

# Little information, efficiency and learning – An experimental study

Atanasios Mitropoulos

Faculty of Economics and Management  
Otto-von-Guericke-University Magdeburg  
PO-box 4120, 39016 Magdeburg, Germany  
e-mail: [atanasios.mitropoulos@ww.uni-magdeburg.de](mailto:atanasios.mitropoulos@ww.uni-magdeburg.de)  
<http://www.uni-magdeburg.de/vwl3/tony/tony.html>

April, 2001

version 1.0

## **Abstract:**

Earlier experiments have shown that under little information subjects are hardly able to coordinate even though there are no conflicting interests and subjects are organised in fixed pairs. This is so, even though a simple adjustment process would lead the subjects into the efficient, fair and individually payoff maximising outcome. We draw on this finding and design an experiment in which subjects repeatedly play 4 simple games within 4 sets of 40 rounds under little information. This way we are able to investigate (i) the coordination abilities of the subjects depending on the underlying game, (ii) the resulting efficiency loss, and (iii) the adjustment of the learning rule.

**Keywords:** mutual fate control, matching pennies, fate-control behaviour-control, learning, coordination, little information

**JEL classification:** C72, C91, D83

I am grateful for financial support by the Deutsche Forschungs Gemeinschaft. I also have to thank seminar participants at the Universities of Magdeburg and Karlsruhe as well as conference participants at the IAREP/SABE meeting in Baden ,the VfS meeting in Berlin, the ESA meeting in Amsterdam, and the GEW meeting in Warberg. I am also grateful to Oliver Kirchkamp, Stefan Napel, Jason Shachat and Georg Weizsäcker for intense discussions, Tilman Slembeck for pointing me to the relevant psychological literature and Alex Alexandersson for helping me with the drawing of the confidence ellipses.

## **1. Introduction**

There are countless experiments on games in which subjects repeatedly play the same game. The vast majority of them involve complete information about the game. These experiments can roughly be further classified into two strands, those that involve a fixed group of players that have to solve a difficult coordination task<sup>1</sup> and those in which players repeatedly play a game against a randomly determined ever changing opponent<sup>2,3</sup>. Both strands of research focus on the investigation of player's adaptation to updated beliefs about opponent's play. The analysis is directed towards the detection of optimal or sub-optimal play among subjects. In order to avoid conflict with issues relating to the search for information these game environments assume away any lack of knowledge about the setting of the game. Especially, subjects are typically provided with all details about the sequence of decision tasks and possible pay-offs. While this approach certainly has the virtue of eliciting adaptive behaviour resulting from adjusted beliefs about other subjects' likely play, it neglects the importance of the informational assumption. Real-world situations are often characterised by uncertainty about the strategic situation the players are faced with. In the computer industry, for example, products that were supposed to compete against each other ended up being used complementary (e.g. computers and mobile phone technology). Others suddenly turned out to be heavy competitors (browsers and office packages). It is therefore an important task to elicit the effect of the informational assumptions on behaviour.

This study is concerned with the analysis of players' behaviour within an environment that is characterised by little information about the underlying strategic relationship among players. Particularly, at the beginning of a repeated task, players do not know whether they have common or opposing interests. In such a setting players are forced to gather and process information during the course of repeated play so that a satisfactory payoff is achieved. However, players are never explicitly told which game they actually played or which they are going to play.

Games which under rich informational conditions produce interesting decision problems typically degenerate to tasks almost impossible to solve when information about the strategic relationship is foreclosed. For example, without knowledge about being involved in a game with multiple Pareto-ranked Nash-equilibria it is hardly conceivable that players reach efficiency

---

<sup>1</sup> Examples for this category are almost all public goods games (see e.g. Ledyard 1995 for an overview) and coordination games (e.g. Van Huyck et al. 1997).

<sup>2</sup> Most bargaining games are played this way (see e.g. Roth 1995 and Güth 1995 for overviews).

<sup>3</sup> There are, however, a number of experimental settings that fall in between the two classes, e.g. the evolutionary games of which Van Huyck (1997) gives an overview.

within a reasonable number of repetitions. Likewise, bargaining games degenerate to “guessing in the dark” if subjects are not provided with any information about the payoff the opponent receives and if additionally the opponent changes from one repetition to the next. For this reason games involving little information have to be somehow “simpler” in nature if one aims at generating behaviour that is driven by the subjects’ hope to achieve a certain goal<sup>4</sup>.

While uncertainty has been studied extensively in single-person decision tasks<sup>5</sup>, there are only few studies dealing with games (of two or more players) involving little information. By little information we mean that neither the payoff matrix, nor its structure, nor feedback about the opponent’s actions and payoffs are provided<sup>6</sup>.

Very recently there have some attempts been made to experimentally investigate behaviour under uncertainty in addition to the lack of knowledge of opponents’ behaviour. This has mainly been done in experiments in which subjects were not given the full description of the game they were going to play. The results from these experiments have been mixed. They show that we are far from understanding what the driving forces of knowledge acquisition are.

In a symmetric Matching Pennies Game, which is characterised by a unique and symmetric mixed strategy Nash equilibrium, Mookherjee and Sopher (1994) show that behaviour under little information shows significantly more and persistent inertia. This result suggests that subjects fail to derive complete knowledge of the game structure by simply playing the game.

The results of Nagel and Vriend (1997) are similar in that subjects obviously do not accord to complex dynamics in order to find the (in their experiment) complicated structural implications. They rather accord to the simple learning direction theory (Selten and Stoecker 1986), which gives them a first approximation of the optimal solution. A similar finding is reported by Huck et al. (1999) for a repeated 4-player oligopoly game under little information. A variant of learning direction theory is found to predict individual behaviour best and outperforms adaptation rules that rely on more information even if this information is actually given.

The sequentially played Ultimatum Game has also been studied under uncertainty. Slembeck (1999) has conducted an experiment under various conditions and has found that behaviour of responders as well as proposers significantly differs depending on how much information responders get. In particular, responders having information about proposers’ payoffs were

---

<sup>4</sup> By restricting the set of possible payoffs to  $\{0,1\}$  we limit the goal to be the maximisation of the own payoff. This argument is explicated in more detail in section 2.

<sup>5</sup> For an overview, see e.g. Camerer (1995).

<sup>6</sup> From a game-theoretic point of view the ranking of payoff entries already contains most of the information necessary to calculate the Nash-equilibrium. A weaker, but still “more-than-little” information is to provide the best-response structure of the game. Interesting results on the ability of subjects to extract the best-response

less inclined to accept offers in later periods than those responders who knew only about their own payoff. Lack of information was also found to cause a significant and lasting loss of efficiency.

Trials to relate behaviour of players under little information to recently developed adaptive rules can be found in Erev and Roth (1998) and Chen and Khoroshilov (2001). While the former find the reinforcement learning rule which was developed by Roth and Erev (1995) to track the data best in several probabilistic two-person games with unique mixed-strategy equilibrium, the latter find the same scheme being outperformed by variants of Camerer and Ho's (1999) and Sarin and Vahid's (1999) adaptation schemes. For one game, however, Chen and Khoroshilov complain about obtaining bad fit for all models.

A similar dissatisfying result can be found in Mitropoulos (forthcoming) in an experiment on *mutual fate control*. He found that learning behaviour is far from efficient. Subjects hardly accord to simple learning dynamics. While the Roth-and-Erev scheme is found to outperform a probabilistic variant of win-stay lose-change, qualitative characteristics of the data are not captured by any learning rule. The present study extends this analysis by giving the same amount of information by way of a probability distribution over games. We, hence, control for the beliefs of subjects on the set of possible games and are able to analyse behaviour on a set of four different games. We further use a within-subject replication so that the effect of experience with a learning task can be studied.

For most of the previous studies a simple learning rule – if applied by all players simultaneously – would have led players to an optimal solution. Moreover, since the ability of players to coordinate plays an important role in achieving a near optimal payoff, group sizes are chosen rather small (two to four players). Our study remains within this general framework. However, we depart from previous studies in that we do not simply foreclose information on the underlying payoff scheme but provide our subjects with the complete set of possible games<sup>7,8</sup>.

Our study also differs from most other approaches in the methods used to study behaviour. We focus on the identification of structural characteristics within subject's play and refrain from ranking theories on purely quantitative grounds, such as the mean squared deviation or

---

structure when repeatedly playing a game under information conditions similar to ours can be found in Oechssler and Schipper (2000).

<sup>7</sup> Mitropoulos (forthcoming) argues that the inability of subjects to figure out the whole set of possible games may, for a rather technical reason, considerably complicate coordination tasks.

<sup>8</sup> Independently, Shachat and Walker (2000) conducted an experiment in which they, too, provide subjects with the whole set of possible games. However, they show subjects their own payoff matrix and provide a distribution over all possible payoff matrices for their opponents only. Similar, but still different is Feltovich (2000).

log-likelihood of predictions. We will use estimation results only for the purpose of tracking behavioural changes as experience with the task rises.

Our main focus will rest on the coordination task subjects have to solve within three of the four possible games. In section 2 we discuss the set of possible games, and show that a simple learning rule (i.e. win-stay lose-randomise) is able to achieve almost complete efficiency. We take this result as our benchmark. Section 3 is devoted to presenting the experimental design while section 4 contains the general data analysis. As it turns out, compared to the benchmark, subjects in our experiments suffer a considerable amount of loss of efficiency. They earn up to 30 percent less than the benchmark payoff. We find the main determinant to be whether subjects succeed in coordinating or not. Since we allowed subjects to repeat the 40-round repeated game with randomly assigned opponents and games, we further investigate for an experience effect and, indeed, we find such an effect: As subjects repeat the learning task they obviously succeed in coordinating more quickly, even though the number of successfully coordinating pairs does not increase. In search for the reason for the improvement in speed of coordination we analyse in section 5 nine different learning rules that have been proposed in the recent, as well as the more distant past. The characteristics of all rules are shown by way of expected motion as well as simulations. We then compare the dynamics with the actual data. We further fit the learning models to individual as well as aggregated data. The discouraging result from our efforts is that none of the rules is able to capture any effect that might be due to experience. Finally in section 6, we show one reason for the failure of all rules to uncover important determinants of behaviour: while adaptive rules always focus on round-by-round adaptation, we show that subjects make frequent use of multi-round patterns. We suspect that such behaviour is driven by a disposition to explore the environment before taking goal-specific action. We conclude in section 7.

## **2. The games**

Consider the situation of subjects not being told which repeated game they are actually going to play. A Bayesian approach would be to first collect the given constraints and subsequently to derive the complete set of possible games. Having determined this set, one is able to construct one's own optimal strategy much in the same way as should be done under complete knowledge of the game, the only difference lying in the consideration of a distribution of games instead of only one game.

A typical situation of uncertainty, however, is that of a subject not having a clue as to what the underlying game is. The game is only revealed gradually by actually playing the game

repeatedly. Usually, such a situation is implemented by way of simply withholding the actual payoff matrix. The drawback of this approach is that beliefs of subjects about the set of possible games may vary and may substantially influence behaviour. Mitropoulos (forthcoming) reports that subjects made various guesses about the actual payoff matrix which rarely coincided with the true payoff scheme applied in that study.

We, therefore, choose to give subjects a complete characterisation of the distribution of possible games while keeping the situation as close as possible to that of uncertainty about the underlying payoff matrix. Our instructions are almost equivalent to saying that the payoff matrix may be any, the only restriction being that payoffs are either 0 or 1. More specifically, the support of the distribution over games consists of all those 2-player, 2-action normal-form games that contain only the values 0 or 1 and do not involve a weakly dominated strategy for either player. What we are left with is a set of twelve different payoff matrices. Due to some matrices being structurally the same as others, the sole difference lying in the permutation of names of actions, we are left with four structurally different payoff matrices, which we subsequently call stage games and which we now describe in more detail.

Table 1 shows Mutual Fate Control (henceforth MFC; for a detailed description of this game, see Rabinowitz et al. 1966). The game represents the situation in which each player’s action determines the payoff of the opponent. Sidowski et al. (1956) emphasise that this is the simplest game involving non-trivial actions and payoffs, and therefore call it the *minimal social situation*.

		player 2	
		A	B
player1	A	0 0	1 0
	B	0 1	1 1

**Table 1: Mutual Fate Control**

Results from earlier work on this game under little information show that subjects had considerable difficulty in coordinating on the efficient cell. In almost all cases of successful coordination pairs either coordinated during the first rounds of play or failed to do so until the very end (of 100 rounds). This was so, despite the fact that the simple learning rule win-stay lose-change would quickly have guided them towards efficiency. However, the analysis was incomplete since the set of possible games had not been controlled for.

		player 2	
		A	B
player1	A	1 0	0 0
	B	0 1	1 1

**Table 2: Fate-Control Behaviour Control**

Fate-control behaviour-control FCBC (for a more detailed description, see Arickx and van Avermaet 1981) is very similar to MFC. Again the actions of both players determine the payoff of the opponent. In FCBC, however, the action with which one player (in table 2 it is player 2) gives the payoff to the opponent is determined by the action of the opponent. So, one player has some influence on his own payoff while the other one has not.

As similar as MFC and FCBC look from the point of view of the stage games, as different they are from the dynamical point of view. The simple rule of win-stay lose-change illustrates this point. Given that both players fully accord to the win-stay lose-change rule, in MFC the pair will converge to the efficient cell within two rounds and stay there forever, no matter where the starting point lies. This is different in FCBC. After starting in the efficient cell, players will stay with their actions forever. If starting in any other of the three remaining cells players will be involved in a continuous cycle. In order to avoid the inefficient cycling a substitution of the lose-change part, for example into a lose-randomise behaviour would be advisable. We will return to this point later in the text when studying efficiency and the adaptation behaviour of the subjects.

		player 2	
		A	B
player1	A	1 1	0 0
	B	0 0	1 1

**Table 3: Coordination**

In the coordination game CO players have simply to coordinate on one or the other combination of actions in order to yield the efficient outcome. In case they do not succeed to coordinate they both get a payoff of zero. To our knowledge this CO game has never been played before. In fact, this game looks more or less trivial. We chose to include it into our set of games because in the debriefing of our earlier experiment (Mitropoulos forthcoming) some of the participants clearly indicated that they assumed this game to have been the underlying

game. Later we will see that a non-trivial number of pairs did not reach coordination until the end, which suggests that, given little information, CO is not as trivial a game as it seems.

		player 2	
		A	B
player1	A	1 0	0 1
	B	0 1	1 0

**Table 4: Matching Pennies**

The matching-pennies game MP has earlier been studied by Mookherjee and Sopher (1994) and Ochs (1995). It is the constant-sum game in which each action allows for both winning and losing. Game Theory implies that subjects will end up playing each strategy with equal probability. However, players have no positive incentive to do so; instead, giving equal probability to both strategies is the sole mixed strategy that impedes exploitation<sup>9</sup>. Because of this, it is far from obvious that subjects will accord to the Nash-equilibrium. The analysis by Mookherjee and Sopher (1994) showed that subjects approximate the Nash equilibrium fairly closely when being informed about the game, but do not conform to stochastic play when left ignorant about the payoff matrix. Ochs (1995) further investigated play under complete information and found that even when knowing the payoff matrix and when being allowed to issue mixed-strategies, the accordance to the Nash equilibrium in the MP game is mainly due to the extreme symmetry of the game.

The distribution over games placed equal probability on each of the four games. Given the game, the distribution placed equal probability on each of the payoff matrices belonging to each game. Subjects were neither informed of which player position (player 1 or player 2) they were assigned. This was again determined via a random draw giving equal probability to each position.

**3. Experimental procedure**

We conducted the experiment at the MaXLab (the Magdeburg Experimental Laboratory) in a fully computerised laboratory. In each session 10 subjects – most of them first-year business and economics students – were randomly allocated to computer terminals, which were separated by mobile cardboard devices. A custom-made computer program, written in Java, helped to make decisions conveniently and to view all information gathered during play. In total, there were 120 subjects participating in the experiments.

---

<sup>9</sup> This point has extensively been discussed by Fudenberg and Levine (1993).

At the beginning of each session, subjects were handed out the instructions on paper. The experimenter (always being the same person) read aloud the instructions. Subjects were then given the opportunity to privately ask questions to the experimenter. Most questions could be answered by simply repeating the corresponding sentences from the instructions. After that, trial rounds were played in which four rounds of each of the four games were played against a computer opponent. Prior to playing the trial rounds, subjects were informed of what the computer was exactly going to play and that these decisions were not going to be paid. They were explicitly told that the trial rounds were intended to make subjects familiar with the computer program, the decision task and the set of possible games. After finishing the trial rounds subjects again were given the opportunity to ask questions. This part of the experiment took 30 to 45 minutes.

By internal computer assignment subjects were then randomly matched to pairs and payoff matrices. Subjects played 40 repetitions of the stage game. They were neither told which of the possible payoff matrices nor which of the possible games they actually played. They were also left ignorant about their player position. In order to leave the focus on payoff maximisation and to avoid subjects playing for the revelation of the underlying game we decided not to tell subjects which game they had been playing even after the games were ended. The only feedback they got was their own payoff of the last round stated in lab dollars. On the screen, a separate window stored all information they got until that round (round number, action, payoff), so subjects did not need to take down any notes. The same procedure was repeated four times. We term these repetitions of the forty-round repeated game as *parts*. Hence, games and opponents were assigned randomly to the subjects in each part. Pairs were matched so that no subject played with the same opponent more than once. Subjects were informed of all this in the instructions.

At the end of the experiment the experimenter separately and privately paid off subjects. Each lab dollar was converted into 0.30 German Marks. A show-up fee of 5 German Marks was added. Money earnings varied between 25 German Marks and 50 German Marks with the average lying at 38 German Marks, while sessions lasted for 90 to 120 minutes. At the time the experiments were run students' wages were approximately 14 German Marks per hour.

In the instructions, subjects were presented all relevant payoff matrices, i.e. all four payoff matrices as described above as well as all payoff matrices with all possible permutations of actions. They were explicitly told that each of the games could occur with equal probability and each of the permutations of a game again could occur with equal probability. We restricted the random assignment only in the way that in each part each game was played by

exactly 15 pairs. In order to make the games more comprehensible we provided a brief verbal description of the games. We took great care to design the description as neutrally as possible.

## 4. Data analysis

### 4.1 Coordination

We start the analysis with the coordination rate within games and parts. The large data set does not allow for a single simple definition to capture all the details of the individual coordination dynamics. There were some pairs that succeeded in coordinating very late in the part, while others coordinated rather soon and exhibited some deviations from the efficient cell towards the end. In order to capture the former as well as the latter we decided to use a step-wise criterion (requiring an appearance rate of roughly 80%) to discriminate between coordinating and non-coordinating pairs:

*Definition 1:*

A pair coordinated on the efficient cell if (i) within the last 7 rounds at least 6 rounds, or (ii) within the last 10 rounds at least 8 rounds, or (iii) within the last 20 rounds at least 16 rounds resulted in the efficient cell.

Implicitly, we exclude consideration of the MP game in which the term “coordination” could be interpreted as joint play of the Nash-equilibrium using mixed strategies giving equal probability to both actions. For the moment, we also do not take into account those pairs that fulfilled the above definition for a different than the efficient cell. We only mention that there were two such pairs (2.2.4 and 12.4.4) that both were FCBC games wherein player 1 for a very long sequence continuously played action A while player 2 showed some variable behaviour.

We now turn to the analysis of the coordination success of the pairs between games and parts. The following Table 5 provides an overview of coordination frequencies, whereby each inner cell contains a set of 15 observations.

	<u>part</u>				
	1	2	3	4	sum
MFC	6	9	9	6	30
FCBC	4	7	12	8	31
CO	12	15	14	12	53
sum	22	31	35	26	114

**Table 5: Frequencies of coordinating pairs**

We notice that, not surprisingly, the coordination rate for CO is very high. However, there is a non-trivial amount of non-coordinating pairs (7 out of 60). For MFC as well as FCBC we find

that the overall coordination rate is almost exactly 50%. However, the coordination rates for MFC and FCBC are a bit lower for part 1 and so are rather similar to what has been observed in previous experiments with complete uncertainty about the set of possible games (Rabinowitz et al., 1966, Mitropoulos, forthcoming). We summarise this as

*Observation 1:*

As compared to complete uncertainty about the underlying game, the coordination rate of pairs playing MFC or FCBC is roughly the same as when prior to playing the game the probability distribution over games is given.

For both MFC and FCBC we further find a single peaked curve of coordination rates over parts; however, neither a chi-square nor a Kolmogorov-Smirnov test can find a significant (to the 10%-level) difference to a flat distribution (i.e. 7.5 coordinations for each game in each part) for either of the games or even for the sum of both. Somewhat astonishing is that even for CO we observe the single-peaked coordination rate; this, however, is a very weak observation and may very well be due to chance events.

We further checked for the correlation of successes among subjects and did not find any sign of concentration of successes. In more detail, we ran Monte Carlo simulations for the success of subjects, once taking the true rate of coordination for each game and each part as underlying distribution and once taking coordination to be a 50-50 chance event, and compared the resulting simulated distributions of the number of subjects over the number of successes with the according actual distribution via a chi-squared test. The results are displayed in table 6.

	number of coordinations				chi-square	p-value
	0	1	2	3 or 4		
actual observation	33	60	19	8		
expectation 1	35.41	54.03	24.28	6.28	3.4517	0.4853
expectation 2	37.55	51.65	24.44	6.37	4.7247	0.3167

**Table 6: Actual numbers of subjects with the according number of successful coordinations and average numbers from simulations**

The comparison of actual observations with the expected numbers calculated from 10,000 simulations of populations using the empirical frequencies (row titled “expectation 1”) is not significant. Likewise the observed distribution is not significantly different to a distribution calculated from 10,000 simulations of populations using the uniform distribution of success (row titled “expectation 2”). We, hence, cannot reject the hypothesis that subjects’ successes to coordinate were determined with equal probability for all subjects. Note that the above tests are rather strict, since we did not take into account that subjects’ successes were determined within pairs.

For the understanding of how people got to coordinate to the efficient cell it is equally important to have a look at the starting round of coordination. Since for some pairs play turned out to be quite variable we had to fix some rule according to which the first encounter of a large number of coordination outcomes can be uniquely identified. Notice that we separate the general definition of coordination from the time at which coordination starts, because a joint definition appeared to result in a very complicated rule that would have impeded interpretation. For the definition of the starting round of coordination we took the following rule (a formal definition requires some notation, and therefore is given in appendix B).

*Definition 2:*

The starting round of successful coordination of a coordinating pair is determined by the first round of the first sequence of successful coordinations that’s length is neither eventually exceeded by a sequence of non-coordination nor is exceeded by 3 times the length of the directly following sequence of non-coordinations.

We are aware of the fact that this definition does not include sequences in which a large number of coordination outcomes have regularly been interrupted by single non-coordinating outcomes. On the other hand, it does include sequences in which pairs had a sequence of 5 or 6 coordination outcomes early on, then had a long sequence of variable outcomes and finally returned to regularly coordinate. Both types of pairs, however, have rarely been observed. Given our above definition table 7 summarises the data.

	part				all
	1	2	3	4	
MFC	11.83	10.44	9.78	2.83	9.00
FCBC	18.75	13.14	10.00	4.25	10.35
CO	9.00	5.07	4.86	7.92	6.55
all	11.55	8.45	7.89	5.60	8.2

**Table 7: Average starting round of coordination**

Recall that from Table 5 we did not get evidence in favour of the hypothesis that with experience subjects learn to coordinate more *often*. As we see from table 7, however, the starting round of coordination is persistently declining over parts; the only exception being part 4 in CO. A Jonkheere test for samples ordered by parts gives us asymptotic two-tailed p-values of <0.05 for each game separately as well as for all games combined. We may, hence, conclude

*Observation 2:*

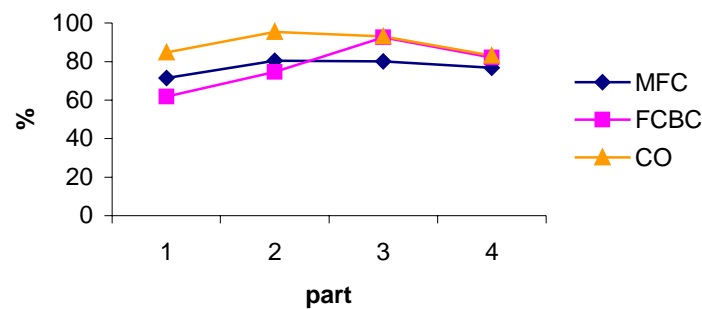
With experience subjects tend to coordinate more *quickly*.

Coordination was clearly easiest in CO. However, contrary to table 5 table 7 suggests that there is a distinction to be made between FCBC and MFC. It looks as though the average

starting round of coordination is lower for MFC. However, a custom-made randomisation test fails to show significance ( $p = 0.27$ ).

## 4.2 Efficiency and payoffs

We now turn our attention to the efficiency issue. We first have to determine what efficiency actually is. Since at the beginning of a game subjects were not aware of how to gain their payoffs and naturally needed time to learn about the incentives of the game, we deem it unjustified to compare actual payoffs with the maximum payoff available. We prefer the benchmark to be defined by the average payoff that is gained when both players apply the most efficient simple adaptive rule, which for our games turns out to be the win-stay lose-randomise dynamics<sup>10</sup>. From this ideal model the benchmark payoffs for the games MFC, FCBC, and CO were found to be 69.8, 68.3, and 78, respectively. This benchmark also functions as a moderator between the games as efficiency requires a higher payoff in those games in which coordination was easier. The average payoffs for pairs over parts are shown in figure 1.



**Figure 1: Efficiency rate over parts based on benchmark payoffs**

We first notice that there are considerable inefficiencies in all three games. As expected, efficiency is generally highest in CO, lying between 83% (part 4) and 96% (part 2). However, due to the higher benchmark the difference to the other games is not very pronounced. We further notice that

*Observation 3:*

The expected increase in efficiency over parts can only be observed for FCBC (Jonkheere-test,  $p < 0.01$ ).

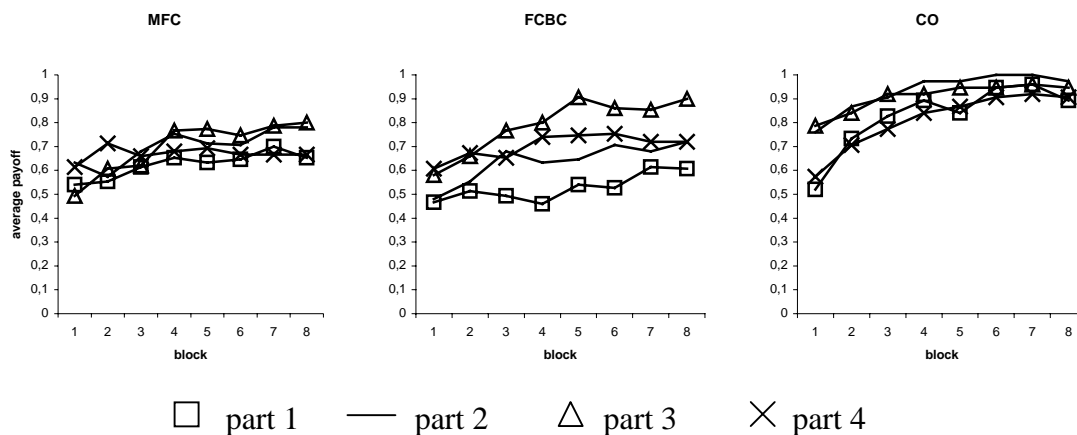
This test includes the data from part 4, meaning that the rise from part 1 to part 3 (62% to 93%) overcompensates for the drop to 83% in part 4. For MFC and CO we cannot find a sig-

<sup>10</sup> A more elaborate justification of the choice of win-stay lose-randomise is given in section 5.1.

nificant trend over all parts. However, for CO we find a significant rise in efficiency over parts 1 to 3 (Jonkheere-test, two-tailed,  $p < 0.01$ ). For MFC even this cannot be found.

When comparing efficiency rates between games we only find significant differences between CO and the other two games, while a difference between MFC and FCBC cannot be confirmed. Separate Mann-Whitney-U tests for parts show that differences between MFC and CO cannot be attributed to single parts as for none of the parts the p-value lies below 0.1. However, the overall significance is  $p = 0.04$ . Tests on the difference between FCBC and CO show that efficiency is significantly higher in CO only for parts 1 and 2 ( $p = 0.000$  and  $p = 0.023$ , respectively).

In addition to the payoff for a whole pair, we are also interested in the payoff development of the individuals. The following figure 2 (a – c) provides information on this issue. Each figure shows the development of average payoffs for one game over 8 blocks of 5-round averages. Each line stands for one part.



**Figure 2: Individual average payoffs over 8 blocks of 5-round averages**

In addition to observation 3 we can state that, not surprisingly, payoffs always slightly increase over time, while the marginal rate of payoff gain roughly declines.

As for the development of average payoffs over parts we checked whether there are differences within the first 10 rounds, rounds 16 to 25 and the last 10 rounds of each part. For MFC we do not find any significant differences at any of the investigated time intervals. Hence, experience does not lead to higher payoffs in this game. In contrast to that, for FCBC we find that, within the first 10 rounds, average payoffs rise with experience (Jonkheere test,  $p < 0.01$ , two-tailed), though differences between successive parts cannot be detected (corresponding Mann-Whitney-U tests always yield  $p > 0.1$ ). Towards the end, however, average payoffs appear to converge. In particular, no significant differences between any two parts can be detected and the corresponding Jonkheere test for a trend with experience yields insignificance

( $p = 0.67$ , two-tailed). In CO differences in average payoffs between parts can only be found for the first 10 rounds, where Mann-Whitney-U tests show differences between parts 1 and 4 on the one hand and parts 2 and 3 on the other hand (significances always with  $p < 0.01$ ). A Jonkheere test for a trend over parts is not significant when considering all parts, but yields significance for parts 1 to 3 ( $p = 0.452$  and  $p < 0.01$ , respectively). Summing up we may state

*Observation 4:*

Significant experience effects on average payoffs are rarely noticeable. The only ones we could detect were limited to either the first rounds (CO) or the intermediate part of play (FCBC). These can be explained by the number of successfully coordinating pairs and the corresponding starting round of coordination.

For FCBC, we additionally examined whether there are specific differences between the player roles, i.e. those players having a unique strategy with which they give a point to the opponent and those whose giving strategy depends on the action played by the opponent. We did not find any such difference.

## 5. Analysis of learning

### 5.1 Basic one-period adaptation behaviour

In terms of learning we first investigate compliance to the win-stay lose-change scheme, which has often been proposed to account best for bivariate decisions under uncertainty (see e.g. Rabinowitz et al. (1966)). The following figure 3 shows aggregate data on the win-stay part of the learning scheme. We separated aggregation between those pairs that did converge and those that did not converge. Win-stay rates for coordinating pairs are calculated only for the rounds until the coordination phase sets in.

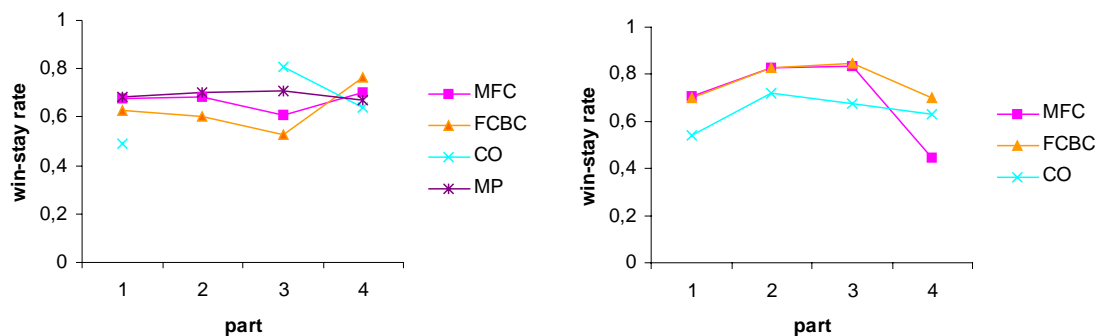


Figure 3 (a and b): Average win-stay rate over parts (left: non-coordinating pairs; right: coordinating pairs)

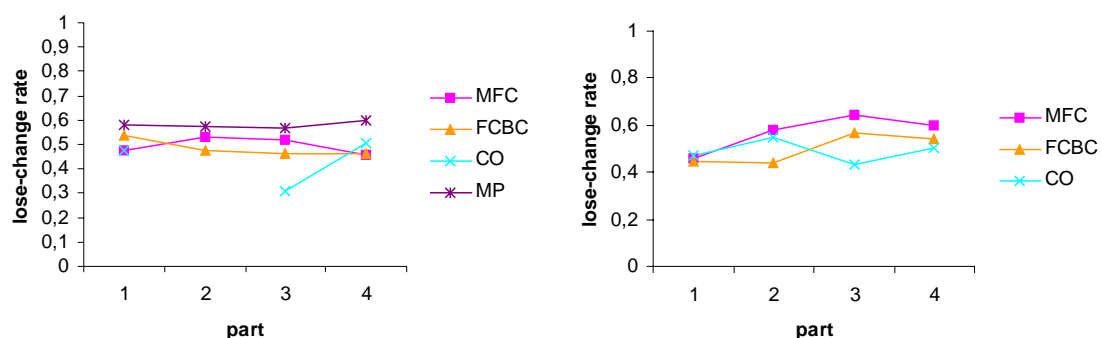
We first note that all win-stay rates (i.e. the rate of staying after succeeding in getting a point) always lie within the range of 0.5 and 0.8, which is rather low as compared to the theoretical prediction of 1. One may argue that since subjects did not know which game they were actually about to play, they may have assumed that the underlying game was MP in which a win-stay rate of 0.5 is just a Nash-equilibrium. However, considering that MP had a 0.25 chance to occur, the win-stay rates are too close to 0.5 to be justified by this argument. We further do not find any correlation between experience (as expressed in the part) and win-stay rates. We, hence, state

*Observation 5:*

Subjects' accordance to stay after winning a point does not depend on the experience gained.

What we do find, however, is that subjects in pairs that eventually coordinate do have a slightly stronger tendency to accord to the win-stay rule. In particular, Mann-Whitney-U tests show that win-stay rates are higher for subjects in coordinating pairs in MFC ( $p < 0.01$ ) and in FCBC ( $p < 0.1$ ). However, a closer look reveals that for both games this significant difference is solely caused by the observations within part 3. We therefore do not dare to state this finding as an assumedly stable observation and, hence, cannot confirm that accordance to the win-stay rule is a major vehicle for coordination. We further note that in MP the rate to stay after winning a point persistently lies at about 0.7. This is much higher than the theoretically predicted 0.5. At this stage we may suspect that this effect is caused by a significant amount of inertia. However, we also have to look at dynamics within each part in order to finally assess this. This is done below.

Before turning to the dynamics within a part, we have a brief look at the aggregate data on the compliance to the lose-change part, which is summarised in figure 4.



**Figure 4 (a and b): Average lose-change rate over parts (left: non-coordinating pairs; right: coordinating pairs)**

Again we find that compliance is rather low, lying at about 0.5 on average. Note that, contrary to the win-stay part of the strategy, randomisation after losing may well be more advisable in all games except MFC. So, we are not surprised to find lose-change behaviour to roughly accord to randomisation. We also do not find any tendency that is dependent on experience. The same is true for subjects in coordinating pairs. However, we are surprised to find the lose-change rate to lie highest for MP (pair wise Mann-Whitney-U tests,  $p < 0.01$ , two-tailed), since the game dynamics of MP should favour randomisation more than in other games. We suspect that this phenomenon is a consequence of over-randomisation on part of the subjects, which has been found to be frequent in games with unique mixed-strategy equilibria (see for example Brown and Rosenthal 1990). We summarise our findings in

*Observation 6:*

Compliance with changing action after losing is rather low, lying at about 0.5. No dependency on the amount of experience can be found. The lose-change rate is highest for MP.

We next examine the win-stay lose-change dynamics within parts. For this purpose we subdivide the forty rounds of play into two blocks of 20 rounds each. We then analyse differences between the first and the second half of a part. We again start with the win-stay analysis (table 8).

	rounds 2 – 20	rounds 21 – 40	observations
MFC	0.67	0.67	60
FCBC	0.67	0.63	58
CO	0.60	0.60	14
MP	0.72	0.67	120

**Table 8: Average win-stay rates over 20-round blocks  
(only non-coordinating pairs)**

The first thing to notice is that none of the games shows a tendency towards more compliance to the win-stay process from one half of the part to the second. This is contrary to what Rabinowitz et al. (1966) have found but in line with the findings of Mitropoulos (forthcoming). The decline is strongest for MP and may partially be due to the incentive to play stochastically. Nevertheless, the findings suggest a disprove of the win-stay assumption. The next table 9 shows the aggregate data for the lose-change part of the theory.

	rounds 2 – 20	rounds 21 – 40	observations
MFC	0.50	0.49	60
FCBC	0.52	0.47	58
CO	0.47	0.46	14
MP	0.59	0.59	120

**Table 9: Average lose-change rates over 20-round blocks  
(only non-coordinating pairs)**

Even for the lose-change part of the theory we cannot find any shift towards more compliance to the simple dynamics. The rates are rather stable. In table 9 the most dramatic decline can be observed for FCBC. This time for MP we find neither more nor less accordance to the lose-change process. We summarise the findings in

*Observation 7:*

None of the games shows an increase in either the win-stay rate or the lose-change rate from one half of the game to the second.

We checked the robustness of this observation in two ways. We, first, replicated table 8 and table 9 using the first and the last 10-round blocks instead of the first and last 20-round blocks. We came to the same conclusion, the only difference lying in the lose-change rate within MP that slightly increases over the two blocks. Second, we also looked at the same rates for each part separately and found only few instances where there were increases in either the average win-stay rate or the average lose-change rate over the two 20-round blocks. Overall, we were confirmed in the finding, that a decline in both win-stay and lose-change rates is the rule.

## 5.2 learning rules

Observation 2 indicates that, as subjects repeat the 40-round task, they obviously improve skills in coordination. The question now is how subjects manage to do this. A natural hypothesis is that subjects in the first part of the experiment start with the application of a certain adaptive rule and adapt to their experiences in the past by changing the shape of adaptation in later parts of the experiment. This calls for an analysis of meta-adaptive behaviour, which, as yet, has rarely been studied in experimental games<sup>11</sup>. Our approach is to take a large set of adaptive rules, all of which are applicable to our low-information environment, and to com-

---

<sup>11</sup> An exception is Stahl (1999), who includes a parameter in his maximum likelihood estimation that measures the extend to which adaptation rules from one part of a repeated game with full information are transferred to another repeated game.

pare characteristics of the rules with actual play. In this section, all adaptive rules are thoroughly introduced. In the next section the results on these rules are presented.

In order to keep our presentation as simple as possible, we decided to use a unified notational framework. In our two-person two-action two-valued payoffs setting we denote players by  $i \in \{1,2\}$  and actions by  $j \in \{A,B\}$ . In case the model is based on probabilistic decision processes, we denote the probability of player  $i$  to play action  $j$  at time  $t$  by  $p_i^j(t)$ . Some models are based on evaluations of approximates to expected payoffs associated with actions. Such approximates for player  $i$  and action  $j$  at time  $t$  are then denoted by  $u_i^j(t)$ . In some models these approximates are somehow processed to  $\tilde{u}_i^j(t)$  which then give the basis for the subsequent decision. The actual choice of player  $i$  at time  $t$  is given by  $a_i(t)$  and the actual payoff to player  $i$  at the end of period  $t$  is given by  $\pi_i(t)$ . Some models further consider aspiration levels that develop over time. They are addressed by  $b_i(t)$ .

Most of the learning rules we present below are parameterised versions of adaptation rules that have been analysed in the past; the only exception being the adaptation rule discussed first which is free of any parameters. We restricted models to have up to 2 parameters, which we call  $\alpha$  and  $\beta$ . Mind that  $\alpha$  and  $\beta$  do have different meanings in each model and therefore are added a suffix for the model they are related to. The reason for restricting the parameterised models in such a way is to keep degrees of freedom constant. We thereby circumvent the usage of information criteria for the comparison of different theories. We are aware of the fact that the strength of certain adaptation rules is their simplicity while others are better in nesting different aspects of cognition via a number of parameters. In this sense, the restriction of the number of parameters to two is a compromise between reducing theories to their cores and allowing for calibration to our data. For the simulations in this section we will use arbitrary chosen values that help to show the basic dynamics of the adaptation rules. In the next subsection, parameters will be determined via maximum likelihood estimations.

### **Win-Stay Lose-Randomise**

Consider first the rule win-stay lose-change. This rule has been shown to be very successful in settings involving dilemma situations, sometimes even outperforming tit-for-tat (see Nowak and Sigmund 1993). This simple learning rule is related to the learning direction scheme developed for larger strategy spaces by Selten and Stoecker (1986) while analysing Prisoner's Dilemma supergames and further investigated by Selten and Buchta (1994) on auction data. The idea is simply that whenever the individual fails to reach the optimal goal she will com-

pute the direction of ascent and adjust her decision accordingly. In our two-strategy two-outcome space this scheme translates into staying with one's strategy if it yielded a payoff of 1 and changing if it resulted in a payoff of zero.

One can easily see that in our game space this adaptation rule is not very efficient. Even when assuming that both players are strictly following the win-stay lose-change rule we get into inefficient cycles of adjustment. Applying this rule to both players of a pair, the MFC game is solved rather quickly (for sure within 3 rounds of play). However, in CO starting in one of the two inefficient cells would lead this process into an infinite cycle between these two cells, and, similarly, in FCBC they are quite likely to get stuck in the inefficient cycle (AA), (AB), (BA).

Such cycles can be avoided by introducing randomisation. Since the win-stay part is necessary for lasting coordination in MFC, FCBC, and CO randomisation is better introduced by replacing the lose-change part with lose-randomise. As a consequence the time needed to coordinate in MFC and CO slightly increases, but in turn FCBC is solved almost as quickly as MFC. In order to accept win-stay lose-randomise as a good benchmark in our setting we still have to clarify the dynamics within MP.

From the outset, it is clear that in MP the sole Nash-equilibrium is for both players in each round to choose each strategy with equal probability. Since in all other games this is not a good rule we have to dismiss this trivial adaptation rule as an optimal strategy for our game-space. Playing best-response strategies to both win-stay lose-change as well as win-stay lose-randomise, however, result in best-response cycles. Using the obvious abbreviations we get (i)  $WSLC \rightarrow WCLC \rightarrow WCLS \rightarrow WSLS \rightarrow WSLC$ , and (ii)  $WSLR \rightarrow WRLC \rightarrow WCLR \rightarrow WRLS \rightarrow WSLR$ . Still, there is good reason to become involved in cycle (ii) rather than in cycle (i). While payoffs from strategies in cycle (i) are completely eradicated by the best-response strategies, payoffs are diminished only partially in strategies involving the lose-randomise part. Hence, win-stay lose-randomise has the virtue of being an almost optimal strategy in MFC, FCBC, and CO whilst also being part of a small best-response cycle within MP.

## **Bush and Mosteller**

Psychologists started to theoretically describe learning processes in the early 20<sup>th</sup> century. Thorndike (1932) and Estes (1950) already formulated basic characteristics of simple learning processes. They drew their findings mainly from laboratory experiments with rats and humans involving single-person decision tasks. Subsequently, a number of stimulus-response models

were developed in order to account for observed behaviour. Among them, the most prominent which is also applicable in settings of multi-person games and therefore is of much interest until today is one variant of the theory developed by Bush and Mosteller (1955).

The basic idea of their model is that an action chosen in one period will be played with higher probability at the next decision node, if it has resulted in a reward and will be played with less probability, if it has resulted in non-reward. This translates to

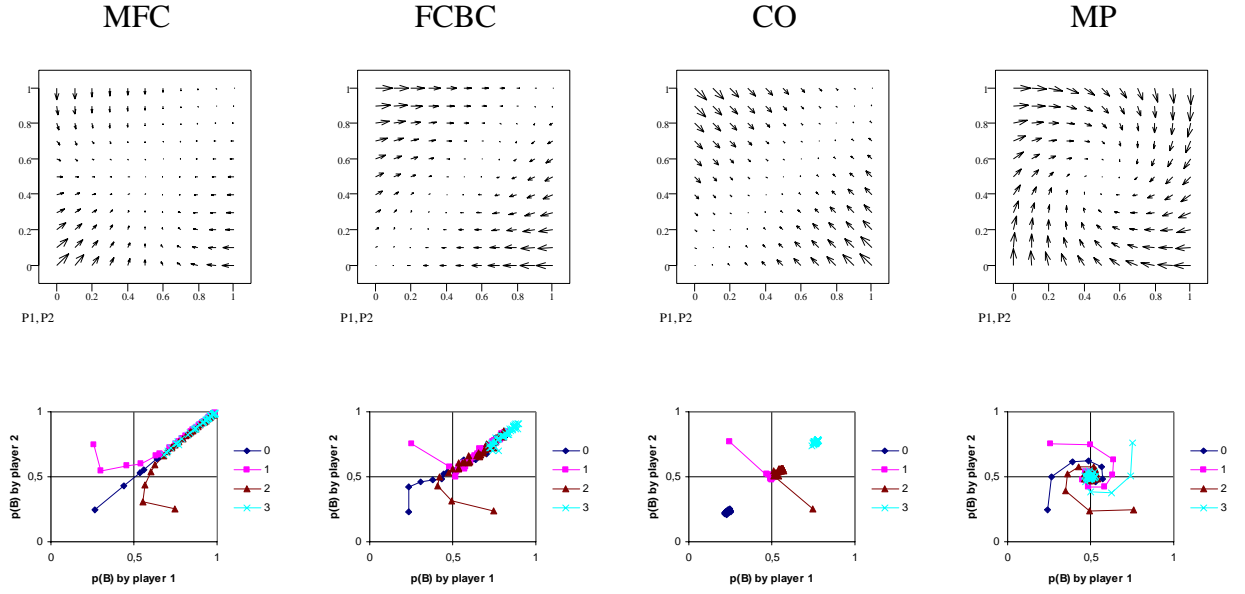
$$p_i^B(t+1) = p_i^B(t) + \alpha^{BM} \cdot (1 - p_i^B(t)) \cdot 1_{\left[ \begin{smallmatrix} a_i(t)=B \\ \wedge \pi_i(t)=1 \end{smallmatrix} \right]} \cdot (-p_i^B(t)) \cdot 1_{\left[ \begin{smallmatrix} a_i(t)=A \\ \wedge \pi_i(t)=1 \end{smallmatrix} \right]} \\ + \beta^{BM} \cdot (1 - p_i^B(t)) \cdot 1_{\left[ \begin{smallmatrix} a_i(t)=A \\ \wedge \pi_i(t)=0 \end{smallmatrix} \right]} \cdot (-p_i^B(t)) \cdot 1_{\left[ \begin{smallmatrix} a_i(t)=B \\ \wedge \pi_i(t)=0 \end{smallmatrix} \right]}$$

An important feature of this rule is that reinforcement of an action after having resulted in success (or reward) and reinforcement of this action after the alternative action has resulted in failure (or non-reward) is larger the less likely the choice of this action has been. Bush and Mosteller assume that there is a linear relation between current choice probability and subsequent choice probability whereby the reinforcement parameter may be different between situations in which the action chosen resulted in success ( $\alpha^{BM}$ ) and situations in which the action chosen resulted in failure ( $\beta^{BM}$ ).

The dynamics resulting from this scheme are shown in figure 5. The first row shows the transition dynamics for each of the four stage games, whereby the x-axis denotes the probability of player 1 playing action B and the y-axis denotes the probability of player 2 playing action B. The arrows indicate the direction and the relative strength of the expected adjustment of the mixed-strategy profile. Note that the sizes of the arrows are normalised, so that no comparison on the strength of adjustment can be made between diagrams. In order to show the predicted speed of adjustments and to capture the impact of time-varying influences we added one more row showing simulations of representative pairs of players. In more detail we first divided the unit space of mixed-strategy profiles into four quadrants and then printed the average path resulting from all pairs starting in the respective quadrant. We used average profiles of 1000 simulated pairs at each round to approximate the path<sup>12</sup>.

---

<sup>12</sup> Similar figures for the average path of play of simulated players can be found in Feltovich (2000).



**Figure 5: Dynamics of the Bush-Mosteller (BM) learning scheme ( $\alpha^{BM}=\beta^{BM}=0.5$ )**

Because of the BM-scheme describing a Markov chain it is no surprise that the dynamics coincide very much with the simulated pairs. In MFC quick adjustment towards the efficient pair of strategies  $(p_1^B, p_2^B) = (1, 1)$  can be observed. In FCBC there is a tendency to adjust much more slowly towards  $(1,1)$ . In the CO game there is quick guidance towards the diagonal from where there is no incentive to adjust further. And in MP pairs cycle clockwise towards the only Nash-equilibrium  $(0.5, 0.5)$  whereby convergence is reached rather quickly.

In order to illustrate the difference to the next learning rule by Mookherjee and Sopher (1994) we also need to look at the dynamics of asymmetric parameters. For this purpose we take  $\alpha^{BM} = 0.8$  and  $\beta^{BM} = 0.2$  and show the corresponding figure 6.

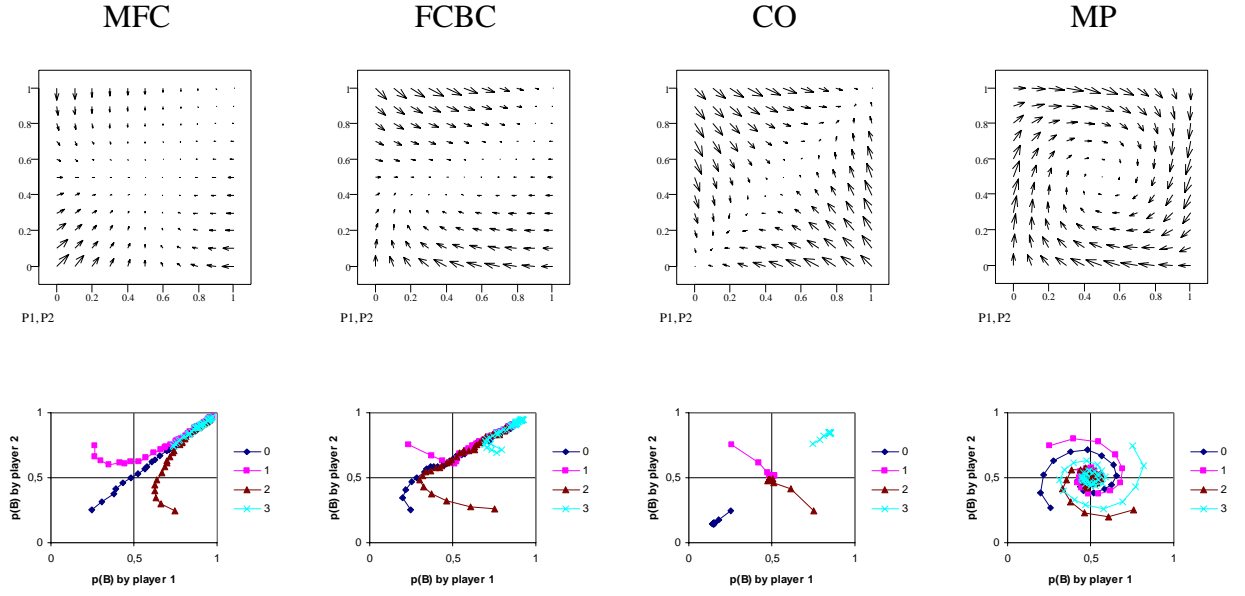


Figure 6: Dynamics of the Bush-Mosteller (BM) learning scheme ( $\alpha^{BM}=0.8, \beta^{BM}=0.2$ )

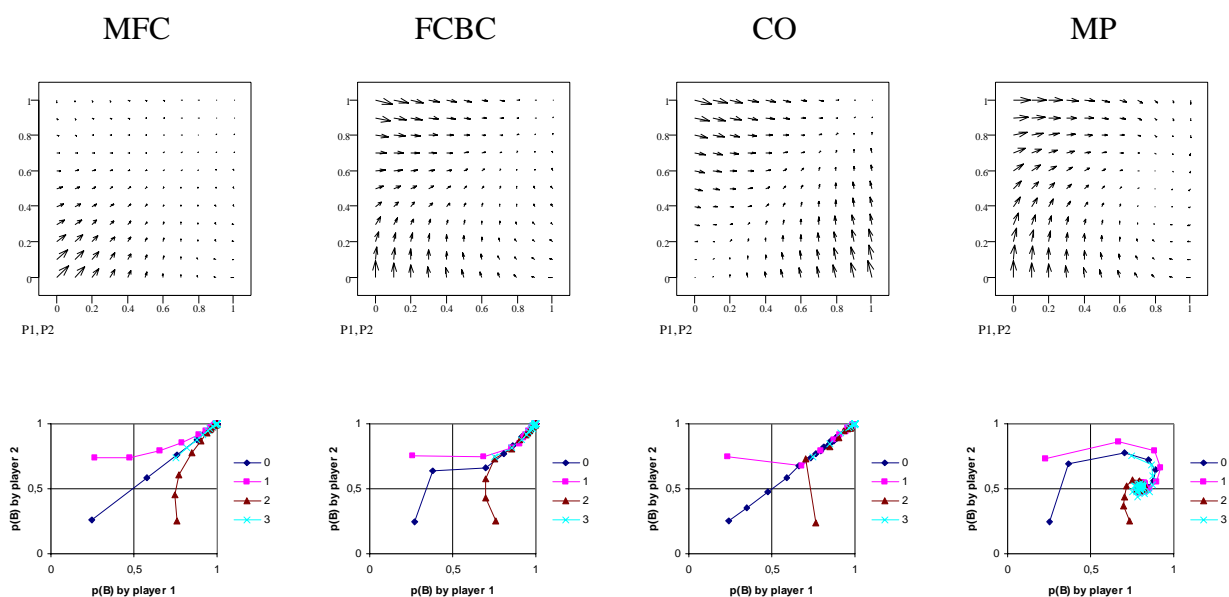
The main difference to figure 5 is that convergence of profiles is reached more slowly. Additionally, in FCBC there is a slight drift towards the upper left quadrant, reflecting the influence by player 1's action on his own payoff. Finally, in CO we observe a slight tendency to drift towards the efficient states (0, 0) and (1, 1).

## Mookherjee and Sopher

A model which is very similar but not quite the same as that by Bush and Mosteller is the one formulated by Mookherjee and Sopher (1994). They also take a linear difference equation but assign different reinforcement parameters to the vindication or refutation of a specific action rather than the confirmation of either action. Mathematically this means

$$\begin{aligned}
 p_i^B(t+1) = & p_i^B(t) + \alpha^{MS} (1 - p_i^B(t)) \cdot 1_{\left[ \begin{array}{l} a_i(t)=B \\ \wedge \pi_i(t)=1 \end{array} \right]} \cdot 1_{\left[ \begin{array}{l} a_i(t)=A \\ \wedge \pi_i(t)=0 \end{array} \right]} \\
 & + \beta^{MS} p_i^B(t) \cdot 1_{\left[ \begin{array}{l} a_i(t)=B \\ \wedge \pi_i(t)=0 \end{array} \right]} \cdot 1_{\left[ \begin{array}{l} a_i(t)=A \\ \wedge \pi_i(t)=1 \end{array} \right]}
 \end{aligned}$$

In case  $\alpha^{BM} = \beta^{BM}$  and  $\alpha^{MS} = \beta^{MS}$  we get identical special cases of the two models. In case we use asymmetric parameters, however, we do observe different dynamics. For illustration we take the case of  $\alpha^{MS} = 0.8$  and  $\beta^{MS} = 0.2$  and show the corresponding figure 7.



**Figure 7: Dynamics of the Mookherjee-Sopher (MS) learning scheme ( $\alpha^{MS}=0.8, \beta^{MS}=0.2$ )**

Compared to the figures within the Bush-Mosteller scheme we see that in MFC incentives to adjust towards the efficient (1,1) are identical in shape but stronger, as the simulations show. A similar, but slightly deranged picture is drawn for FCBC. In the CO game, despite its symmetry concerning the two actions, the Mookherjee-Sopher scheme suggests a drive towards (1,1). And for the MP game we again observe a clockwise cycle, but one that is shifted towards player 1 playing B with high probability, while player 2 keeps up playing both strategies with equal probability.

The dynamics obviously result from the parameters simply implying stronger adjustment for vindictive experience with B than with A, which is a rather unreasonable assumption in a situation in which the names for the actions are nothing but labels. In our environment players start off playing the repeated game without knowing the labels of one's actions. Unless there is evidence that subjects try to match labels (e.g. due to focal points) which they quickly identify while playing there is little sense in assuming anything else than  $\alpha^{MS} = \beta^{MS}$ . From this point of view, the Bush-Mosteller scheme looks richer in structure, since it allows for different reinforcements depending solely on the feedback rather than on combinations of feedback and label.

## Cross

Drawing from the same stimulus-response concept, Cross (1973) formulated another similar model. The difference to the two models above lies in the different interpretation of payoffs.

$$p_i^B(t+1) = p_i^B(t) + \alpha(\pi_i(t))(1 - p_i^B(t)) \cdot 1_{a_i(t)=B} - \alpha(\pi_i(t))p_i^B(t) \cdot 1_{a_i(t)=A}$$

To Cross the reinforcement parameter for the linear difference equation is a monotonic function  $\alpha(\cdot)$  of the payoff resulting from the respective action. A payoff of 0, hence, does not necessarily mean that the probability of repeating the same action will be reduced. Cross even considers  $\alpha(\cdot)$  to be the identity, so that  $\alpha(0)=0$ , meaning that a payoff of 0 results in the continuation of the given choice probabilities. We follow Cross and use  $\alpha(\pi) = \alpha^{CR} \cdot \pi + \beta^{CR}$ . In figure 8 we use  $\alpha^{CR} = 0.9$  indicating a rather strong tendency to adjust, and  $\beta^{CR} = 0$ .

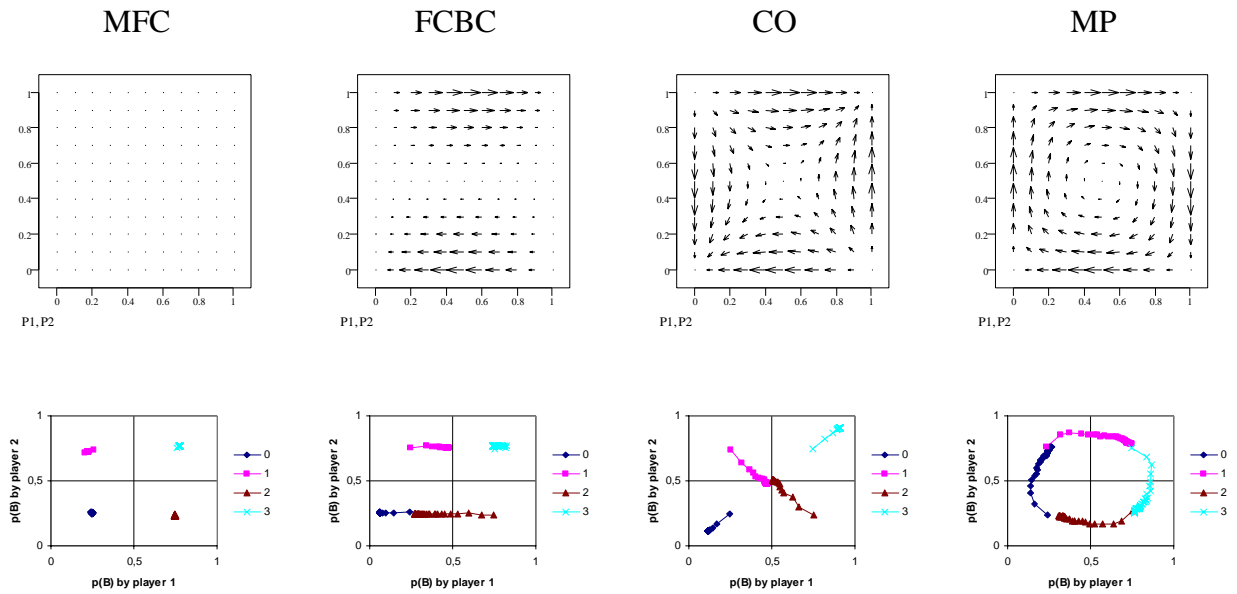


Figure 8: Dynamics of the Cross (CR) learning scheme ( $\alpha^{CR}=0.9, \beta^{CR}=0$ )

Again, the dynamics can be seen to differ from the learning rules above. While in MFC there is no drive towards any direction, i.e. pairs will wander around according to chance events, in FCBC there will be a simple adjustment by player 1 only, the one player who has some influence on his own payoff. In CO there are symmetric drifts towards the efficient pairs (0,0) and (1,1), while the strength of the drifts diminishes the closer the pair gets to the diagonal or the anti-diagonal. Finally, in MP one again observes a clockwise cycling around the unique Nash-equilibrium. However, differently to Bush-Mosteller and Mookherjee-Sopher the cycling does hardly show any attraction towards the center.

## Börgers and Sarin

A model that is still in the tradition of reinforcement rules but introduces aspiration is that by Börgers and Sarin (1997). They let the agent's aspiration evolve according to a simple linear

process, while choice probabilities are reinforced by an amount proportional to the distance between the aspiration level and last period's payoff.

$$p_i^B(t+1) = p_i^B(t) - |\pi_i(t) - b_i(t)| p_i^B(t) + |\pi_i(t) - b_i(t)| \cdot 1_{\left[ \begin{array}{l} a_i(t)=B \\ \wedge \pi_i(t) > b_i(t) \end{array} \right]} \vee 1_{\left[ \begin{array}{l} a_i(t)=A \\ \wedge \pi_i(t) \leq b_i(t) \end{array} \right]}$$

The aspiration level of individual  $i$  evolves according to

$$b_i(t+1) = \beta^{BS} b_i(t) + (1 - \beta^{BS}) \pi_i(t).$$

The second parameter is given by the initial aspiration level, i.e.  $\alpha^{BS} = b_i(1)$ . The evolving aspiration level  $b$  causes the stationary look at the dynamics to neglect certain time-dependent aspects. For illustration we, hence, provide dynamics and simulations for two distinct cases, namely  $\alpha^{BS} = 0.9$  (i.e. initially optimistic agents) in figure 9 and  $\alpha^{BS} = 0.1$  (i.e. initially pessimistic agents) in figure 10.

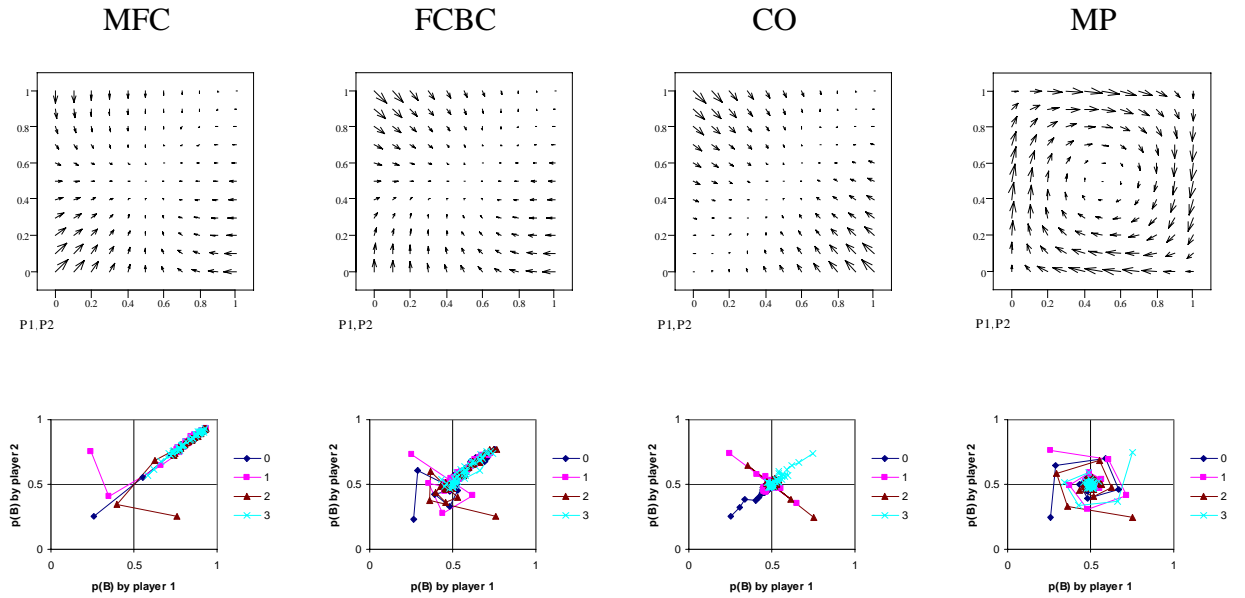
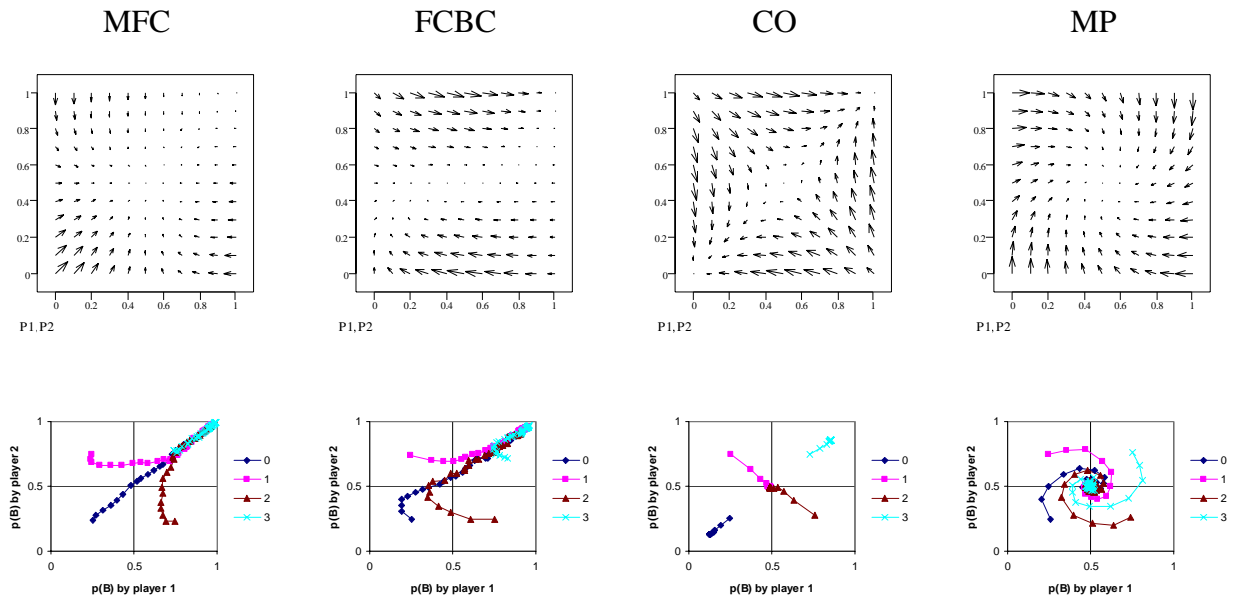


Figure 9: Dynamics of the Börgers-Sarin (BS) learning scheme ( $\alpha^{BS}=0.9, \beta^{BS}=0.1$ )



**Figure 10: Dynamics of the Börgers-Sarin (BS) learning scheme ( $\alpha^{BS}=0.1, \beta^{BS}=0.1$ )**

The main difference between optimistic and pessimistic players is that pessimists adjust more smoothly to vindictive experiences. For pessimists, vindictive experience causes the chosen strategy to be reinforced strongly, while refuting experience causes small adjustments of choice probabilities. For optimists the opposite is true; vindication has small effects and refutation has large effects on choice probabilities. As a consequence, the volatile behaviour of optimists in the game FCBC causes them to get stuck in intermediate choice probabilities until the aspiration level has decreased to values in the order of expected payoffs. To the contrary, pessimists show steady adaptation in the direction of efficient play. The rising aspiration level causes the reinforcing effect of vindication to decrease at much higher choice probabilities for action B than is observed for optimists.

Similarly for the game MP we find that the Börgers-Sarin scheme with optimistic aspiration levels resembles that of the cycling Cross dynamic, while the Börgers-Sarin scheme for pessimists looks more similar to the converging symmetric Bush-Mosteller scheme. The differences vanish as the aspiration levels approach levels near 0.5.

## Roth and Erev's reinforcement learning

The above reinforcement learning rules are all restricted to process only current choice probabilities and last periods payoffs. In this respect they all share the feature of complete loss of memory, i.e. they can be visualised as infinite-set Markov Chains. A model still confined to the reinforcement idea but with an additional consideration of memory is that by Roth and

Erev (1995)<sup>13</sup>. The idea of this process is that probabilities are indirectly determined by cumulated propensities. The propensities evolve via

$$u_i^j(t+1) = \beta^{RE} \cdot u_i^j(t) + \pi_i(t) \cdot 1_{a_i(t)=j},$$

i.e. the propensity of an alternative is first discounted by a depreciation parameter  $\beta^{RE}$  (Roth and Erev 1995 call it the “forgetting parameter”). If this alternative has been chosen in the last round then its propensity is added the actual payoff that resulted from that decision, otherwise the propensity stays at the discounted level. The second parameter of the model determines the strength of initial experience, which is defined as the sum of initial propensities, i.e.

$$\alpha^{RE} = u_i^A(1) + u_i^B(1).$$

The choice probabilities within a round are then simply proportional to the sum of propensities:

$$p_i^j(t) = \frac{u_i^j(t)}{\sum_{k \in \{A,B\}} u_i^k(t)}.$$

The basic dynamics are depicted in figure 11. In fact, the diagrams show little difference to the ones for the Cross scheme, the reason being that the only difference lies in the introduction of diminishing effects of feedback via increasing values for the propensities. Hence, the basic difference lies in the process tending to slow down after a while, which has been termed by psychologists the *power law of practice*.

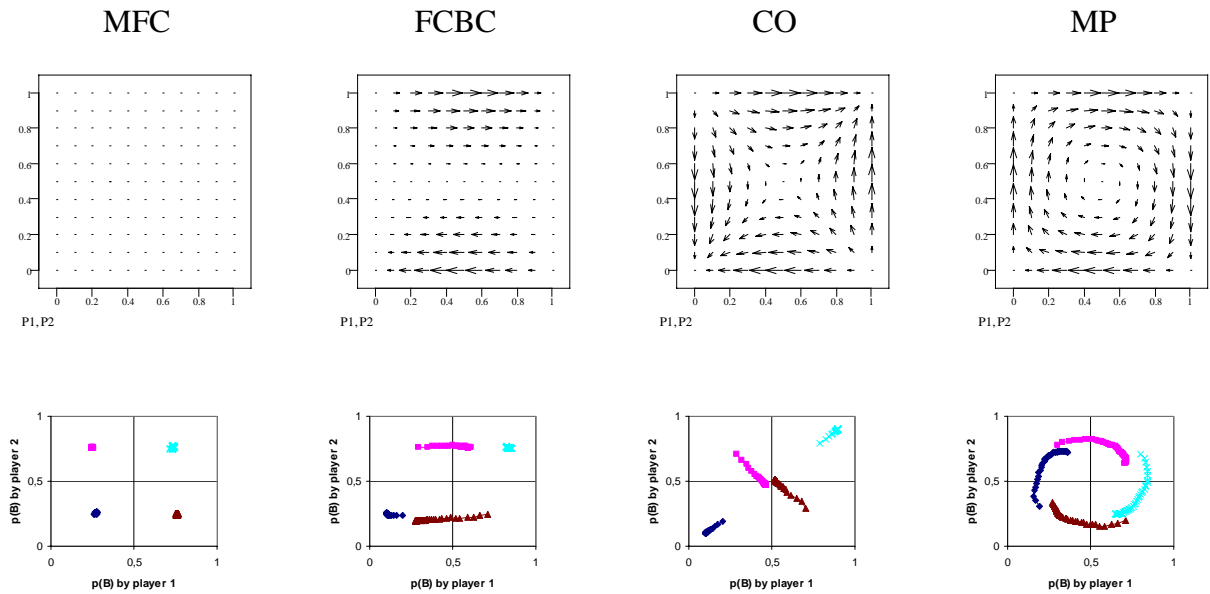


Figure 11: Dynamics of the Roth-Erev (RE) learning scheme ( $\alpha^{RE}=1, \beta^{RE}=0.95$ )

<sup>13</sup> This process has first been presented by Harley (1981) as a solution to the problem of finding an evolutionary stable learning rule.

Apart from the apparent slowing down of the process as play advances, the process shares all characteristics of the Cross scheme.

### **Karandikar et al.**

Karandikar et al. (1998) also consider evolving aspirations but give up the idea of reinforced choice probabilities. Rather, they vary the idea of win-stay lose-randomise by specifying the randomisation process. The scheme win-stay lose-randomise as discussed above can be approximated by special cases of the Karandikar et al. scheme. Their approach is not constructive in the sense that they only stated necessary conditions but did not explicate the randomisation rule. The model they present looks as follows.

$$p_i^B(t+1) = 1 \cdot \mathbb{1}_{\left[ \begin{array}{l} a_i(t)=B \\ \wedge \pi_i(t) \geq b_i(t) \end{array} \right]} + h(b_i(t) - \pi_i(t)) \cdot \mathbb{1}_{\left[ \begin{array}{l} a_i(t)=B \\ \wedge \pi_i(t) < b_i(t) \end{array} \right]} + (1 - h(b_i(t) - \pi_i(t))) \cdot \mathbb{1}_{\left[ \begin{array}{l} a_i(t)=A \\ \wedge \pi_i(t) < b_i(t) \end{array} \right]},$$

whereby  $h(\cdot)$  fulfils (i)  $h(0)=1$ , (ii)  $-\infty < \tilde{g} < h'(x) < 0$  for all  $x$  and some  $\tilde{g}$ , and (iii)  $h(x) \geq \tilde{p} > 0$  for all  $x$  and some  $\tilde{p}$ , so that the probability of change is bounded away from 0. They also assume that the aspiration levels evolve probabilistically with

$$P[b_i(t+1) = (1 - \beta^{KA})b_i(t) + \beta^{KA}\pi_i(t)] = 1 - \eta^{KA} \text{ and}$$

$$P[b_i(t+1) = (1 - \beta^{KA})b_i(t) + \beta^{KA}\pi_i(t) + G(b_i(t))] = \eta^{KA},$$

whereby  $G(b(\cdot))$  is a random variable with density  $g(\cdot|b)$ .

Since the perturbation of the aspiration level is important only for games involving Pareto-ordered Nash-equilibria (such as in the stag-hunt game or the Prisoner's Dilemma) we assume deterministically evolving aspiration levels, i.e.  $\eta^{KA} = 0$ . For further analysis we specify

$$h(x) = \arctan\left(\frac{\gamma^{KA}}{x^2} + \tan\left(\frac{\pi}{2} \cdot \delta^{KA}\right)\right) \cdot \frac{2}{\pi}$$

whereby  $\gamma^{KA}$  indicates the speed of decline as the gap between aspiration level and actual payoff rises and  $\delta^{KA}$  denotes the asymptotic bound from below. Just as the win-stay lose-randomise rule, but contrary to all other schemes, the Karandikar et al. scheme suggests a probability assessment that is solely based on the previous action and payoff and ignores previous choice inclinations. The dynamics for this scheme are depicted in figures 12 and 13. For the purpose of exposition, we fix  $\gamma^{KA} = 0.2$  and  $\delta^{KA} = 0.1$ , so that  $h(\cdot)$  shows reasonable sensitivity as its argument varies between 0 and 1. The basic characteristics of the process do not depend on slight variations of these parameters. Similar to the Börgers-Sarin scheme we show the dynamics for players with initially optimistic aspiration levels and the dynamics for players with initially pessimistic aspiration levels.

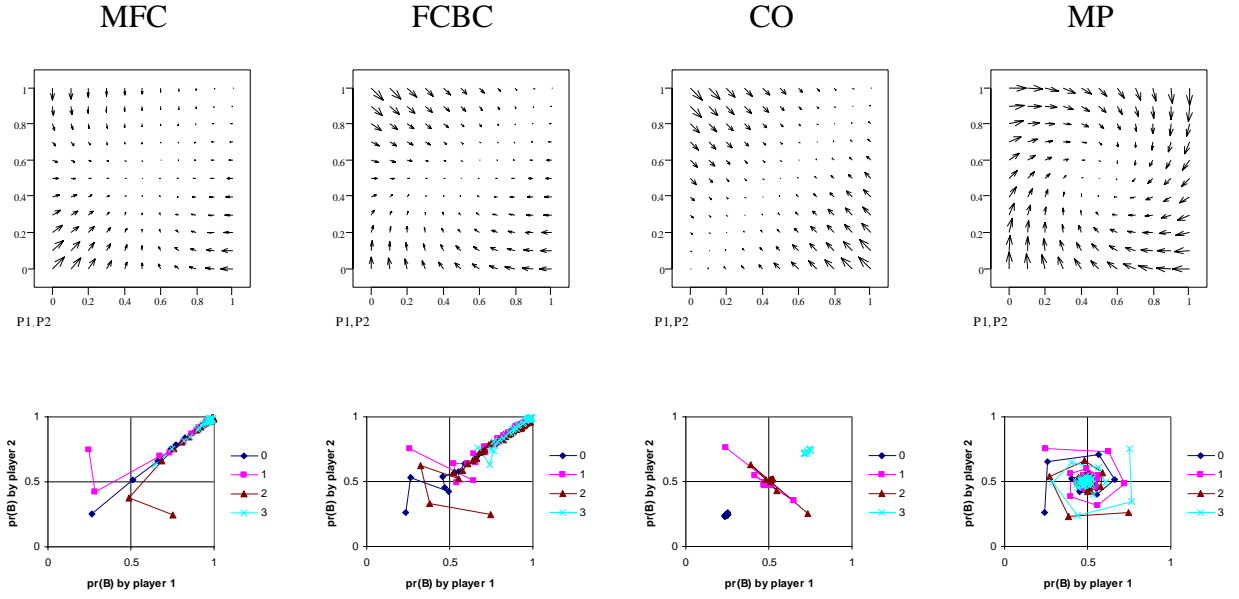


Figure 12: Dynamics of the Karandikar et al. (KA) learning scheme ( $\gamma^{\text{KA}}=0.2, \delta^{\text{KA}}=0.1, \eta^{\text{KA}}=0, \alpha^{\text{KA}}=0.9, \beta^{\text{KA}}=0.1$ )

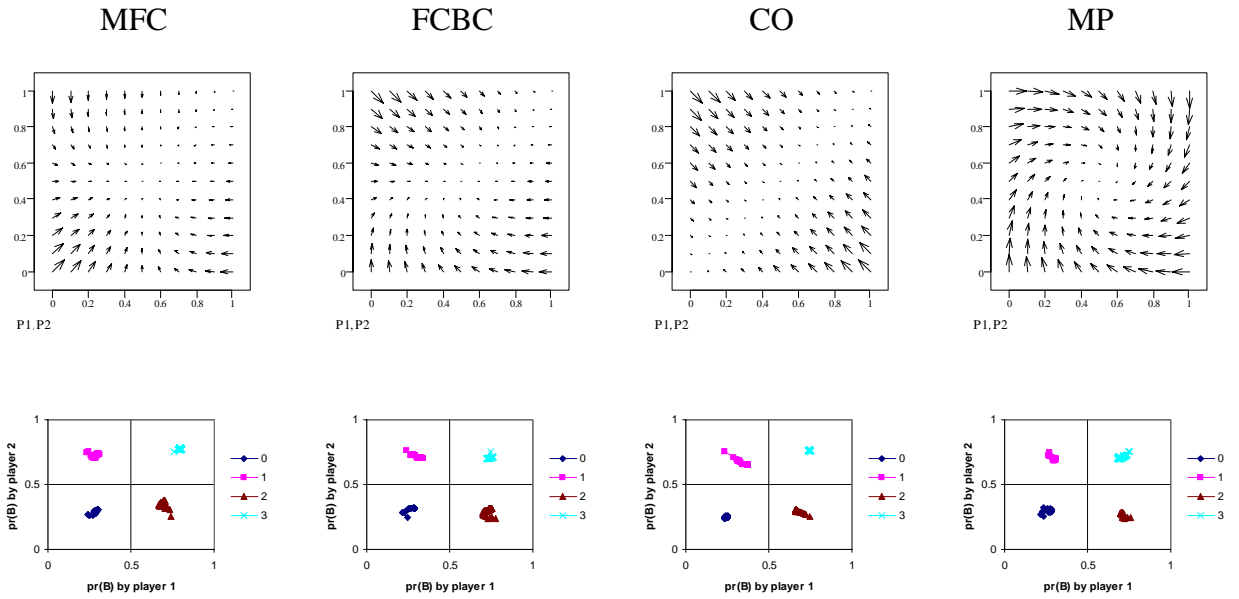


Figure 13: Dynamics of the Karandikar et al. (KA) learning scheme ( $\gamma^{\text{KA}}=0.2, \delta^{\text{KA}}=0.1, \eta^{\text{KA}}=0, \alpha^{\text{KA}}=0.1, \beta^{\text{KA}}=0.1$ )

From the figures we see that for the Karandikar process the initial aspiration level has a markedly different impact than on the Börgers-Sarin scheme. Here, starting with a low initial aspiration level easily causes players to get stuck in actions that regularly yield zero payoff, because the difference between the low aspiration level and the payoff of zero is not large enough to induce a change in choice. This observation uncovers one of the weaknesses of the scheme; even if we introduce small shocks in aspiration levels the process would need a large amount of time to leave a strategy that returns the expected low payoff. Experimentation with

the parameter  $\alpha^{KA}$  showed that for our game environment values of about 0.3 suffice to make lock ins sufficiently improbable so that the dynamics follow the trajectory.

Somewhat surprising is the fact that the Karandikar scheme, when initialised with optimistic players, roughly yields the same dynamics as the Bush-Mosteller scheme, even though it does not make use of last round's choice probabilities.

## Sarin and Vahid

A completely different approach is that of using past payoffs to generate approximates for the expected payoffs of the respective alternatives. Not the first, but probably the simplest approach in this line is the one by Sarin and Vahid (1999). They assume that the expected payoffs follow a reinforcement process with the difference equation being linear in the difference between previous expectation and current payoff. The decision is then a stochastic process depending on these payoff expectations.

Formally stated, the payoff expectation for action  $j$  evolves according to

$$u_i^j(t+1) = u_i^j(t) + \alpha^{SV} (\pi_i(t) - u_i^j(t)) \cdot 1_{a_i(t)=j}.$$

Before players choose their action these assessments are perturbed by a random variable  $Z$  that may depend on time, i.e.

$$\tilde{u}_i^j(t) = u_i^j(t) + Z(t).$$

Players then choose the one alternative that promises the maximum payoff according to the perturbed assessments, i.e.

$$a_i(t) = \arg \max_{j \in \{A, B\}} \tilde{u}_i^j(t).$$

Matsushima (1998) simultaneously analysed such processes in a much broader context. Concerning the random variable  $Z(t)$  we will restrict ourselves to the normal distribution with zero mean and standard deviation  $\sigma^{SV}$ , whereby the second parameter will be the standard deviation of the perturbation, i.e.  $\beta^{SV} = \sqrt{\sigma^{SV}}$ . For the dynamics diagrams we had to restrict expected payoffs so that  $u_i^A + u_i^B = 1$  for all  $i$ . Since the  $u_i^j$ 's evolve over time, the dynamics diagrams again do not tell us the whole story of the process. In fact, the simulation results turn out to be strikingly similar to the ones of the Bush-Mosteller scheme with asymmetric parameters. We do not have a proper account for this phenomenon.

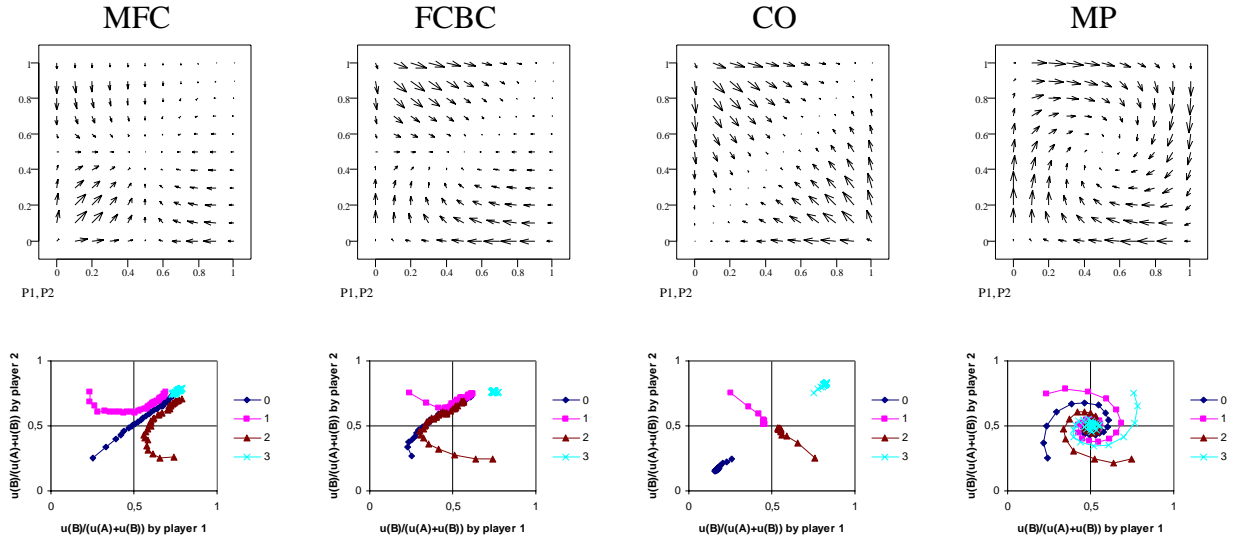
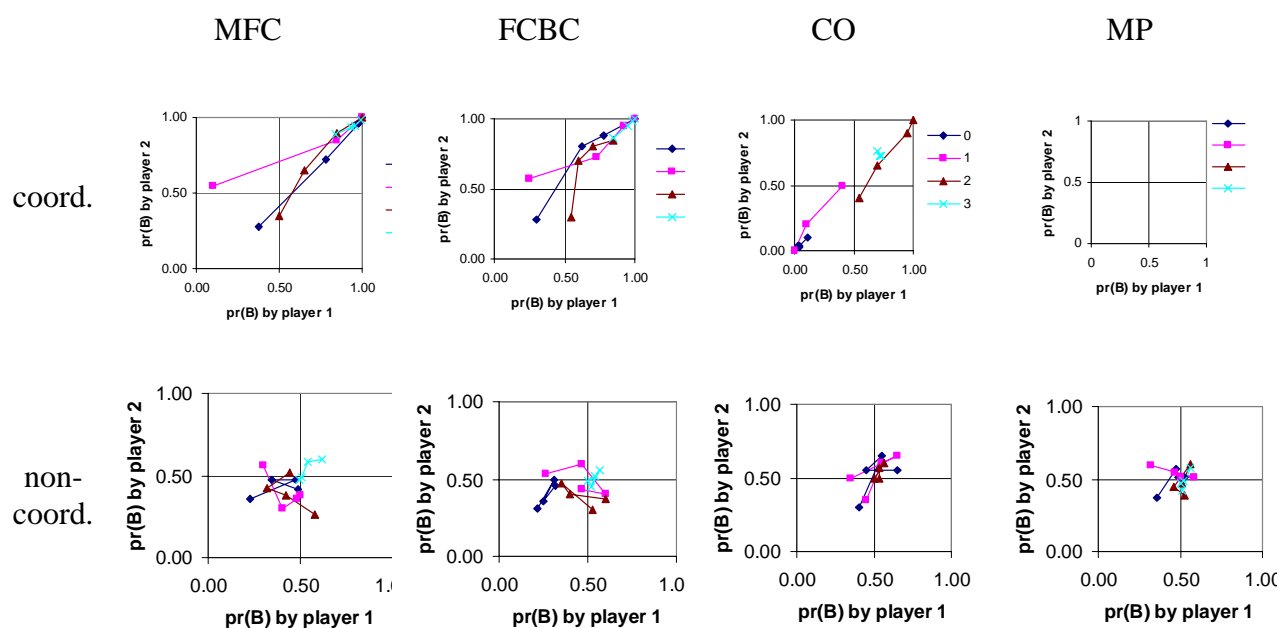


Figure 14: Dynamics of the Sarin-Vahid (SV) learning scheme ( $\alpha^{SV}=0.5, \beta^{SV}=0.2$ )

### 5.3 Aggregate paths and results from player-based estimations

Our approach is different from many other studies on adaptation behaviour in that it does not focus on quantitative comparisons between learning rules, such as the goodness of fit or the predictive success. We prefer to use methods that reveal structural characteristics of learning rules and compare them to the characteristics of the data. The first such analysis concerns the path of play. For this purpose we redraw the figures that show the average paths of pairs for actual data. We proceeded in the following way: We, first, classified each pair of players according to the choice frequencies within the first 10 rounds. Pairs for which both players played action B four times or less were classified into quadrant 0, pairs for which player 1 played B four times or less and player 2 played action B five times or more often were classified into quadrant 1, likewise classification for quadrants 2 and 3 were determined. By this procedure we classified each pair into one starting quadrant. Figure 15 shows, for each game separately, the average paths of all pairs that fall into the respective starting quadrant. Since variance of play was quite large, the pictures show only four data points per path, each one relating to the average choice within a 10-round block.



**Figure 15: 40-round path of average choice proportions by game and coordination**

The figures clearly show that subject behaviour of coordinating pairs was quite different from behaviour of non-coordinating pairs. For the game MFC it is no surprise to see all paths of coordinating pairs to lead to the efficient outcome (1,1), i.e. (B,B) for sure, since we selected the pairs to contain only those that did approach this outcome. The same is true for the graphs of coordinating pairs for the game FCBC. The figure for MP by definition does not contain any data. However, the figure for the CO game is worth to be looked at in more detail. Those pairs that start in quadrant 0 obviously quickly make their way through to the NE at (A,A). In contrast to that, pairs that start in quadrant 3 do not always reach the coordination at (B,B). 3 pairs out of 15 starting in quadrant 3 finally coordinate on (A,A). Finally, there is only one coordinating pair that starts in quadrant 1 and only two coordinating pairs that start in quadrant 2. Thus, it is no surprise, either, that paths for both classes of pairs follow a straight line towards one of the pure strategy NE.

Overall, the essence of figure 15 is that for the non-coordinating pairs there is no trend towards efficient play to be observed whatsoever. On average, play always amounts to roughly allocate equal frequency to each strategy. This is independent of the game and the number of rounds already played. Variance looks smaller for the MP game, but this is mainly due to the significantly larger number of pairs over which aggregation has been performed. Our conclusion is

*observation 8:*

Subjects within pairs that do not succeed in coordination show no sign of a trend towards efficient play<sup>14</sup>.

One way to find out whether subjects show a difference in adaptation behaviour when they have experienced the learning task as compared to part one when they are not familiar with the learning environment is to estimate the parameters of learning rules for the four different parts separately. Since different games give rise to different dynamic processes we also perform estimations for each game separately. Each point in each diagram represents one estimation performed on all data from coordinating pairs until coordination sets in. The underlying idea is that the learning rules may be able to capture qualitative characteristics of play until coordination sets in by way of a corresponding sensitivity of the parameters. Figure 16 shows the path of estimated parameters over parts for each learning rule and each game. The game MP is not included since coordination is not an issue there.

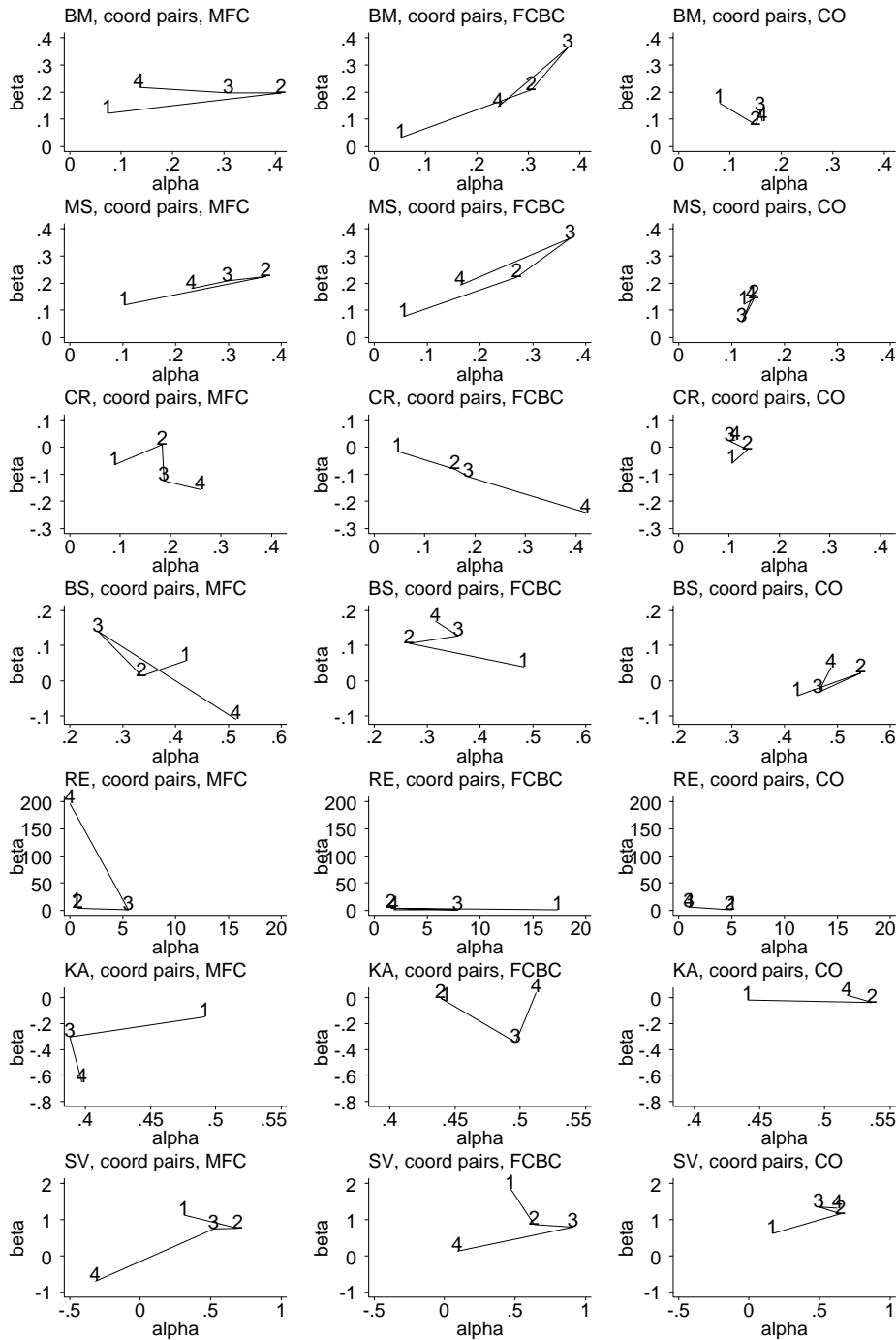
The results are rather discouraging. None of the learning rules is able to elicit any steady trend of parameters. No conclusions from estimations can be drawn on the cognitive processes of subjects over parts. We also did the same analysis with all non-coordinating pairs as well as with estimations for all pairs combined. The same conclusion was reached. One criticism concerning the method of analysis used concerns the representative agent approach. There may very well be a trend over parts, which, however, is obscured by the heterogeneity of subjects' behaviour. From a static point of view the maximum likelihood procedure weighs those predictions more for which prediction is difficult to achieve (i.e. to get a high likelihood). Likewise those subjects are given more weight whose path is difficult to fit by the respective learning rule. A more rigorous analysis would thus involve estimation of parameters for individual players, and this is what we present next.

We, hence, leave the standard approach of the analysis of subject play which is usually based on the representative agent assumption. We rather estimate parameters of learning rules for each agent separately. This is one of few studies to perform this task<sup>15</sup>.

---

<sup>14</sup> The author found this statement also to be true independent of the experience gained over parts.

<sup>15</sup> In fact, the author only knows of one other study that does individual estimations of updating behaviour and this is the, as yet, unpublished work by Shachat and Walker (2000).



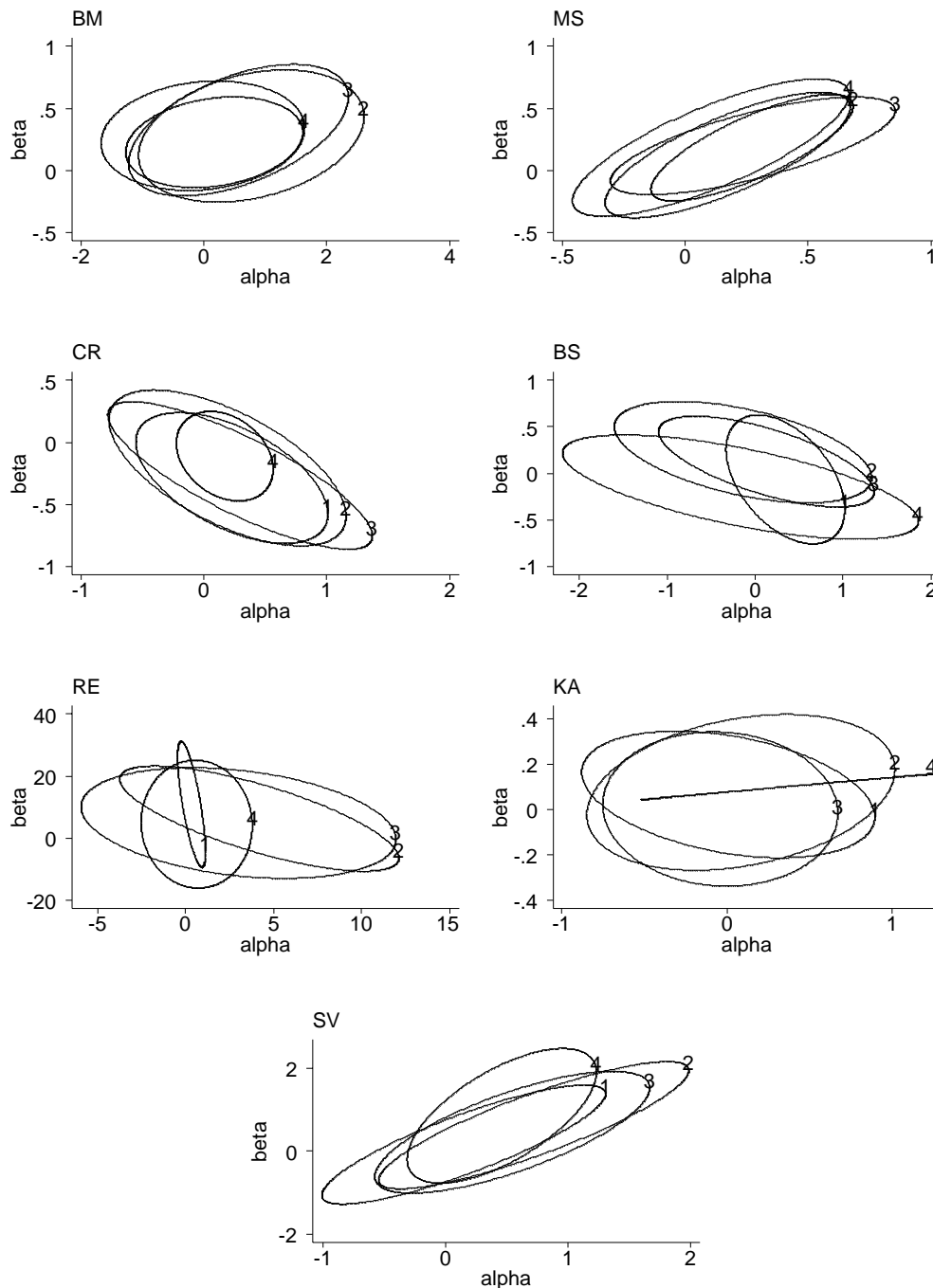
**Figure 16: Paths of estimation parameters over parts by learning rule and game using data from coordinating pairs until coordination sets in**

Our hope to get more insight as to which changes in adaptive behaviour cause the improved ability to coordinate on the efficient cell was disappointed. Figure 17 gives an account of individual estimations for coordinating pairs. Each graph shows confidence ellipses around the mean for individual parameter estimations of one learning rule. Ellipses are drawn for estimations of a single part over all games<sup>16</sup>. Each ellipse depicts the 90 percent confidence ellipse<sup>17</sup>

<sup>16</sup> There are not enough data points to perform the same analysis for each game separately.

<sup>17</sup> As approximated by the ellipse with two times the standard deviations of the parameters as respective radii and the correlation coefficient indicating the angle.

around the mean of all estimated parameter points excluding outliers according to Hadi's (1992, 1994) algorithm and excluding players for which the estimation algorithm did not converge<sup>18</sup>. Estimations were done via a maximum-likelihood estimation for one-period ahead predictions<sup>19</sup>.



**Figure 17: Confidence ellipses for estimation parameters by parts and rules**

<sup>18</sup> The exclusion of those players for which the estimation procedure did not lead to sensible results leads to a slight bias in data selection. Typically, those players were excluded for which few changes in behaviour were recorded.

<sup>19</sup> Apart from the renunciation of the representative agent assumption this procedure is fairly standard in the literature; see e.g. Camerer and Ho (1999) and references therein.

Similar to figure 16, figure 17 shows that none of the learning rules is able to capture any behavioural trend over parts. The estimations fail to answer the question as to why or how subjects were able to coordinate more quickly as they gained experience with the task. This overall picture does not change when depicting the same graphs for non-coordinating pairs. The ellipses for non-coordinating pairs typically show slightly smaller variance, which is probably due to the larger number of observations, but are very similar otherwise. We summarise these results in

*observation 9*

None of the learning rules is able to capture a change of adjustment behaviour as is suggested by observation 2. This is independent of whether estimations of parameters are done using the representative agent approach or for each individual separately.

It is informative to look at the number of observations that comprise a confidence ellipse. Since estimations for different learning rules were performed for the same data the numbers reflect the ability of the respective learning rule to produce reasonable estimation results. After eliminating all those subjects that coordinated with their partner before period 10, those for which the estimation procedure did not converge and those who were classified as outliers according to Hadi's criterion, we got the following numbers of successful estimation of individual learning parameters, which are depicted in table 10.

rule	part			
	1	2	3	4
BM	19	17	22	7
MS	19	14	19	8
CR	13	9	14	6
BS	15	12	21	6
RE	6	7	6	5
KA	9	10	15	2
SV	13	17	17	9

**Table 10: Number of successful estimations for players in coordinating pairs by parts and rules**

The results clearly depend on the way parameters were specified for the learning rules. They also depend on the particular estimation procedure<sup>20</sup>. Nevertheless, the numbers give a rough impression of how stable estimations were. From table 10, and much more so when including the numbers on non-coordinating pairs, we get the impression that the RE and the KA schemes were much worse in producing useful results than were the other learning rules.

<sup>20</sup> Estimations were performed using Stata 6.0. According to its manual the procedure uses the BFGS method with some modifications as to the rescaling and the initialisation of the search algorithm. Control estimations using the implementation of the BHHH and BFGS methods within TSP 4.5 produced similar results.

## 6. Patterns

All the adaptation rules considered so far only capture round-by-round adjustments. However, they fail to account for another aspect that has been put forward to be crucial in any learning process, namely the investigation and understanding of the game environment. As has already been observed by Mitropoulos (forthcoming), subjects appear to frequently be involved in stationary patterns that supposedly serve to find an understanding of the feedback subjects get from the environment. We systematically investigate subjects' use of patterns by using the following

### *Definition 3:*

A pattern played by a subject is identified by one of the following sequences of actions that occurs before and until coordination of the corresponding pair sets in:

- i) a sequence of at least six consecutive rounds with the same action, i.e.
  - a) ...AAAAAA... or
  - b) ...BBBBBB...
- ii) a sequence of at least five consecutive alternations of actions, i.e.
  - a) ...ABABABA... or
  - b) ...BABABAB...
- iii) a sequence of at least seven consecutive alternations between staying and changing, i.e.
  - a) ...ABBAABBA... or
  - b) ...BAABBAAB...

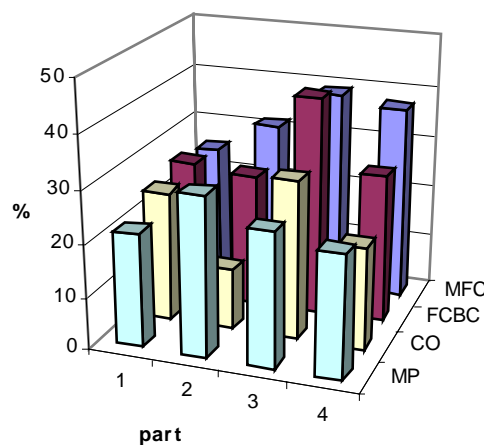
In order to assess whether subjects used these patterns more often than randomly playing agents we followed the following procedure. We, first, added up all rounds that lie within a sequence identified as a pattern. We then used simulated agents to produce the corresponding sequences of play. Each agent corresponded with a subject playing a forty-round repetition. The agent chose his actions with fixed probabilities that were determined by the relative frequency of play by the actual subject. Of course, we limited the simulation to the number of rounds until coordination set in in the actual play. Finally, we calculated the number of rounds captured by one of the above patterns just in the same way as we did for the actual subjects. In order to approximate the distribution of the simulated results we repeated simulations for the

whole of the 12 sessions 1000 times. The results are given in table 11. They show that each pattern is played significantly more often than is suggested by random play.

pattern	observed	simulated mean	p-Value
1	1941	1823.95	0.000
2	633	528.93	0.000
3	683	405.63	0.000
sum	3257	2758.50	0.000

**Table 11: Observed and simulated mean number of rounds captured by patterns (p-values are determined using t-tests)**

The following figure 18 shows that the occurrence of patterns is a rather stable phenomenon. Playing patterns is common in all games and over all parts.



**Figure 18: Frequency of rounds falling into patterns in percent of all rounds before coordination**

Patterns always account for more than 20% of the total number of rounds before coordination sets in (except for part 2 in CO where there are only few observations) and reach up to 41.8% in part 3 with FCBC as the underlying game. There is neither a clear difference between the games nor a trend over parts to be observed.

We further investigated whether playing patterns is typical for a rather small subpopulation of the subjects and found that 113 out of the 120 subjects that participated in our experiments showed an identified pattern within the 160 rounds of play. Summarising we state

*Observation 10*

The use of patterns is a stable phenomenon throughout all games and independent of the previous experience with the forty-round task. 94% of all subjects eventually play an identified pattern.

## **7. Conclusion**

Our experiment was designed to shed light on the acquisition of information in environments in which players are provided little information about the underlying game. We confronted subjects with a set of simple 2x2-games and let them play anonymously with the same opponent for forty rounds. By this means subjects had the opportunity to figure out the incentives of the underlying game without knowing the exact shape from the beginning. Three of the four games had a single efficient and dominant outcome. Despite the large number of repetitions, subjects had considerable difficulty in coordinating. In the simple coordination (CO) game there was a non-trivial number of pairs that did not reach coordination after forty rounds. In the more subtle games of Mutual Fate Control (MFC) and Fate-Control Behaviour-Control (FCBC) only half of the pairs finally reached coordination. Foreclosure of structural information obviously has a considerable effect on the success of efficient coordination.

Experience is the second variable we investigate in our experiment. We let subjects play four repetitions of the coordination task and found that experience does not significantly increase the number of successful coordinations. However, there was a significant experience effect on the time of coordination. Experience helps reducing the time needed to coordinate. This is a clear sign in favour of the hypothesis that with experience subjects gain sophistication. Efficiency, however, is not significantly affected. To the contrary, we observe a clear downward trend in total payoff for pairs from the third to the fourth repetition of the task. Consequently, individual payoffs do not show a steady increase over repetitions.

In terms of learning we find that the most popular approach to explaining learning behaviour in such bivariate decision tasks, the win-stay lose-change strategy does not predict behaviour very well. After winning a point, subjects repeat the same action only 50 to 80 % of the time. An analysis between the first half and the second half of each part shows that even in the course of learning subjects do not increasingly conform to the win-stay lose-change rule. In some cases rather the opposite can be observed. Changing the action after experiencing a payoff of 0 is even less likely. In this case, subjects appear to rather accord to randomisation, which in three out of the four games is indeed superior to the strict change of action.

A more elaborate consideration of adaptation rules that have been put forward for accounting for behaviour in more recent studies shows that these learning rules are neither able to reveal structural differences of adaptation behaviour. Thus, no learning rule was able to give us theoretical insights into what drives subjects to coordinate more efficiently as experience increases.

The last section gives a hint as to why adaptation rules may not be able to tell the most important part of actual behaviour under little information. Subjects appear to use patterns of play that last for considerably more than one period. These patterns seem to fulfil an exploratory purpose much in the way Kalai and Lehrer (1995) have proposed that subjects approach tasks of decision making involving an uncertain environment. Subjects seem to study the reaction of the environment to their specific patterns and draw some conclusions after terminating their exploration time. For our environment, in which subjects often had to coordinate on an efficient cell, this complicated cognitive approach leads to a disability of the timing of exploratory phases. The result is a massive loss of efficiency as compared to a simple adaptive rule such as win-stay lose-change.

The conclusion that we draw for future research is that for the study of behaviour under little information it is necessary to investigate cognitive processes in more detail. In particular, it is necessary to find out more about which exploratory patterns are being used, how long they last and what inferences are drawn from the reaction of the environment by the subjects. By this means we may well be able to shed much more light on coordination failure and its economic implications that are due to the lack of information.

## **8. References**

- Arickx, M. and E. Van Avermaet, (1981) "Interdependent Learning in a Minimal Social Situation" *Behavioral Science* 26, 229 – 242.
- Börgers, Tilman and Rajiv Sarin (1997) "Learning Through Reinforcement and Replicator Dynamics" *Journal of Economic Theory* 77, 1 – 14.
- Brown, James N. and Robert W. Rosenthal (1990) "Testing the minimax hypothesis: A re-examination of O'Neills game experiment" *Econometrica* 5, 1065 – 1081.
- Bush, Robert R. and Frederick Mosteller (1955) "Stochastic Models for Learning" New York, John Wiley and Sons
- Camerer, Colin (1995) "Individual Decision Making" in *The Handbook of Experimental Economics* by John H. Kagel and Alvin E. Roth (eds.), Princeton University Press.
- Camerer, Colin and Teck-Hua Ho (1999) "Experienced-Weighted Attraction Learning in Normal Form Games" *Econometrica* 67, 827 – 874.
- Chen, Yan and Yuri Khoroshilov (2001) "Learning Under Limited Information" mimeo, University of Michigan, January 2001.
- Cross, John G. (1973) "A Stochastic Learning Model of Economic Behavior" *Quarterly Journal of Economics* 87, 239 – 266.

- Estes, William K. (1950) "Toward a Statistical Theory of Learning" *Psychological Review* 57, 94 – 107.
- Feltovich, Nick (2000) "Reinforcement-Based vs. Belief-Based Learning Models in Experimental Asymmetric-Information Games" *Econometrica* 68, 605 – 641.
- Fudenberg, Drew and David Levine (1993) "Learning Mixed Equilibria" *Games and Economic Behavior* 5, 320 – 367.
- Gilboa, Itzhak and David Schmeidler (1996) "Case-Based Optimization" *Games and Economic Behavior* 15, 1 – 26.
- Güth, Werner (1995) "On ultimatum bargaining experiments – A personal review" *Journal of Economic Behavior and Organization*, 329 – 344.
- Hadi, A.S. (1992) "Identifying multiple outliers in multivariate data" *Journal of the Royal Statistical Society, Series B* 54, 761 – 771.
- Hadi, A.S. (1994) "A modification of a method for the detection of outliers in multivariate samples" *Journal of the Royal Statistical Society, Series B* 56, 393 – 396.
- Harley, Calvin B. (1981) "Learning the Evolutionarily Stable Strategy" *Journal of Theoretical Biology* 89, 611 – 633.
- Huck, Steffen, Hans-Theo Normann and Jörg Oechssler (1999) "Learning in Cournot Oligopoly – An Experiment" *Economic Journal* 109, C80 – C95.
- Kalai, Ehud, and Ehud Lehrer (1995) "Subjective Games and Equilibria" *Games and Economic Behavior* 8, 123 – 163.
- Karandikar, Rajeeva, Dilip Mookherjee, Debraj Ray, and Fernando Vega-Redondo (1998) "Evolutionary Aspirations and Cooperation" *Journal of Economic Theory* 80, 292 – 331.
- Ledyard, John O. (1995) "Public Goods: A Survey of Experimental Research" in *The Handbook of Experimental Economics* by John H. Kagel and Alvin E. Roth (eds.), Princeton University Press.
- Matsushima, Hitoshi (1998) "Learning about Stochastic Payoff Structures" Faculty of Economics, University of Tokyo, discussion paper no. CIRJE-F-7
- Mitropoulos, Atanasios (2000) "Learning under minimal information – an experiment on mutual fate control", Discussion Paper No. 10/2000, Faculty of Economics and Management, University of Magdeburg, Germany
- Mookherjee, Dilip and Barry Sopher (1994) "Learning Behavior in an Experimental Matching Pennies Game" *Games and Economic Behavior* 7, 62 – 91.

- Nagel, Rosemarie and Nicolaas J. Vriend (1997) "An Experimental Study of Adaptive Behavior in an Oligopolistic Market Game" working paper, Universitat Pompeu Fabra, Barcelona, Spain.
- Nowak, Martin and Karl Sigmund (1993) "A Strategy of Win-Stay, Lose-Shift that Outperforms Tit-for-Tat in the Prisoner's Dilemma Game" *Nature* 364, 56 – 58.
- Ochs, Jack (1995) "Games with Unique, Mixed Strategy Equilibria: An Experimental Study" *Games and Economic Behavior* 10, 202 – 217
- Oechssler, Jörg and Burkhard Schipper (2000) "Can you guess the game you're playing?" mimeo, Rheinische Friedrich-Wilhelms-Universität Bonn, June 2000.
- Rabinowitz, L., Harold H. Kelley, and R.M. Rosenblatt (1966) "Effects of Different Types of Interdependence and Response Conditions in the Minimal Social Situation" *Journal of Experimental Social Psychology* 2, 169 – 197.
- Roth, Alvin E. (1995) "Bargaining Experiments" in *Handbook of Experimental Economics*, J. Kagel and A. Roth (eds.) Princeton: Princeton University Press
- Roth, Alvin E. and Ido Erev (1995) "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term" *Games and Economic Behavior* 8, 164 – 212.
- Sarin, Rajiv and Farshid Vahid (1999) "Payoff Assessments without Probabilities: A Simple Dynamic Model of Choice" *Games and Economic Behavior* 28, 294 – 309.
- Selten, Reinhard and Joachim Buchta (1994) "Experimental Sealed Bid First Price Auctions With Directly Observed Bid Functions" In *Games and Human Behavior: Essays in Honour of Amnon Rapoport*, D. Budescu, I. Erev and R. Zwick (eds.) Lawrence Erlbaum Ass., Mahwah, NJ.
- Selten, Reinhard and Rolf Stoecker (1986) "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames" *Journal of Economic Behavior and Organization* 7, 47 – 70
- Shachat, Jason and Mark Walker (2000) "Unobserved Heterogeneity and Equilibrium: An Experimental Study of Bayesian and Adaptive Learning in Normal Form Games" mimeo, May 30, 2000.
- Sidowski, J.B., L.B. Wyckoff, and L. Tabor (1956) "The Influence of Reinforcement and Punishment in a Minimal Social Situation" *Journal of Abnormal and Social Psychology* 52, 115 – 119.
- Slembeck, Tilman (1999) "Low Information Games – Experimental Evidence on Learning in Ultimatum Bargaining" University of St. Gallen, discussion paper no. 9903

- Stahl, Dale. O. (1999) "Population Rule Learning in Symmetric Normal-Form Games: The Model and Estimation Algorithm" mimeo, April 1999.
- Thorndike, Edward L. (1932) "The Fundamentals of Learning" New York, AMS Press
- Van Huyck, John B. (1997) "Emergent Conventions in Evolutionary Games" draft paper cc2\_0\_4, December 1997.
- van Huyck, John B., Joseph P. Cook and Raymond C. Battalio "Adaptive Behavior and coordination failure" *Journal of Economic Behavior and Organization* 32, 483 – 503.

### **Technical note**

Statistical analyses have been done using SPSS 9.0 and Stata 6.0 and 7.0. Stata was also used for drawing the confidence ellipses in section 5.3. Other graphics were drawn using MathCad 6.0 and Excel 97 and 2000.

## ***Appendix A: Instructions (translated from German)***

### ***General instructions***

Welcome to today's experiment of the Magdeburg Experimental Laboratory **MAXLAB**. The experiment in which you are participating is part of a research project on experimental economics and is being financially supported by public institutions.

The experiment is going to be performed using the computers in this laboratory. The computer programme is easy to handle, so that no prior knowledge or practice is required in order to use it. All actions can be performed using the mouse. During the whole experiment complete anonymity among participants will be ensured. That is, no player will ever get to know the actions of any other player. In order to make sure of this, we urge you to remain quiet and not to try to contact any other participant. Any kind of communication between participants is forbidden.

Some players were led to cabins while others are separated only by cardboard walls. This arrangement has no consequences for the present experiment. The cabins do have importance in other experiments. In the present experiment, however, all players are treated in the same way. All players are handed out identical instructions and all players use the same computer programme with which he/she can submit his/her decision.

The whole experiment will probably last for less than two hours. It is divided into three phases. You are already in the first phase in which the instructions are read aloud by the experimenter. After that, you will have the opportunity to ask questions and to clarify possible misunderstandings. Please do not speak aloud. Please do not speak to any of the other participants, either. Instead, please raise your hand, so that a member of our team can approach you and privately answer your question.

In the second phase you will play a fixed number of trial rounds, within which the partner is simulated by the computer. These trial rounds are intended to make you, first, accustomed to the computer programme, and second, get to know the payoff schemes that will also be relevant in the third phase. These decisions serve only as trials and are not paid out. More details on the trial rounds are given below. After the trial rounds you will again have the opportunity to ask questions. As before, we urge you to raise your hand, so that a member of our team can approach you.

Finally, in the third phase for a fixed number of rounds you will face the same decision task as has been shown to you in the trial rounds. Please note two important differences to the trial rounds: First, now you will not play together with a player simulated by the computer, instead your partner is a real player chosen among the other participants within this room. Second, each round is payoff relevant, that is the points from each round that are assigned to you as a result of the payoff scheme, your decision, and the decision of your partner, determine how much money you will get at the end of the experiment. More details concerning the rounds of this third phase follow later on.

At the end of this payoff relevant third phase each player will be paid out in cash. For this purpose each player is separately called to the experimenter's desk. There, he/she will get the money in exchange for a signature.

## The payoff schemes

In each round two persons play with each other. Each of the two persons has a choice between two actions, A or B. The actions of the two players determine the payoff of the two players. For this purpose various payoff schemes are being used. Altogether, there are four payoff schemes. In each round one of the payoff schemes is actually being used. In general a payoff scheme looks as follows:

		player 2	
		A	B
player 1	A	a w	b x
	B	c y	d z

This payoff scheme shows for all possible combinations of actions how many points the two players get. Points for player 1 are given in the upper left corner of each cell (a, b, c, and d), while points for player 2 are given in the lower right corner of each cell (w, x, y, and z). For example: In case player 1 chooses action A and player 2 chooses action B, then player 1 gets exactly b points and player 2 gets x points.

The four actually used payoff schemes are the following. In each row one of these payoff schemes is given. The corresponding payoff matrices within a row are generated by exchange of actions. All matrices within a row, thus, share the same structure. What changes is only the meaning of actions A and B. For each game the structure is briefly explained in words.

### Game 1

		player 2	
		A	B
player 1	A	1 1	0 1
	B	1 0	0 0

		player 2	
		A	B
player 1	A	1 0	0 0
	B	1 1	0 1

		player 2	
		A	B
player 1	A	0 1	1 1
	B	0 0	1 0

		player 2	
		A	B
player 1	A	0 0	1 0
	B	0 1	1 1

Each player has one action with which he can give a point to the other player. Using the other action will result in zero points for the other player.

### Game 2

		player 2	
		A	B
player 1	A	1 1	0 1
	B	0 0	1 0

		player 2	
		A	B
player 1	A	0 0	1 0
	B	1 1	0 1

		player 2	
		A	B
player 1	A	0 1	1 1
	B	1 0	0 0

		player 2	
		A	B
player 1	A	1 0	0 0
	B	0 1	1 1

Player 1: One action gives one point to player 2. The other action gives zero points to player 2.

Player 2: Depending on which action player 1 chooses, one action by player 2 gives one point to player 1 and the other action gives zero points to player 1.

Example: Given the payoff matrix at the far left, player 1 gives a point to player 2 by choosing action A, while player 1 gives zero points to player 2 by choosing action B. In case player 1 chooses action A, player 2 gives a point to player 1 by choosing action A. If player 1 chooses action B then player 2 gives a point to player 1 by choosing action B.

### Game 3

		player 2	
		A	B
player 1	A	1 0	0 1
	B	0 1	1 0

		player 2	
		A	B
player 1	A	0 1	1 0
	B	1 0	0 1

One combination of actions yields one point for each player. In case both players choose the alternative action they again get one point each. If, however, their actions do not “match” with each other then none gets any point.

### Game 4

		player 2	
		A	B
player 1	A	1 0	0 1
	B	0 1	1 0

There is always only one player who gets a point. Player 1 gets the point, if both players choose the same action. Player 2 gets the point, if the players choose differing actions.

### **The trial rounds**

You will be given the opportunity to get accustomed to the computer programme and the payoff schemes. In more detail this trial phase is divided into four parts. Within the first part you will face game 1. In the second part game 2 will be the underlying game. In the third part this will be game 3, and in the fourth part this will be game 4. Within each part you will play exactly four rounds of the corresponding game. Only the first of the alternative matrices will be relevant, that is within the trial rounds only the matrices at the far left of each row will be used. You will always take the role of player 2. Your partner will be simulated by the computer. The computer, thus, in each part will always be player 1 and will play the sequence (A, A, B, B).

Using the mouse pointer you submit your choice by clicking on one of the two buttons, A or B. After that, the choice of your partner, who is simulated by the computer, and your payoff will be shown to you.

Please note that the trial rounds are **not** going to be paid out, that is your actions in this phase do not yield actual payoffs. This trial phase only serves the purpose to make you accustomed to the computer programme and the payoff schemes.

### **The experiment**

The actual experiment is also divided into **4 parts**. Each part now consists of **40 rounds**. Hence, you will play 160 rounds altogether. In each part your partner will be a fixed person. This person will be **randomly assigned** to you at the beginning of each part and will then stay the same person throughout the 40 rounds. We will make sure that you will face a different person in each of the four parts. You will never get to know who this person is.

In each part the **payoff matrix also remains fixed**. The payoff matrix will be **randomly assigned** to you at the beginning of each part and will remain the same throughout the 40 rounds. Each game can occur with equal probability. According to the same random process it will be determined which of the possible matrices will be used. Furthermore, it is equally ran-

domly determined whether you will be player 1 or player 2. Finally, each round is **payoff relevant**, that is for each point you gain you will get exactly **0.30 DM**.

In comparison to the trial rounds you will get less information. In more detail you will **not** be told:

- who your partner is,
- which of the payoff schemes is being used,
- whether you are player 1 or player 2 within the payoff scheme, and
- which actions your partner has chosen in the last rounds.

You will get to know:

- your payoff from the last round, and
- your total payoff.

All information you gather within a part is shown to you in a separate window on the computer screen. At the end of each part this window will be cleared.

### ***Cash payoff***

After the 160 payoff relevant rounds of the third phase you will be separately called to the experimenter's desk. The experimenter will pay you out according to your total payoff. Specifically, you will get a fixed amount of 5 DM for your participation plus 0.30 DM times your total payoff. We will round the amount up to the next integer value. Please sign the receipt before you leave the room.

### **Final remarks**

- Please do not start any programmes, except of those named by the experimenter.
- In case of any kind of problems during the experiment, please contact a member of our team by raising your hand.
- After the experiment, please do not talk to anybody about the content of this experiment, so that future groups start their experiment with the same informational status as you did.

Thank you for your participation. Good luck.

Your **MAXLAB**-Team

## Appendix B

In order to give a formal definition of the starting round of convergence, we first have to introduce some notation. Let the set of possible outcomes of the stage game be denoted by

$$C = \{c_{AA}, c_{AB}, c_{BA}, c_{BB}\}.$$

Then the history of outcomes of 40 rounds of play by a pair  $i$  is denoted by the vector

$$o_i = (o_i^1, \dots, o_i^{40}),$$

whereby each  $o_i^j \in C$ .

Distinguishing only between sequences of successive play of the coordination cell ( $c_{BB}$ ) and sequences of successive non-coordination we can rewrite the history of outcomes in a vector of sequence lengths,

$$s_i = (s_i^0, \dots, s_i^{T_i}),$$

whereby for all  $j$  even all  $s_i^j$  denote the length of the corresponding sequence of successive non-coordinations, and for all  $j$  odd all  $s_i^j$  denote the length of the corresponding sequence of successive coordinations. Note, that only  $s_i^0$  (the first non-coordination sequence) is allowed to have length 0, all other sequences necessarily having length greater than zero. Note also that  $T_i$  denotes the number of sequences in the history of pair  $i$  (including the possibly empty initial sequence of non-coordinations).

The central definition is the first coordination sequence of a coordinating pair  $i$ :

$$j_i^* = \arg \min_{j \in \{1, \dots, T_i\}} \left\{ s_i^j \left| \begin{array}{l} (j \text{ even}) \\ \wedge (\forall \gamma > j, \gamma \text{ even} : s_i^j > s_i^\gamma) \\ \wedge (s_i^j > 3s_i^{j+1}) \\ \wedge (s_i^j \geq 3) \end{array} \right. \right\}.$$

The starting round of coordination (SRC) of a pair  $i$  is then given by

$$SRC_i = \sum_{j=0}^{j_i^*-1} s_i^j + 1.$$