SONDERFORSCHUNGSBEREICH 303
„Information und die
Koordination wirtschaftlicher Aktivitäten“

Projektbereich B
Discussion Paper No. B-379

Simultaneous Evolution of Learning Rules and Strategies

Oliver Kirchkamp[1]

June 1996

---

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

An electronic version of the paper is available at http://www.econ3.uni-bonn.de/~oliver/endogLea.html

Financial support from the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn is gratefully acknowledged. I thank Karl Schlag and Avner Shaked for helpful and stimulating comments.

RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT BONN

D-53113 Bonn, Lennéstraße 37

Abstract

We study a model of local evolution. Agents are located on a network and interact strategically with their neighbors. Strategies are chosen with the help of learning rules that are based on the success of strategies observed in the neighborhood.

The standard literature on local evolution assumes these learning rules to be *exogenous* and fixed. In this paper we consider a specific evolutionary dynamics that determines these learning rules *endogenously*.

We find with the help of simulations that in the long run learning rules behave deterministically but are asymmetric in the sense that while learning they put more weight on the learning players' experience then on the observed players' one. Nevertheless stage game behavior under these learning rules is similar to behavior with symmetric learning rules.

Keywords: Evolutionary Game Theory, Networks. JEL-Code: C63, C72, D62, D63, D73, D83, R12, R13.

Contents

1 Introduction

In many models on local evolution (see Axelrod (1984, p. 158ff), Lindgren and Nordahl (1994), Nowak and May (1992, 1993), Nowak, Bonnhoeffer and May (1994), Eshel, Samuelson and Shaked (1996), Kirchkamp (1995)) the 'evolving part' of players' characteristics are players' strategies. How these strategies evolve is given exogenously by a so called 'learning rule'. There are of course learning rules which lead to different results. In this paper we want to find out, whether the above learning rules or the properties implied by the above rules can be justified by an evolutionary argument. We will analyze a model where players' learning rules are not given exogenously but instead are chosen by the players themselves. In other words, we replace a fixed learning rule by a dynamics which selects learning rules.

Such a dynamics will yield a set of learning rules which we can compare with the exogenously given learning rules from the literature. Further we can compare the stage game behavior of a population using endogenous learning rules with the stage game behavior of a population with fixed learning rules.

We want to ask two questions: First we will investigate whether the *learning rules* discussed in the above literature are likely to be selected by evolution. Second we want to know whether the *behavior* of a society with endogenous learning rules is different from the behavior of one with a fixed learning rule. In this paper we will present simulation results to give an answer to this question.

We find it particularly interesting to analyze evolution of learning rules in a *local* framework because it is a 'fragile' framework. Small changes of players' learning rules are more influential in a local framework than in a global one. With local evolution some reasonable learning rules imply cooperation in prisoners' dilemmas[1] others do not.[2] In a global framework both kinds of learning rules imply the same (noncooperative) result.

Binmore and Samuelson (1994) study endogenous learning rules in a global framework. In contrast to the model we present below, their players vary only a single parameter of their learning rules, an aspiration level. In the model discussed below players will vary an aspiration level and two other parameters which characterize sensitivity to the player's own and an observed neighbor's payoff.

2 The Model

2.1 Overview

The environment We will consider a population of players each occupying one cell of a torus of size 50 × 50. Players will play games with their neighbors on this network, learn repeated game strategies from their neighbors and update their learning rule using information from their neighbors.

---

[1] E.g. the rules that we have studied in Kirchkamp (1995).

[2] E.g. learning rules where switching probabilities are linear in payoffs and always stochastic.

Simulations will start from random initial configurations which are described in detail in section 2.6 on page 10. The games that players play within these neighborhoods will change from time to time. A more detailed description of these games is given in section 2.2 on page 4.

Players' characteristics Players are described by three characteristics: Stage game strategies, a repeated game strategy, and a learning rule.

We visualize the three parameters with the help of figure 1 on the following page.

- *Learning rules* of a player are influenced by three factors: A selection dynamics which is given exogenously, information on the players' own learning rule and its payoffs, and information on her neighbors' learning rules and their payoffs. One might interpret the selection dynamics as a function that takes information on learning rules and payoffs as arguments and, thus, determines a new learning rule for a player. The selection process will be perturbed by mutations. How learning rules are determined by the selection dynamics is discussed in more detail in section 2.5 on page 8.

- A player's *repeated game strategy* is influenced by following factors: A learning rule, information on the player's own payoff and repeated game 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. The process will again be perturbed by mutations. How repeated game strategies are selected by learning rules is discussed in more detail in section 2.4 on page 5.

- 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 2.3 on page 5.

  Stage game strategies determine interactions among players. Given a game which is specified exogenously and which changes from time to time, stage game strategies of two players determine stage game payoffs. These payoffs contribute to the payoffs of the repeated game strategies and the learning rules. On the basis of the latter payoffs, then new repeated game strategies and learning rules are selected.

The above properties change stochastically at different speeds. Players *interact* with a high probability and change their *repeated game strategy* with a low probability. The probability that the underlying *stage game* changes is smaller than the individual probability to change a repeated game strategy and even smaller is the probability that a *learning rule* is updated.

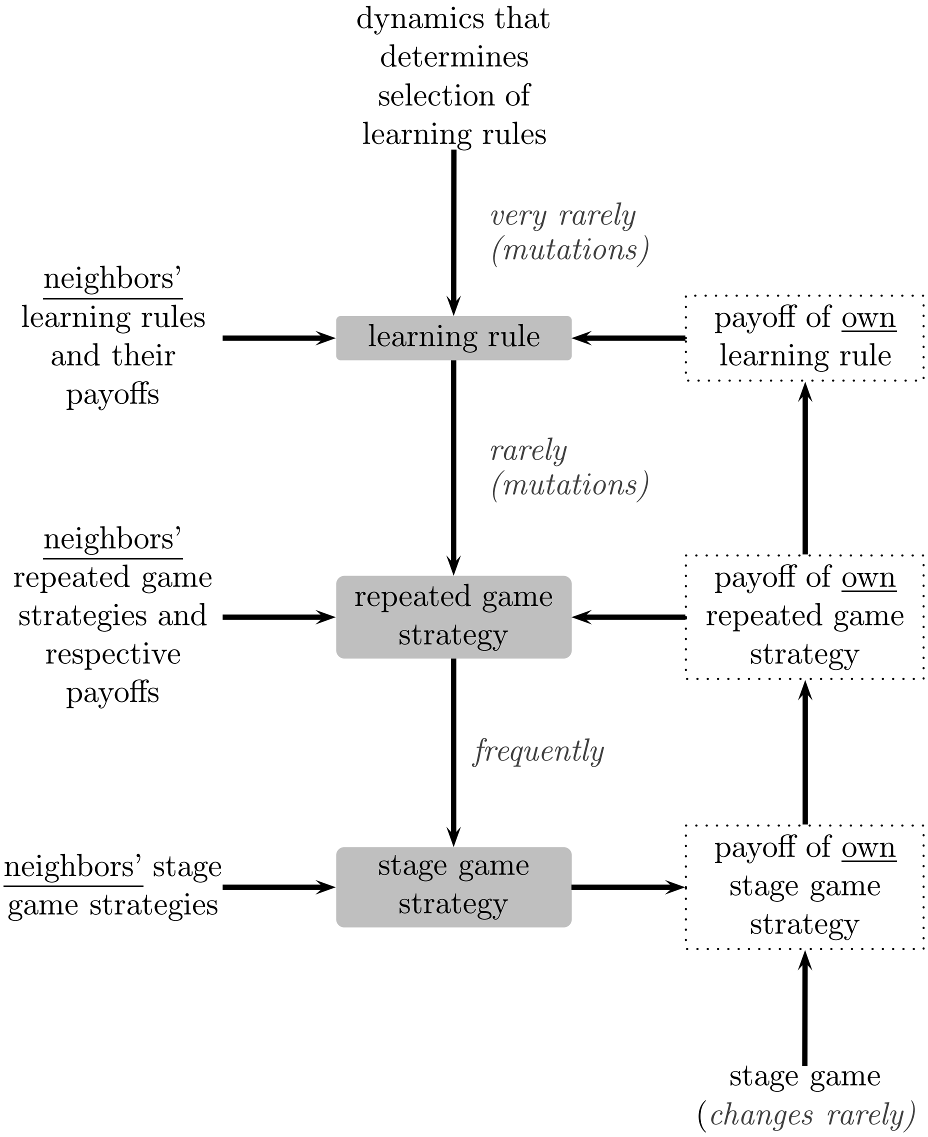


Figure 1: Properties of a player.

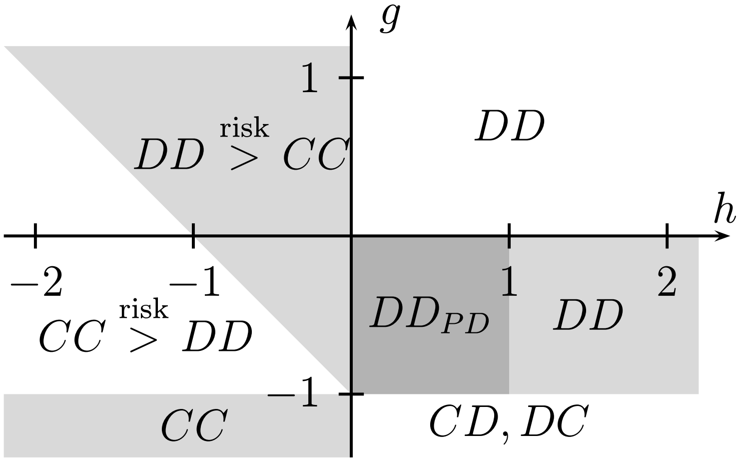
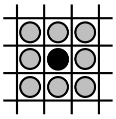


Figure 2: The space of considered games.

2.2 Stage Games

Players play games within a neighborhood. In the simulations that we will discuss in the following such a neighborhood has the following shape:



A player (marked as a black circle) plays against those eight neighbors (gray) which live no more than one cell horizontally or vertically apart. In each period a random draw will decide for each neighbor separately whether an interaction will take place. In the simulations that we will discuss in the following, each possible interaction will take place with probability 1/2. This probability is low enough to avoid synchronization among neighbors, it is still high enough to make simulations sufficiently fast.

We will assume that games change from time to time. In the simulations that we will discuss in the following, every 2000 periods a new game will be selected for the whole population. This game will be a symmetric 2 × 2 game of the following form:

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

(1)

The parameters g and h in the above game will be selected randomly following an equal distribution over the intervals $-1 < g < 1$ and $-2 < h < 2$. We can visualize the space of games in a two-dimensional graph (see figure 2).

The range $-1 < g < 1$ and $-2 < h < 2$ includes both prisoners' dilemmas and coordination games. All games with $g \in (-1, 0)$ and $h \in (0, 1)$ are prisoners' dilemmas (DD_{PD} in figure 2), all games with $g > -1$ and $h < 0$ are coordination games.

In Kirchkamp (1995) we have found that, given the learning rule 'copy best strategy', players cooperate at least in some prisoners' dilemmas. Further they did not always coordinate on the risk dominant equilibrium but followed a criterion which put also some weight on Pareto dominance. In the following we want to analyze whether this behavior persists also with endogenous learning rules.

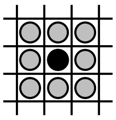
2.3 Repeated Game Strategies

We assume that each player uses a single repeated game strategy against all her neighbors. Repeated game strategies will be chosen (see section 2.4 for details of the choice procedure) from the set of (Moore) automata with less than three states. Table 1 on the next page gives a list of all these automata.

2.4 Learning Rules

From time to time a player has the opportunity to revise her repeated game strategy. In our simulations we will assume that this opportunity is a random event that occurs for each player independently with probability 1/24. Thus, learning will be a rare event, as compared to interaction.

If a player updates her repeated game strategy she samples randomly one member of her neighborhood and then applies her individual learning rule.[3] Neighborhoods contain again the eight immediate neighbors:



The learning rules that we study in the following use the following information:

1. The learning player's repeated game strategy.
2. The payoff u_{own} of the player's repeated game strategy, i.e. the average payoff per interaction that the player received while she used this repeated game strategy.
3. A sampled player's repeated game strategy.

---

[3] Notice, that this learning rule uses information on a *single* sampled player. The learning rules discussed in Axelrod (1984, p. 158ff), Lindgren and Nordahl (1994), Nowak and May (1992, 1993), Nowak, Bonnhoeffer and May (1994), Eshel, Samuelson and Shaked (1996), Kirchkamp (1995) use information on *all* neighbors from the learning neighborhood simultaneously. We assume here that only a single player is sampled to simplify our learning rule in the sense that only a single alternative to the players current repeated game strategy is present. In section 3.3 on page 13 we will see that, at least for cooperation in prisoners' dilemmas and behavior in coordination games, this simplification has almost no effect on the properties. E.g. cooperation in prisoners' dilemmas occurs with the learning rule 'copy best player' for almost the same range of games, regardless whether only one or all neighbors are sampled.

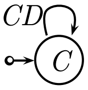
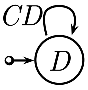
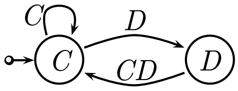
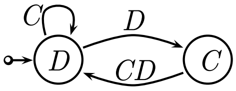
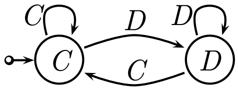
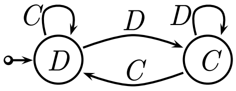
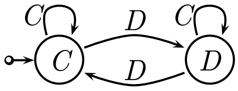
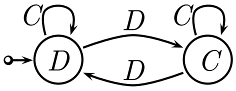
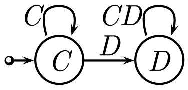
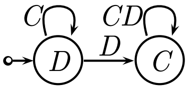
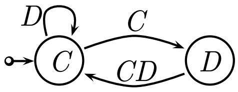
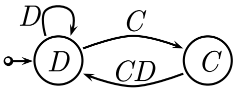
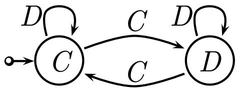
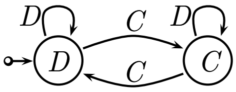
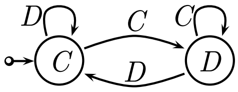
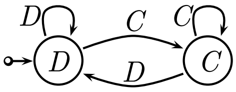
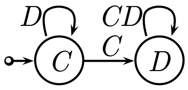
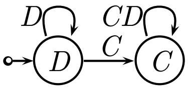
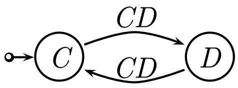
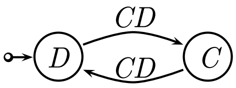
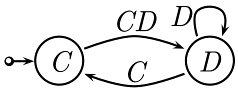
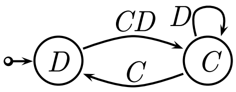
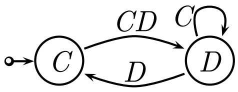
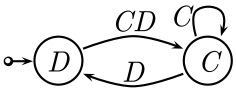
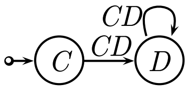
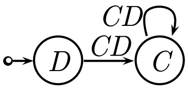
always cooperate		always defect	
forgive quickly			
tit-for-tat			
		tat-for-tit	
grim			
			
reverse tat-for-tit			
		reverse tit-for-tat	
		reverse grim	
blinker		reverse blinker	
			
			
cooperate once, defect forever		defect once, cooperate forever	

Table 1: All 26 automata with less than three states

4. The sampled player's repeated game strategy payoff u_{samp}, i.e. the average payoff per interaction that the player received while she used this repeated game strategy.

Learning rules are characterized by a vector of three parameters $(\hat{a}_0, \hat{a}_1, \hat{a}_2) \in \mathbb{R}^3$. Given a learning rule $(\hat{a}_0, \hat{a}_1, \hat{a}_2)$ a learning player samples one neighbors' strategy and payoff and then switches to her strategy with probability

$$p(u_{\text{own}}, u_{\text{samp}}) = \langle \hat{a}_0 + \hat{a}_1 u_{\text{own}} + \hat{a}_2 u_{\text{samp}} \rangle \tag{2}$$

where

$$\langle x \rangle := \begin{cases} 1 & \text{if } x > 1 \\ 0 & \text{if } x < 0 \\ x & \text{otherwise} \end{cases} . \tag{3}$$

u_{own} and u_{samp} denote the player's and her neighbor's payoff respectively.

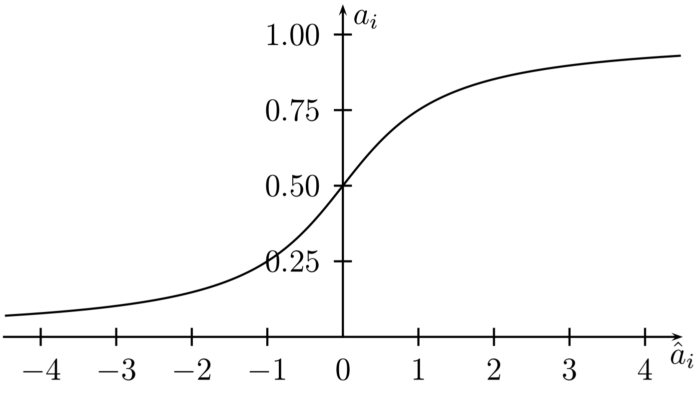
Thus, the two parameters $\hat{a}_1$ and $\hat{a}_2$ reflect sensitivities of the switching probability to changes in the player's and the neighbor's payoff. The parameter $\hat{a}_0$ reflects a general readiness to change to new strategies, which can be interpreted as a higher or lower inclination to make an experiment or to try something new.

Note that with $(\hat{a}_0, \hat{a}_1, \hat{a}_2) \in (-\infty, +\infty)^3$ we can specify both stochastic and deterministic rules. An example for a deterministic rule ('switch if better') is $(\hat{a}_0, \hat{a}_1, \hat{a}_2) := (0, -\bar{a}, \bar{a})$ with $\bar{a} \to \infty$.

Notice also that our parameter $\hat{a}_0$ is similar to the aspiration level Δ from the global model studied in Binmore and Samuelson (1994). However, the learning rules studied in Binmore and Samuelson are not special cases of our learning rules, since their decisions are perturbed by exogenous noise. For cases where this noise term becomes small our rule can approximate the rule of Binmore and Samuelson (1994) with $(\hat{a}_0, \hat{a}_1, \hat{a}_2) := (\Delta, -\bar{a}, \bar{a})$ with $\bar{a} \to \infty$.

Normalization To represent $(\hat{a}_0, \hat{a}_1, \hat{a}_2) \in \mathbb{R}^3$ we will use in the following normalized values $(a_0, a_1, a_2) \in [0, 1]$ such that

$$\hat{a}_i \equiv \tan\left(\pi a_i - \frac{\pi}{2}\right) \qquad \forall i \in \{0, 1, 2\} . \tag{4}$$



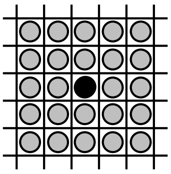
We use normalized values for the following reason: The learning rules from the literature are often deterministic, which means that they can be represented as extreme points of our parametrization. If we want to allow for deterministic rules, we have to allow for values of $\hat{a}_1$ and $\hat{a}_2$ that tend to $-\infty$ and $+\infty$ respectively. In the following (see section 2.5) we will analyze a dynamics where players try to choose parameters of their learning rule optimally. If the long run learning rule is, as in the literature, a deterministic one, thus optimal parameter values are infinite, optimization becomes a problem. We therefore map the unbounded space of parameters of our learning rules into a bounded space.

Mutations When a player learns a repeated game strategy, with probability $\frac{1}{10}$ learning fails and she learns a random strategy. In this case, any repeated game strategy, as described in section 2.3, is selected with equal probability.

2.5 Exogenous Dynamics that Select Learning Rules

From time to time a player also has the opportunity to revise her learning rule. In our simulations we will assume that this opportunity is a random event that occurs for each player independently with probability 1/4000.

Thus, such an update occurs very rarely. We find it justified that for these rare events players make a larger effort to select a new learning rule. If a player updates her learning rule she samples all members of a neighborhood. Here we assume that this neighborhood is larger than the one used for interaction and learning of repeated game strategies. Players learn learning rules from neighbors which are no more than two cells horizontally or vertically apart:[4]



For all their neighbors and for themselves they have the following information:

1. The normalized parameters of the respective learning rule a_0, a_1, a_2.
2. The average payoff per interaction that the respective player received while this learning rule was used, $u(a_0, a_1, a_2)$.

Figure 3 shows (only for one dimension) an example for a sample of several pairs of a parameter a_i and a payoff u (black dots) together with the respective estimation of the functional relationship (gray line) between a_i and u.

Since in our model learning rules are updated very rarely, we want to implement a learning rule which makes some effort to evaluate available information efficiently. We assume that players make an estimation of a model that helps them explaining

---

[4] Results do not change substantially if this neighborhood is smaller or larger.

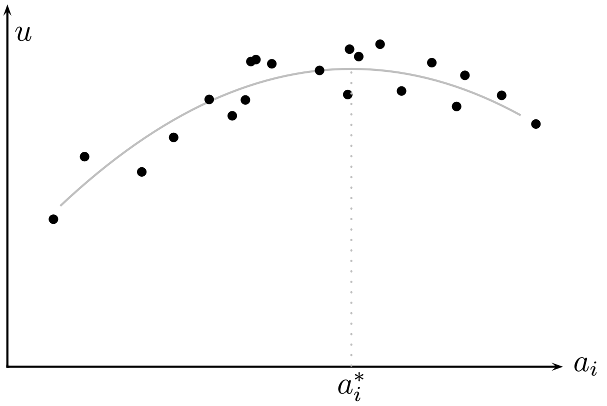


Figure 3: An example for samples of pairs of parameters and payoffs (black) which are used to estimate a functional relationship (gray) between a_i and u. Given this relationship an optimal value a_i^* is determined.

their environment, in particular their payoffs. Players use such a model to choose an optimal learning rule.

To model such a decision process we assume that players use a quadratic function of the learning parameters to explain success of a learning rule. Formally the quadratic function can be written as follows:

$$u(a_0, a_1, a_2) = c + (a_0, a_1, a_2) \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} + \\ +(a_0, a_1, a_2) \begin{pmatrix} q_{00} & q_{01} & q_{02} \\ q_{01} & q_{11} & q_{12} \\ q_{02} & q_{12} & q_{22} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} + \epsilon \qquad (5)$$

Players make an OLS-estimation to derive the parameters of this model (ϵ describes the noise).[5]

The OLS-Regression determines the parameters $(c, b_0, b_1, b_2, q_{00}, q_{01}, q_{02}, q_{11}, q_{12}, q_{22})$ of the above model. Given this model, the player can determine the combination of a_0, a_1, a_2 that maximizes $u(a_0, a_1, a_2)$. Of course, such a maximum need not exist. For our simulations we find that a unique interior maximum can be found for more than 99% of all updates. In the rare case where we find no unique interior maximum, the most successful neighbor is copied.

Mutations We will also introduce mutations for players' learning rules. When a player updates her learning rule, with probability $\frac{1}{10}$ learning fails and she learns

---

[5] We have chosen a quadratic function because it is one of the simplest models which still has an optimum. Similarly we assume that players derive this model using an OLS-estimation because this is a simple and canonical way of aggregating the information players have. We do not want to be taken too literally: We want to model players that more or less behave *as if* they would maximize a quadratic model which is derived using an OLS-estimation.

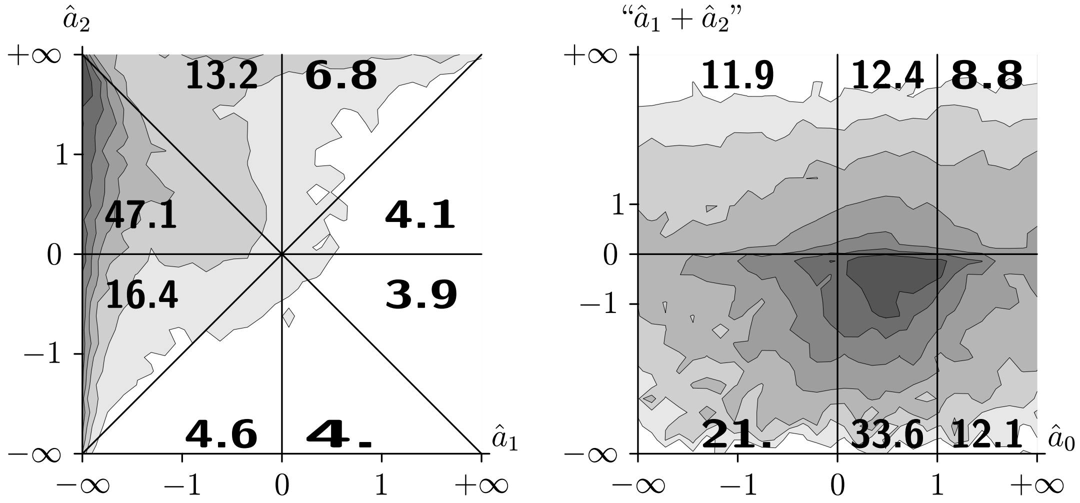


Figure 4: Long run distribution over parameters of the learning rule (a_0, a_1, a_2). Average over 188 simulations runs, each lasting for 400 000 periods. Relative frequencies are given as percentages.

a random learning rule that is chosen following an equal distribution (for the normalized parameters) over $(a_0, a_1, a_2) \in [0, 1]^3$, which is equivalent to a random and independent selection of $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$ following each a Cauchy distribution.

2.6 Initial Configuration

At the beginning of each simulation each player starts with a random learning rule that is chosen following an equal distribution over $(a_0, a_1, a_2) \in [0, 1]^3$. Thus the parameters $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$ are distributed independently following a Cauchy distribution. Also each player starts with a random repeated game strategy, again following an equal distribution over the available strategies.

3 Results with Endogenous Learning Rules

3.1 Distribution over Learning Parameters

We have run several simulations on a 50×50 grid, lasting 400 000 periods each. Figure figure 4 displays averages over several simulations.

Since we can not display a distribution over the three-dimensional space (a_0, a_1, a_2) we will analyze two different projections into subspaces. The left part of figure 4 displays the distribution over (a_1, a_2), the right part over $(a_0, a_1 + a_2)$ respectively. Axes range from 0 to 1 for a_0, a_1 and a_2 and from 0 to 2 in the case of $a_1 + a_2$. Labels on the axes do not represent the normalized values but instead $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$ which range from $-\infty$ to $+\infty$.[6]

---

[6] The figure was derived from a table of frequencies with 30×30 cells. The scaling of all axes follows the normalization given in equation 4 on page 7. To be precise, the value "$\hat{a}_1 + \hat{a}_2$" represents

Both pictures are simultaneously a density plot and a table of relative frequencies:

Density plot: Different densities of the distribution are represented by different shades of gray. The highest density is represented by the darkest gray.[7]

Table of relative frequencies: The pictures in figure 4 also contains a table of relative frequencies. The left picture is divided into eight sectors, the right picture is divided into six rectangles. The percentages within each sector or rectangle represent the amount of players that use a learning rule with parameters in the respective range.

The left part of figure 4 shows two interesting properties of endogenous evolution: First, with endogenous evolution learning rules are sensitive to a player's own payoff. Second, they are substantially less sensitive to observed payoffs. We call this latter property *suspicion.*

Sensitivity to own payoffs: Remember that the initial distribution over a_1 and a_2 is an equal distribution. Thus, had we drawn the right part of figure 4 in period one, the result would have been a smooth gray surface without any mountains or valleys. Starting from this initial distribution our learning parameters have changed substantially. Even if not in all cases $\hat{a}_1$ became $-\infty$, the distribution over learning parameters puts most its weight on small values of a_1.

Insensitivity to sampled payoffs: In the left part of figure 4 we see that 71.5% of all players use a learning rule with $|\hat{a}_2| < |\hat{a}_1|$, i.e. a learning rule which puts more weight on the player's own payoff than on the sampled payoff.

If we restrict ourselves to 'reasonable' learning rules with $\hat{a}_1 < 0$ and $\hat{a}_2 > 0$ then 76.1% of all these rules have the property that $|\hat{a}_2| < |\hat{a}_1|$.

Notice that in both cases the initial distribution over parameters of the learning rule implies that 50% of all rules fulfill $|\hat{a}_2| < |\hat{a}_1|$.

We call this kind of behavior 'suspicious' in the following sense: A sampling player may realize that an observed learning rule is successful for her neighbor. Nevertheless she does not know whether the same rule is equally successful at her own location. Perhaps the success of her neighbor's rule is related to the particular behavior of players which are neighbors of her neighbor, but not her own. Thus, our player might fear that the sampled neighbor's experience can not be generalized for her own case.

---

actually $\tan(\pi \cdot (\hat{a}_1 + \hat{a}_2) - \pi/2)$ and not $\hat{a}_1 + \hat{a}_2$. In the current context this difference should be negligible.

[7] Densities are derived from a table of frequencies with a grid of size 30×30 for each picture. We actually map logs of densities into different shades of gray. The interval between the log of the highest density and the log of 1% of the highest density is split into seven ranges of even width. Densities with logs in the same interval have the same shade of gray. Thus, the white area represents densities smaller than 1% of the maximal density while areas with darker shades of gray represent densities larger than 1.9%, 3.7% 7.2%, 14%, 27% and 52% of the maximal density respectively.

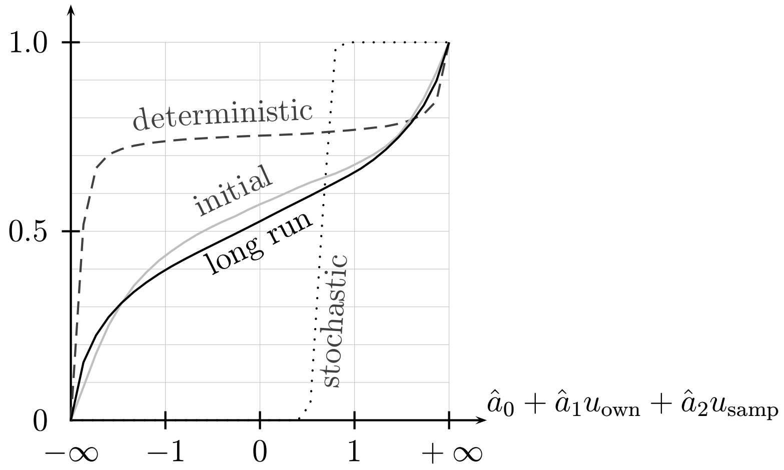


Figure 5: Long run distribution over switching probabilities

Our aim is to compare properties of endogenous learning rules with properties of those studied in parts of the literature on local evolution.

In this section we have seen that endogenous learning rules may be *similar* to the above fixed learning rules in the sense that small changes in the player's own payoff may lead to drastic changes in the probability to adopt a new strategy.

Endogenous learning rules *differ* from those studied in parts of the literature on local evolution in the sense that changes in an observed player's payoff only lead to small changes in the probability to adopt a new strategy.

In the next section we want to investigate how these properties are reflected in actual switching probabilities.

3.2 Probabilities to Switch to a Sampled Learning Rule

When a player follows her learning rule as specified in equation 2 on page 7 she determines a probability to switch to the observed repeated game strategy. Figure 5 shows the distribution of these switching probabilities.[8] In addition to the distribution of switching probabilities in the *long run*, figure 5 also shows three reference distributions:

initial: The distribution of switching probabilities, given the initial distribution over learning rules.

stochastic: The distribution of switching probabilities, given a fixed learning rule $(\hat{a}_0, \hat{a}_1, \hat{a}_2) = (1/2, -1/8, 1/8)$.

deterministic: The distribution of switching probabilities, given a fixed learning rule $(\hat{a}_0, \hat{a}_1, \hat{a}_2) = (1/2, -100, 100)$.

---

[8] This figure was derived from a table of frequencies with 30 cells. The scaling of the horizontal axis follows the normalization given in equation 4 on page 7.

The horizontal axis represents $\hat{a}_0 + \hat{a}_1 u_{\text{own}} + \hat{a}_2 u_{\text{samp}}$. Following the learning rule 2 a player will switch *stochastically* with probability $\hat{a}_0 + \hat{a}_1 u_{\text{own}} + \hat{a}_2 u_{\text{samp}}$ if this expression is between zero and one. Otherwise the player will either switch *with certainty* or not at all. Figure 5 shows that in the long run more than 50% of all learning events a player will not switch at all and in more than 30% of all events she will switch with certainty. Only in about 12% of all learning events her decision will be a stochastic one. Thus, we might be tempted to describe switching behavior in the long run as mainly deterministic. But we have to be careful: Comparison with the other three distributions leads to the conclusion that this degree of deterministic behavior is already present in the initial distribution. Switching probabilities of endogenous learning rules still seem to be significantly different those of deterministic rules.

In the previous two sections we have found that some properties of endogenous learning rules are different from those of fixed deterministic rules. Endogenous learning rules are less sensitive to changes in a sampled player's payoff and switching is stochastic at least sometimes.

In the next section we will study whether these differences carry over to stage game behavior.

3.3 Stage Game Behavior

Figure 6 on the next page shows proportions of stage game strategies for various games both for endogenous and for fixed learning rules. Remember that in our simulations the underlying game changes every 2000 periods. Just before the game changes we determine the proportion of stage game strategies C and D. These proportions are represented in figure 6 as circles. The position of the circle is determined by the parameters of the game, g and h. The size of the circle is proportional to the proportion of C or D, whichever is the larger. The color of the circle is white if the proportion of Cs is larger and black otherwise.

Figure 6 compares two cases: An exogenously given learning rules of the 'switch if better' type, approximated as $(\hat{a}_0, \hat{a}_1, \hat{a}_2) = (0, -100\,000, 100\,000)$ and the case of endogenous learning rules.

In both pictures two areas can be distinguished. One area where most of the simulations lead to a majority of C and another one where most simulations lead to a majority of D. We make two observations:

- The fixed learning rule 'switch if better', which is only an approximation of one of the learning rules studied in Kirchkamp (1995), leads to very similar results.
  - There is cooperation for a wide range of prisoners' dilemmas.
  - In coordination games players do not follow the principle of risk dominance but another principle which is between risk dominance and Pareto dominance.

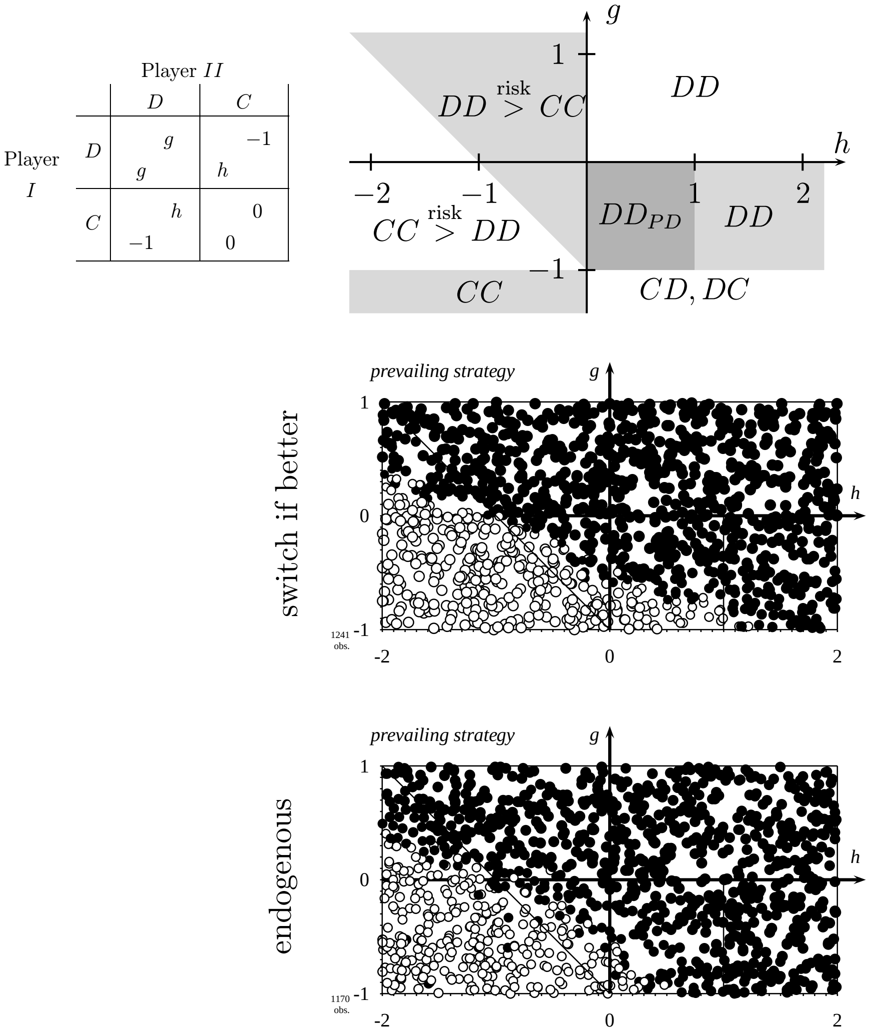


Figure 6: Stage game behavior depending on the game. (○=most players play C, ●=most players play D).

- Endogenous learning rules lead to a stage game behavior which is similar to the one achieved with the fixed rule ‘switch if better’. There is still some cooperation for prisoners’ dilemmas (however, less cooperation than with the ‘switch if better’ rule) and behavior in coordination games does not follow risk dominance.

The first point is interesting to note, because it shows that the model that we have studied in this paper is comparable with the model analyzed in Kirchkamp (1995) at all.

The second point shows that properties of network evolution discussed in Kirchkamp (1995) do not depend on the choice of the particular learning rule. They persist even with endogenous evolution of learning rules. Since Axelrod (1984, p. 158ff), Lindgren and Nordahl (1994), Nowak and May (1992, 1993), Nowak, Bonnhoeffer and May (1994), Eshel, Samuelson and Shaked (1996) all use learning rules which are exogenously given and which are similar to ‘switch if better’ we can hope that part of their results still hold, even if we allow players to choose their own learning rules — at least as long as the set of learning rules they may choose from is as limited as in our above discussion.

4 Conclusions

In the previous two sections we have followed two questions.

The first one was whether learning rules, which are used in Axelrod (1984, p. 158ff), Lindgren and Nordahl (1994), Nowak and May (1992, 1993), Nowak, Bonnhoeffer and May (1994), Eshel, Samuelson and Shaked (1996), in Kirchkamp (1995) can be justified as the outcome of an evolutionary selection dynamics.

We have found that the dynamics that we have analyzed here selects rules which are different from the ones commonly assumed in the literature. In particular the learning rules which are selected following our dynamics are much less sensitive to changes in a sampled player's payoff.

The second question we asked was whether endogenous learning rules lead at least to a similar stage game behavior. Here we have found that indeed important properties of stage game behavior, like cooperation for some prisoners' dilemmas and coordination not on risk dominant equilibria, is present both with our endogenous learning rules and with fixed learning rules specified above and in the literature.

Besides the selection dynamics that we have presented above we have also analyzed other selection dynamics. In Kirchkamp and Schlag (1995) we study dynamics where players use less sophisticated update rules than the OLS-model used in this paper. We have analyzed models where players move only in the direction of the maximum of the OLS model, but do not adopt the estimate of the optimal rule immediately. Further we have analyzed models where players do not estimate any model at all but instead copy successful neighbors. Both alternative specifications lead to similar properties of learning rules: Switching probabilities are less sensitive to changes in payoff of the neighbor and more sensitive to changes in payoffs of the player herself. Also properties of the induced stage game behavior are similar: Both alternative specifications lead to cooperation for some prisoners' dilemmas and coordination not on risk dominant equilibria. Thus, we can regard the above results as fairly robust.

References

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

Binmore, K., and L. Samuelson (1994): "Muddling Through: Noisy Equilibrium Selection," Discussion Paper B–275, SFB 303, Rheinische Friedrich Wilhelms Universität Bonn.

Eshel, I., L. Samuelson, and A. Shaked (1996): "Altruists, Egoists and Hooligans in a Local Interaction Model," Tel Aviv University and University of Bonn.

Kirchkamp, O. (1995): "Spatial Evolution of Automata in the Prisoners' Dilemma," Discussion Paper B–330, 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.

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.