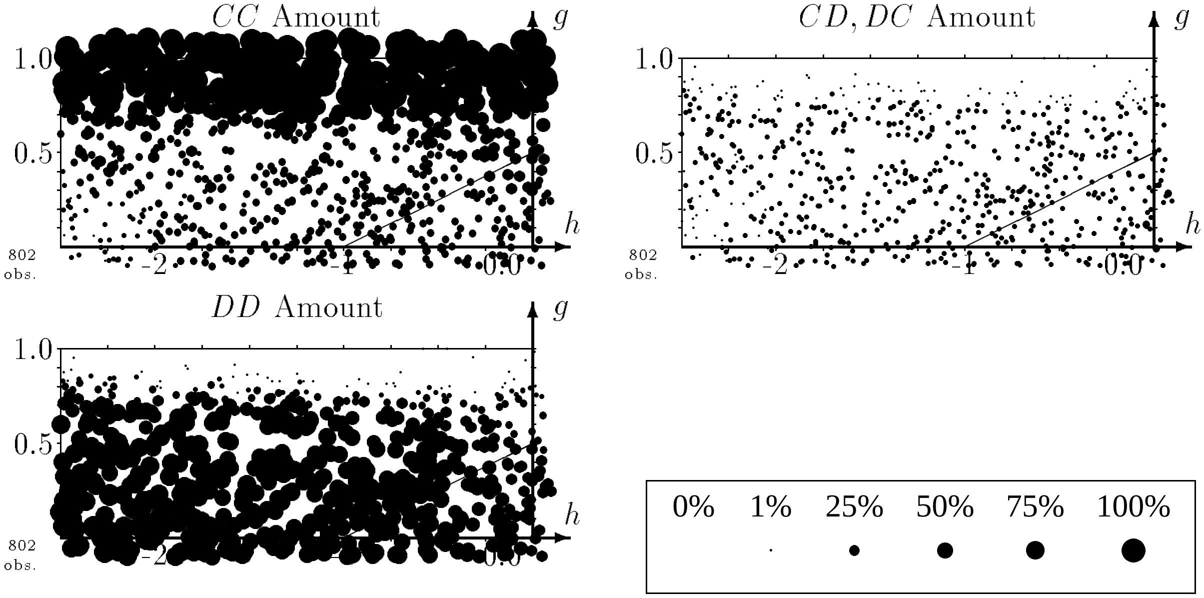


Figure 6: Complex (two-state) strategies induce cooperation: 2-state-strategies, $p_{\mathrm{I}} = 1$, $t_{\mathrm{L}} \in \{10, \ldots, 14\}$, copy-best-strategy, $E^{i,t} := \{t\}$, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, grid=$80 \times 80$.

the cluster are wildly exploited by their D-playing neighbors. Thus, C dies out if rewards from cooperation g are substantially smaller than 1. This is what we have summarized in observation 1 on page 17.

To motivate the larger cooperative area with automata as repeated game strategies, replace the cluster of Cs by a cluster of e.g. tit-for-tat-playing automata. Remember that with two-state automata initially tit-for-tat is present in the network. A tit-for-tat playing automaton is able to cooperate with other tit-for-tat playing automata inside the cluster, but cannot be exploited by Ds outside the cluster. Thus, for the same region of payoffs where the Cs had to give up, the tit-for-tat can still survive. Only if gains from cooperation are substantially lower also cooperation of two-state automata breaks down.

Problems with deterministic interaction: We have mentioned in observation 2 on the page before that for one-state automata switching from deterministic to stochastic modelling has not much impact on the results (if the initial configuration is chosen randomly). This changes with the introduction of discriminatory behavior. One problem is that deterministic interaction may favor automata that appear to be strange. An example is given in figure 7 on the following page. Figure 7 displays on the left the proportions of two automata in the $g \times h$-space, similar to the way we represented the proportion of pairs of strategies for the stage-game in section 3.3. On the right side we display the relative success of the strategies. Simulations that lead to a more-than-average payoff of a given strategy are marked with a /, simulations that lead to a less-than-average payoff are marked with a \. In cases where all automata use the same stage-game strategy average payoffs for all automata are therefore identical and are marked with a –.

We see that both the strategies 'grim' and 'blinker' are in the long run present for a huge set of payoffs. We also see that 'grim' receives in most of the cases more than average

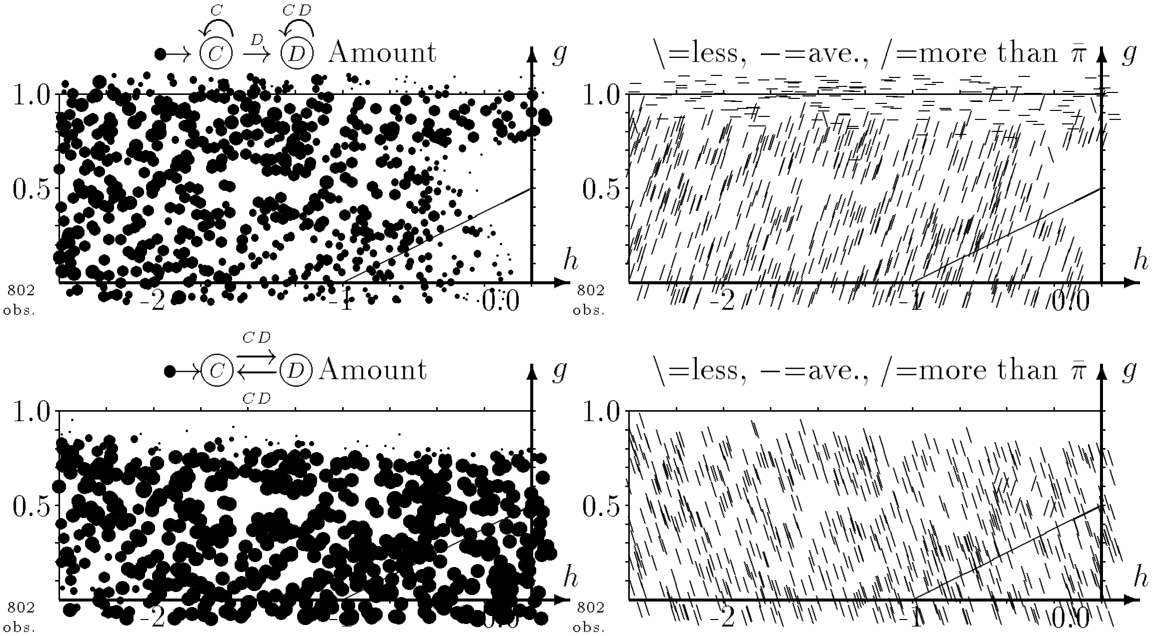


Figure 7: Deterministic interaction may favor strange strategies: 2-state-strategies, $p_{\mathrm{I}} = 1$, $t_{\mathrm{L}} \in \{10, \ldots, 14\}$, copy-best-strategy, $E^{i,t} := \{t\}$, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, grid=$80 \times 80$.

payoff whereas 'blinker' earns less than average payoff. So we come to the following observation:

Observation 4 *With complex (two-state) strategies payoff among strategies may differ for a wide range of payoffs.*

This was not the case for one-state automata, where for most payoffs either only D or only C survived. Thus, the whole population had the same payoff there.

While we introduce this observation in a somewhat arbitrary environment, it remains true for all other parametrizations (in particular for stochastic interaction) which we study in the following.

Not only do the two automata 'grim' and 'blinker' give us a nice example for unequal payoffs of strategies that survive in the long run, further 'blinker' gives a good example for an unreasonable strategy that is eliminated once we add stochastic interaction to the already present stochastic learning. Figure 10 on page 25 show again population shares and relative payoffs for 'grim' and 'blinker'. 'Grim' is still present for most of the population and achieves more than average payoff (if we are not in a payoff range where everybody cooperates). But now 'blinker' is almost completely eliminated:

Observation 5 *Deterministic interaction favors the appearance of 'strange' strategies like the 'blinker'.*

How can it occur that at least in the deterministic model 'blinkers' replicate so successfully? Closer inspection shows that deterministic interaction allows them to form clusters that 'blink' synchronously. Let us assume that one 'blinker' got replicated. This is most likely to happen if during replication the ‘blinker’ was in his second state and was thus supported by his neighborhood. The next period the descendant will start in his first state as the parent does. From now on, when the parent is in his first state the child will be in the first state as well. The same is true for all further descendants of this cluster.

Now, since all elements of the cluster are synchronized, if they play against any other strategy they either *all* do the right thing (they all play D and thus exploit all others simultaneously) or they all do the wrong thing (they play C and support all others with their joint forces). Growth during D-periods more than outweighs losses during C-periods. During periods of low payoff a player has a low chance to change his own strategy because updates do not happen too often. During periods of high payoff he has a higher chance that somebody in the neighborhood copies his strategy. As payoffs of the ‘blinkers’, *if* they are successful, are *really* high, the ‘blinker’ reproduces quickly in this situation.

Mutual exploitation and favorable environments: Further inspection shows that ‘blinkers’ are part of a *favorable environment* that is created by ‘grim’. The latter ‘feeds’ sometimes (when he is born) his environment with a C. This facilitates reproduction of ‘blinkers’ which quickly turn ‘grim’ into his second state and which are successfully exploited from now on.

The kind of ‘symbiosis’ of two automata, where one gains substantially more payoff, but sometimes feeds the other and thus causes it to replicate is typical for this spatial model.

This case gives a nice example for the common fact that exploitation is always mutual. Figure 8 on the following page shows the relative amount of pairs of strategies that are encountered by ‘grim’. The size of the circles is again proportional to the amount pairs of strategy.

The range where DC is played (i.e. a D playing ‘grim’ exploits a C playing opponent) coincides with the range where CD is played. As ‘grim’ has more than average payoff (see figure 7 on the page before) we should not be surprised that DC happens more often than CD. However, ‘grim’ cannot play always DC because then his opponent would encounter low payoffs and copy next time a more successful strategy (e.g. grim).

Let us summarize this in the following observation:

Observation 6 *An automaton that exploits others has to support its victims at least sometimes.*

3.5.2 Stochastic interaction, learning from short memory

Above we argued that deterministic interaction is an unrealistic assumption which leads to questionable results, e.g. the appearance of strange automata. If we disturb the dynamics introducing stochastic interaction the large cooperative payoff region persists, but strange automata disappear.

In the following we assume that in each period each single interaction takes place with probability $p_{\mathrm{I}} = 1/2$. Thus, players interact in each period mostly against half of their neighbors approximately.

Let us first check the amount of cooperation that is displayed in figure 9 on the following page.

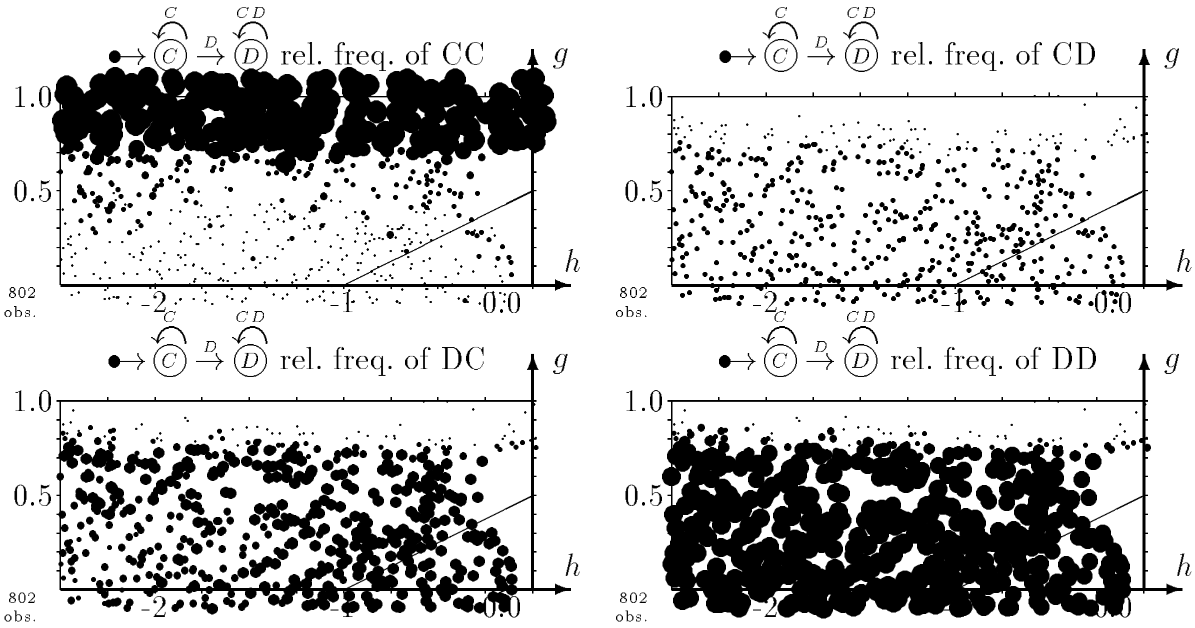


Figure 8: Exploiters have to ‘feed’ their victims at least sometimes: The pairs of strategies that are encountered by ‘grim’.
2-state-strategies, $p_{\mathrm{I}} = 1$, $t_{\mathrm{L}} \in \{10, \ldots, 14\}$, copy-best-strategy, $E^{i,t} := \{t\}$, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, grid=80 × 80.

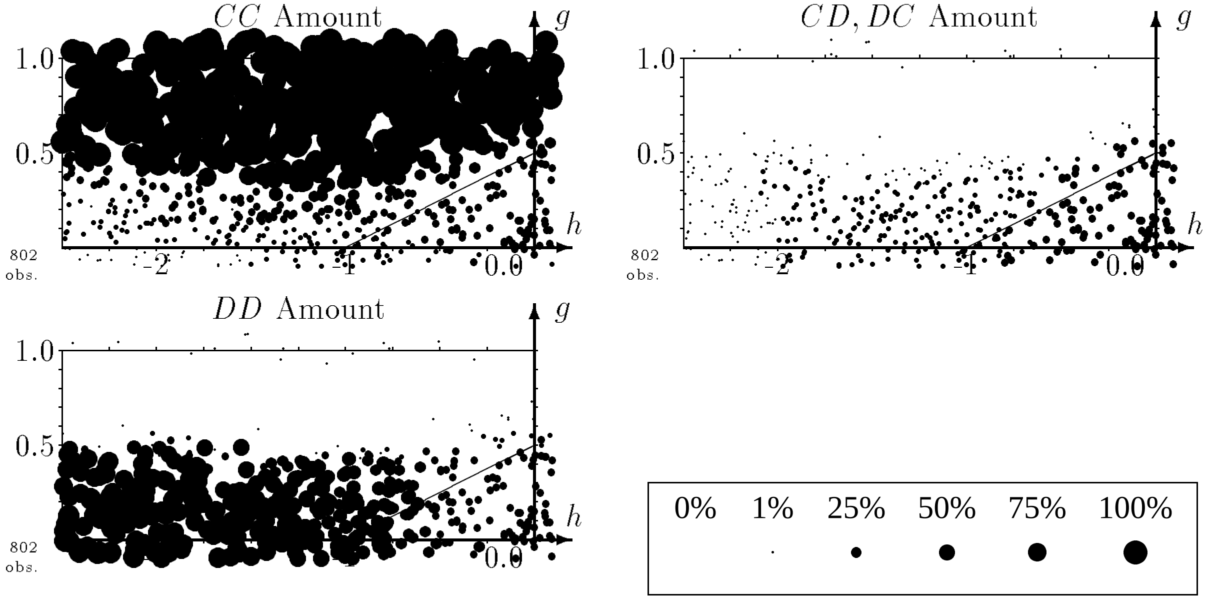


Figure 9: Complex (two-state) strategies induce cooperation also with stochastic interaction: 2-state-strategies, $p_{\mathrm{I}} = \frac{1}{2}$, $t_{\mathrm{L}} \in \{20, \ldots, 28\}$, copy-best-strategy, $E^{i,t} := \{t\}$, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, grid=80 × 80.

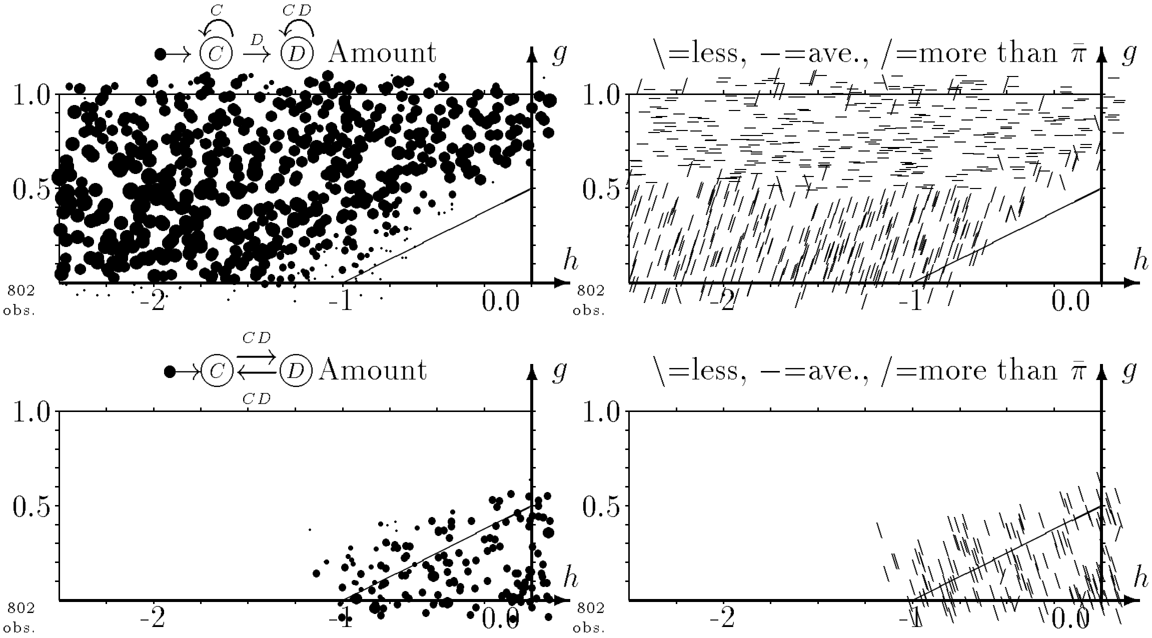


Figure 10: Stochastic interaction eliminates strange strategies: 2-state-strategies, $p_{\mathrm{I}} = \frac{1}{2}$, $t_{\mathrm{L}} \in \{20, \ldots, 28\}$, copy-best-strategy, $E^{i,t} := \{t\}$, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, grid=$80 \times 80$.

Observation 7 *With stochastic interaction complex (two-state) automata induce an even larger cooperative payoff region than with deterministic interaction.*

The argument is here the same as for observation 3 on page 20.

Elimination of strange strategies: Now we can ask what happens to 'blinkers' if their synchronization is disturbed by stochastic interaction. Let us compare deterministic interaction (as described in figure 7 on page 22) with stochastic interaction (in figure 10).

'Blinkers' grow in a deterministic setting because they are synchronized with their neighbors. Introduction of stochastic interaction disturbs the synchronization. Therefore, with stochastic interaction 'blinkers' are mostly seen in a payoff region outside the prisoner's dilemma. This triangle where we find 'blinkers' is exactly the region where playing a correlated pair of strategies that puts weights $\frac{1}{2}$ both on CD and DC pareto dominates all other strategies. Once synchronized it is very easy for the 'blinker' to follow this strategy.

Payoff of 'blinkers' is again lower than average, whereas the payoff of 'grim' is higher. In this setting 'grim' has found other neighbors to exploit.

Problems with learning from short memory: Up to now we have always assumed that the learning decision is based on the *short memory* (as described in section 2.6). We made this assumption to allow for a comparison with May and Nowak's results.

A critique against this approach can be motivated with figure 9 on the preceding page. Let us consider games with $g < 0$. These games are no prisoners' dilemmas anymore. We have still carried out some simulations with games of this type which are displayed at the lower edges of the diagrams in figure 9

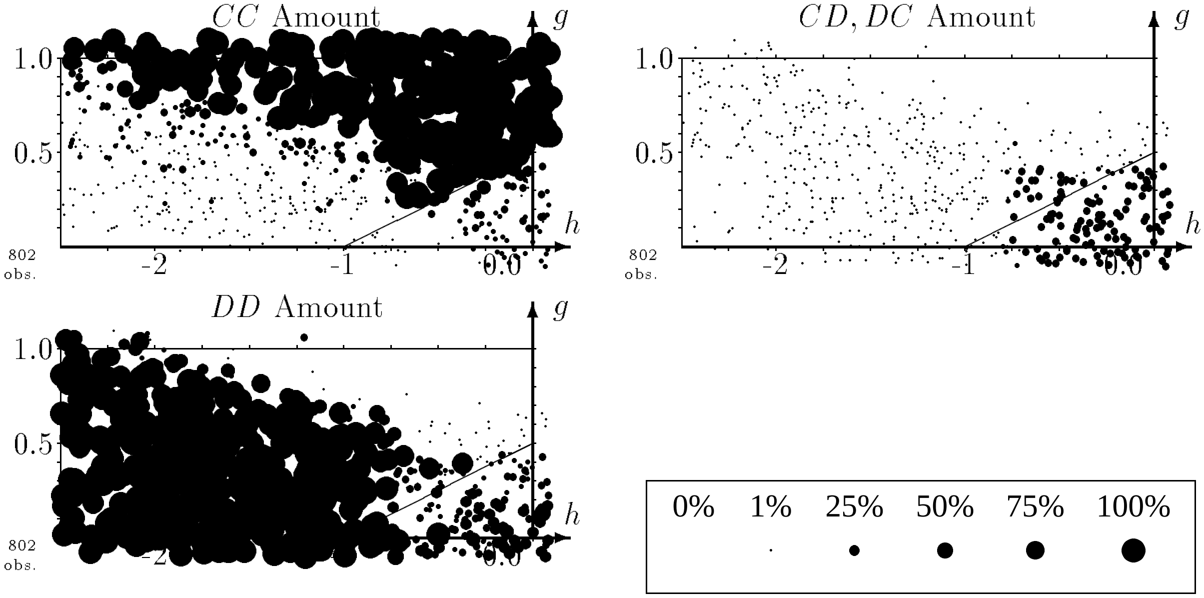


Figure 11: 2-state-strategies, $p_{\rm I} = \frac{1}{2}$, $t_{\rm L} \in \{20, \dots, 28\}$, copy-best-strategy, $E^{i,t} := \{t' | \forall_{\tau \geq t'} : x^{i,\tau} = x^{i,t}\}$, $r_{\rm L} = r_{\rm I} = 1$, grid=80 × 80.

Observation 8 *Lack of memory leads to the strange occurrence of pairs of strategies CD and DC even in games were DD is the pareto-dominant strategy pair.*

In games with $g < 0$ irrationality of C-playing players is even harder to justify than in the prisoners' dilemma. There C is already irrational on the *individual* level. But in the prisoners' dilemma CC was at least *socially* desirable. Thus, CD could be explained as a failed attempt to achieve mutual cooperation. With $g < 0$ the CC-payoff is smaller than the DD-payoff, thus, C is neither *individually* nor *socially* optimal.

However, automata that play both C and D can successfully replicate while they are in their D state, given that simultaneously enough neighbors are playing C in the stage game. This behavior is again supported by a symbiosis of automata where one tries to exploit the other by simultaneously motivating him to reproduce.

In many situations it might be questionable to assume that players copy a strategy due to a one-period success, while the same strategy receives on average (over a complete cycle with his opponent) less payoff. In the following we will therefore prefer the assumption that players have 'long memory' as described in section 2.6 on page 10.

3.5.3 Stochastic interaction, learning from long memory

Compare learning from *short memory* (as shown in figure 9 on page 24) with learning from *long memory* (figure 11).

If we look at the frequency of CD in both cases we find the following:

Observation 9 *Mutual exploitation is much more likely if learning depends only on short memory than if learning relies on long memory.*

This is particularly true for the region $g < 0$, i.e. the case where DD pareto dominates CC. With long memory exploiters cannot take advantage of that fact that yesterday's losses will be forgotten when their victim replicates. Closer inspection of the data shows the following:

Observation 10 *The variance of payoffs over different automata is smaller with learning from long memory.*

To explain this observation we assume an environment with two strategies. Strategy A receives always $\dot{x}$, strategy B gains either x_h or x_l with probability $\frac{1}{2}$ respectively. Say that $x_h > \dot{x} > x_l$ and $\dot{x} > (x_h + x_l)/2$. As the average payoff of B is low, we should expect B to die out. Let us assume that a neighborhood has size n and we observe only a few isolated Bs, each of them surrounded by n players of type A. The probability to learn is assumed to be $1/t_{\mathrm{L}}$ in each period.

With learning from *short memory* the probability that the last B in a neighborhood dies is therefore $1/(2t_{\mathrm{L}})$. On the other hand the probability that an A copies B's strategy is $n/(2t_{\mathrm{L}})$. Therefore, the number of Bs will rise in this situation. Here it is sufficient that from time to time individual payoff is higher than average payoff, no matter how low x_l really is.

With learning from *long memory* things change. Let us assume that learning appears equally likely after t_{L} (even) and $t_{\mathrm{L}} + 1$ (odd) periods. Then the average payoff of B is either $(x_l + x_h)/2$ (with probability $\frac{1}{2}$) or $(tx_l + (t + 2)x_h)/(2t + 2)$ (with probability $\frac{1}{4}$) or $((t + 2)x_l + tx_h)/(2t + 2)$ (with probability $\frac{1}{4}$). So Bs can only reproduce if the weighted average $(tx_l + (t+2)x_h)/(2t+2)$ is larger than $\dot{x}$. For any larger t_{L} this implies that $(x_h + x_l)/2$ has to be very close to $\dot{x}$ if learning uses long memory.

The shape of the cooperative region: If we compare figure 9 on page 24 with figure 11 on the page before we note that the shape of the cooperative region is different. The value of h has more influence on the amount of cooperation:

Observation 11 *With increasing losses from exploitation ($-h$) gains from cooperation (g) have to be higher to induce cooperative behavior with learning from long memory.*

This is different from learning with short memory. There we observed that h had not much influence. In the discussion following observation 10 we explained that with learning from short memory 'bad' experiences (CD) have less influence on reproductive success. With learning from long memory on the other hand their influence matters. Since CD-payoff is determined by the value of h figure 11 on the page before indeed indicates that the shape of the cooperative region is influenced by h.

Observation 12 *CD and DC is played primarily in the region $g < \frac{1}{2} + \frac{1}{2}h$.*

Games in this region are not prisoners' dilemmas, here alternation between CD and DC pareto-dominates the Nash Equilibrium DD.

3.6 Coordination games

We can apply the same analysis not only to the prisoners' dilemma, but also to other games. The game

$$\begin{array}{cc|c|c|} \multicolumn{2}{c}{} & \multicolumn{2}{c}{\text{Player } II} \\ \multicolumn{2}{c|}{} & C & D \\ \cline{2-4} \text{Player} & C & \begin{matrix} & g \\ g & \end{matrix} & \begin{matrix} & -1 \\ h & \end{matrix} \\ \cline{2-4} I & D & \begin{matrix} & h \\ -1 & \end{matrix} & \begin{matrix} & 0 \\ 0 & \end{matrix} \\ \cline{2-4} \end{array} \tag{16}$$

is a 'coordination game' for $g > -1$ and $h < 0$ (see figure 3 on page 16). The game has two pure equilibria, one where both players play C, and another where both players play D. Both players would prefer to coordinate on one of the two equilibria.

Coordination problems appear often in reality and it is not trivial to explain why society solves a coordination problem in a particular way. Cars could either drive on the left or on the right side of a road, coauthors may either use wordprocessor A or B, buyers and sellers may agree upon using either currency A or B, distances may be given in miles or kilometers, travelers may meet at the station or at the airport. . .

Asking which equilibrium players might choose in the above game one might argue that one equilibrium leads to a payoff of 0, the other to g for both players respectively — hence they should coordinate on CC if $g > 0$ and on DD if $-1 < g < 0$. Following this argument is called following the principle of *pareto dominance*.

Consider now a game where $g = 1$ (hence CC pareto-dominates DD) and $h = -100$. Here coordinating on CC involves more risk, because a mistake of the opponent would lead to a painful payoff of -100. On the other hand coordinating on DD leads to less payoff (if coordination succeeds), but is less painful for a player if the opponent fails to coordinate. Therefore we might advise players to coordinate on CC, which means following the principle of *risk dominance*. In the games described in figure 3 on page 16 the equilibrium CC risk dominates DD if and only if $g > -1 - h$. A thorough discussion of risk-dominance is e.g. given in Harsanyi and Selten (1988).

Recent work of Kandori, Mailath and Rob (1993) or Young (1993) suggest that global evolution might select the risk-dominant equilibrium. Our experiences with prisoners' dilemma games above have already shown that local evolution has a tendency to follow the principle of pareto optimality. We will see that coordination games have a similar property.

In figure 12 on the following page we observe that in coordination games players almost always meet on one equilibrium. The amount of (miscoordinating) CD's is extremely rare. We further see that evolution selects neither always the risk dominant nor always the pareto dominant solution. If we draw an imaginary line that separates a region where mostly CC is played from the region where mostly DD is played, this line does not coincide with the line that divides the region of CC pareto dominant equilibria from the DDs nor does it coincide with the line the divides CC risk dominant equilibria from DDs. Instead the evolutionary separation lies between these two lines. We can interpret this as a kind of 'aggregation' of the two principles of risk dominance and pareto dominance.

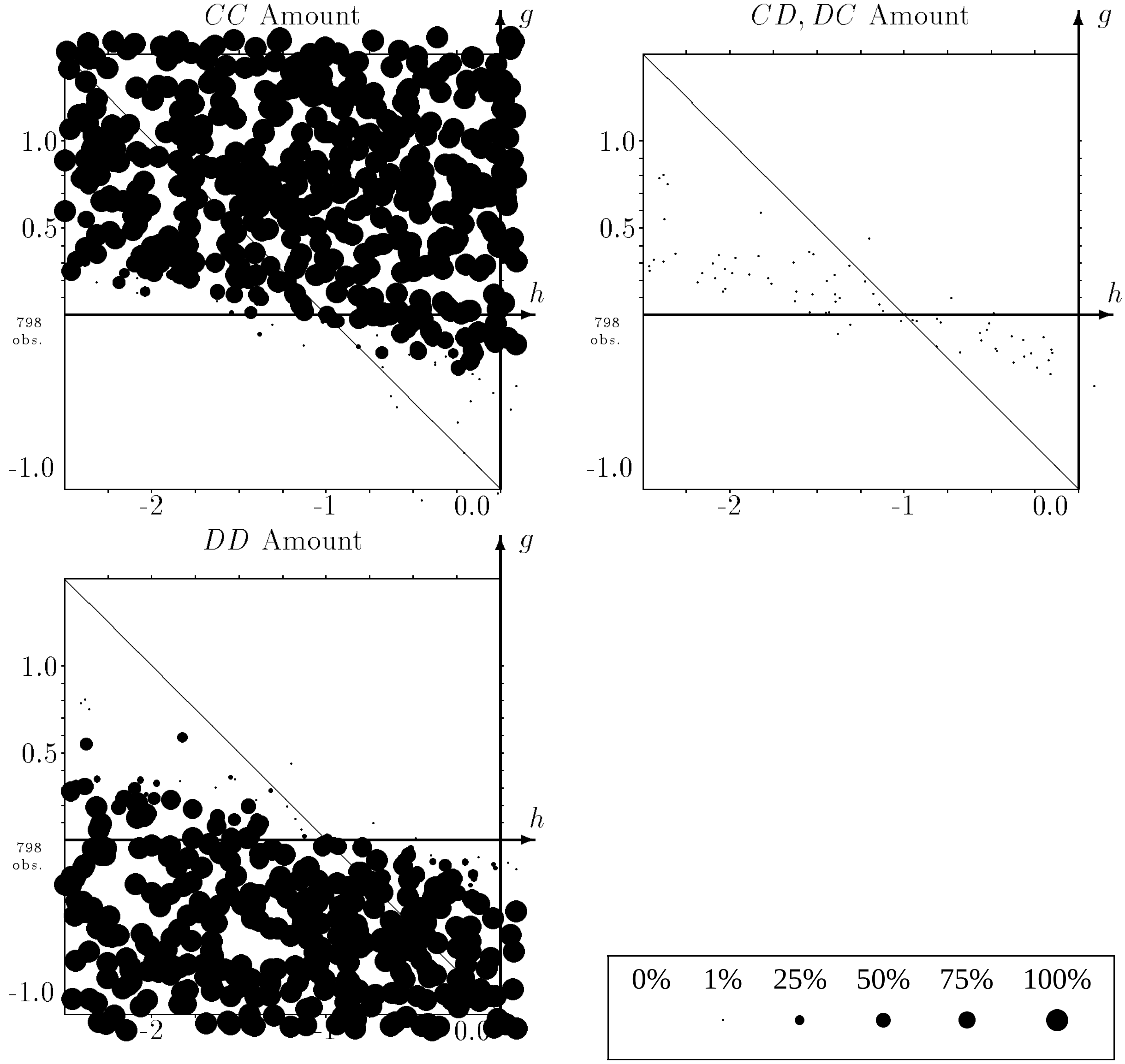


Figure 12: Coordination games with two-state-strategies: 2-state-strategies, $p_{\mathrm{I}} = \frac{1}{2}$, $t_{\mathrm{L}} \in \{20, \dots, 28\}$, copy-best-strategy, $E^{i,t} := \{t' | \forall_{\tau \geq t'} : x^{i,\tau} = x^{i,t}\}$, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, grid=$80 \times 80$.

Observation 13 *In a spatial environment both pareto dominance and risk dominance matter for selection of equilibria.*

The difference between the model of global evolution of Kandori, Mailath, Rob and Young which predict risk dominance and the local evolution that we study here is explained easily. The main argument of Kandori, Mailath, Rob and Young is that once the whole population has settled in the risk dominant equilibrium a proponent for the pareto dominant equilibrium would have to convince a relatively higher part of the population of the superiority of the pareto dominating strategy. In the spatial environment matters change: If even only a small part of the network starts coordinating on the pareto dominant equilibrium their adherents have higher payoffs than their neighbored adherents of risk dominant equilibria. In some circumstances they will convince their neighbors to play also the pareto dominant equilibrium.

3.7 The influence of local learning

A particular feature of the model analyzed so far was the assumption that evolution (or learning) operates on a local level. Figure 13 on the next page shows what happens if we move gradually from this assumption to the assumption of global evolution (or global learning). As larger neighborhoods lead to timeconsuming simulations we have chosen here a smaller torus of size only 21×21. A learning radius r_{L} of 10 means here that a player learns from the whole population, smaller values for r_{L} stand for more local evolution. To save space we have further only displayed the amount of mutual cooperation.

We observe first that the smaller grid size (only 21×21 compared with 80×80 in the rest of this paper) does not affect the results. Comparison of the top left diagram in figure 13 on the following page ($r_{\mathrm{L}} = 1$) with figure 11 on page 26 ($r_{\mathrm{L}} = 1$, but a larger grid) shows no influence of the size of the grid.

We further see that the amount of cooperation decreases gradually while evolution becomes more global. Figure 13 on the following page shows that it does not matter whether the learning radius is e.g. three or four, but that it matters whether evolution operates on a local level at all. The bottom right diagram in figure 13 shows global evolution and, here the cooperative region has become substantially smaller.

Thus, it is not only locality of *interaction*, it is also locality of *evolution* that explains cooperation in this environment.

We have carried out the same exercise keeping the learning radius r_{L} fixed and varying the interaction radius r_{I}. Again we find that moving from a local model with a small interaction radius to a more global model with a larger interaction radius eliminates cooperation gradually and smoothly.

3.8 Other dimensions

Beyond the results that we presented above we have carried out many further simulations, each showing that the effects that we pointed out here persist for several modifications of the model. One possible modification is e.g. the change from a two-dimensional torus to tori with other dimensions. Figure 14 on page 32 shows the amount of mutual cooperation

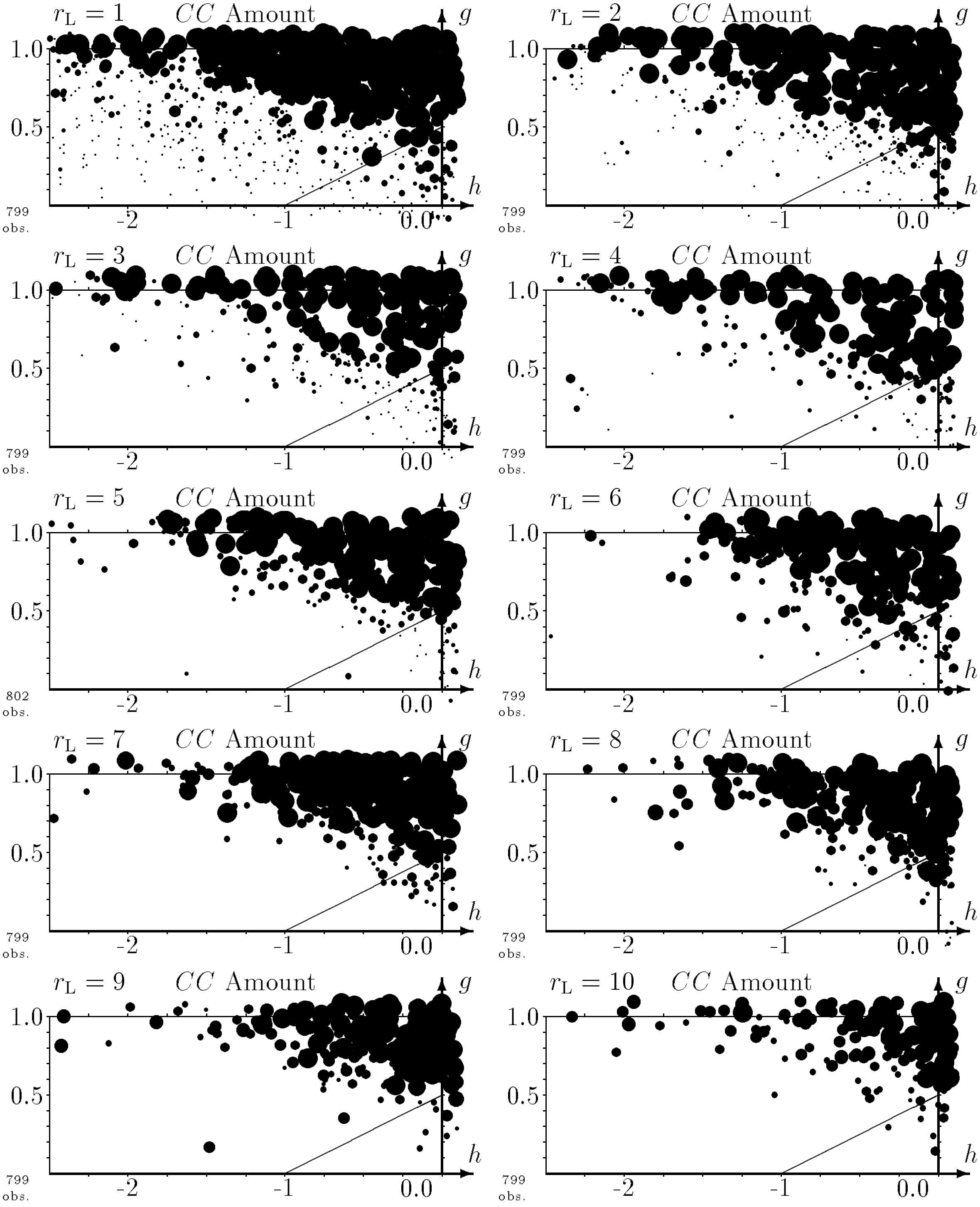
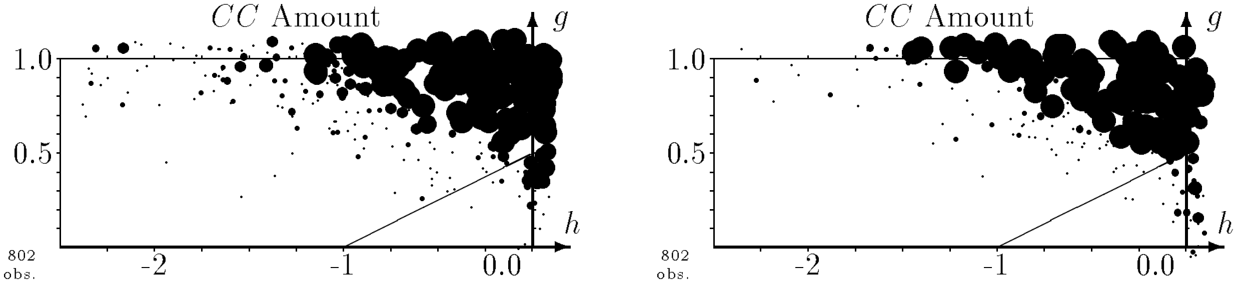
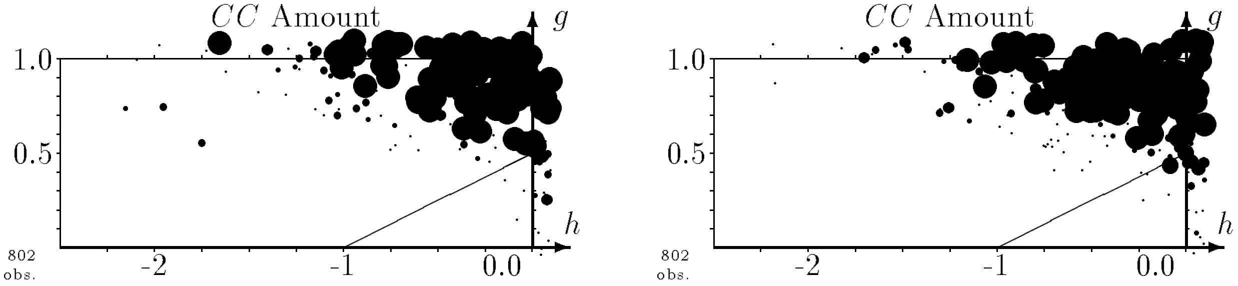


Figure 13: The influence of local learning, two-state-strategies, $p_I = \frac{1}{2}$, $t_L \in \{20, 28\}$, copy-best-strategy, $E^{i,t} := \{t' | \forall_{\tau \geq t'} : x^{i,\tau} = x^{i,t}\}$, $r_I = 1$, grid=21×21.



line : 4096, $r_{\rm I} = r_{\rm L} = 40$ square : 64×64, $r_{\rm I} = r_{\rm L} = 4$



cube : $16 \times 16 \times 16$, $r_{\rm I} = r_{\rm L} = 2$ hypercube : $8 \times 8 \times 8 \times 8$, $r_{\rm I} = r_{\rm L} = 1$

Figure 14: Amount of mutual cooperation in 1. . . 4-dimensional networks: 2-state-strategies, $p_{\rm I} = \frac{1}{2}$, $t_{\rm L} \in \{20, \dots, 28\}$, copy-best-strategy, $E^{i,t} := \{t' | \forall_{\tau \geq t'} : x^{i,\tau} = x^{i,t}\}$.

for networks in several dimensions. All networks have 4096 players and all neighborhoods have almost the same number ($80 \dots 125$) of opponents. Notice that the cooperative area is slightly smaller than the one of figure 11 on page 26. This is due to the fact that the neighborhoods in figure 11 contain a much smaller number of players. The main point here is to note that the dimension of the network has almost no effect on the size of the cooperative area.

4 Conclusions

Among the questions that we followed above we would like to summarize three points: First we analyzed the effect of introducing more or less complex strategies into a population. Let us concentrate on how complexity of strategies affects the amount of cooperation in the prisoners' dilemma. One might have guessed that cooperation might be more stable if players' structure is very simple. Players that are capable of using more complex strategies move quicker to an individually rational solution. Instead it turns out that more complexity fosters cooperation.

Second we find it interesting to note that payoffs vary among strategies in the long run. While inequality might be a realistic phenomenon, global evolution would not be able to explain it. In particular with local evolution and interaction we found symbioses of strategies where one partially feeds but mainly exploits the other.

Third, among several other modeling issues we have compared stochastic and deterministic timing of evolution and interaction. Nowak and May have used a model with deterministic timing and found cooperation for a prisoners' dilemma. Huberman and Glance then argued that stochastic timing might extinguish cooperation totally. While this is doubtlessly true for particular parametrizations of the model we found that stochastic timing can be introduced in a reasonable way and not eliminate cooperation at all.

We have seen that with deterministic interaction the evolutionary system might get stuck into strange patterns of behavior. This behavior of the population is often not robust against stochastic disturbances. We think that stochastic timing of the update of the strategy and stochastic interaction are preferable and avoid these artificial effects.

We haven't investigated many further learning rules yet. In particular the assumption that any information found in the neighborhood is equally valuable, whether it comes from a distant player (who might face a significantly different environment) or from a player nearby, might be regarded as very strong. One problem that comes with the analysis of these learning rules is that their numerous parameters are difficult to endogenize. We have done some simulations that use variants of genetic algorithms (sometimes players copy a repeated game strategy, sometimes they copy a learning rule — see Kirchkamp and Schlag (1995)) in order to determine useful learning rules but still a lot of work has to be done.

We have hopefully convinced the reader that despite the lack of analytical sharpness evolution of networks is a sufficiently interesting field to justify the application of less sharp means like simulations. We tried to show that it is possible to replace analytical generality at least partially by very extensive simulations. In particular we tried to make the results as robust as possible introducing random initial configurations and random proportions of automata.

References

Ashlock, D., E. A. Stanley, and L. Tesfatsion (1994): "Iterated Prisoner's Dilemma with Choice and Refusal of Partners," in *Artificial Life III*, ed. by C. G. Langton, no. XVII in SFI Studies in the Sciences of Complexity, pp. 131 – 167. Addison-Wesley.

Axelrod, R. (1984): *The evolution of cooperation*. Basic Books, New York.

Eshel, I., L. Samuelson, and A. Shaked (1994): "Altruists, Egoists and Hooligans in a Local Interaction Model," Tel Aviv University and University of Bonn.

Glance, N. S., and B. A. Huberman (1993): "Evolutionary games and computer simulations," *Proceedings of the National Academy of Sciences*, 90, 7716–7718.

Harsanyi, J. C., and R. Selten (1988): *A General Theory of Equilibrium Selection in Games*. The MIT Press, Cambridge, Massachusetts.

Hegselmann, R. (1994): "Zur Selbstorganisation von Solidarnetzwerken unter Ungleichen," in *Wirtschaftsethische Perspektiven I*, ed. by K. Homann, no. 228/I in Schriften des Vereins für Socialpolitik, Gesellschaft für Wirtschafts- und Sozialwissenschaften, Neue Folge, pp. 105–129. Duncker & Humblot, Berlin.

Kandori, M., G. G. Mailath, and R. Rob (1993): "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica*, 61, 29–56.

Kirchkamp, O., and K. H. Schlag (1995): “Endogenous Learning Rules in Social Networks,” Rheinische Friedrich Wilhelms Universität Bonn, Mimeo.

May, R. M., and M. A. Nowak (1992): “Evolutionary Games and Spatial Chaos,” *Nature*, 359, 826–829.

——— (1993): “The Spatial Dilemmas of Evolution,” *International Journal of Bifurcation and Chaos*, 3, 35–78.

Maynard Smith, J., and G. R. Price (1973): “The Logic of Animal Conflict,” *Nature*, 273, 15–18.

Sakoda, J. M. (1971): “The Checkerboard Model of Social Interaction,” *Journal of Mathematical Sociology*, 1, 119–132.

Schelling, T. (1971): “Dynamic Models of Segregation,” *Journal of Mathematical Sociology*, 1, 143–186.

Taylor, P. D., and L. B. Jonker (1978): “Evolutionary stable strategies and game dynamics,” *Math. Biosc.*, 40, 145–156.

Wolfram, S. (1984): “Universality and Complexity in Cellular Automata,” *Physica 10D*, pp. 1–35.

Young, H. P. (1993): “The Evolution of Conventions,” *Econometrica*, 61, 57–84.

Zeeman, E. (1981): “Dynamics of the evolution of animal conflicts,” *Journal of Theoretical Biology*, 89, 249–270.