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]

October 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.html

Support by the Deutsche Forschungsgemeinschaft, SFB 303, is gratefully acknowledged. I thank George Mailath, Georg Nöldeke, 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 or coordination games are played repeatedly within neighborhoods where players do not optimize but instead copy successful strategies.

Discriminative 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 cooperation persists even in a stochastic environment that players do not always coordinate on risk dominant equilibria in 2×2 coordination games and that success among surviving strategies may differ.

We also present two analytical models that explain some of these phenomena.

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

Contents

1 Introduction

In this paper we will discuss a model which uses an evolutionary approach within a spatial model. We will concentrate on simple strategic situations: prisoners' dilemmas and coordination games. For these games we want to study conditions that lead to cooperation and coordination. Can players sustain cooperation in prisoners' dilemmas? Does the introduction of repeated game strategies affect the amount of cooperation in prisoners' dilemmas? Will all strategies that survive in equilibrium achieve equal payoffs? How do players coordinate in coordination games?

To give an example for a spatial prisoners' dilemma, consider a world with several firms located in space and competing for customers which are located among these firms. Firms could set high prices, hoping for a monopoly profit if neighboring firms set high prices as well. Firms could also set low prices hoping to undercut prices of other firms and, thus, stimulating demand for their product.[1]

In a world with global interaction the above story would be a trivial example of Bertrand competition. This is a standard example where we expect firms to set low prices in the long run. In a (local) world where firms do not all share the same market but where instead any two firms have a different set of customers in common, firms' behavior might change. Since Hotelling (1929) a lot of spatial models have been specified and solved. Nevertheless spatial models which assume that players' rationality is limited have been neglected for a long time. Sakoda (1971) and Schelling (1971) were possibly among the first to present results of simulations of spatial evolution, later followed by Axelrod (1984, p. 158ff), May and Nowak (1992, 1993), Bonnhoeffer, May and Nowak (1994) and Lindgren and Nordahl (1994). All of them have studied models where a population is represented as a cellular automaton. Players are represented as single cells which are in interaction with neighboring cells and which are learning from neighboring cells. Ellison (1993) and Eshel, Samuelson and Shaked (1996) have analytically derived properties of particular cellular automata.

Some of the above (Sakoda (1971), Schelling (1971) and Ellison (1993)) assume that players optimize myopically. Others (Axelrod (1984, p. 158ff), May and Nowak (1992, 1993), Bonnhoeffer, May and Nowak (1994), Lindgren and Nordahl (1994) and Eshel, Samuelson and Shaked (1996)) assume that players learn through imitation of successful strategies. We will follow the latter approach.

Common assumptions of the literature that studies local evolution are *synchronous interaction and learning* (Axelrod (1984), May and Nowak (1992, 1993), Lindgren and Nordahl (1994), Eshel, Samuelson and Shaked (1996)) and *players which are restricted to use stage game strategies only* (May and Nowak (1992, 1993), Eshel, Samuelson and Shaked (1996)). In the current paper we will relax these assumptions. A common property of standard evolutionary models, the fact that all surviving strategies achieve equal payoffs in the long run, may vanish in our framework. We try to investigate this phenomenon more deeply. We present simulation results

[1] An example for a spatial coordination game can be found similarly: Neighboring firms could be interested in finding common standards, meeting in the same market etc.

and try to give a motivation with simpler models which can be solved analytically.

In contrast to all of the above, except Ellison (1993) and Bonnhoeffer, May and Nowak (1994) we will model a population where players' behavior is not synchronized through an external clock.[2]

In contrast to all of the above, except Axelrod (1984, p. 158ff) and Lindgren and Nordahl (1994) we will model players who are able to distinguish between their neighbors.

We find regions of prisoners' dilemmas where evolution leads to cooperation. We consider a simpler evolutionary dynamics which captures features of our simulation model, which approximates its properties, but which can be studied using simple analysis. This model also gives an accurate description of the conditions that lead either to the selectioction of risk dominant or Pareto dominant equilibria in coordination games.

In our simulations we find that local evolution can lead to the coexistence of strategies which achieve different payoffs. We are tempted to describe this phenomenon as 'exploitation'. We give an example of a simpler evolutionary dynamics which allows us to study this effect analytically.

Various meanings of 'evolution': Recently evolutionary models have become popular among game theorists and economists. Economists use the expression 'evolution' to describe any kind of dynamic process. 'Evolutionary economics' in a Schumperterian (Schumpeter 1934) tradition analyzes the dynamics of a process where individuals find continuously new knowledge, often in the form of new technologies. The process *how* a new technology is found is often defined as being exogenous to the model and not specified in detail. Evolutionary economics concentrates more on the effects newly developed technologies have on the evolution of growth of other technologies and of the economy as a whole.[3]

'Evolutionary game theory' does not explain the complex process of how new technologies are invented either. Still models in evolutionary game theory are sometimes simple enough such that it is possible to specify precise assumptions on the behavior of individuals. Assumptions on individuals' rationality in this context are often different. Popular restrictions are e.g. the assumption of myopia — individuals optimize under the (wrong) assumption that they are the only ones who will change their behavior — limited access to information which may further be unreliable and furthermore inertia that gives individuals only very rarely the opportunity to update their strategy.

Other models in evolutionary game theory (which are often inspired by biology) exclude optimizing behavior, in the sense that individuals try to evaluate fictitious

[2]With 'synchronization' we mean that it is *predetermined* whether in any given interval certain interactions or learning events take place. The most common specification is then that during each period *all* possible interactions and learning events take place. We call an event (like interaction or learning) not synchronized if e.g. for each possible interaction a random draw decides whether the interaction takes place.

[3]See Richard Nelson (1995) for an exhaustive overview over recent developments in evolutionary economics.

situations, altogether. Successful strategies will grow as long as they are successful, unsuccessful strategies will vanish from the population.[4] The underlying story might (in biological models) model a birth process where users of successful strategies produce more offspring and a death process where users of unsuccessful strategies are more likely to die soon. In social contexts common assumptions are that successful strategies are more likely to be copied by other members of a society while users of unsuccessful strategies are likely to abandon their strategy.

The evolutionary model that we will present in the current paper will be of the latter kind: Players follow a specific rule that imitates a successful strategy among several strategies whose 'success' can at least partially be observed by the player.

The meaning of 'space': Spatial models are also common among economists. If several individuals form together a population, each member may interact with the other members of the population in a different way. Instead of allowing any possible structure of different relationships among individuals it is convenient and often realistic to assume that interactions of the members of a society can be explained in certain dimensions.

Firms which are located in space could be in competition for customers which are located between these firms. Thus, each firm influences only nearby competitors and is not in interaction with other firms which are far away. Furthermore space need not be geographic space, also qualities of a product can explain differentiated interaction. Producers of small cars might interact with other producers of small and medium sized cars but will possibly not be in interaction with completely different producers.

In the current paper we study a model where a population is represented as a cellular automaton. In the following we summarize some of the literature on spatial evolution which is based on cellular automata.

All players are connected through a chain of neighborhoods but these connections can be very diverse. This contrasts with the model that we present in Kirchkamp (1995) where two players are either members of the same pair or only connected through the (homogeneous) learning process. The disadvantage of the cellular automaton model is that due to the diversity of connections among players this structure is substantially harder to analyze. Therefore cellular automata are often analyzed using simulations.

Related literature on spatial evolution: 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), May and Nowak (1992), Bonnhoeffer, May and Nowak

[4]See e.g. Maynard Smith and Price (1973) for static concepts and Taylor and Jonker (1978) and Zeeman (1981) for a dynamic model)

(1994), Ellison (1993), Lindgren and Nordahl (1994) and Eshel, Samuelson and Shaked (1996)) take players' positions as fixed but allow players to change their states. Furthermore there are models where both players are allowed to move and to change their state (see Hegselmann (1994)).

Another distinction is that some authors (like Sakoda, Schelling and Ellison) assume that players optimize myopically while others (Axelrod, May and Nowak, Bonnhoeffer, May and Nowak, Lindgren and Nordahl and Eshel, Samuelson and Shaked) assume that players learn through imitation.

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

Cellular automata with migrating players: 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.[5] He 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 how segregations appears via unorganized individual behavior without any collective enforcement or economic need.

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 the payoffs of the prisoners' dilemma to be played depend 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. During all their choices players optimize myopically. Hegselmann finds convergence both to clusters of players that belong to a similar risk class and cooperation among members of the same class.

Endogenous networks: 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.

[5]Sakoda himself names these attitudes Crossroads, Mutual Suspicion, Segregation, Social Climber, Social Worker, Boy-Girl, Couples, Husband-Wives.

Conway's life: 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 game of life, 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 her state according to the state of her eight neighbors. A living player remains alive only if she has two or three living neighbors, otherwise she dies of loneliness or overpopulation. A dead player becomes a living player only if she 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.[6]

Models with myopic optimization: Ellison (1993) presents an analytical model of a population whose members are distributed on a line. Members of this population have rarely the opportunity to change their state. In this case they optimize myopically. Further there are few mutations of players' strategies. Except for the local structure Ellison follows a model of Kandori, Mailath and Rob (1993) and Young (1993). Both Kandori, Mailath and Rob and Young find that in a global model a population which plays a coordination game finds in the very long run the risk dominant equilibrium. The argument in the global model is both with Kandori, Mailath and Rob and with Young that it takes fewer mutations to move from the risk dominated equilibrium to the risk dominant than vice versa. Still it might take a long time to reach the risk dominant equilibrium since the part of the population which has to mutate simultaneously to initiate such a move can be almost half the population. Ellison points out that in a local model it is sufficient to begin with a very small cluster of mutants to immediately start the move towards the risk dominant equilibrium.

Models with learning through imitation: While in the local models of Schelling, Sakoda and Ellison players are at least capable to optimize myopically, Axelrod (1984, p. 158ff), May and Nowak (1992), Bonnhoeffer, May and Nowak (1994), Lindgren and Nordahl (1994) and Eshel, Samuelson and Shaked (1996) consider a model where players learn through imitation.

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 (which consists of 200 repetitions of the underlying stage game) 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 Axelrod's global

[6] A finitely large cellular automaton can of course never generate chaos since it has only a limited number of states — chaos occurs only with automata of infinite size.

model. These are strategies that get along well with 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.

A similar model has been studied by Lindgren and Nordahl (1994), who in contrast to Axelrod do not analyze evolution of a set of prespecified strategies, but instead allow strategies to mutate and thus consider a larger number of potential strategies. Similar to Axelrod, Lindgren and Nordahl assume that reproductive success of a strategy is determined by the average payoff of an infinite repetition of a given game against a player's neighbors.

While Axelrod and Lindgren and Nordahl study a population with a large number of possible strategies, May and Nowak (1992) analyze a population playing a prisoners' 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. Bonnhoeffer, May and Nowak (1994) consider several modifications of this model: They compare various learning rules, discrete versus continuous time and various geometries of interaction.

Eshel, Samuelson and Shaked (1996) derive analytically the behavior of a particular cellular automaton. The price they have to pay for the beauty of the analytical result is the restriction to a small set of parameters. So they have to assume a particular topology and neighborhood structure. They find that there is a range of prisoners' dilemmas where increasing the gains from cooperation reduces the amount of players who actually cooperate. It is not clear how their result depends on the particular set of parameters they analyze.

The model that we are going to present in this paper: In the following we will study a model where players are not allowed to travel but where they change their states while learning successful strategies. Thus, our model is more related to Axelrod (1984, p. 158ff), May and Nowak (1992, 1993), Bonnhoeffer, May and Nowak (1994), Lindgren and Nordahl (1994) and Eshel, Samuelson and Shaked (1996)

In contrast to May and Nowak and Bonnhoeffer, May and Nowak we allow for different actions against different opponents if histories against these opponents are different. We introduce discriminative 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 and in Lindgren and Nordahl, there are some important differences. Axelrod and Lindgren and Nordahl assume that players simultaneously participate in a tournament and then synchronously update their strategies. If a player preserves her repeated game strategies in the next round this strategy is 'reset' during the learning stage. Thus players' behavior previous to a learning step influences players' behavior after the learning step only through the learning process and *not* due to properties of the repeated game strategy. We see two possibilities to interpret Axelrod's model:

It is possible to understand the evolutionary process described by Axelrod and

Lindgren and Nordahl 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. A justification for this *synchronization* is hard to find. We will explain below how properties of a synchronized model differ from an asynchronous model.

Another possible interpretation of Axelrod's and Lindgren and Nordahl's model would be to represent the choice of a repeated game strategy for a single tournament (which consists of playing the stage game repeatedly for a given number of periods) as the choice of a stage game strategy for a coordination game. If we follow this interpretation a major difference between our's and Axelrod's respective Lindgren and Nordahl's model is that we assume neighbors of learning players to preserve their memory.

A substantial difference to the models of Axelrod, May and Nowak, Lindgren and Nordahl and Eshel, Samuelson and Shaked is that not everybody learns at the same time. While the *synchronization of learning and interaction* simplifies the analysis a lot, it is hard to justify. We explain below in detail how 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. Bonnhoeffer, May and Nowak (1994) argue on the other hand that introducing stochastic behavior matters only little. While we completely agree with the spirit of both argument we, nevertheless, try to point out that not the mere fact of introducing stochastic behavior determines the amount of cooperation. It is the way *how* stochastic behavior is introduced that determines persistence or breakdown of cooperation.

While the cellular automaton model of a population and the introduction of discriminative behavior incorporates more real life flavor, it is on the other hand more 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 1000 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 sections 2.1 to 2.9. Section 3.1 discusses which properties of our simulations are stable in the long run, section 3.2 describes the space of games that we consider, section 3.3 studies a simplified model and gives a first impression of what we can expect in the cellular automaton model. Section 3.4 discusses the representation of our results. Section 3.5 connects our results to the findings of May and Nowak (1992, 1993) investigating several models with simple repeated game strategies that are more similar to their model. Section 3.6 extends the model to more complex (two-state) repeated game strategies, section 3.8 extends the analysis to coordination games and sections 3.9 and 3.10 discuss robustness of the results. Section 4 finally draws some conclusions.

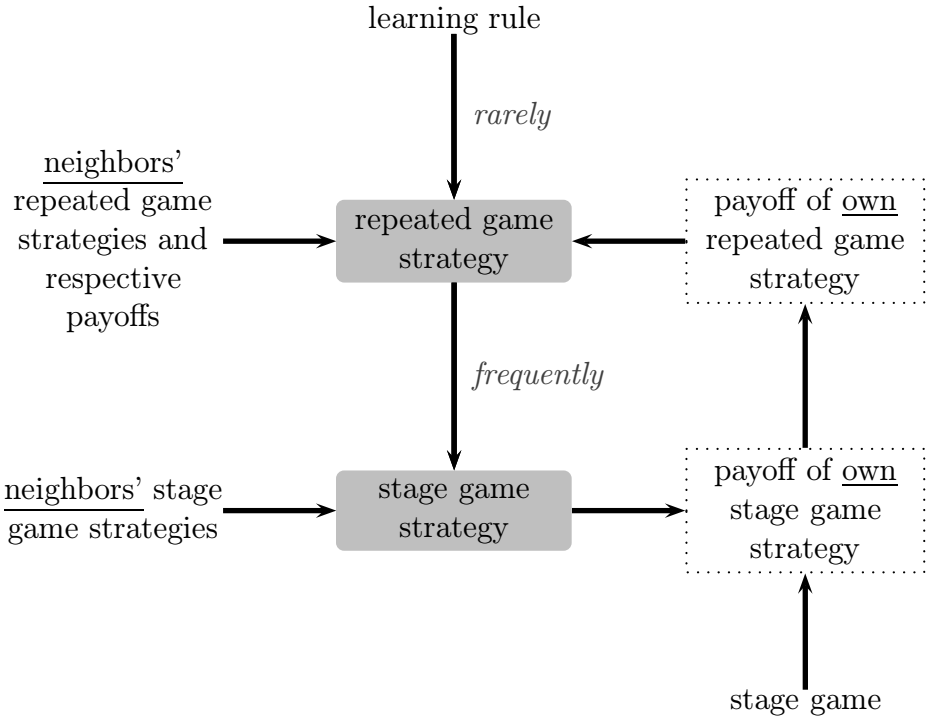


Figure 1: Properties of a player.

2 The Model

2.1 Overview

The environment We will consider a population of players each occupying one cell of a torus of size 80×80. Details of the spatial structure are described in section 2.2, the neighborhoods are described in section 2.3 and timing will be discussed in section 2.4. Players will play games with their neighbors on this network and learn repeated game strategies from their neighbors. The games and strategies are described in sections 2.5 and 2.6. The learning behavior of the players will be described in sections 2.7 and 2.8.

Simulations will start from random initial configurations which are described in detail in section 2.9 on page 20. Simulations will be repeated over and over again[7] for different games.

Players' characteristics Players are described by two characteristics: Stage game strategies and a repeated game strategy.

We visualize the two parameters with the help of figure 1 on the page before.

- A player's *repeated game strategy* is influenced by following factors: A fixed learning rule, information on the player's own payoff and repeated game

[7] Simulations will be repeated at least 800 times.

strategy and information on her neighbors' payoff and their repeated game strategy. One might interpret the learning rule as a function that takes a player's and her neighbors' payoff and repeated game strategy as arguments, thus, determining a new repeated game strategy for the player. We will study two learning rules, one that copies always the repeated game strategy of the most successful *player* in the neighborhood and the other which copies the most successful (on average) *repeated game strategy* in the neighborhood. More details on selection of repeated game strategies can be found in section 2.8 on page 17.

- A player's *stage game strategies* are determined by her repeated game strategy and her neighbors' stage game behavior. One might interpret a player's repeated game strategy as a function that takes her neighbors' stage game behavior as arguments to determine a player's stage game strategies. How stage game strategies are determined by repeated game strategies is discussed in more detail in section 2.6 on page 12.

  Stage game strategies determine interactions among players. Given an exogenously specified stage game stage game strategies of two players determine stage game payoffs. These payoffs contribute to the payoffs of the repeated game strategies. On the basis of the latter payoffs new repeated game strategies are selected. Payoffs of the repeated game strategy may be the payoffs players received in the current period or may be averages over several periods (see section 2.7 for details).

The above two properties change randomly and at different speeds. Players interact with a high probability and change their repeated game strategy with a low probability. Details of the timing are described in section 2.4 on the following page.

2.2 Spatial Structure

Players are assumed to live on a torus[8]. We represent the torus as a rectangular network, e.g. a huge checkerboard. The edges of this network are pasted together. Each cell on the network is occupied by one player. We use a torus instead of a simple checkerboard to avoid boundary effects. Thus, the neighborhood of all players has the same structure.

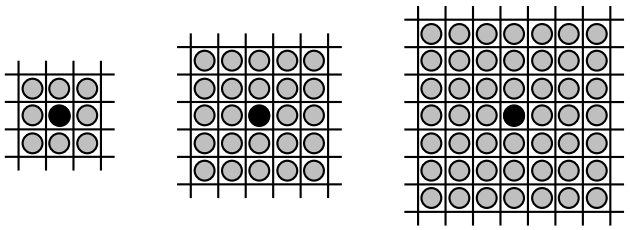
A location may represent a geographical position and interaction only with geographical neighbors. However, location can also be interpreted as producing a certain differentiated product and interaction only with producers that manufacture a similar product. Further, in the context of a model with overlapping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

[8] In section 3.10 we will investigate some other topologies where players are located on a circle, on a cube, or on a hypercube in four dimensions.

Most simulations were carried out on a square of size 80×80. The results presented in sections 3.9 and 3.10 show that for sufficiently large networks neither the exact size of the network nor the dimension matter significantly.

2.3 Neighborhoods

In contrast to models of global evolution and interaction we will consider players who interact only with their neighbors and who learn only from their neighbors. Below we sketch some neighborhoods characterized by various interaction radii r_{I}. The possible interaction partners of a player are described as gray circles while the player itself is represented as a black circle.



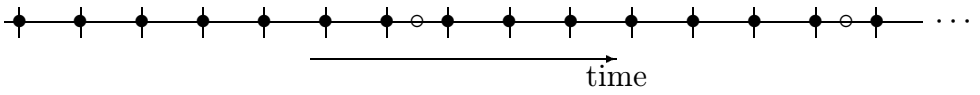
$r_{\mathrm{I}} = 1$ $r_{\mathrm{I}} = 2$ $r_{\mathrm{I}} = 3$

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 she has information on *all* members of N_{L}^i.

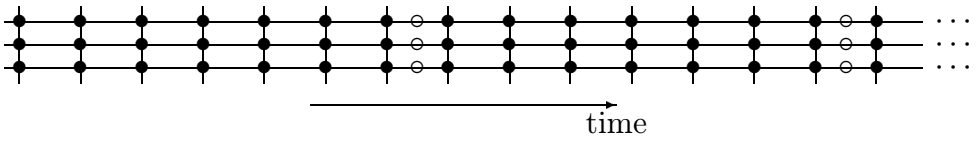
2.4 The Role of Time

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 repeated game strategy about once a week. We describe interaction of a player with her top and bottom neighbor by ⊕ and learning by ∘. Two weeks in the life of a member of this population could be represented as follows.



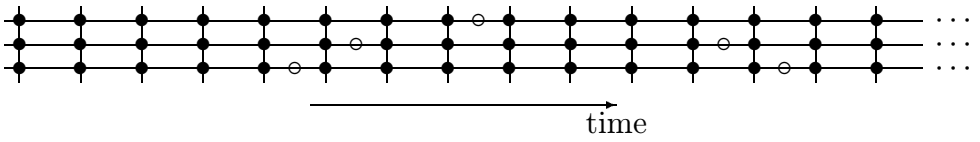
(1)

For simplicity we could model this behavior assuming that all possible interactions take place daily while learning and change of repeated game 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:



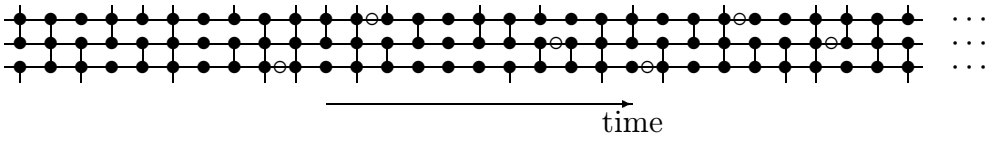
(2)

We will see in sections 3.5 and 3.6 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 repeated game 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 her neighbor while on other days she might not interact at all. Given that she 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[9]. 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 her 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 her 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. Even smaller probabilities for interactions lead to similar properties, but with smaller probabilities p_{I} our simulations need more time to approach their long run behavior.

Since we also want the timing of *learning* to be stochastic, we will assume that repeated game strategies have something like a stochastic 'lifetime' t_{L} that in our simulations is typically distributed equally between 20 and 28 periods. Once the

[9]May and Nowak (1992, 1992) and Eshel, Samuelson, and Shaked (1996) assume that learning takes place after all possible interactions took place exactly once, Axelrod (1984, p. 158ff) assumes that learning takes place after players interacted synchronously for a large number of periods.

'lifetime' expires players 'learn'. They possibly change to a new repeated game 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 more complex (two-state) repeated game strategies an opportunity to unfold. Thus, learning is a rare event as compared with interaction. An even larger lifetime leads to similar properties, but with larger lifetime our simulations need more time to approach their long run behavior.

2.5 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. Stage game strategies will be named C and D.

Notice that all the dynamics of population behavior that will be discussed in this and the following chapter 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	
		a	b
Player I	a	g g	1 h
	b	h 1	0 0

(5)

where $0 < g < 1$, $h < 0$ and $g > \frac{1}{2} + \frac{1}{2}h$.[10]

2.6 Repeated-Game Strategies

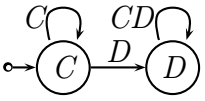
May and Nowak (1992, 1993) consider a model of spatial evolution of repeated game strategies which are not discriminative. A player could either always play C against all her neighbors or always play D. A player having e.g. one neighbor playing always C and another neighbor playing D might, however,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

One way to model discriminative behavior is to assume that players use repeated game strategies. Models which allow for repeated game strategies in the context of spation evolution are Axelrod (1984, p. 158ff) and Lindgren and Nordahl (1994). Both assume that players either simultaneously play tournaments (which consist of a given number of repetitions of the stage game) or synchronously update their strategies. Even if a player decides to keep her old strategy (which happens often once the evolutionary process has approached a more 'long run' state) Axelrod and Lindgren and Nordahl force her to forget what has happened in the previous period.

[10] We require $g < 1$ and $h < 0$ to make C a dominated stage game strategy. Then $0 < g$ and $g > \frac{1}{2} + \frac{1}{2}h$ ensure that CC Pareto dominates both DD and cycling between CD and DC.

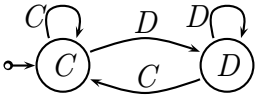
In the following we also model discriminative behavior as repeated game strategies. In contrast to Axelrod and to Lindgren and Nordahl we introduce repeated game strategies such that players may remember their opponents' behavior even after learning takes place. We will see below that this assumption together with the fact that players learn asynchronously fosters an interesting property: Heterogenous payoffs.

Notation for Moore automata: 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':



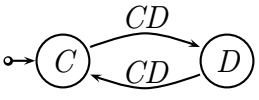
Grim is a repeated game strategy that has two 'states'. The stage game strategy that is actually used by grim in either of the two states is shown within the two circles. In the left state grim plays C, in the right state it plays D. The little arrow to the left of grim that points to the initial state 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 repeated game strategy reacts upon 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 see that once grim is in its second state then there is only one arrow to be 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 commonly considered repeated game 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 3.6 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. However, in the following we will focus on populations where only repeated game strategies with less than three states are present. Table 1 on the next page gives a list of all these automata.

We make this restriction mainly to limit the number of possible automata (there are only 26 automata with less than three states). We think that this is not a severe restriction since many interesting repeated game 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 think that evolution of repeated game 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' repeated game strategies to be less sophisticated is only consistent. We do not claim that *all* automata with less than three states are sensible repeated game strategies, but we expect that in a sensible model odd repeated game strategies should be eliminated through evolution.

Asynchronous learning of automata: We will now give some examples how strategic interaction among two players may evolve given that their repeated game strategies can be characterized by automata. We try to be very detailed in our examples to make clear that strategic interaction depends on the learning behavior of our players. In particular it may be important whether players learn synchronously or asynchronously.

Let us first 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	$\cdots$

Above we have only considered interaction of no more than two players. In the networks that we will discuss below a single player will have several neighbors. For

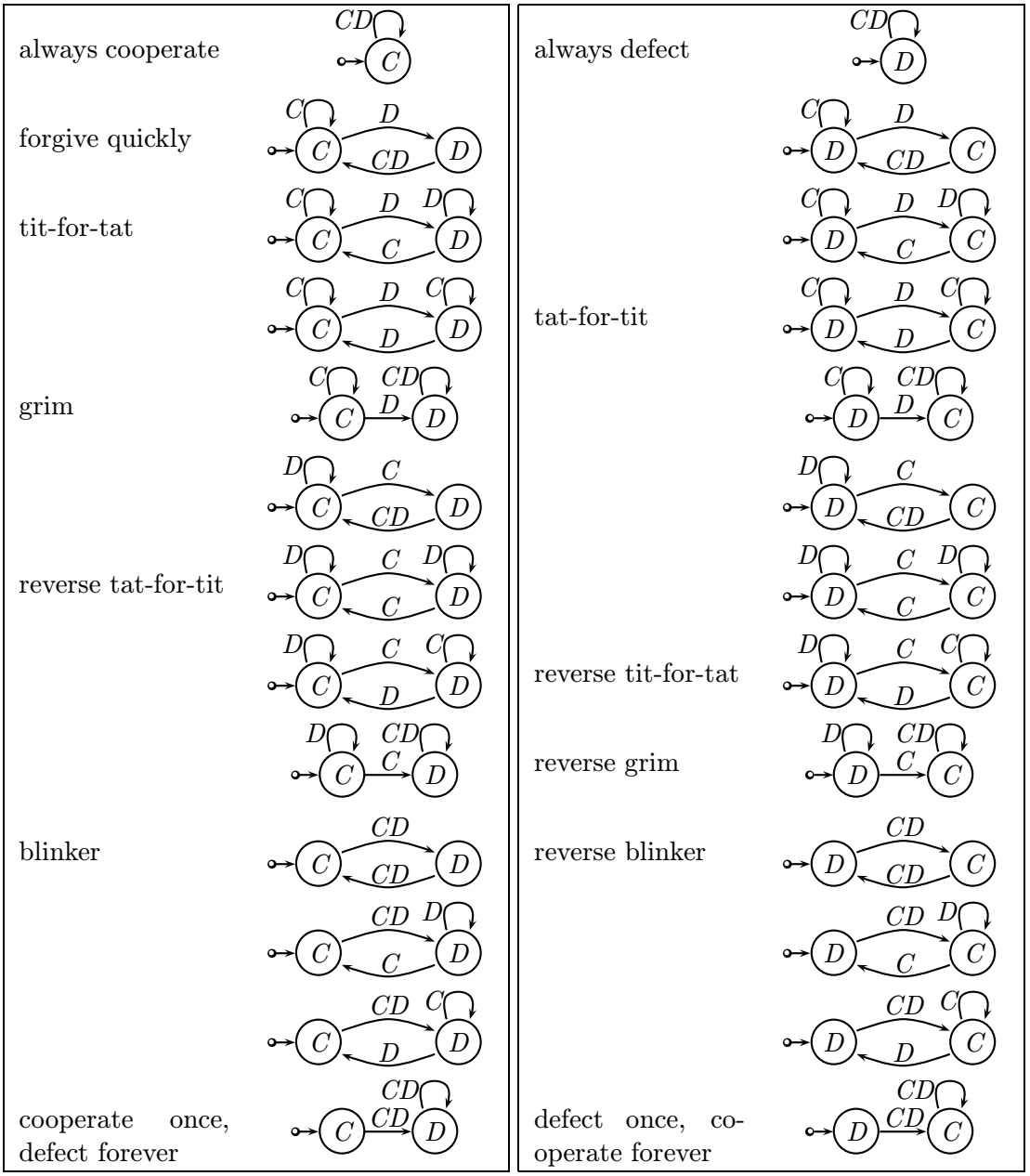


Table 1: All 26 automata with less than three states

each of her neighbors she has a copy of her automaton. While all automata of one player are identical for all of her 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 repeated game strategy. Let us assume this happens in period 15 where she 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 newborn grim starts cooperatively in period 15, but finds out that its opponent from the second floor already defects and therefore switches to her second state as well. Thus, we see two different pairs of grim here. The top pair cooperates all the time while the bottom pair defects from period 16 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.

2.7 Relevant history

When players learn, one source of information they will use will be average payoffs (per interaction) of their neighbors. In the following we will describe how these average payoffs are calculated. We will denote the set of periods that player i considers as relevant or that she 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 repeated game strategy without

interruption are relevant for her learning decision. Thus, if a player used repeated game 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 repeated game 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 adequate in the context of the repeated game strategies that we analyze. We have seen in section 2.6 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 a seriously wrong perception of a repeated game strategy's performance. Averaging over several periods avoids this problem.

We sum up the total number of player i's interactions with her current repeated game 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 sum up player i's payoff during $E^{i,t}$ as

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

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

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

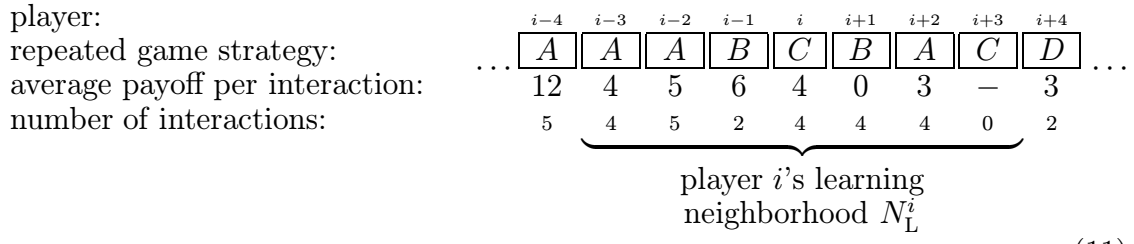
2.8 Update of Repeated Game Strategies

In section 2.4 we have already discussed various assumptions concerning *when* players could get an opportunity to update their repeated game strategies. In the current section we will discuss what kind of information players have when they get a learning opportunity and how they will use it.

We will assume players whose capabilities are restricted in several ways. They are not fully rational, they are not able or not willing to analyze games, and they do not try to predict their opponents' behavior. Nevertheless, players' behavior will have some structure since it is guided by imitation of successful repeated game strategies. Often players simply copy good examples without knowing why the example was so successful and without spending much effort in checking whether this repeated game strategy might be as promising for the copying player herself.

We assume that such an imitating player has incomplete information about the total population. All she can observe are repeated game strategies and their respective average payoffs per interaction in her neighborhood N_{L}^i. We may imagine that

when in period t a player learns she asks her neighbors $j \in N_{\mathrm{L}}^i$ for their strategy $x^{j,t}$ and their respective average payoff per interaction $u^{j,t}$. As an example let us consider the following population that lives on a line:[11]



(11)

Assume that player i learns and that she can see three players to the left and three players to the right. Thus, she does not realize that farther to the left there is an A with a high average payoff of 12. Nor can she see that there even exists a repeated game strategy D. She 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 (she herself) 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:

2.8.1 Copy Best Player

A learning player could simply look around in the neighborhood which she observes and determine the *player* with the highest average payoff per interaction. In our example she 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 repeated game 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. In this case players need a tie breaking rule. We will assume that if a players' current repeated game strategy is already among the repeated game strategies of the best players, then she keeps her current repeated game strategy. Otherwise she randomizes among the repeated game strategies of the most successful players. We assume here that the probability to adopt a certain players' repeated game strategy will be proportional to the number of interactions that led to the payoff of the respective player.[12] This can be formalized as follows: The set of most successful players that player i observes at time t will be called $M^{i,t}$.

$$M^{i,t} := \arg\max_{j \in N_{\mathrm{L}}^i} \left(\frac{u_{\mathrm{e}}^{j,t}}{n_{\mathrm{e}}^{i,t}} \right) \tag{12}$$

[11] Notice that most of our simulations below assume that players live on a torus and *not* on a line. We give an example for a player's learning behavior on a line just because it is easier to visualize.

[12] We have done simulations with other tie-breaking rules and got the impression that the particular choice of the tie-breaking rules has no influence.

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

$$P(x^{i,t+1}=s) := \begin{cases} 1 & \text{if } x^{i,t} \in \{x^{j,t} | j \in M^{i,t}\} \text{ and } s = x^{i,t} \\ 0 & \text{if } x^{i,t} \in \{x^{j,t} | j \in M^{i,t}\} \text{ and } s \neq x^{i,t} \\ \dfrac{\sum\limits_{j \in M^{i,t} \wedge x^{j,t}=s} n_{\mathrm{e}}^{j,t}}{\sum\limits_{j \in M^{i,t}} n_{\mathrm{e}}^{j,t}} & \text{otherwise} \end{cases} . \quad (13)$$

2.8.2 Copy Best Strategy

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

$$f_s^{i,t} := \begin{cases} \dfrac{\sum\limits_{j \in U_s^{i,t}} u_{\mathrm{e}}^{j,t}}{\sum\limits_{j \in U_s^{i,t}} n_{\mathrm{e}}^{j,t}} & \text{if } \sum\limits_{j \in U_s^{i,t}} n_{\mathrm{e}}^{j,t} > 0 \\ -\infty & \text{otherwise} \end{cases} \quad (14)$$

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

In example 11 on the preceding page the learning player would find out that strategy A has an average payoff per interaction of 5, strategy B has an average payoff per interaction of 2, and strategy C has an average payoff of 4.

A learning player that uses the rule 'copy best strategy' switches to the *repeated game strategy* with the highest average payoff, thus, in our example she will become an A. Again there could be more than one repeated game strategy with maximal payoff. The tie breaking rule will be similar to the one we assumed in section 2.8.1. If the current repeated game strategy of the player is among the most successful repeated game strategies then the learning player keeps her current repeated game strategy. Otherwise she adopts one of the best repeated game strategies randomly with probabilities proportional to the number of interactions the users of the respective repeated game strategies had. This can be formalized as follows: Define the set of most successful repeated game strategies as

$$N^{i,t} := \arg\max_s \left(f_s^{i,t}\right) . \quad (15)$$

The probability that player i uses repeated game strategy s in the next period is then

$$P(x^{i,t+1}=s) := \begin{cases} 1 & \text{if } x^{i,t} \in N^{i,t} \text{ and } s = x^{i,t} \\ 0 & \text{if } x^{i,t} \in N^{i,t} \text{ and } s \neq x^{i,t} \\ \dfrac{\sum\limits_{j \in U_s^{i,t}} n_{\mathrm{e}}^{j,t}}{\sum\limits_{\sigma \in N^{i,t}} \sum\limits_{j \in U_\sigma^{i,t}} n_{\mathrm{e}}^{j,t}} & \text{otherwise} \end{cases} . \quad (16)$$

2.8.3 Symmetry of Learning Rules

Notice that both learning rules described above are *symmetric* in the sense that a player puts the same weight on her own experience (payoff) and on the experience of the observed players.[13] Kirchkamp and Schlag (1995) find that once evolution selects 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.

2.8.4 Learning Repeated Game Strategies with Multiple States

If players learn repeated game strategies with a *single state* (like, play always D) they obviously have to use this state from the next period on. If players on the other hand learn repeated game strategies with *multiple states* (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. When a player has the opportunity to learn, but does not change to a different automaton we assume that she continues to use the same automata in whatever states it actually is against her different neighbors.

2.9 Initial State of the Population

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 with Fixed Learning Rules

3.1 Convergence

Before we talk about results of simulations, which properties of our simulated populations behave stable in the long run. In the following paragraph we give an example for the fact that the state of single players may be constantly changing even in the long run. Still some statistics, like proportions of stage game strategies, approach a behavior which is stable in the long run.

[13] To be precise, due to our tie-breaking rule, learning rules behave *not* completely symmetric in the case when several repeated game strategies achieve maximal payoff and the player's current repeated game strategy is among these strategies.

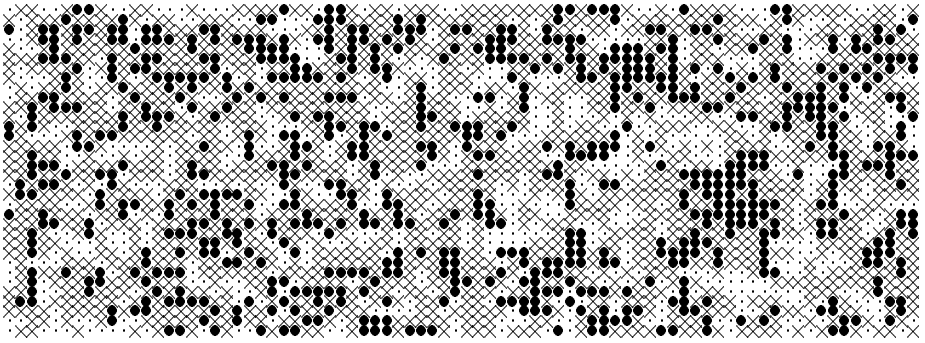
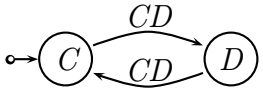
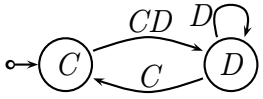
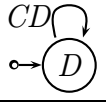


Figure 2: Part of an example net after 2000 periods.

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

		Player II	
		a	b
Player I	a	0.3 0.3	1 −0.7
	b	−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 $t_{\mathrm{L}} \in \{10 \ldots 14\}$. The network is initialized randomly. For simplicity of graphic representation we assume that only the following three automata are present in the network. They are denoted with the following symbols:

Automata	Symbols
	×
	·
	•

A typical state of a network after 2000 generations is displayed in figure 2 on the next page. Here players are still permanently changing their repeated game strategy. Thus, at least for this example, even after 2000 periods the population did not converge to a state where each individual's behavior remains constant. Nevertheless the some properties of the system seem to be stable in the long run. 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 3 on the following page shows the typical development of the frequency of the pair of stage game strategies CD and DC on the vertical axis. The horizontal

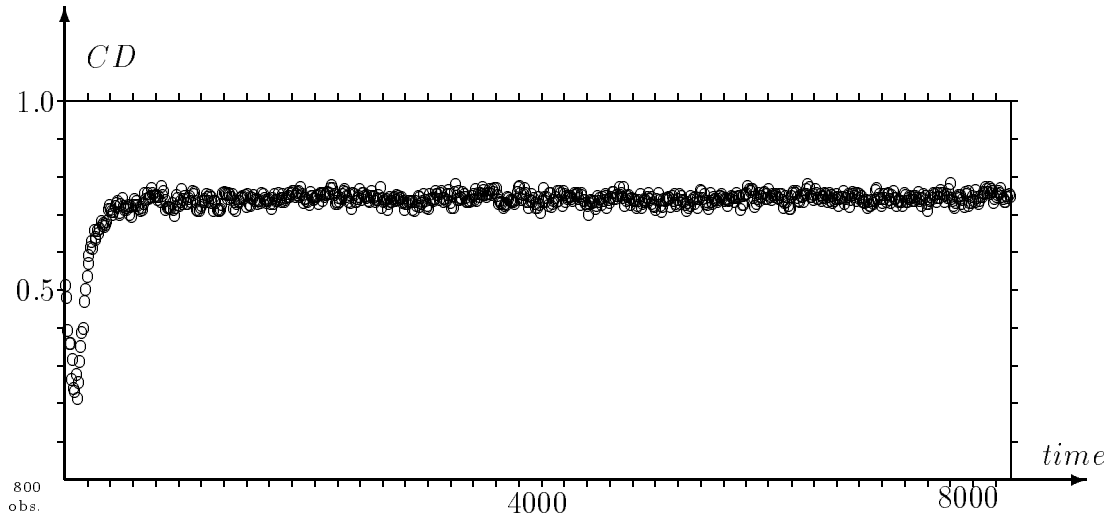


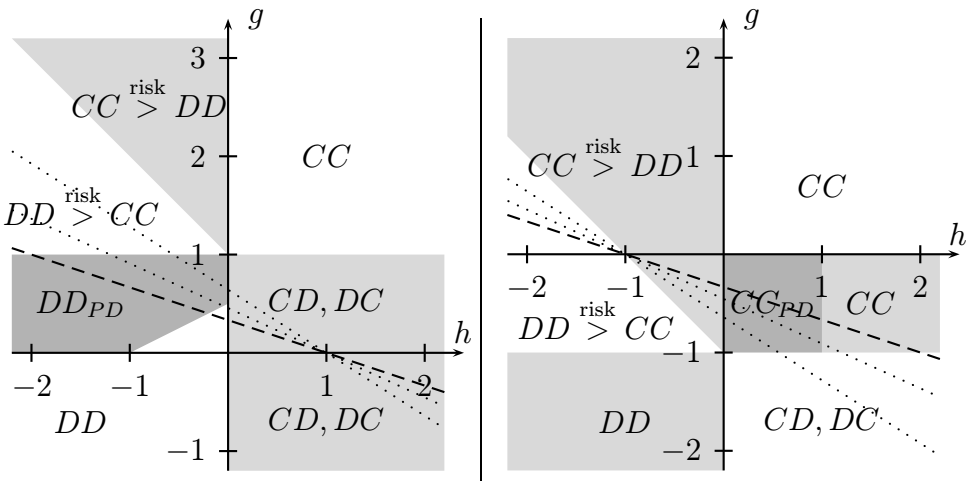
Figure 3: Convergence: Two-State-Strategies, $p_{\mathrm{I}} = \frac{1}{2}$, $t_{\mathrm{L}} \in \{20, \ldots, 28\}$, copy-best-strategy, short memory, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, network=$80 \times 80$, $g = 0.3$, $h = -0.7$,

axis shows time. While after a period of stabilization a more or less constant value for the proportion of CD is achieved the proportion seems not to converge. The observed values oscillate within a small radius. This behavior persists even during very long simulations. Most of our simulations continued for 1000 to 2000 periods. We have done 2000 simulations that continued for 20 000 periods and even some lasting for up to 1000 000 periods. These longer simulations lead exactly to the same property. For all the simulations we did, the proportions of repeated game strategies or the proportions of pairs of stage game strategies oscillate within a radius which is small as compared to changes of the same statistics that result from changes of the underlying game or other parameters.

We will observe later that proportions of repeated game strategies and proportions of combinations of stage game 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 2 on the page before 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, *which* players play a certain combination of stage game strategies, (which repeated game strategy a player uses) and *why* they do it, but we know at least *what* pairs of stage game strategies are played. In sections 3.6.1 and 3.7.1 we will also look at repeated game strategies.



Player II

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

Figure 4.a

Player II

Player I	C	D
C	g, g	-1, h
D	h, -1	0, 0

Figure 4.b

Figure 4: The space of considered stage-games.

3.2 Representation of the Space of Considered Stage Games

In this paper we look at symmetric 2×2 games. These games are parameterized by four payoffs. Given our learning rules several of these games behave equivalently. It is sufficient to study a very small subset of of all symmetric 2×2 games which can be parameterized by only two payoffs.

Evolutionary dynamics as given by the learning rules 'copy best player' and 'copy best strategy' (see sections 2.8.1 and 2.8.2) 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. All generic[14] symmetric 2×2 games can be derived from the games given in figure 4 on the following page. These latter games are described by only two parameters g and h, thus, they are easily represented in a plane.

Figure 4 on the next page shows for both types of stage games the regions of different equilibria. CC and DD denote regions of games which are not prisoners' dilemmas and which have only one Nash equilibrium. CC_{PD} and DD_{PD} denote prisoners' dilemmas with one equilibrium, CD, DC is a region where CD and DC

[14]Generic in payoff space.

are symmetric equilibria of the stage game, 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. The dashed and dotted diagonal lines will be described in section 3.3 below.

Some symmetric 2×2 games are contained both in the representation of figure 4.a and figure 4.b. Games from figure 4.b can be transformed to the shape of figure 4.a by subtracting g from all payoffs, then dividing by $h - g$ and finally exchanging the names of the stage game strategies. Similarly games from figure 4.a can be transformed into the shape of figure 4.b. As long as $h > g$ (given the representation of figure 4.b) this transformation does not change the best reply structure of the game. This means in particular that the same prisoners' dilemmas are contained both in figure 4.a and 4.b. Whenever we study prisoners' dilemmas we can therefore without loss of generality restrict ourselves to the DD_{PD}-section of figure 4.a.

3.3 A Model of Clusters

Before we turn to the simulation results of the complex model that we have described in section 2, let us make an estimation of how such a model might behave. To do this, we will study a simpler model in a continuous framework An argument which is similar to the following but which applies to a discrete framework has been made by Eshel, Samuelson and Shaked (1996). We will study a situation which occurs frequently in a model with local learning: players learning between two clusters. The reason to study the situation at the border between two clusters it that due to the local learning strategies can not appear in an isolated spot. When a player adopts a new strategy we will find the same strategy already in the neighborhood. Thus, we should expect strategies to appear in homogeneous clusters. Changes are most likely to appear at the border between two clusters. We will study the situation at this border. We will first find that if clusters are already large then the behavior of the system can be estimated easily.

Then we look at smaller clusters. There we will find that systems where small clusters are present have the tendency that some clusters will vanish completely creating, thus, new large clusters. The behavior of a large cluster system can then be studied easily with the model of large clusters.

Large clusters: Let us assume (we make this assumption *only* for the *current* section) that players are continuously distributed along a line and that both learning and interaction radius are r. Players copy the strategy with the highest average payoff in their learning neighborhood. Let us consider a player at position $\hat{x}$ between two of these clusters. Let us first assume that the world consists of only two clusters, each of infinite length. We will later see that it is enough to assume that clusters have a length n which is larger than $2r$.

All players at position $x < \hat{x}$ play C, all players at $x > \hat{x}$ play D. Since the interaction radius is r, a player that lives at position x has the probability $p_C(x)$ to meet another C.

$$p_C(x) = \begin{cases} 1 & \text{if } x \leq \hat{x} - r \\ \frac{r+\hat{x}-x}{2r} & \text{if } x \in (\hat{x} - r, \hat{x} + r) \\ 0 & \text{if } x \geq \hat{x} + r \end{cases} \tag{17}$$

The probability to meet a D is $1 - p_C(x)$. Since the learning radius is r as well, average payoffs u_C and u_D for the two strategies can be calculated for the player at $\hat{x}$ as

$$u_C = \frac{3g + h}{4} \qquad u_D = \pm\frac{1}{4} \tag{18}$$

for the games in figure $4.^{\text{a}}_{\text{b}}$ respectively. Given the game in figure 4^{a}_{b} our player at position $\hat{x}$ will imitate a C if $g > (-h \pm 1)/3$ and a D if $g < (-h \pm 1)/3$. The dashed lines in figure 4 represent the games where our player at $\hat{x}$ is indifferent which clusters' strategy she should follow. Above she will become a C, below she will become a D. In these cases we will, starting from sufficiently large clusters, observe that either only the C or only the D clusters will grow — depending on whether we are above or below the dashed line.

Small clusters: In the above paragraph we have assumed that the homogeneous clusters which contain only a single strategy are large. To be precise, their diameter had to be larger than $2r$. The same calculation we have done above can also be done for clusters which are smaller. In the following we will assume that clusters are not smaller than the learning and interaction radius. Let us assume the world consists of a sequence of alternating clusters whose members play C and D respectively. The C playing clusters have radius $n_C r$ where $n_C \in [1, 2]$ and the D playing clusters have radius $n_D r$ where $n_D \in [1, 2]$.[15] Then a player who lives precisely between two clusters sees for both clusters a proportion of

$$F(n) = \frac{1}{4}(5 - 4n + n^2) \tag{19}$$

(for $n = n_C$ and $n = n_D$ respectively) playing against the respective opposite strategy. The remaining $1 - F(n)$ play against their fellows which are in the same cluster. Then average payoffs are

$$u_C = (1 - F(n_C))g + F(n_C)h \qquad u_D = \pm F(n_D) \tag{20}$$

for the games in figure $4.^{\text{a}}_{\text{b}}$ respectively. The difference $u_C - u_D$ is a quadratic function of n_C and n_D

$$u_C - u_D = g + \frac{1}{4}(-g + h)(5 - 4n_C + n_C^2) \mp \frac{1}{4}(5 - 4n_D + n_D^2) \tag{21}$$

depending on the type of the game (figure $4.^{\text{a}}_{\text{b}}$). It is now interesting to know for which values of n_C and n_D the expression $u_C - u_D$ is positive or negative. If $u_C - u_D$ is positive then n_C will grow while n_D will shrink and vice versa. For a given prisoners' dilemma the region where $u_C - u_D$ is positive is an ellipsis around $(n_C, n_D) = (2, 2)$. For a given coordination game this region is bounded by a hyperbola with center

[15] The precise analysis of even smaller clusters becomes really tedious. We think that it is reasonable to assume that our approach estimates even the behavior of smaller clusters where $n < 1$.

$(n_C, n_D) = (2, 2)$ as long as $g > h$ otherwise it is again an ellipsis around $(n_C, n_D) = (2, 2)$.

For coordination games where $g > h$ and for a given combination (n_C, n_D) one of n_C or n_D will grow until the boundary is reached. Thus, clusters become larger and larger until for all clusters $n > 2r$. But then we can follow the simple analysis for large clusters given above and conclude that in a coordination game C wins if $g > (-h - 1)/3$.

For prisoners' dilemmas things are slightly more complicated. Assume that we start inside the ellipsis where $u_C > u_D$. Then n_C will grow until the boundary of the ellipsis is reached. At this stage the boundary between the two clusters may stop moving. We have now a small cluster of Ds next to a large cluster of Cs.

This situation is stable because the smaller the cluster of Ds becomes the less important the *negative* influence among the Ds will be but the more important the gain they have from being close to many Cs.

Therefore in a prisoners' dilemma small clusters of Ds may survive together with Cs. This means that in a prisoners' dilemma our above estimate (for large) clusters was possibly too optimistic. Instead it might happen that size of clusters remains small and that the region where cooperation is stable is smaller than with large clusters.

Given the game in figure 4 player $\hat{x}$ will become a C if

$$g > \frac{\mp(5 - 4n_D + n_D^2) + h(5 - 4n_C + n_C^2)}{1 - 4n_C + n_C^2} \tag{22}$$

for the games in figure 4.$^{\text{a}}_{\text{b}}$ respectively. The dotted lines in figure 4 represent the games where our player at $\hat{x}$ is indifferent which cluster's strategy she should follow, given that clusters have a diameter of $3r/2$ or $5r/4$, i.e. the case $n_C = n_D = 1.5$ and $n_C = n_D = 1.25$. In this case our player will become a C above the respective dotted line, below she will become a D. If clusters become smaller, then n_C and n_D shrink and the set of games where a player $\hat{x}$ is indifferent between the left and the right cluster moves (for prisoners' dilemmas) upwards. For prisoners' dilemma games this means in particular that the smaller clusters are, the smaller the set of prisoners' dilemmas where we should expect cooperation will be. As long as cluster size remains constant for all games, we should expect the borderline of cooperative behavior to have a linear shape as described by inequality 22. Our simulations show that this borderline has for a reasonable parametrization indeed a linear shape.

In the remainder of the paper we will consider the discrete model described in section 2. Players will be discrete, we will most of the time consider a two dimensional network and the size of the clusters will be determined endogenously. We will see that the continuous model of the current section gives a surprisingly good approximation of the discrete model that we will study in the following.

3.4 Representation of the Results

To explain the representation of the results take for example figure 6 on page 30. The figure shows the results of 800 different simulations. We choose 800 times randomly

different combinations of g and h. Taking the structure of game 5 on page 12 as given, specific values of g and h define a specific game. g and h were chosen such 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. As already mentioned in section 3.1, after 2000 periods we can expect that population statistics like proportions of stage game strategies or automata have approached their long run behavior.

Figure 6 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 stage game 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.

Following the discussion in section 3.3 we have also indicated in figure 6 the set of games where inequality 22 becomes binding for various values of n. The solid gray line corresponds to the case $n_C = n_D = 2$, the dashed gray lines represent $n_C = n_D = 1.5$ and $n_C = n_D = 1.25$.

3.5 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) repeated game strategies 'always C' and 'always D' (in the following denoted with 'C' and 'D') are allowed.

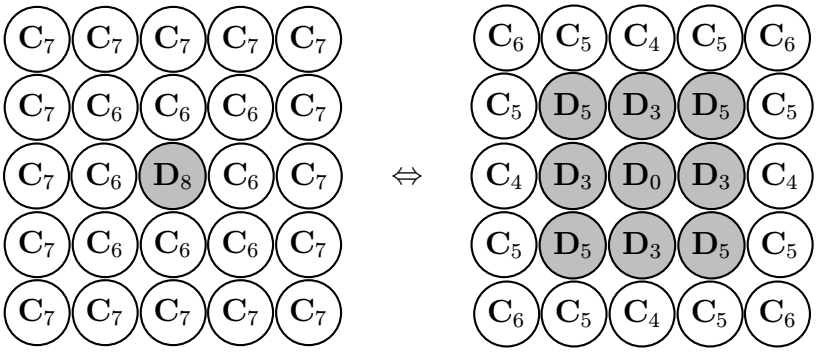
We know from global models that, whatever the rest of the population does, D is *always* more successful than C, thus, we should expect C to die out in the long run. The following paragraph sketches the idea why in a local model under certain circumstances C can survive.

Certainly a *single* C-playing automaton cannot survive if it is surrounded and exploited by Ds. However, we may imagine 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 high 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. A C that is situated close to the borderline between C and D is likely to change to a D. Its payoff from interaction with a C becomes low if gains g from cooperation are low. Furthermore its 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.

Figure 5 on the following page gives an example for the above argument. Assume that a large population plays the game given in figure 5. Most of the population cooperates but let us assume that somewhere a single player plays D. The left part of figure 5 displays the population around this player. When her cooperating

Player I \ Player II	C	D
C	7/8 7/8	1 −1/8
D	−1/8 1	0 0

Neighborhoods for learning and interaction contain the eight immediate neighbors. Players interact with certainty. Only today's payoffs matter. The learning rule is 'copy best strategy'. Strategies are: C and D. The population cycles between the following two states:



odd periods even periods

Figure 5: An example for the survival of C.

neighbors calculate average payoffs per interaction they will find that D is more successful than C and become a D in the next period. In the next period we have already nine Ds in the population. This state of the population is displayed in the right part of figure 5. But now the situation of the D is different. The initial D sees nothing but Ds in its neighborhood, thus, it has no opportunity to change its strategy. The newborn Ds on the other hand will find that C receives on average a higher payoff than D. Thus, all of them will become a C in the next period. Now we are back again in the left part of figure 5.

May and Nowak (1992, 1993) consider a similar spatial model where only Cs and Ds play a prisoners' dilemma. They assume the learning rule 'copy best player' (see section 2.8.1 on page 18), synchronous interaction ($p_{\mathrm{I}} = 1$), and synchronous evolution ($t_{\mathrm{L}} = 1$). One of their results is that for certain initial configurations and for certain payoffs cooperation may indeed persist in a given 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 6 on the following page shows a 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. The borderline between cooperation and defection follows the prediction given in section 3.3 for a cluster size of about $1.2r$.

We also see that *some* simulations lead to mutual cooperation for smaller values

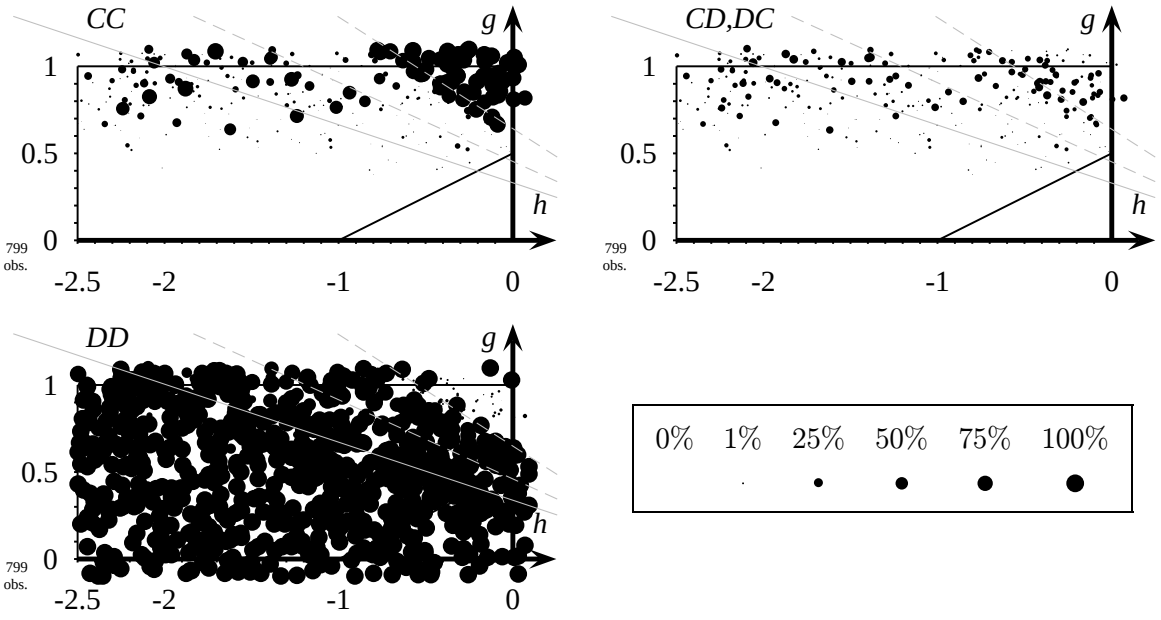


Figure 6: One-State-Strategies, $p_{\mathrm{I}} = 1$, $t_{\mathrm{L}} = 1$, copy-best-player, short memory, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, network=$80 \times 80$.

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 6).

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 suggested random sequential learning to model more realistic timing. While in each period all possible interactions take place, only one single player learns in a given period. The order in which players learn is determined randomly. In a particular setting[16], where May and Nowak found cooperation, random sequential learning leads to general defection. We agree with their finding that for a given initial state cooperation might vanish once stochastic timing is introduced. We have analyzed the influence of stochastic behavior in the framework that was described in section 2.4 on page 10. The precise way to introduce stochastic behavior differs slightly from Glance and Huberman's model[17] but, as we observe in figure 7 on the following page, the precise way how a stochastic

[16]The initial configuration, game, learning rule etc. corresponds to figure 3.a–c of May and Nowak (1992).

[17]Glance and Huberman give each period at most one player an opportunity to change her strategy and players interact each period deterministically, i.e. each possible interaction takes place with certainty. In our setting each possible interaction takes place with probability 1/2 and players learn randomly after 20...28 periods.

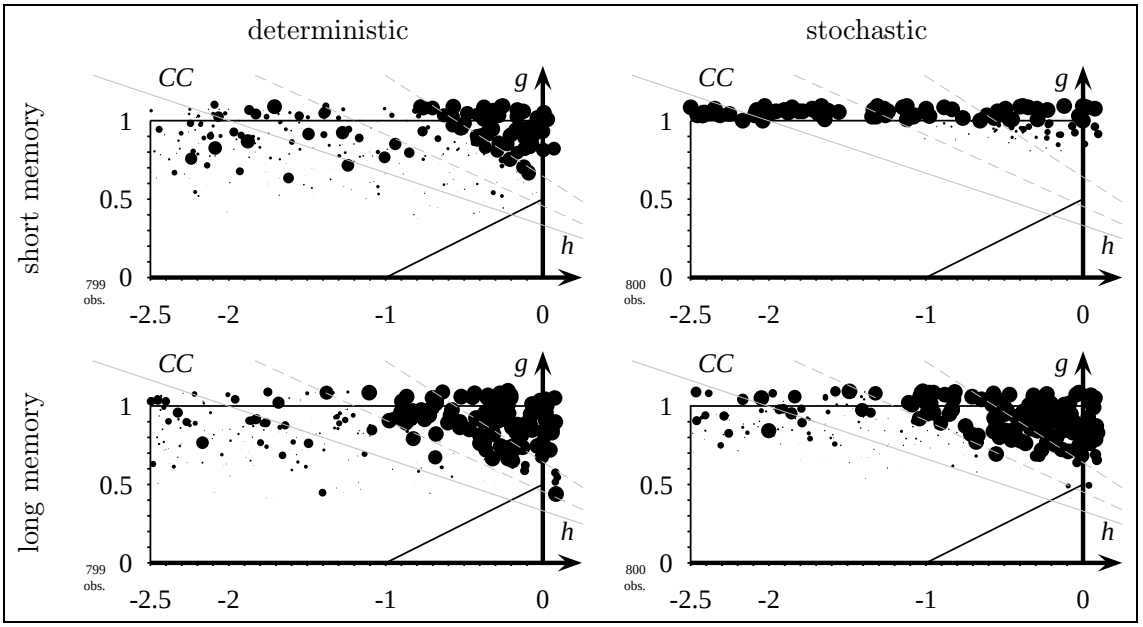


Figure 7: Stochastic learning eliminates cooperation only with 'short memory' ('copy best player' learning rule, network=80×80, $r = 1$).

environment is modeled does not matter. Cooperation dies out both in Glance and Huberman's and in our simulations.

To save space we display only the proportion of mutual cooperation for a variety of models. The top left diagram in figure 7 displays the proportion of cooperation in May and Nowak's model. We have seen this diagram already in figure 6 on the page before. The top right diagram in figure 7 displays the same model, but now with stochastic learning and interaction.

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 2.7) as a base for learning decisions. Matters change if players use *long memory*. This case is displayed in the bottom like of figure 7. Here we see that introducing stochastic behavior has almost no influence on the proportion of cooperation.

Further we see that the model described in section 3.3 hardly explains the results of stochastic interaction and *short memory*. At least it is not possible to give a certain cluster size that explains cooperative and non-cooperative area. Both *long memory* cases and the deterministic *short memory* result seem to be more compatible with a fixed cluster size in the model of section 3.3.

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 figure 8 we show the results for the 'copy best strategy' learning rule. It turns out that changing the learning rule has little impact on the proportion of mutual cooperation.

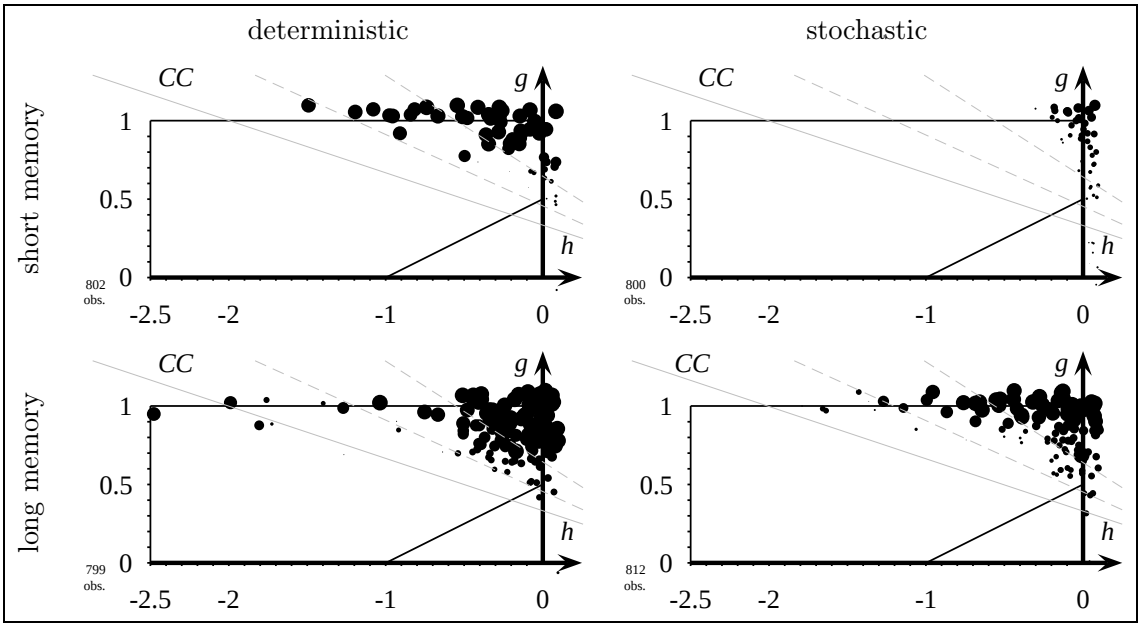


Figure 8: Also with the 'copy best strategy' learning rule, we have that stochastic learning eliminates cooperation only with 'short memory'.

Observation 2 *With simple (one-state) strategies introduction of stochastic behavior may eliminate cooperation if players' memory is short.*

We have the following intuition for this property: With our evolutionary dynamics players situations are changing permanently. Sometimes a cooperative player may be surrounded by other Cs, sometimes she may be at the borderline of a cluster of Cs. Obviously in the latter case she is more likely to learn to become a D. If players' memory is long, the payoffs that she uses for her learning decision are not influenced by her current situation. Given that cooperative players spend a larger part of their lifetime within clusters of other Cs her payoff will be relatively high. However, if a player's memory is short, the payoffs that she uses for her learning decision are influenced by her current situation. Only when she is close to some D players she may learn to become a D but this is exactly the situation where her payoffs decrease. With short memory this decrease can not be balanced by positive experiences with C in the past. Thus, with short memory, cooperation is more vulnerable.

3.6 Introducing Discriminative Behavior

The above discussed simple (one-state) repeated game 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 discriminative behavior, assuming that players' repeated game strategies can be represented as small (two-state) automata as described in section 2.6. Each player uses only one automaton, but this automaton can be in different states against different neighbors. Results change substantially now.

An alternative way to introduce automata which we will not use in the remainder of the paper would be the following: A single automaton reacts on stage game behavior in the players' neighborhood and determines a single stage game strategy that is played against *all* neighbors without discrimination. We will not follow such an approach for two reasons:

- We want to study discriminative behavior, which requires individual stage games strategies against any neighbor.
- Automata that react on the state of a whole neighborhood are more complex than automata which react on the state of a single player. Related to this argument is the fact that it would be much harder to select a small subset of repeated game strategies for our simulations based on criteria such as small complexity.

As mentioned above we assume now that players use 'copy best strategy' as defined in section 2.8.2 as a learning rule. Most of the results we find hold for 'copy best player' (see section 2.8.1) too.

We will start in sections 3.6.1 to 3.7.1 with simpler models where players learn from short memory (as described in 6 on page 17) and where interaction is deterministic. Analyzing these simple models makes it already possible to understand some important properties of spatial models which are still present in more complex models. One of these properties that we want to discuss within the simple framework is exploitation. Studying these simple models makes it also possible to see their limitations. Sometimes deterministic interaction and short memory leads to the growth of strange stage game and repeated game strategies. We will therefore study in section 3.7.2 on page 43 a model where players learn from long memory (as described in 7 on page 17) and where interaction is stochastic. Such a model is sometimes harder to understand, but its properties are more robust and sometimes more convincing that those of the simple models that we study in the following paragraphs.

3.6.1 Deterministic Interaction and Stochastic Learning

In the following we study a model with deterministic interaction and stochastic learning. Within this framework we already see that introduction of automata leads to more cooperation.

We furthermore give a simple example of how exploitation and different payoffs of coexisting repeated game strategies may survive in the population. This feature persists also with stochastic interaction, but can be understood easier in the deterministic context.

The way we introduce stochastic behavior is similar to Glance and Huberman (1993) who, too, analyzed a model where interaction was deterministic, but learning was stochastic. Our model which in contrast to Glance and Huberman, allows for automata as repeated game strategies shows that this might be not enough. We will see that with stochastic learning, as long as interaction is still deterministic,

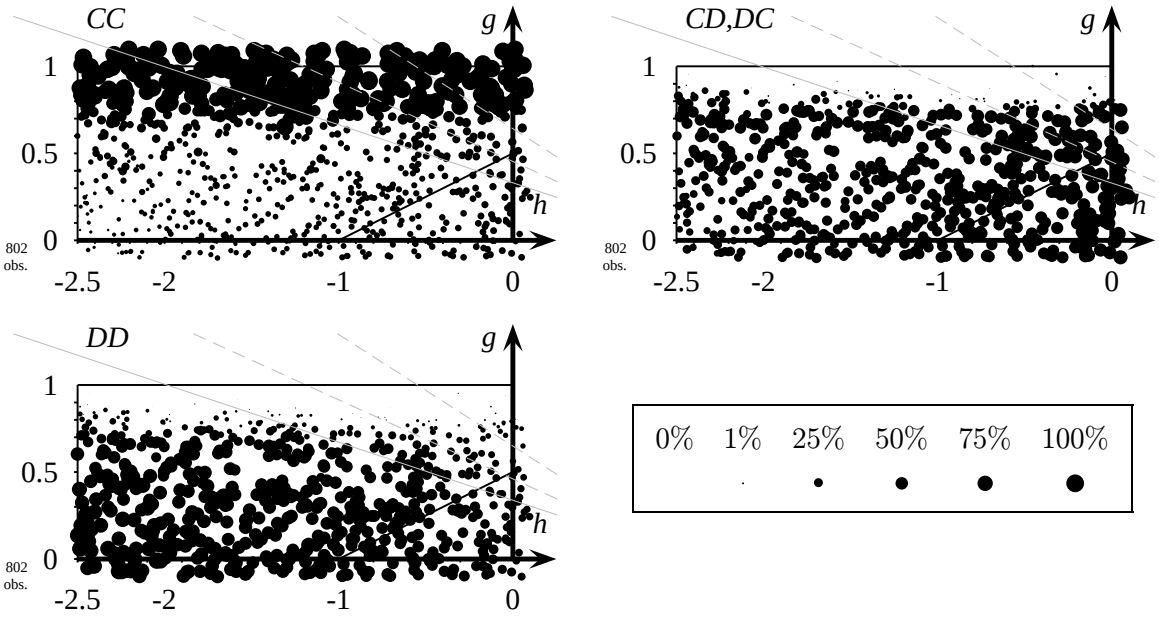


Figure 9: Complex (two-state) strategies induce cooperation: 2-state-strategies, $p_{\mathrm{I}} = 1$, $t_{\mathrm{L}} \in \{10, \ldots, 14\}$, copy-best-strategy, short memory, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, network=$80 \times 80$.

funny strategies are surprisingly successful. These strategies disappear as soon as interaction becomes stochastic.

More cooperation with complex automata: Let us first have a look at figure 9 on the next page where we display simulation results for a population that 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 stage game strategies CC is played.*

This observation is explained easily. For one-state automata (see section 3.5) 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 the cluster are 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.

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 have to give up, the

Player I \ Player II	C	D
C	7/8 7/8	1 −1/8
D	−1/8 1	0 0

Neighborhoods for learning and interaction contain the eight immediate neighbors. Players interact with certainty. Only today's payoffs matter. The learning rule is 'copy best strategy'. Strategies are Tit-for-Tat and D. Payoffs which are slightly higher than 6 are denoted with T_6^+.

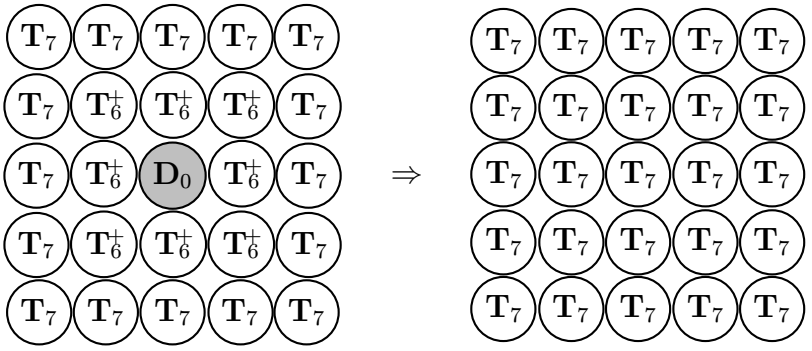


Figure 10: Why cooperation is more successful among automata.

tit-for-tat can still survive. Only if gains from cooperation are substantially lower also cooperation of two-state automata breaks down.

Figure 10 on the following page gives an example. The example is similar to the example we gave in the discussion of figure 5 but here we have replaced the Cs by tit-for-tat automata which are denoted with T. Those Ts which are surrounding the D defect against the D but cooperate against other Cs. Thus, their payoff is substantially higher than the payoff of the Cs in figure 5.[18] When the D compares payoffs of repeated game strategies that it can observe it finds that T is more successful and will become a T. Since the only remaining D dies out the population will remain forever in this state.

The above argument does, however, not imply that for all games and for all dynamics tit-for-tat is more successful than D. Tit-for-tat may loose payoffs for two reasons: First a newborn tit-for-tat can be exploited because it has not recognized its defective opponent yet. Second automata like tit-for-tat might have difficulties to synchronize with other tit-for-tats. Two tit-for-tats need not play CC all the time. They might also alternate in playing CD and DC. For the payoffs of the game in figure 10 this makes no difference. But if cooperation becomes more risky, e.g. playing C against D results in a lower payoff, then Ts could be substantially worse off.

[18]Since the payoff is slightly higher than 6 in this example, we call them T_6^+ in this example.

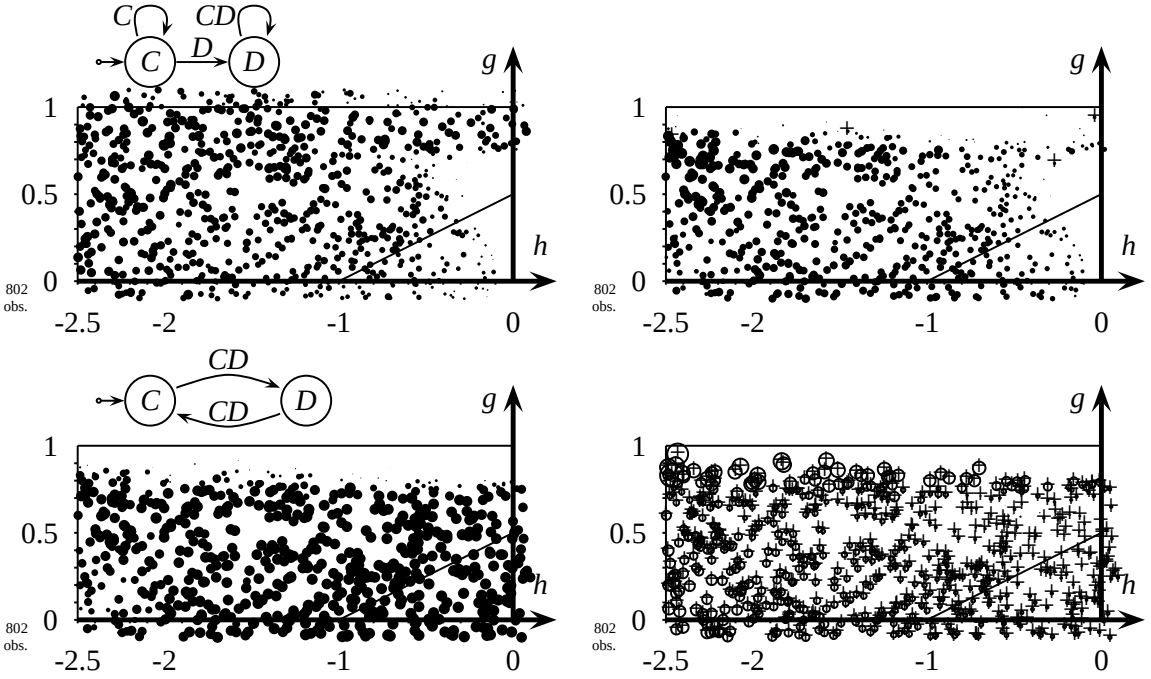


Figure 11: Deterministic interaction may favor odd strategies: 2-state-strategies, $p_{\mathrm{I}} = 1$, $t_{\mathrm{L}} \in \{10, \ldots, 14\}$, copy-best-strategy, short memory, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, network=$80 \times 80$.

Thus, it is no surprise to find games where even with automata cooperation breaks down.

Exploitation In the following paragraph we will discuss how our evolutionary dynamics may lead to inequalities — several repeated game strategies which all survive in the long run but which achieve different payoffs.

In figure 9 on page 34 we have displayed how the proportions of stage game strategies depend on the payoffs of the underlying game. Likewise we can analyze the proportion of automata which are generating these stage game strategies. We show here only the proportions of two automata, *grim* and *blinker*, which appear very often. Figure 11 on the following page displays on the left the proportions of grim and blinker in the $g \times h$-space, similar to the way we represented the proportion of pairs of strategies of the stage-game in section 3.4. We see that both repeated game strategies are present in the long run for a wide range of payoffs.

The right part of figure 11 displays the relative success of the repeated game strategies. Simulations that lead to a payoff of the repeated game strategy which is higher than the average payoff of the whole population[19] of a given repeated game strategy are represented as a black circle. The position of this circle is given by the payoffs g and h of the game. Circles are larger if the difference between a strategies

[19]Averages are taken over the last 50 periods of a simulation in order to exclude the influence of cycles. To calculate the average the total payoff achieved with the respective repeated game strategy was divided by the number of respective interactions.

payoff and population payoff are larger. Simulations that lead to an average payoff (per interaction) of the strategy which is smaller than the average payoff of the population are represented as a crossed white circle. Circles are again larger the larger the difference between average population payoff and strategies' payoff. If this difference becomes small the circle may become so small that only the cross (which has a constant size) is visible. In cases where all automata use the same stage game strategy average payoffs per interaction for all automata are identical and get no circles and no crosses at all.

Grim receives in most of the cases an average payoff per interaction which is higher than the average population payoff whereas blinker earns an average payoff which is smaller than the average population payoff. So we come to the following observation:

Observation 4 *With complex (two-state) strategies heterogeneity (over strategies) of average strategy payoff per interaction is high for a wide range of game-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 (in particular with deterministic interaction), it remains true for all other parametrizations which we study in the following.

Notice that the above described heterogeneity of payoffs differs from the heterogeneity of strategies pointed out by Lindgren and Nordahl (1994). They find that spatial evolution yields in contrast to global evolution starting from a homogeneous initial state in the long run a heterogoneous state in the sense that several different strategies are distributed in 'frozen patchy states' over the network. In the current and in the following section we try to show two related things: First even a population which has not reached a static state (and will possibly never reach one) can exhibit some stable properties, e.g. proportions of different strategies that may achieve a stable level. Second, not only that strategies are heterogeneously distributed over the network in such a state, in particular *payoffs* of strategies are different.

3.7 A Simple Model of Exploitation and Support

In the following paragraph we will consider a simplified model, to explain why and how blinkers are exploited by grim with synchronous interaction.

We will denote possible states of grim with G^C and G_D, depending on whether grim is in its cooperative or its defective state. Similarly we denote possible states of blinker with B^C or B_D. The average payoff per interaction of grim and blinker in a given period will be denoted with $\bar{u}_{\mathrm{G}}$ and $\bar{u}_{\mathrm{B}}$ respectively.

In the following we will describe an evolutionary dynamics of a population that has a cyclical structure. We define a cyclical equilibrium as follows:

Definition 1 (cyclical equilibrium) *A cyclical equilibrium is a sequence of states of a population that, once such a sequence is reached, given an evolutionary dynamics, repeats forever.*

In the model that we consider below, all of our cycles will be of length one or two. We will not only consider cycles of a complete state of the population, but also cycles of average payoffs or cycles of proportions of strategies or pairs of strategies. We will denote a cycle of any two objects a and b with the symbol $\lceil ab \rceil$. When we mention several cycles in the same context, like $\lceil ab \rceil$ and $\lceil cd \rceil$ then we want to say that a happens in the same period as c and b happens together with d. A cycle of length one can be denoted as e.g. $\lceil a \rceil$. When we mention $\lceil a \rceil$ together with $\lceil bc \rceil$ we want to say that a population alternates between a state where a happens together with b and another state where a happens together with c.

For the simplified model that we discuss in this section we make the following assumptions:

1. There are only two repeated game strategies present in the population: Grim and blinker.

2. Players interact with each of their opponents in each period exactly once.

3. Players do not observe their neighbors' payoffs but average payoffs of repeated game strategies over the whole population. They use the 'copy best strategy' learning rule, i.e. when they learn they copy the strategy with the highest average payoff (per interaction) in the population.

4. Players will learn very rarely. The individual learning probability in each period is ϵ. We assume that $\epsilon \to 0$.

The most crucial departure from our simulation model is assumption 3, i.e. that learning is now based on average payoffs of the whole population and not of the individual neighborhood. This simplifies the analysis drastic- ally.

Proposition 1 *Under the above assumptions there are four possible classes of cyclical equilibria.*

1. *A population that contains any proportion of* $\lceil \langle G^C, G^C \rangle \rceil$ *and* $\lceil \langle G_D, G_D \rangle \rceil$ *is in a cyclical equilibrium.*

2. *A population that contains any proportion of the cycles of pairs* $\lceil \langle B^C, B^C \rangle \langle B_D, B_D \rangle \rceil$, $\lceil \langle B_D, B_D \rangle \langle B^C, B^C \rangle \rceil$ *and* $\lceil \langle B^C, B_D \rangle \rceil$ *is in a cyclical equilibrium.*

3. *There are cyclical equilibria where in both periods average payoffs of grim and blinker are equal.*

4. *There is a single cyclical equilibrium both grim and blinker are present and have different payoffs at least in some periods.*

   This equilibrium has the property that payoffs alternate between $\bar{u}_{\mathrm{G}} = \epsilon \cdot (1 + 2h)/2 < \bar{u}_{\mathrm{B}} = \epsilon/2$ *and* $\bar{u}_{\mathrm{G}} = 1/2 > \bar{u}_{\mathrm{B}} = (g + h)/2$.[20] $1/4$ *of the population*

[20] g and h are parameters of the underlying game 5.

consists of pairs $\lceil\langle G_D, G_D\rangle\rceil$, $1/2$ *of the population consists of a cycle of pairs* $\left\lceil\langle G_D, B_D\rangle\langle G_D, B^C\rangle\right\rceil$ *and* $1/4$ *of the population consists of a cycle of pairs* $\left\lceil\langle B_D, B_D\rangle\langle B^C, B^C\rangle\right\rceil$.

The proof is given in appendix A on page 50.

The case that we have observed in the previous simulation corresponds to case 4 in the above simplified model. We see how it can happen that one repeated game strategy earns substantially less than average payoff but still does not die out. In our simplified example blinker could replicate in exactly the same number of periods and at the same speed as grim. It could do so, because each second period blinker had more than average payoff. Still, while blinker is sometimes more successful than grim it is only an ϵ more successful. When each second period grim is more successful than blinker, grim does not only earn an epsilon more but substantially more. Thus, over time grim earns more than blinker but still keeps blinker alive, due to a few newborn and cooperating blinkers which give blinker a slight advantage which is enough to make it survive.

Exploitation and favorable environments: We can interpret the above phenomenon as blinkers being part of a *favorable environment* that is created by grim. The latter 'feeds' sometimes (when it is born) its environment with a C. This facilitates reproduction of blinkers which quickly turn grim into its 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 reciprocal. Figure 12 on the following page shows the relative proportions of pairs of stage game strategies that are encountered by grim. The size of the circles is again proportional to the proportion of the respective pair of stage game strategies.

One could have expected that grim *never* experiences CD. In figure 12 we see that indeed grim does not play very often C against a D. Still CDs occur in the same range of games where we also find DC. If we interpret CD as 'support' and DC as 'exploitation', we can formulate the following observation:

Observation 5 *An automaton that exploits others has to support its victims at least sometimes.*

Grim cannot experience DC always because then its opponent would encounter low payoffs and copy next time a more successful repeated game strategy (e.g. grim). Instead in equilibrium it has to play CD sufficiently often to keeps its opponent alive.

Odd strategies Not only do the two automata grim and blinker give us a nice example for unequal payoffs of repeated game strategies that survive in the long run, furthermore blinker gives a good example for an unreasonable repeated game

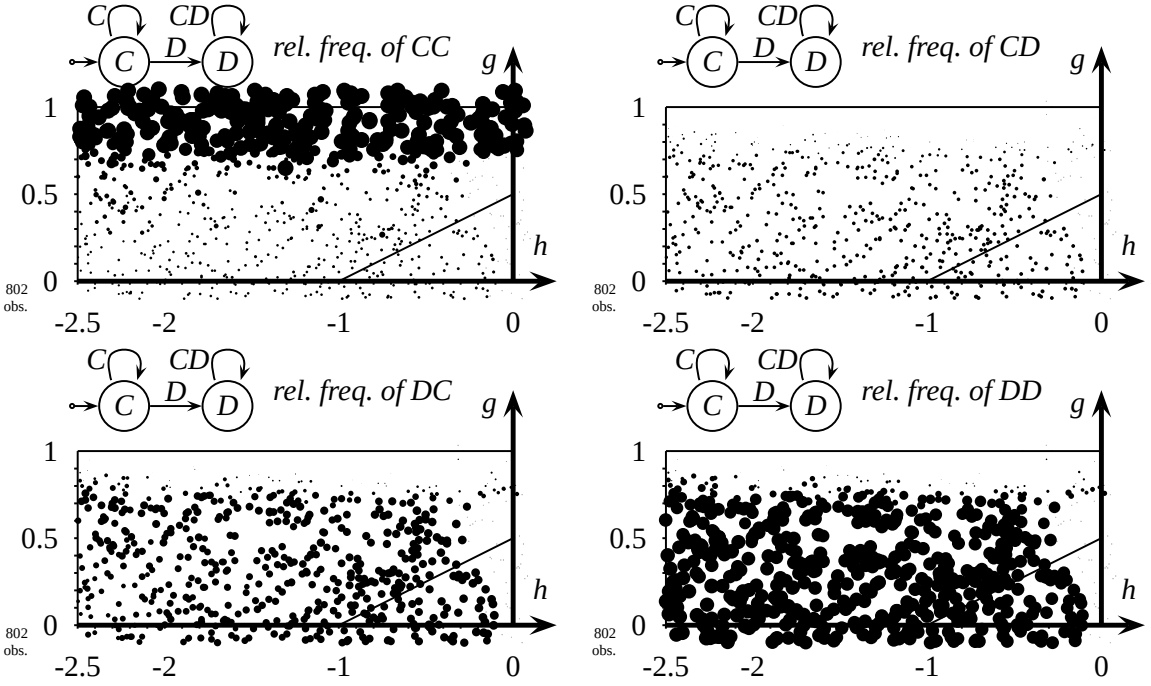


Figure 12: Exploiters have to 'feed' their victims at least sometimes: The pairs of stage game strategies that are encountered by grim.
2-state-strategies, $p_{\mathrm{I}} = 1$, $t_{\mathrm{L}} \in \{10, \ldots, 14\}$, copy-best-strategy, short memory, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, network=$80 \times 80$.

strategy that is eliminated once we add stochastic interaction to the already present stochastic learning. Figure 14 on page 42 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 6 *Deterministic interaction favors (together with stochastic learning) the appearance of 'odd' repeated game strategies like the blinker.*

In the following paragraph we will consider a simplified model to explain why and how blinkers are exploited by grim with synchronous interaction.

3.7.1 Stochastic Interaction, Stochastic Learning, Learning from Short Memory

Above we tried to illustrate in which way deterministic interaction is an abstract assumption and that properties of a dynamics with deterministic interaction vanish with synchronous interaction. If we disturb the dynamics introducing stochastic interaction the large cooperative payoff region persists, but odd automata disappear.

In the following we assume that in each period each single interaction takes place with probability $p_{\mathrm{I}} = 1/2$. Interactions are independent events, i.e. the probability that a certain interaction takes place is not influenced by the fact that other interactions take place.

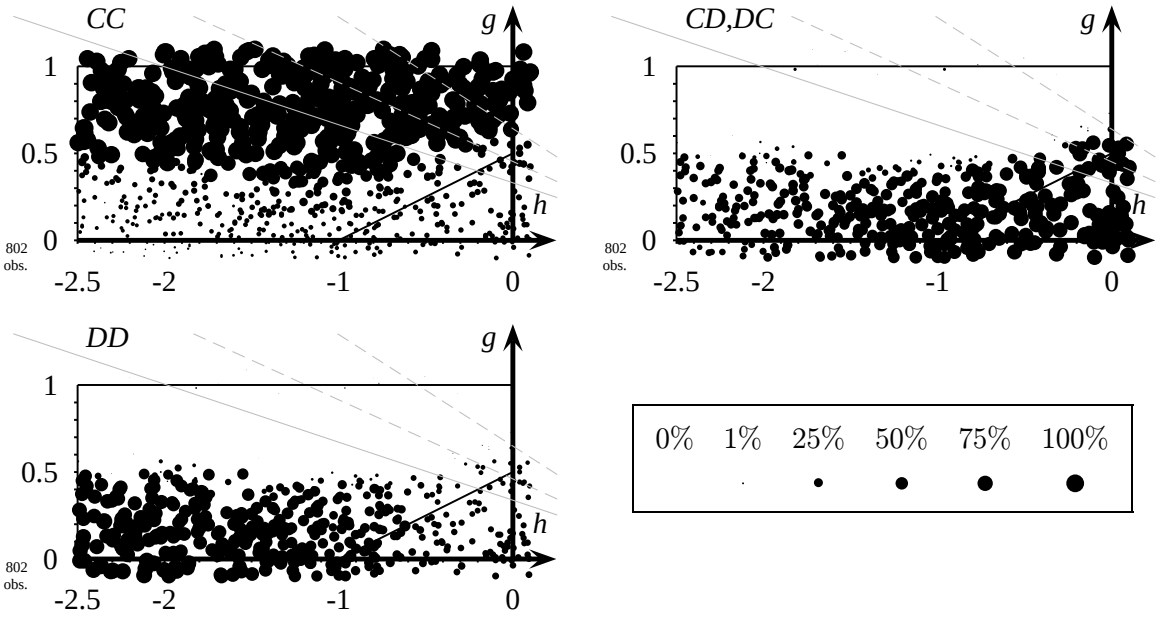


Figure 13: Complex (two-state) strategies induce cooperation also with stochastic interaction: 2-state-strategies, $p_{\mathrm{I}} = \frac{1}{2}$, $t_{\mathrm{L}} \in \{20, \ldots, 28\}$, copy-best-strategy, short memory, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, network=$80 \times 80$.

Let us first check the proportion of cooperation that is displayed in figure 13.

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

The argument for a large cooperative area here is similar to the one we gave for observation 3 on page 34. Again repeated game strategies like grim protect themselves against defectors and may be able to cooperate with other cooperators. Thus mutual cooperation can be sustained more easily.

Elimination of odd strategies: Above we have found that deterministic interaction gives rise to odd repeated game strategies like the blinker. In figure 14 we see that stochastic interaction eliminates this funny property.

Blinkers grow in a deterministic setting because they can be 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 prisoners' dilemma. The triangle where we find blinkers is exactly the region where playing a correlated pair of stage game strategies that puts weights 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.

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

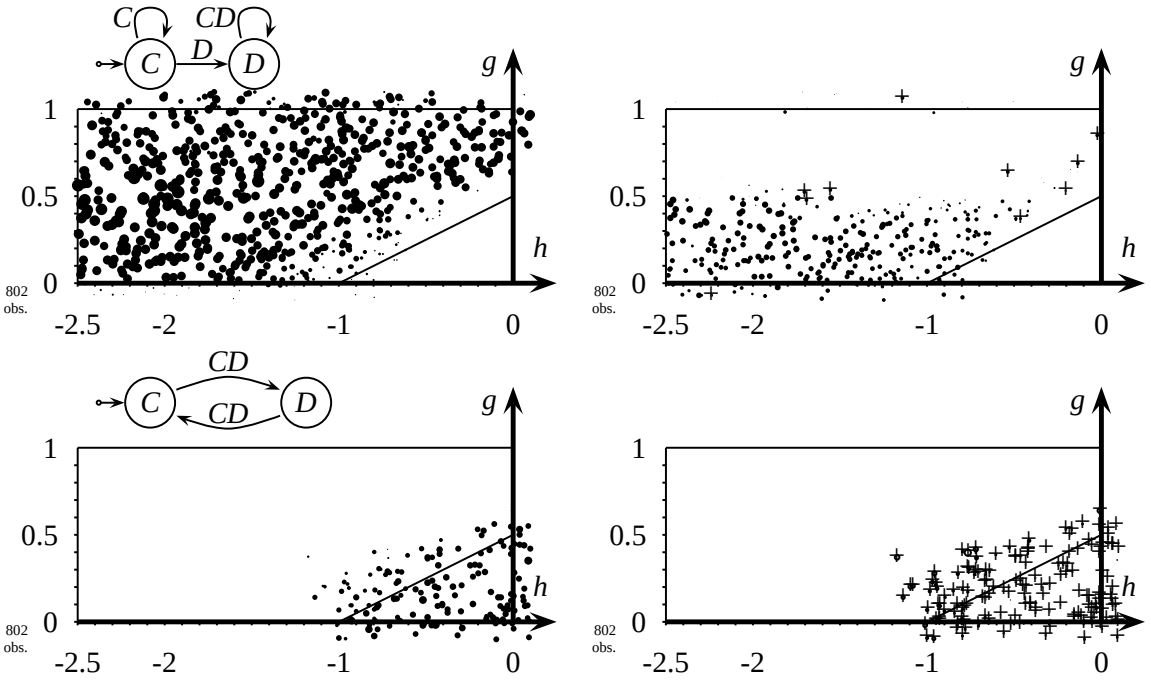


Figure 14: Stochastic interaction eliminates odd repeated game strategies: 2-state-strategies, $p_{\mathrm{I}} = \frac{1}{2}$, $t_{\mathrm{L}} \in \{20, \ldots, 28\}$, copy-best-strategy, short memory, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, network=$80 \times 80$.

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 2.7). 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 13 on the page before. 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 13

Observation 8 *Lack of memory leads to pairs of stage game strategies CD and DC, even in games were DD is the Pareto dominant pair of strategies.*

In games with $g < 0$ irrationality of C-playing players is even harder to justify than in the prisoners' dilemma ($0 < g < 1$). For a prisoners' dilemma C is irrational on the *individual* level but still *socially* desirable. In the context of prisoners' dilemmas CD can be explained as a failed attempt to achieve mutual cooperation.

Outside the range of prisoners' dilemmas, with $g < 0$ the CC-payoff is smaller than the DD-payoff, thus, C is neither *individually* nor *socially* desirable.

However, C can still appear in games with $g < 0$ if automata play both C and D. They 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 it to reproduce.

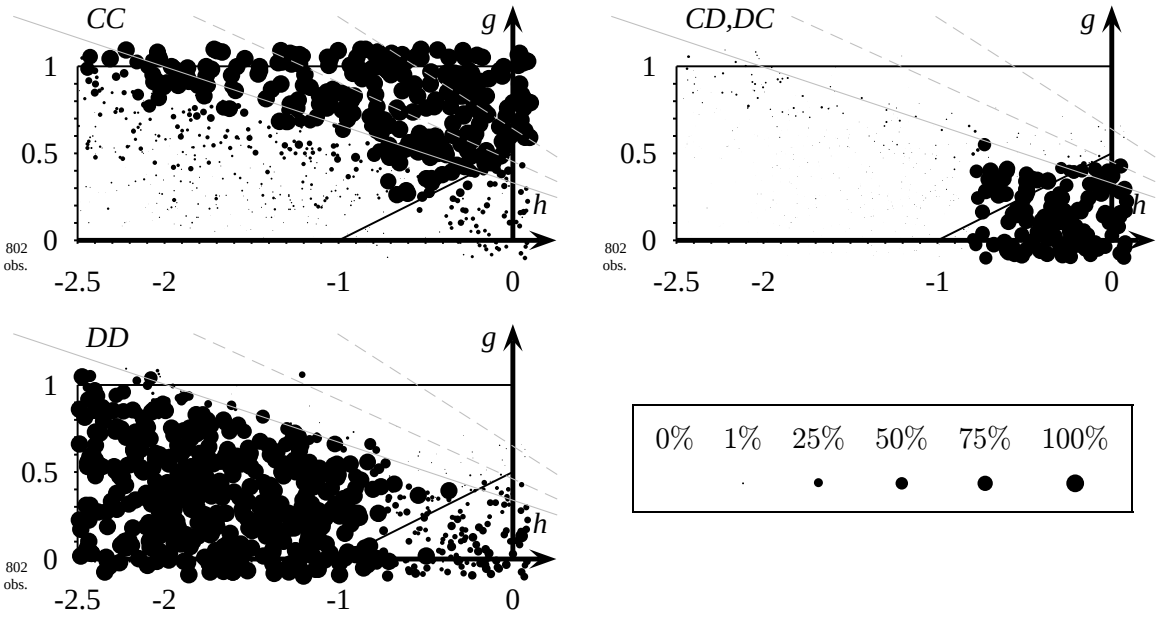


Figure 15: 2-state-strategies, $p_{\mathrm{I}} = \frac{1}{2}$, $t_{\mathrm{L}} \in \{20, \ldots, 28\}$, copy-best-strategy, long memory, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, network=$80 \times 80$.

The above observation shows that our assumption that players learn from short (one-period) memory may have questionable implications. In many situations it might be problematic to assume that players copy a repeated game strategy due to a one-period success, while the same repeated game strategy receives on average (over a complete cycle with its opponent) less payoff. In the following we will therefore make the assumption that players have 'long memory' as described in section 2.7 on page 16.

3.7.2 Stochastic Interaction, Stochastic Learning, Learning from Long Memory

In section 3.7.1 we have analyzed an evolutionary process with learning from *short memory* (as shown in figure 13 on page 41). In the current section we will assume that players learn from *long memory* (see figure 15).

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

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

This is also true for the region $g < 0$, i.e. the case where DD Pareto dominates CC. With long memory exploiters cannot take advantage of the fact that yesterday's losses will be forgotten when their victim replicates. Comparing grim's payoff in figure 11 on page 37 with figure 14 on the page before we see that in both cases grim receives more than average population payoff for a wide range of games. However, the difference between grim's payoff and average population payoff is larger with

short memory. Grim is a typical case. For the other automata the following holds as well:

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

To give an intuition for this observation assume an environment with only two repeated game strategies, A and B. Call A's average payoff (over time) $\bar{u}_A$ and the variance of A's payoff $\sigma^2_{u_A}$. Similar B's payoff has average $\bar{u}_B$ and variance $\sigma^2_{u_B}$. What we stated in observation 10 was that there is some variance in payoffs over automata. Let us assume that $\bar{u}_A > \bar{u}_B$. Still B may be able to survive, if either $\sigma^2_{u_B}$ or $\sigma^2_{u_A}$ is large enough. In this case there will always be some periods where B's current payoff is larger than A's current payoff. Since 'long memory' reduces the variance of payoffs *over time* for a given automaton the gap between $\bar{u}_A$ and $\bar{u}_B$ must be smaller to allow for survival of B. But then the variance of average payoffs *over automata* must be smaller, too.

Since 'long memory' reduces the variance of payoffs the gap between $\bar{u}_A$ and $\bar{u}_B$ must be smaller to allow for survival of B.

The shape of the cooperative region: If we compare figure 13 on page 41 with figure 15 we note that the shape of the cooperative region is different. The value of h in figure 15 has more influence on the proportion 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 than with learning from short memory.*

This is different from learning with short memory. There we observed that h had not much influence. In the discussion following observation 10 on the preceding page 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. The CD-payoff is given by the value of h. Thus, figure 15 shows 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.

3.8 Coordination Games

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

		Player II	
		a	b
Player I	a	g g	-1 h
	b	h -1	0 0

(23)

is a 'coordination game' for $g > -1$ and $h < 0$ (see figure 4 on page 24). 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.

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 $-1 < g < 0$. We say that the respective equilibrium is the 'Pareto dominant'.

Consider now a game where $g = 1$ (hence CC Pareto-dominates DD) and $h = -100$. Here coordinating on CC involves more risk, because a deviation 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 DD, which means following the principle of *risk dominance*. In the games described in figure 4 on page 24 the equilibrium CC risk dominates DD if and only if $g > -1 - h$. A thorough discussion of risk dominance is given in Harsanyi and Selten (1988).

Recent work of Kandori, Mailath and Rob (1993) or Young (1993) suggest that in a global model where players optimize myopically, evolution selects the risk dominant equilibrium in the very long run. Ellison (1993) has studied a spatial model where players optimize myopically and found that there the risk dominant equilibrium is selected even faster than in the model of Kandori, Mailath and Rob (1993) or Young (1993).

In section 3.3 we have already studied a simplified model of a continuum of players and considered the situation of a learning player whose left neighbors all play C and whose right neighbors all play D. In Ellison's model a *myopically optimizing* player will calculate that chances to meet either a C or a D are equally $1/2$ and therefore select the risk dominant equilibrium.

In the continuous model of section 3.3 a player who *imitates* successful strategies behaves differently than Ellison's myopically optimizing player. She will in the same situation become a C if

$$g > (-1 - h)\frac{4 - n}{4 + n} \tag{24}$$

where $n \leq 2$ describes the size of the cluster. Clusters with a diameter larger than 2 times the neighborhood radius can be treated as $n = 2$. While in a prisoners' dilemma clusters are often small[21] we have large clusters in coordination games. Thus we can assume $n = 2$. The gray line in figure 16 describes the set of games where $g = (-1 - h)/3$, i.e. the case where for $n = 2$ the expressions on both sides of inequality 24 are equal.

Observation 13 *It is almost precisely the line $g = (-1 - h)/3$ which divides the region where only C is played from the region where only D is played. CD's which fail to coordinate are extremely rare.*

[21]Small groups of defectors are successful. As soon as they start growing they kill themselves.

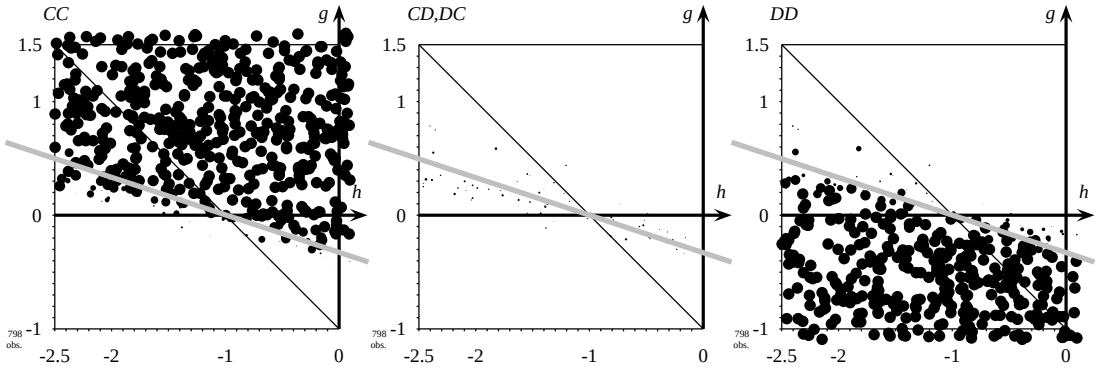


Figure 16: Coordination games with two-state-strategies: $p_{\mathrm{I}} = \frac{1}{2}$, $t_{\mathrm{L}} \in \{20, \ldots, 28\}$, copy-best-strategy, long memory, $r_{\mathrm{L}} = r_{\mathrm{I}} = 1$, network=$80 \times 80$.

In particular the continuous model described in section 3.3 explains the behavior of the discrete population substantially better than Pareto dominance or risk dominance.

3.9 The Influence of Locality

A particular feature of the model analyzed so far was the assumption that learning and interaction was local. Figure 17 on the next page shows what happens if we move gradually to a global model. We have chosen here a smaller torus of size only 21×21 because we want to move all the way from a local to a global model. A learning and interaction radius r of 10 means here that a player learns from the whole population and interacts with the whole population, smaller values for r stand for more local evolution and interaction. To save space we have only displayed the proportion of mutual cooperation.

We first observe that the smaller network size (only 21×21 compared with 80×80 in the previous examples) does not affect the results. Comparison of the top left diagram in figure 17 on the preceding page ($r_{\mathrm{L}} = 1$) with figure 15 on page 44 ($r_{\mathrm{L}} = 1$, but a larger network) shows no influence of the size of the network.

We furthermore see that the proportion of cooperation decreases gradually while the model becomes a more global one. Figure 17 on the preceding page shows that it does not matter whether the radius of the neighborhood 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 17 shows the global model with 441 players which all learn from each other and all interact with each other. Here the cooperative region has become substantially smaller.

We have carried out two similar exercises. In the first we kept the learning radius r_{L} fixed and varied the interaction radius r_{I}. In the second we varied the learning radius r_{L} while the interaction radius r_{I} was held constant. The results are similar to the above. Increasing either the learning or the interaction radius reduces gradually the size of the cooperative area. Thus, both locality of evolution and locality of interaction are responsible for cooperation.

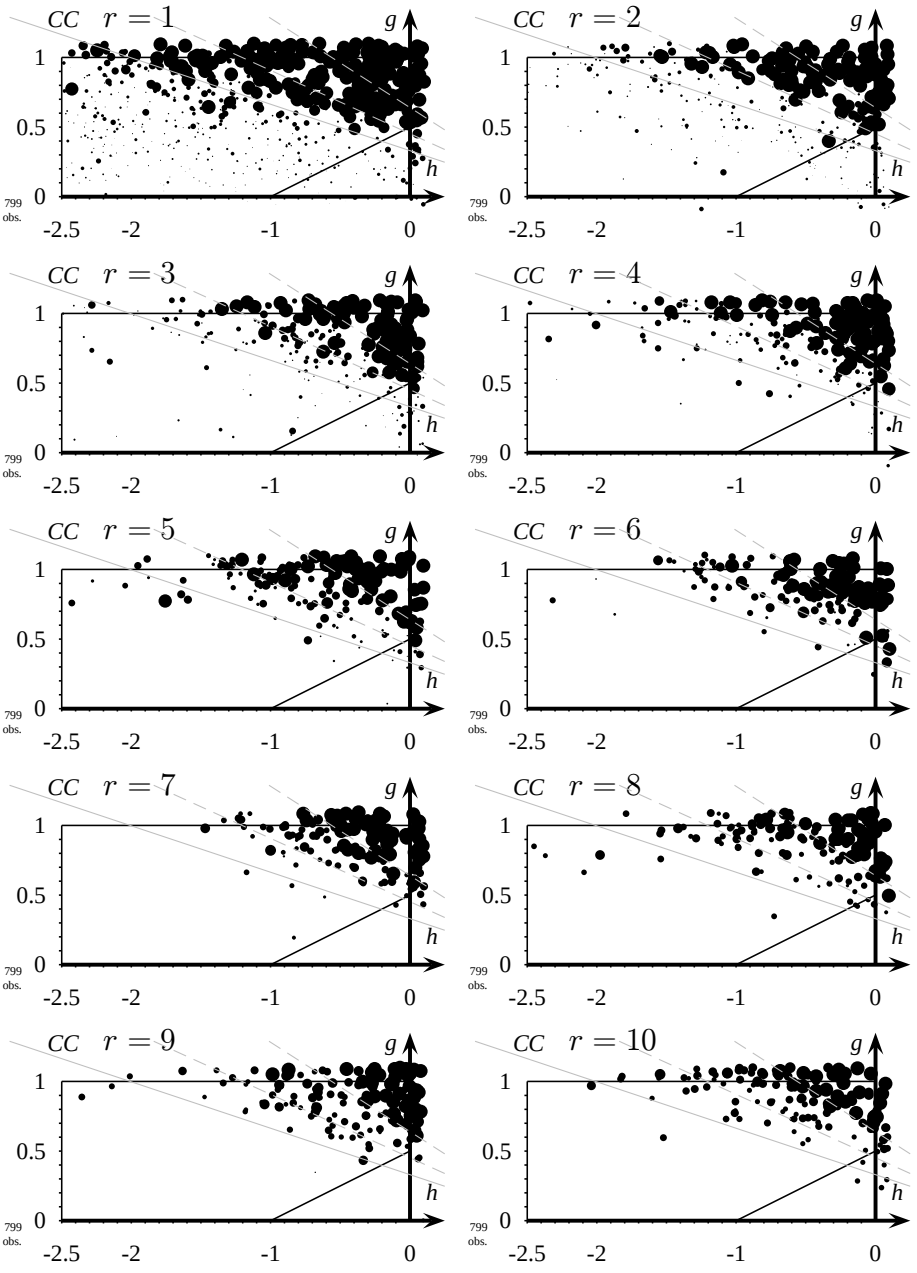


Figure 17: The influence of locality, two-state-strategies, $p_{\mathrm{I}} = \frac{1}{2}$, $t_{\mathrm{L}} \in \{20, 28\}$, copy-best-strategy, long memory, network=21×21.

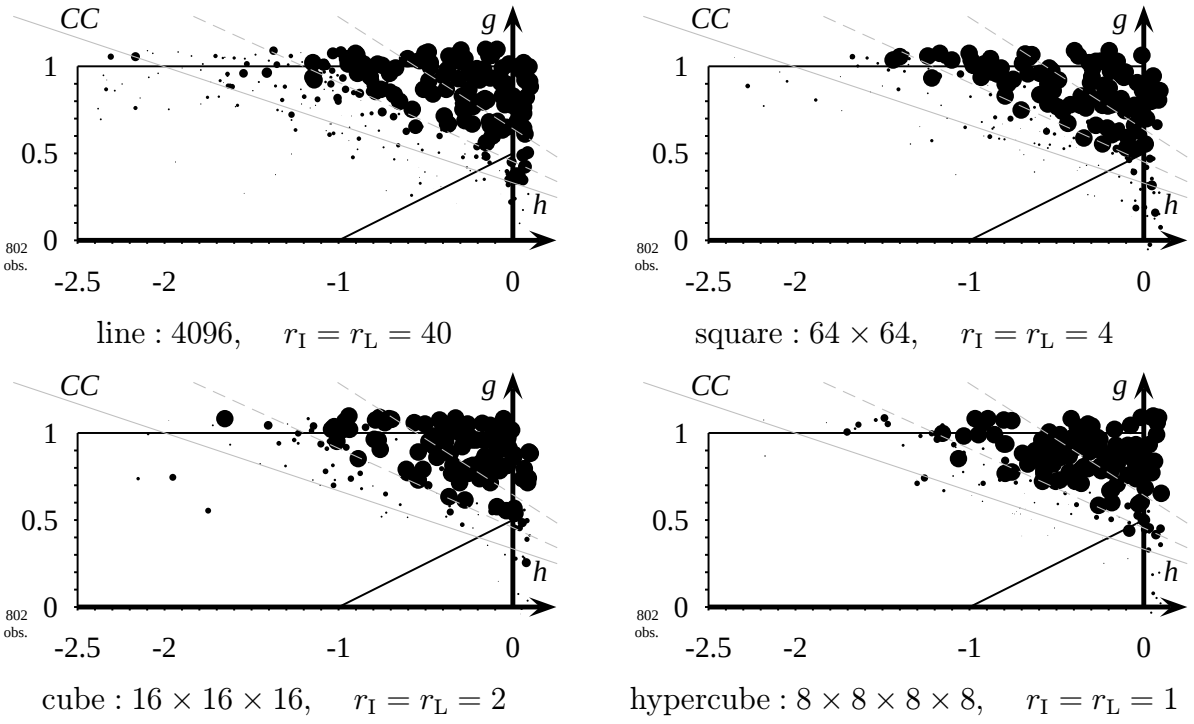


Figure 18: Proportion of mutual cooperation in 1...4-dimensional networks: 2-state-strategies, $p_I = \frac{1}{2}$, $t_L \in \{20, \ldots, 28\}$, copy-best-strategy, long memory.

3.10 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.[22] Figure 18 on the next page shows the proportion of mutual cooperation for networks in several dimensions. All networks have 4096 players and all neighborhoods have almost the same number (80...125) of opponents. Notice that the cooperative area is slightly smaller than the one of figure 15 on page 44. This is due to the fact that the neighborhoods in figure 15 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 the following points:

[22]We are grateful to George Mailath who suggested to carry out the following simulations.

A continuous approximation of a discrete network In section 3.3 we have sketched a model of a continuous population distributed on a line, divided into clusters of different strategies. This approach is similar to an idea used by Ellison (1993). We then assume that the size of the clusters is constant. Thus, the behavior of players which are located on the borderlines of the clusters can be analyzed easily. The behavior of our simulated networks fits with the predictions given by the continuous population model if timing is stochastic. In particular this model gives an accurate description of the conditions that lead either to the selection of risk dominant or Pareto dominant equilibria in coordination games.

More cooperation with discriminative behavior We analyzed the effect of introducing discriminative behavior, modeled as automata, into a population. One might have guessed that cooperation would be more stable if players use simple strategies. Players that are capable of using more complex strategies might move quicker to a more rational solution. Instead it turns out that more complexity fosters cooperation.

Payoffs vary among strategies in the long run While inequality might be a common phenomenon, global evolution can hardly explain it. Local evolution and interaction allows for symbioses of strategies where one partially feeds but mainly exploits the other. In section 3.7 we have explained with the help of a simple model that different payoffs appear together with a cyclical structure of the payoffs: In some periods strategy a is more successful, in other periods strategy b. If during periods when b is more successful the payoff difference between a and b is only small, while during periods when a is more successful the payoff difference is large, then a's payoffs will be higher on average but bs may still survive.

Stochastic timing does not necessarily eliminate cooperation We have tried to respond to a criticism raised by Huberman and Glance (1993) who argued that stochastic timing might eliminate cooperation completely. While this is doubtlessly true for the model studied by Huberman and Glance, we found that for other reasonable models stochastic timing has no influence on the amount of cooperation at all. In particular stochastic timing seems not to affect cooperation if we allow for more complex strategies (automata), regardless whether we are in the Huberman and Glance framework or not.

Stochastic timing not only for learning but also for interaction Huberman and Glance assume that only learning is a stochastic event. In their model interaction is still deterministic. We have given an example for stochastic learning and deterministic interaction where the evolutionary system got stuck into odd patterns of behavior. We think that stochastic timing of the update of the strategy and stochastic interaction are preferable and avoid these artificial effects.

We have not investigated many other learning rules yet. The learning rules that we use in this paper have two important properties:

- Our learning rules are *deterministic* in the sense that players always learn a 'best' strategy, no matter *how much* better this 'best' strategy is compared to other strategies. This is certainly a property that stabilizes long-run unequal payoffs. Long-run inequalities persist in a system since some strategies grow in periods when they are only slightly better than other strategies, while others grow in other periods, when they are much better than the former.

- Our learning rules are *asymmetric* in the sense 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 the player herself.

We have done some simulations that use a variant of genetic algorithms (sometimes players copy a repeated game strategy, sometimes they copy a learning rule — see Kirchkamp and Schlag (1995)) in order to let players choose themselves their preferred degree of asymmetry of learning rules. In these simulation learning rules evolve which are both asymmetric and stochastic. They lead to less cooperation than the learning rules which we investigated in the current paper. We suspect that it is not the property to be asymmetric but the property to be be stochastic that lead to less cooperation in these simulations. Still in this area a lot of work has to be done.

A Proof of Proposition 1

In the following we will denote a pair of two automata a and b with the symbol $\langle a, b\rangle$. For notational convenience we will make no distinction between $\langle a, b\rangle$ and $\langle b, a\rangle$

1 If there are only grims in the population, then only $\langle G^C, G^C\rangle$, $\langle G^C, G_D\rangle$ and $\langle G_D, G_D\rangle$ are possible. $\langle G^C, G_D\rangle$ is unstable, thus, only $\langle G^C, G^C\rangle$ and $\langle G_D, G_D\rangle$ will remain in the population. Given any proportion of $\langle G^C, G^C\rangle$ and $\langle G_D, G_D\rangle$ no player will ever learn, since there are no alternative repeated game strategies to observe.

2 If there are only blinkers in the population, then only $\lceil\langle B^C, B^C\rangle\langle B_D, B_D\rangle\rceil$, $\lceil\langle B_D, B_D\rangle\langle B^C, B^C\rangle\rceil$ and $\lceil\langle B^C, B_D\rangle\rceil$ are possible. All of these cycles of pairs are stable. No player will ever learn, since there are no alternative repeated game strategies to observe.

3 Trivially a state where possibly payoffs cycle, but nevertheless in all periods grim and blinker achieve the same payoff is an equilibrium, since no player ever has any reason to learn. An example for such an equilibrium might be a population where 1/4 are cycles $\lceil\langle B^C, B^C\rangle\langle B_D, B_D\rangle\rceil$, 1/4 are cycles $\lceil\langle B_D, B_D\rangle\langle B^C, B^C\rangle\rceil$, $(1+h)/(4g)$ are $\lceil\langle G^C, G^C\rangle\rceil$ and $(2g-1-h)/(4g)$ are

$\lceil\langle G_D, G_D\rangle\rceil$ where g and h are parameters of the underlying game 5 on page 12 (of course not for all, but at least for some values of g and h the above proportions are all in $[0,1]$). In this case average payoffs of blinkers and grims is in each period $(1+h)/2$.

This equilibrium is not stable. Change the proportion of a cycle that contains blinkers only slightly and payoffs of blinkers start cycling around the average population payoff. In the discussion of case 4 which is described below we will see that once such a state is reached, the cycling becomes stronger and stronger until the equilibrium of case 4 is reached.

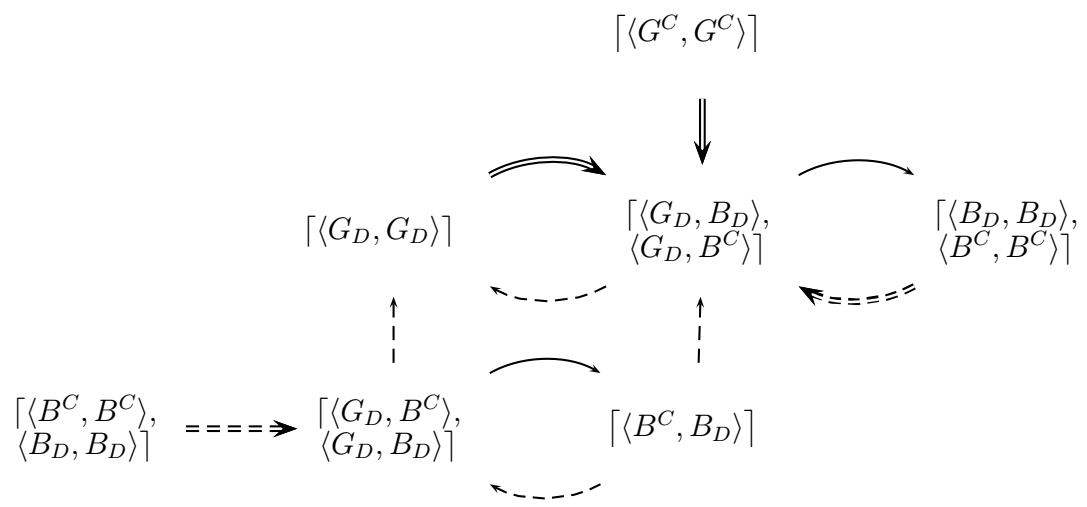
4 $\langle G^C, G^C\rangle$, $\langle G_D, G_D\rangle$, $\langle G_D, B_D\rangle$, $\langle B_D, B_D\rangle$, $\langle B^C, B^C\rangle$ all belong to a stable cycle in the sense that after a limited number of transitions each pair will return to its original state. Unstable pairs are the remaining ones $\langle G^C, G_D\rangle$, $\langle G^C, B^C\rangle$ and $\langle G^C, B_D\rangle$. Unstable pairs can only be present in a cyclical equilibrium if they are introduced through learning. Since we take the limit of the the learning rate $\epsilon \to 0$, the proportion of unstable pairs in the population will be 0 as well. Thus, in a cyclical equilibrium we have only stable pairs.

All stable pairs belong to cycles of length either one or two. Therefore average payoffs in a cyclical equilibrium also form a cycle of length either one or two.

A cyclical equilibrium with both grim and blinker present in the population must have average payoffs of grim and blinker respectively belonging to a cycle of length two. Otherwise one repeated game strategy would have *always* more than average payoff, which together with our learning rule, copy always the best average strategy, leads to the extinction of the other strategy. Let us without loss of generality assume that average payoffs form a two cycle with $\lceil \bar{u}_{\mathrm{B}} > \bar{u}_{\mathrm{G}}, \bar{u}_{\mathrm{G}} > \bar{u}_{\mathrm{B}}\rceil$. Thus, in the first period of each cycle $\bar{u}_{\mathrm{B}} > \bar{u}_{\mathrm{G}}$, in the second $\bar{u}_{\mathrm{G}} > \bar{u}_{\mathrm{B}}$.

Figure 19 on the following page shows now the dynamics of stable pairs. Let us give one example how to determine the transitions. There is some flow away from the cycle $\lceil\langle B^C, B^C\rangle\langle B_D, B_D\rangle\rceil$. In the second period of the cycle, when $\bar{u}_{\mathrm{G}} > \bar{u}_{\mathrm{B}}$, a $\langle B_D, B_D\rangle$ has with probability ϵ an opportunity to learn, which will then lead (in the first period of the next cycle) to a pair $\langle G^C, B^C\rangle$. The pair $\langle G^C, B^C\rangle$ is an unstable pair which becomes in the second period of the current cycle a $\langle G_D, B_D\rangle$. The latter is a member of the stable cycle $\lceil\langle G_D, B^C\rangle\langle G_D, B_D\rangle\rceil$. Since we have two blinkers in the pair $\langle B_D, B_D\rangle$ the total flow away from the cycle $\lceil\langle B^C, B^C\rangle\langle B_D, B_D\rangle\rceil$ has speed 2ϵ. Since we consider the case $\epsilon \to 0$ we can neglect the case that both members of a pair learn simultaneously.

When we consider all transitions given in figure 19 we see that only the three cycles $\lceil\langle G_D, G_D\rangle\rceil$, $\lceil\langle G_D, B_D\rangle\langle G_D, B^C\rangle\rceil$ and $\lceil\langle B_D, B_D\rangle\langle B^C, B^C\rangle\rceil$ will remain in the population in the long run. Given the flow between these three cycles we see that 1/4 $\lceil\langle G_D, G_D\rangle\rceil$, 1/2 $\lceil\langle G_D, B_D\rangle\langle G_D, B^C\rangle\rceil$ and 1/4 of $\lceil\langle B_D, B_D\rangle\langle B^C, B^C\rangle\rceil$ are a stable equilibrium.



Transitions with speed 2ϵ are shown with double lines. Speed ϵ is shown with single lines. Transitions that take place in the second period are shown with dashed (single or double) lines. The first period is shown with solid lines.

Figure 19: Transitions of cycles of stable pairs if payoffs cycle like $\lceil \bar{u}_{\mathrm{B}} > \bar{u}_{\mathrm{G}}, \bar{u}_{\mathrm{G}} > \bar{u}_{\mathrm{B}} \rceil$.

Now let us look at the payoffs. In the first period of a cycle almost all pairs mutually defect, i.e. they get, given the game 5 on page 12, a payoff of zero. Nevertheless a few pairs are always present which actually learned a new strategy. Thus, we have $\epsilon/2$ of $\langle G^C, G_D\rangle$ and $\epsilon/2$ of $\langle G^C, B_D\rangle$. Therefore payoffs are in the first period of a cycle as follows:

$$\bar{u}_{\mathrm{G}} = \frac{\epsilon}{2}(1+2h) \qquad < \qquad \bar{u}_{\mathrm{B}} = \frac{\epsilon}{2}$$

In the second period of a cycle learning players are negligible:

$$\bar{u}_{\mathrm{G}} = \frac{1}{2} \qquad > \qquad \bar{u}_{\mathrm{B}} = \frac{1}{2}(g+h)$$

Thus, indeed a cycle of payoffs like $\lceil \bar{u}_{\mathrm{B}} > \bar{u}_{\mathrm{G}}, \bar{u}_{\mathrm{G}} > \bar{u}_{\mathrm{B}} \rceil$ exists. ■

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.

Ellison, G. (1993): "Learning, Local Interaction, and Coordination," *Econometrica*, 61, 1047–1071.

Eshel, I., L. Samuelson, and A. Shaked (1996): "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.

Hotelling, H. (1929): "Stability in Competition," *Economic journal*, 39, 41–57.

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

Kirchkamp, O. (1995): "Asynchronous Evolution of Pairs, How spatial evolution leads to inequality," Discussion Paper B–329, SFB 303, Rheinische Friedrich Wilhelms Universität Bonn.

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

Lindgren, K., and M. G. Nordahl (1994): "Evolutionary dynamics of spatial games," *Physica D*, 75, 292–309.

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

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

Nelson, R. (1995): "Recent Evolutionary Theorizing about Economic Change," *Journal of Economic Literature*, 33, 48–90.

Nowak, M. A., S. Bonnhoefer, and R. M. May (1994): "More Spatial Games," *International Journal of Bifurcation and Chaos*, 4, 33–56.

Nowak, M. A., and R. M. May (1993): "The Spatial Dilemmas of Evolution," *International Journal of Bifurcation and Chaos*, 3, 35–78.

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.

Schumpeter, J. (1934): *The Theory of Economic Development*. Harvard University Press, Cambridge.

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.