Emergence of Coordination in Evolutionary Games[1]

Claudia Lawrenz[2]
University of Osnabrueck

May 2000

[1] Prepared for the 6th International Conference of the Society for Computational Economics on Computing in Economics and Finance, July 6-8, 2000, Universitat Pompeu Fabra, Barcelona, Spain. This is a preliminary version of May 2000, Version 1.

[2] University of Osnabrueck, Department of Economics, Rolandstr. 8, D-49069 Osnabrueck, Germany, EMail: clawrenz@oec.uni-osnabrueck.de

Abstract

This paper presents a computational model of learning which is intended to capture some basic observations of recent studies of game experiments. Furthermore it should give a satisfactory explanation for the convergence of coordinating behavior in the context of evolutionary games. The model combines elements from replicator dynamics and exponential fictitious play; every player has a collection of behavioral rules and updates it by a genetic algorithm. The game environment is chosen following the experimental settings of Van Huyck, Battalio and Rankin (1997). Depending on the matching protocol and the label treatment distinct levels of coordination emerge, but even within the same settings quite different courses of the experimental sessions can be observed. The presented model is able to capture the qualitative properties of the experimental behavior, especially it can explain all the courses in terms of varying the size of agents' memory and the experimentation rate.

1 Introduction

How do player behave in evolutionary games if there is no unique rational strategy? Individual rationality is not sufficient to attain a coordination in this context. The required rationality is a social not merely an individual phenomenon (Arrow 1986). A game is social rational, if players expect the game to be consistent with the equilibrium. In the presence of anonymous opponents and random matching protocol the players have to construct a strategic model of their environment in response to their experiences.

The model presented here combines experimental and theoretical results of learning in games. We explicitly model the out-of-equilibrium-behavior of heterogeneous agents. Furthermore we assimilate elements of real decision-making-processes and explain the coordinating processes in unknown environments with a computational model of learning.

Especially we investigate a simple asymmetric coordination game. Without fitting the parameters of the model we explain the observed degrees of coordination under different experimental settings in terms of the agent 's memory and the experimentation rate.

This work is in a line with other papers comparing their learning models to observed experimental behavior (see e.g. Cheung and Friedman (1998), Erev and Roth (1999)). However, here we make use of genetic operators (Dawid 1996) combined with elements of exponential fictitious play(Fudenberg and Levine 1998).

In section 2 we introduce the model environment and the design of the algorithm, in section 3 we discuss the simulation results and compare them with experimental findings of (Van Huyck, Battalio and Rankin 1997). Finally we give some conclusions.

2 A Computational Model of Learning in Games

2.1 The Model Environment

The simulated model environments, 2x2 matrix games, are chosen as simple as possible. We concentrate on the explanation of the coordination behavior and the construction of beliefs concerning the opponent's actions. Not captured are moral or psychological aspects as fairness, altruism or envy. Furthermore there is no incentive or possibility to cooperate or to play multistage strategies.

We rather explicate stylized facts from recent game experiments, such as that subjects act non-deterministically, response to changes in opponent´s behavior even if they occur in very late periods(Erev and Roth 1998), they form beliefs about their opponent(Cheung and Friedman 1998), use the expected future performance of their strategies as selection criterion (Van Huyck forthcoming) and play at least temporarily suboptimal strategies.

These observations violate game theoretical predictions and suggestions from other learning and evolutionary concepts as replicator dynamics.

2.2 Learning by Experimentation and Recombination

At the beginning of a game sequence every player has a possibly different collection of ideas how to react to observed opponent´s behavior. In the first period she chooses one rule per chance and applies it. Then she notices her opponent´s action and her yielded payoff. She updates her assessment of behavior and strengthens or weakens the importance of the applied rule. Through this the probability with that the rule is played in the next period is altered. After a round the agent revises her rules. At the end of the revision process she checks whether the newly created rules promises better results than the old ones. If this is the case she maintains the changes, otherwise she cancels them.

2.3 Evolutionary Algorithm Application

Eight subjects are playing a 2x2 matrix game against each other in four simultaneous games per trial. The players know that they are matched randomly and don't know their opponent. Every player is totally defined by an array of six behavioral rules and her data collection of opponent´s behavior. A rule i of player j is encoded by a string $\rho_t^{j,i}$ of length 6, its structure is shown in table 1.

c^0	c^1	c^2	c^3	c^4	c^5
δ_i	p_0^i	$p_{0.2}^i$	$p_{0.4}^i$	$p_{0.6}^i$	$p_{0.8}^i$

table 1: rule structure

All elements c^k are drawn from the interval [0,1]. The player uses this rule to determine her next action. Firstly she transforms the perceived history of her opposing players into an assessment of the opponent´s actual behavior y_t.

$$y_t^{(j,i)} = \frac{a_{t-1}^{(-j)} + \sum_{m=1}^{m_{\max}} \delta_i^m a_{t-1-m}^{(-j)}}{1 + \sum_{m=1}^{m_{\max}} a_{t-1-m}^{(-j)}} \tag{2.1}$$

with

$$a_t^{(-j)} = \begin{cases} 1 & \text{if j´s opponent takes strategy 1 in period t} \\ 0 & \text{otherwise} \end{cases}$$

The discount factor δ_i is prescribed by subject´s j ith rule. This assessment is widely used and a generalization of the assessment in fictitious play (Cheung and Friedman 1998). But in addition we impose a memory constraint of $m_{\max}$ periods. Through this procedure player j using strategy i has formed the expectation that her opponent will play strategy 1 with the probability $y_t^{(j,i)}$.

In the next step the agent chooses a response $p_t(y_t)$[1] that specifies the probability strategy 1 is played by agent j and is determined by the behavioral rule i.

$$p_t(y_t) = \begin{cases} c_1 & \text{für } y_t \epsilon [0; 0,2) \\ c_2 & \text{für } y_t \epsilon [0,2; 0,4) \\ c_3 & \text{für } y_t \epsilon [0,4; 0,6) \\ c_4 & \text{für } y_t \epsilon [0,6; 0,8) \\ c_5 & \text{für } y_t \epsilon [0,8; 1] \end{cases} \tag{2.2}$$

The player reacts in the same way if she assesses the opponent is playing strategy 1 with 0 or 19%[2]. After executing a rule i the agent collects the action $a_t^{(-j)}$ and updates the weight of rule i

$$\bar{u}_t^j(i) = \frac{1}{q_t^i w_{t-1}(i)} \left[\tilde{u}_t^j - \bar{u}_{t-1}^j \right] + \bar{u}_{t-1}^j \tag{2.3}$$

This fitness definition is deducted from the analysis of smooth fictitious play(Fudenberg and Levine 1998), $\tilde{u}_t^j$ is the payoff of player j in period t, $\bar{u}_{t-1}^j$the average payoff realized by j during the course of the game. $w_{t-1}(i)$ is the number of times rule i has been applied during the last $m_{\max}$ periods, q_t^i is the ex-ante-probability of rule i, i.e., the probability with that i has been

[1] If the context is clear, for simplicity indices are omitted.

[2] Consequently an assessment of the opponent playing strategy 1 of "almost sure" (80-100%) to "unlikely" (0-20%) would be sufficient.

chosen before the round. This probability equals the relative fitness of rule i according equation (2.4)

$$q_t^i = \frac{\bar{u}_t^j(i)}{\sum_{k \epsilon S^j} \bar{u}_t^j(k)} \tag{2.4}$$

where S^j is the set of rules of subject j. After the weighting up of the executed rule all rules are subjected to a mutation, crossover and election operator.

Because of the coding of real numbers the mutation operator has to be altered compared to the standard genetic operator (see as a reference Goldberg (1986)). A mutation of an element of the string representing the rule occurs with probability p_{mut}. If mutation takes place in c_{t+1} at position k, the element is transformed into

$$c_{t+1}^k = \begin{cases} 0 & \text{falls } c_t^k + m \cdot 0,1 < 0 \\ 1 & \text{falls } c_t^k + m \cdot 0,1 > 1 \\ c_t^k + m \cdot 0,1 & \text{sonst} \end{cases} \tag{2.5}$$

where m is random variable with following distribution

m	-5	-4	-3	-2	-1	1	2	3	4	5
P	0.0375	0.0625	0.0625	0.0375	0.25	0.25	0.125	0.0625	0.0625	0.0375

The mutations are blind and not endogenously determined.

Subsequently the crossover operator is applied. We realized uniform crossover; the rules are randomly paired. With probability $p_{cross} = 0.7$ crossover takes place and every component is exchanged with the corresponding component of the juxtaposed rule with a probability of 50%.

ρ_t^j	$0,7$	$0,8$	$0,3$	$0,3$	$0,1$	$0,3$	$\Rightarrow$	ρ_{t+1}^j	$0,7$	1	$0,3$	$0,4$	$0,6$	$0,1$
ρ_t^k	1	1	$0,5$	$0,4$	$0,6$	$0,1$		ρ_{t+1}^k	1	$0,8$	$0,5$	$0,3$	$0,1$	$0,3$

An example for uniform crossover

At the end of the revision process the player judges whether the newly created rules have a higher expected future performance than the old ones. For this she calculates the potential of the rule $\pi(\rho^i)$ that is the weighted expected return of the rule over all possible probability distributions over the opponent´s strategies.

	A	B
A	$0,0$	$1,1$
B	$1,1$	$0,0$

table 2: a simple coordination game

For the simple coordination game (table 2) we will investigate later on, the potential is computed by

$$\pi(\rho^i) = \frac{1}{2}\left[\, y_t^i \cdot (1 - c_k^i) + (1 - y_t^i) \cdot c_k^i \right] + \frac{1}{8}\left[\sum_{l=1, l \neq k}^{5} c_l^i \cdot \left(1 - \frac{1}{5}\frac{2l-1}{2}\right) + (1 - c_l^i) \cdot \frac{1}{5}\frac{2l-1}{2} \right]$$

where c_k^i is the component of the rule that prescribes the reaction to y_t^i. This is somehow the expected profit of the next period; the component of the rule, that is expected to be the most relevant, gets a higher weight than the others. If the potential of the created strings is larger than that of the old ones, the changes by mutation and crossover are maintained, otherwise the player keeps the old rules unchanged. The new rules inherit the fitness from the old ones.

3 Simulation results

We compare our simulation results with the experimental findings of Van Huyck et al. (1997). They investigate two sources of mutually consistent behavior, population and label treatment. A group of eight to 14 players is randomly paired and plays the matrix game shown in table 3.

		No labels				Labels	
		Other participant´s choice				Column choice	
		1	2			1	2
Your Choice	1	0	40	Row Choice	1	0	40
	2	40	0		2	40	0

table 3: payoff matrix in the experiments of (Van Huyck et al. 1997)

The experimental game sequences last for 30 to 75 periods. An overview over the results is given in figure 8. In the simulation studies first beliefs are always set according to the first choices in the experimental observations.

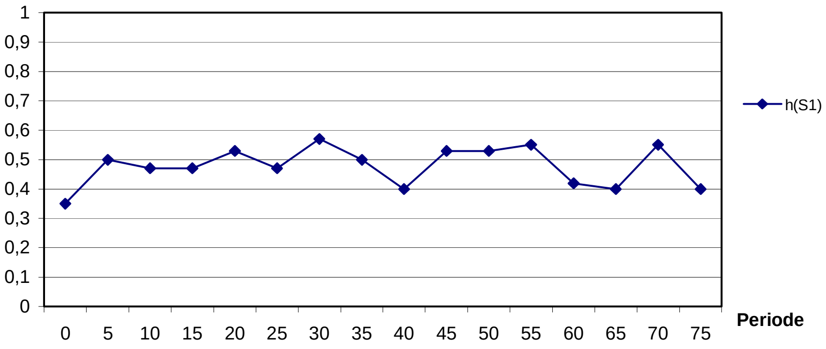


Figure 1: frequency of strategy 1, 1-population-game

3.1 A 1-Population Game

In this setting the players don´t know if they are row or column players. There is no chance to coordinate. Consequently the analysis of the Nash equilibria and replicator dynamics suggests that players use the asymptotically stable equilibrium mixed strategy p(strategy 1)=0.5.

The simulation results in figure 1 report the development of the frequency of strategy 1 in the population. Every data point is an aggregation over five periods and all players. The simulation course exhibits a persistent fluctuation around the predicted equilibrium frequency. The average earning per period of 0.25$ equals almost exactly the predicted earning under the equilibrium strategy.

Testing the null hypothesis on population level that the players don´t use the equilibrium mixed strategy can be rejected in both, experimental and simulation data. But investigating the transitions between states reveal that states beyond the equilibrium strategy are more likely to persist than below , in the simulation data states below the equilibrium strategy are more likely to persist. Consequently regarding the transition states in both, experimental and simulation data, the null hypothesis that players use the equilibrium mixed strategy can be rejected. Furthermore there doesn´t exist a purification of the equilibrium strategy in which one half of the population

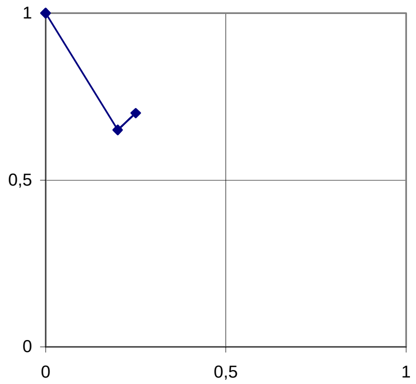


Figure 2: frequency of strategy 1, 2-population game, course 1

is always playing strategy 1 and one half always playing strategy 2.

The deviations from equilibrium strategy become greater if the memory decreases. Coinciding with this the subjects favor rules that specify a discount factor near 1.

3.2 A 2-Population Game

Using a 2-population protocol the players are divided into two subgroups. Every round one member of a subgroup is randomly paired with a member of the other subgroup. Replicator dynamics predict an unstable fixed-point $(0.5, 0.5)$ and two asymptotically stable fixed-point $(1, 0)$ and $(0, 1)$.

Quite different courses can be observed here. The figures 2-4 the frequency of strategy 1 in population 1 on the horizontal axis and of population2 at the vertical axis. Each data point is an aggregation over all players of the subgroup and five periods.
With the usual setting of a memory constraint of 20 periods and a mutation rate of 3 % the course shown in figure 2 is created. Players of population 1 stick on rules that impose a high probability of playing strategy 2 and consequently strengthen this rule. A convention is totally established after 14 periods.

However, if the experimentation rate is increased to 10 % and the memory size shrinks down to ten periods the course of figure 3 arises. Caused by the higher degree of randomization in player´s behavior the coordination sequences are interrupted several times. After 45 periods a degree of coordination of 85 % in the simulated and 89% in the experimental session is

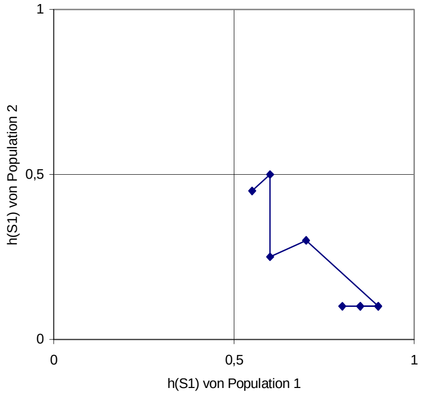


Figure 3: frequency of strategy 1, 2-population game, course 2

attained.
The third course shown in figure 4 results from players starting with misspecified rules. The populations get caught by the mixed equilibrium and after 45 periods still 65% of population 1 and 60 % of population 2 play strategy 1.

This is the only case –in the simulation and in the experiment– when the average earning stays with 0.17 $ distinctly under the average earning that is attained if the equilibrium mixed strategy is taken. However, in the simulation player starts to coordinate after about 100 periods towards (0,1).

3.3 Label Treatment

Another possibility to allow for coordination processes is to tell a player at he beginning of every round whether he is row or column player. Van Huyck et al. test if the players realize that their payoff matrix has changed according to table 4.

	σ_{11}	σ_{12}	σ_{21}	σ_{22}
σ_{11}	0,0	20,20	20,20	40,40
σ_{12}	20,20	40,40	0,0	20,20
σ_{21}	20,20	0,0	40,40	20,20
σ_{22}	40,40	20,20	20,20	0,0

table 4: payoff matrix, if strategies depend on labels

σ_{ij} means "play strategy i as row player and j as column player".

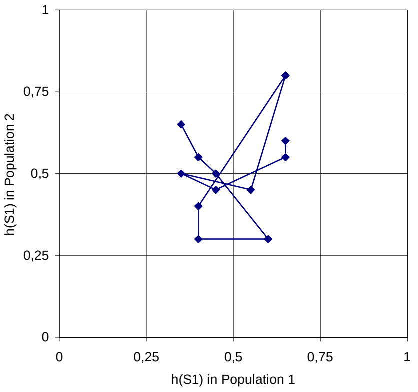


Figure 4: frequency of strategy 1, 2-population protocol, course 3

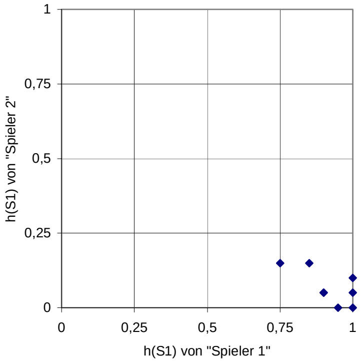


Figure 5: frequency of strategy 1, 1-population game with labels, course 1

To deal with the label treatment our model design has to be extended. A rule is now encoded by a string of length 11, the elements 2 to 6 (7 to 11) encode the player´s response if she is row (column) player. She collects the data about the opponent´s actions separated for each label.

The experimental and the simulation results demonstrate that the subjects generally make use of the additional information given by the labels. However, the achieved degree of coordination is significantly lower than in the two-population treatment. The courses in figure 5-7 show the frequency of strategy 1 token by players labeled "row player" on the horizontal and of "column players" on the vertical axis. Data points are aggregations over five periods and all row and column players respectively .

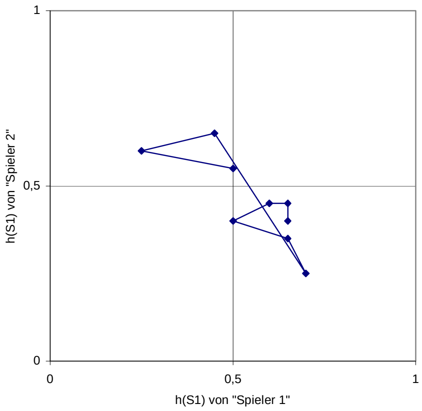


Figure 6: frequency of strategy 1, 1-population game with labels, course 2

The first course in figure 5 is somehow atypically because there is a straight forward development to the equilibrium $(1, 0)$. The average earning of 0.37$ is the highest achieved in all courses. This can be explained by the high degree of coordination (0.75, 0.15) established already at the beginning of the game sequence.

If there isn´t a straight development to an efficient equilibrium in the very first periods, there persists a high rate of experimentation for a couple of rounds. The figures 6 and 7 show courses in which the experimentation rate is 30%.

Depending on the first choices the players get caught by the equilibrium mixed strategy in figure 6 or they circle around a degree of coordination of 65% in figure7. In both courses the players don´t learn to make use of their labels to establish a convention. There doesn´t evolve a unidirected structure in the rules.

3.4 Summary of results

Figure 8 sums up the results of the experimental sessions by Van Huyck et al. and our simulation studies. During the course of the experimental sessions a full degree of coordination can only be observed in one session of the 1-population-label treatment and in one of the 2-population sessions.

However, the simulation results show that in the intermediate term the equilibrium mixed strategy is not stable in the 2-population-protocol.

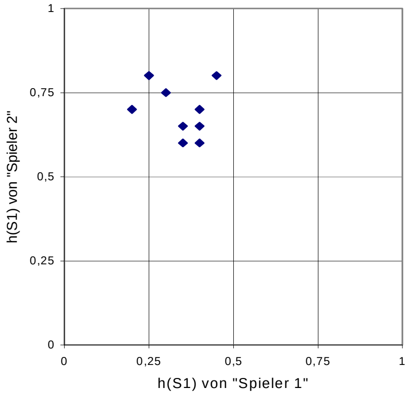
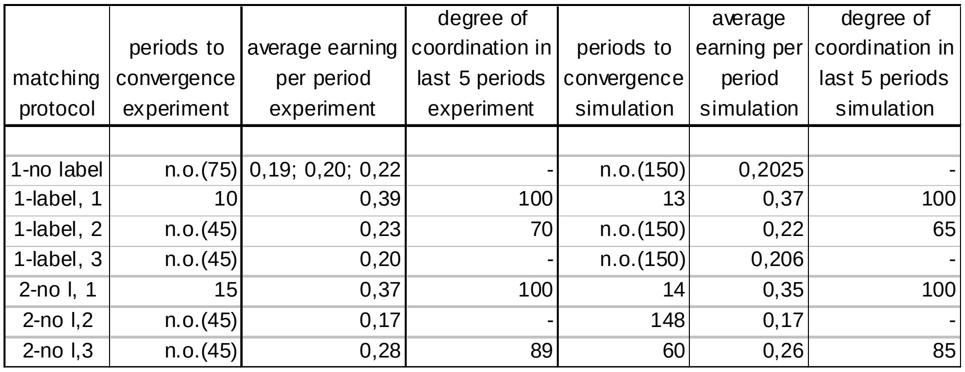


Figure 7: frequency of strategy 1, 1-population protocol with labels, course 3



matching protocol	periods to convergence experiment	average earning per period experiment	degree of coordination in last 5 periods experiment	periods to convergence simulation	average earning per period simulation	degree of coordination in last 5 periods simulation
1-no label	n.o.(75)	0,19; 0,20; 0,22	-	n.o.(150)	0,2025	-
1-label, 1	10	0,39	100	13	0,37	100
1-label, 2	n.o.(45)	0,23	70	n.o.(150)	0,22	65
1-label, 3	n.o.(45)	0,20	-	n.o.(150)	0,206	-
2-no l, 1	15	0,37	100	14	0,35	100
2-no l,2	n.o.(45)	0,17	-	148	0,17	-
2-no l,3	n.o.(45)	0,28	89	60	0,26	85

Figure 8: experimental results by Van Huyck et al. and simulation results

Some more general conclusions are that the short- and intermediate-term learning depend highly on the first choices. A development to a non-equilibrium-state is possible, also to a non-efficient equilibrium.

The players processes additional information about the population treatment more efficiently than the information about labels.

4 Conclusions

We presented a computational model of learning in evolutionary games that is able to capture the qualitative and aggregated quantitative experimental observations by (Van Huyck et al. 1997). The model design overcomes several shortcomings of the application of genetic algorithms and of analytical learning theories as replicator dynamics.

Our main results are that first choices, the experimentation rate and the agents´memory size are crucial for the course of the sessions. Additional informations concerning labels or populations are not processed efficiently. Experimentation, recombination, the future performance and the diversity of rules are identified as important elements of the decision making process.

References

Arrow, Kenneth J., "Rationality of self and others in an economic system," in R. Hogart and M. Reder, eds., *Rational Choice. The contrast between exconomics and psychology*, University of Chicago Press, 1986, pp. 201–216.

Cheung, Ying-Wong and Daniel Friedman, "A comparison of learning and replicator dynamics using experimental data," *Journal Of Economic Behavior And Organization*, 1998, *35*, 263–280.

Dawid, Herbert, *Adaptive Learning by Genetic Algorithms*, Springer, 1996.

Erev, Ido and Alvin E. Roth, "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Journal*, 1998, *88* (4), 848–881.

___ **and** ___ , "The effect of adding a constant to all payoffs: experimental investigation, and implications for reinforcement learning models," *Journal Of Economic Behavior And Organization*, 1999, *39* (1), 111–128.

Fudenberg, Drew and David K. Levine, *The Theory of Learning in Games*, The MIT Press, 1998.

Goldberg, David E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1986.

Huyck, John B. Van, "Emergent Conventions in Evolutionary Games," in Charles R. Plott and Vernon L. Smith, eds., *Handbook of Experimental Economics Results*, forthcoming.

___ **, Raymond C. Battalio, and Frederick W. Rankin**, "On the Origin of Convention: Evidence From Coordination Games," *The Economic Journal*, 1997, *107*, 576–597.