Chmura, Thorsten; Pitz, Thomas

**Working Paper**

# An Extended Reinforcement Algorithm for Estimation of Human Behaviour in Congestion Games

Bonn Econ Discussion Papers, No. 24/2004

**Provided in Cooperation with:**
Bonn Graduate School of Economics (BGSE), University of Bonn

*Suggested Citation:* Chmura, Thorsten; Pitz, Thomas (2004) : An Extended Reinforcement Algorithm for Estimation of Human Behaviour in Congestion Games, Bonn Econ Discussion Papers, No. 24/2004, University of Bonn, Bonn Graduate School of Economics (BGSE), Bonn

This Version is available at:
https://hdl.handle.net/10419/22901

Bonn Econ Discussion Papers

Discussion Paper 24/2004

# An Extended Reinforcement Algorithm for Estimation of Human Behaviour in Congestion Games

by

**Thorsten Chmura, Thomas Pitz**

December 2004


Bonn Graduate School of Economics
Department of Economics
University of Bonn
Adenauerallee 24 - 42
D-53113 Bonn

The Bonn Graduate School of Economics is
sponsored by the

Deutsche Post World Net
MAIL EXPRESS LOGISTICS FINANCE

# An Extended Reinforcement Algorithm for Estimation of Human Behaviour in Congestion Games

**T. Chmura**[a,1] **, T. Pitz**[a,2]

[a] Laboratory of Experimental Economics
Adenauerallee 24-42, 53113 Bonn, Germany

**Abstract:** The paper reports simulations applied on two similar congestion games: the first is the classical minority game. The second one is a asymmetric variation of the minority game with linear payoff functions. For each game simulation results based on an extended reinforcement algorithm are compared with real experimental statistics. It is shown that the extension of the reinforcement model is essential for fitting the experimental data and estimating the players behaviour.

**Key Words:** congestion game, minority game, laboratory experiments, reinforcement algorithm, payoff sum model

**Acknowledgements:** We are grateful to the BMBF for financial support.

## 1 The Investigated Games

### 1. 1. Concestion Game I (CI) – The Minority Game

The first discussed congestion game (CI) is a well known minority game. The set-up of the minority game introduced by W. Brian Arthur is the following: a number of players n have to choose in several periods whether to go to a place A or B. Those players who have chosen the less crowded place win, the others lose. The number of players in each Simulation was 9, the number of periods was 100. The players get an payoff $t_A$ and $t_B$ depending on the numbers $n_A$ and $n_B$ of participants choosing *A* and *B*, respectively:

$$t_A = 1,\ t_B = 0 \ \Leftrightarrow \ n_A < n_B$$

$$t_B = 1,\ t_A = 0 \ \Leftrightarrow \ n_A > n_B .$$

[1] Email: chmura@wiwi.uni-bonn.de URL: http://www.wiwi.uni-bonn.de/labor/

The period payoff was $t_A$ if *A* was chosen and $t_B$ if *B* was chosen. There are no pure equilibria in this game. The pareto-optimum can be reached by 4 players on one and 5 players on the other pace.

### 1.2. Assymetric Congestion Games (CII)

The second congestion game (CII) is a variation of the minority game: the number of Agents in this game was 18, 36, 54, 72 and 90. The number of played periods was 100.

The period payoff for the 18 player setting was 40 – *t* with *t* = $t_A$ if *A* was chosen and *t* = $t_S$ if *B* was chosen, where $t_A$ and $t_B$ depends on the numbers $n_A$ and $n_B$ of participants choosing *A* and *B*, respectively: $t_A = 6 + 2n_A$ and $t_B = 12 + 3n_B$.
All pure equilibria of the game are characterized by $n_A$ = 12 and $n_B$ = 6. The equilibrium payoff is 10 units per player and period. The pareto-optimum can be reached by $n_A$ = 11 and $n_B$ = 7.

The modified payoff functions for the experiments with 36, 54, 72 and 90 agents are $18\lambda$, $\lambda$=2,…,5:

$$p_A = 40\lambda - [6\lambda + 2n_A]$$
$$p_B = 40\lambda - [12\lambda + 3n_B]$$

Table 1 shows all pure equilibria in the CII depending on the number of players.

| Number of Players | Equilibrium | |
|---|---|---|
| | A | B |
| 18 | 12 | 6 |
| 36 | 24 | 12 |
| 54 | 36 | 18 |
| 72 | 48 | 24 |
| 90 | 60 | 30 |

**Table 1:** Pure equilibria in CII depending on the number of participating agents.

CI and CII are frequently interpreted as traffic scenarios [Schreckenberg, Selten, Pitz, Chmua (2003)]. In case of CII place A and place B are understand as road with high (main road) and road with low (side road) capacity road and $t_A$ and $t_B$ as travel times

### 1.2. Experimental Set-up of CI and CII

Each of the games CI and CII with 9 and 18 persons were played 6 times with students at the Laboratory of Experimental Economics in Bonn. Additionally CII was played 1 time with 36, 54, 72 and 90 students. Subjects are told that in each period they have to make a choice between A and B. The subjects of CII set-up did not know the payoff function. They were told that if there A and B is choosen by the same number of people, subjects who had choosen A get a better payoff than subjects who had choosen B. At the end of an experiment each patricians was played an amount in Euro proportional to his cumulated payoff sum he had reached over the 100 periods.

The experimental data statistics is listed and compared with simulation results in chapter 4.

## 2. Reinforcement Learning

### 2.1. Reinforcement Algorithm with Pure Strategies

The reinforcement algorithm with pure strategies already described by Harley (1981) and later by Arthur (1991) has been used extensively by Erew and Roth (1995) in the experimental economics literature. Figure 1 explains the original reinforcement algorithm.

We are looking at player *i* who has to choose among *n* pure strategies 1,…,*n* over a number of periods *t*, *t*=1..*T*. The probabilities of "strategy *x* is chosen by player *i*" is proportional to its

"propensity" $q_{i,x}^t$. In period 1 these propensities are exogenously determined parameters. Whenever the strategy *x* is used in period *t*, the resulting payoff $a_x^t$ is added to the propensity if this payoff is positive. If all payoffs are positive, then the propensity is the sum of all previous payoffs for this strategy plus its initial propensity. Therefore one can think of a propensity as a payoff sum.

**Initialisation**: For each player *i* let $[\, q_{i,1}^1,...,q_{i,n}^1 \,]$ the initial propensity, where *n* is the number of strategies, which are used in the simulations.

**1. period:** Each player *i* chooses strategy *x* with probability $\frac{q_{i,x}^1}{\sum_{y=1}^{n} q_{i,y}^1}$.

**t+1. period:** For each player *i*, let $a_i^t$ the payoff of player *i* in period *t,*

*x* the number of the chosen strategy in period *t*.

CASE I:: $a_i^t \geq 0$:

$$q_{i,y}^{t+1} := \begin{cases} q_{i,y}^t + a_i^t & \text{für } y = x_0 \\ q_{i,y}^t & \text{für } y \neq x_0 \end{cases}$$

CASE II $a_i^t < 0$:

$$q_{i,y}^{t+1} := \begin{cases} q_{i,y}^t & \text{für } y = x_0 \\ q_{i,y}^t + \left| a_i^t \right| & \text{für } y \neq x_0 \end{cases}$$

Each player *i* chooses strategy *x* with probability $\frac{q_{i,x}^{t+1}}{\sum_{y=1}^{n} q_{i,y}^{t+1}}$.

**Figure 1 :** The Reinforcement Algorithm

## 2.2. The Empirical Foundation for an Extended Reinforcement Model

The only pure strategies in CI and CII are "place A" and "place B". This strategies do not represent a players belief about the other participants behaviour. In our extended model we add two further strategies which include the consideration of players about the others based on last periods payoff.

*Direct:* A participant who had a good (bad) payoff may stay on the last periods place (change his last choice). We call this *direct response mode*. A change is the more probable the worse the payoff was. The *direct* response mode is the prevailing one but there is also a *contrarian* response mode.

*Contrarian:* Under the *contrarian response mode* a change of the last choice is more likely the better the payoff was. The contrarian participant expects that a high payoff will attract many others in the next period.

In CI a "*bad*" payoff could obviously defined by 0 a "*good*" payoff by 1. In CII with $18\lambda$, $\lambda$=1,…,5 players the pure equilibrium payoff is $\varepsilon=10\lambda$. Payoffs perceived as "*bad*" tend to be below $\varepsilon$ and payoffs perceived as "*good*" tend to be above $\varepsilon$. Accordingly we classified the strategie of a subject as *direct* if there is a change (stay) after a payoff smaller (greater) than 10. The opposite strategy is classified as *contrarian*.

## 2.3 Measuring Direct and Contrarian Strategies

For each subject let $c_-$ ($c_+$) be the number of times in which a subject changes from A to B or from B to A when there was bad (good) payoff in the period before. And for each subject let $s_-$ ($s_+$) be the number of times in which a subject stays on the same place when there was bad (good) payoff in the period before.

For each subject in the experiments CI and CII a Yule coefficient $Q$ has been computed as follows:

$$Q=\frac{c_-\cdot s_+ - c_+\cdot s_-}{c_-\cdot s_+ + c_+\cdot s_-}\ ,\ \ c_-\cdot s_+ + c_+\cdot s_- \neq 0\,.$$

The Yule coefficient has a range from –1 to +1. In the rare cases subject never (in each period) changes his last choise, we defined Q = 0 because the decision of such a subject

depends not on the last period payoff. Subject with Yule coefficients below –.5 could be understand as classified as direct and subjects above +.5 as contrarian.

## 2.4. Extended Reinforcement Learning

In our simulations of CI 9 Agents respectively of CII 18,46, 54, 72, 90 Agents interact for 100 periods just like in our experiments described in section 3. In CI and CII each player has two pure strategies:

***Place A*:** This strategy consists in taking A.

***Place B*:** This strategy consists in taking B.

After the first period in each of the two games (CI) and (CII) the two extended strategies direct and contrarian are available additionaly:

**(CI) *direct*:** The payoff of a player is 1, then the player stays on the same place last chosen. If his payoff is 0 the players changes (from A to B or from B to A).

**(CI) *contrarian*:** The payoff of a player is 1, then the player changes (from A to B or from B to A). If his payoff is 0 the players will stay on the same place.

**(CII) *direct*:** This strategy corresponds to the direct response mode. The payoff of a player is compared to his median payoff among his payoffs for all periods up to now. If the present payoff is lower then this median payoff, then the place is changed. If the payoff is greater than this median payoff, the player stays on the same place as before. It may also happen that the current payoff is equal to the median payoff. In this case, the place is changed if the number of previous payoffs above the median is greater than the number of previous payoffs below the

median. In the opposite case, the place is not changed. In the rare cases where both numbers are equal, the place is changed with probability ½.

**(CII) *contrarian*:** A player who takes this strategy stays on the last chosen place if his current payoff is smaller then the median payoff among the payoffs for all previous periods and he changes the place in the opposite case. If the current payoff is equal to this median payoff, then he changes the place if the number of previous payoff below the median payoff is greater then the number above the median payoff. If the numbers of previous payoff below and above the median payoff are equal, the place is changed with probability ½.

### 2.4. Initial Propensity

The difficulty arises that the initial propensities must be estimated from the empirical data. For each game CI and CII we did this by varying the initial propensities for the strategies *place A* and *place B* over all integer values from 1 to 10 and the initial propensities for the strategies *direct* and *contrarian* over all integer values from 0 to 10.

For each initial propensity we tested 1000 Simulations. To show the general behaviour of the simulations in Figure: 3-5 one could see several selected statistical parameters depending on the initial propensities listed in figure 2. The numbers refer to the strategies *place A, place B, direct* and *contrarian* in this order.

$$I_0 := \{[q,q,0,0] : 1 \le q \le 120\},\ \ I_1 := \{[q,q,q,0] : 1 \le q \le 120\}, I_2 = \{[q,q,q,q] : 1 \le q \le 120\}$$
$$I_3 := \{[q,q,0,q] : 1 \le q \le 120\},\ I_4 := \{[0,0,q,q] : 1 \le q \le 120\}$$

**Figure 2**: Initial Propensities.

One could see in figure 3 that for each simulation run and each initial propensity the mean number of agents on place A is close to 4.5. The convergence to the theoretical mixed equilibrium was already observed in the simulation data of Roth & Erev.

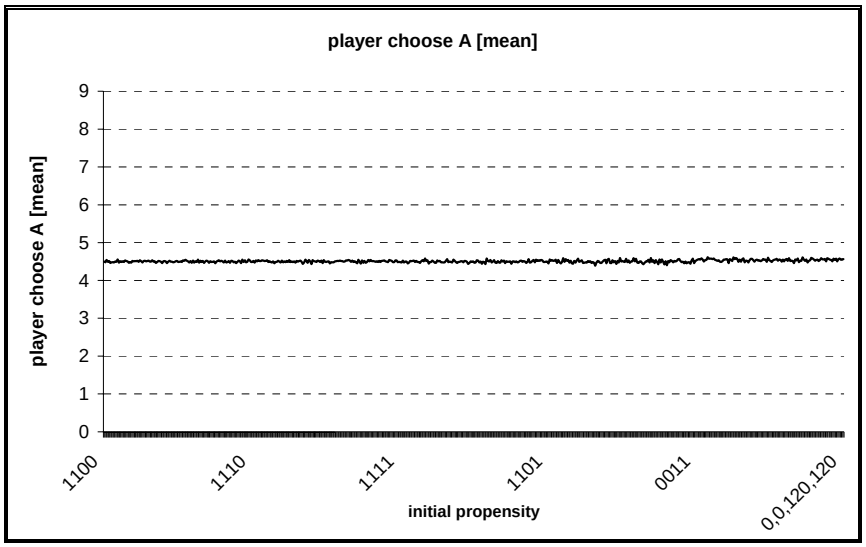


**Figure 3**: Players on A.

The standard deviation of number of players on place A per period (figure 4) is correlated to the number of changes (figure 5) per periods. It got the highest values with propensities from the set $I_1$. In this cases the strategy direct is present and contrarian is absent. The strategy direct forces changes after a the bad payoff 0 which is the most frequent in the majority game.

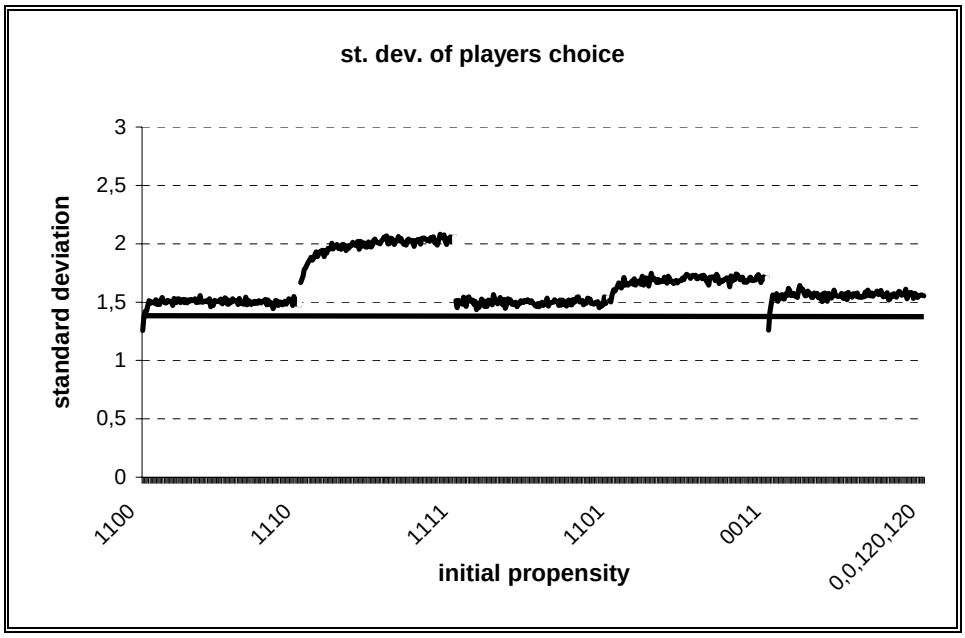


**Figure 4**: Standard Deviation of Number of Players on Place A

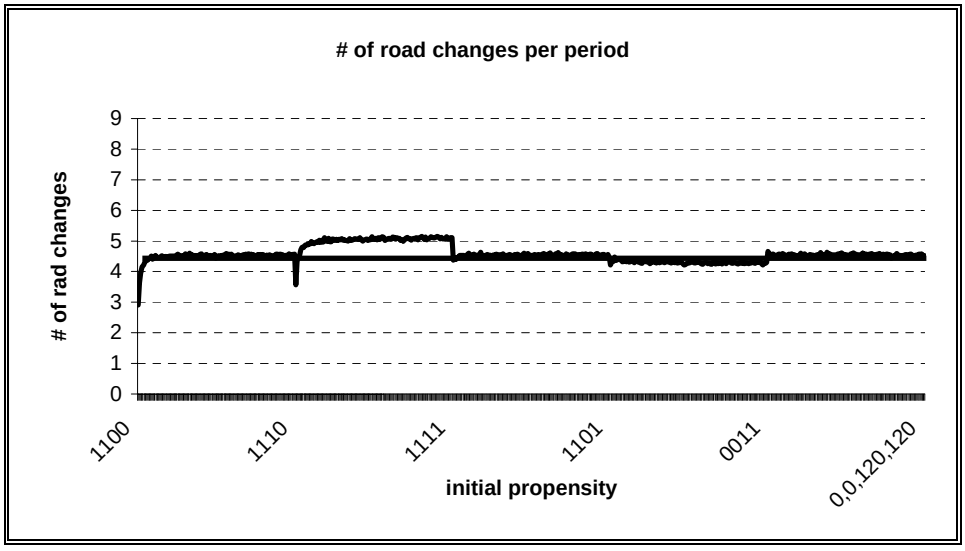


**Figure 5**: Number of Road Changes per Period.

Players with high Yule-coefficients in the experiments are assigned to the direct type, this appears also in the figure 6. For the initial propensities, in which no contrarian change behaviour is implemented for example (1110) step high Yule-coefficients up. For the initial

propensities, in which the contrarian behaviour is favoured for example (1101) all values the of the Yule-coefficients are negative.

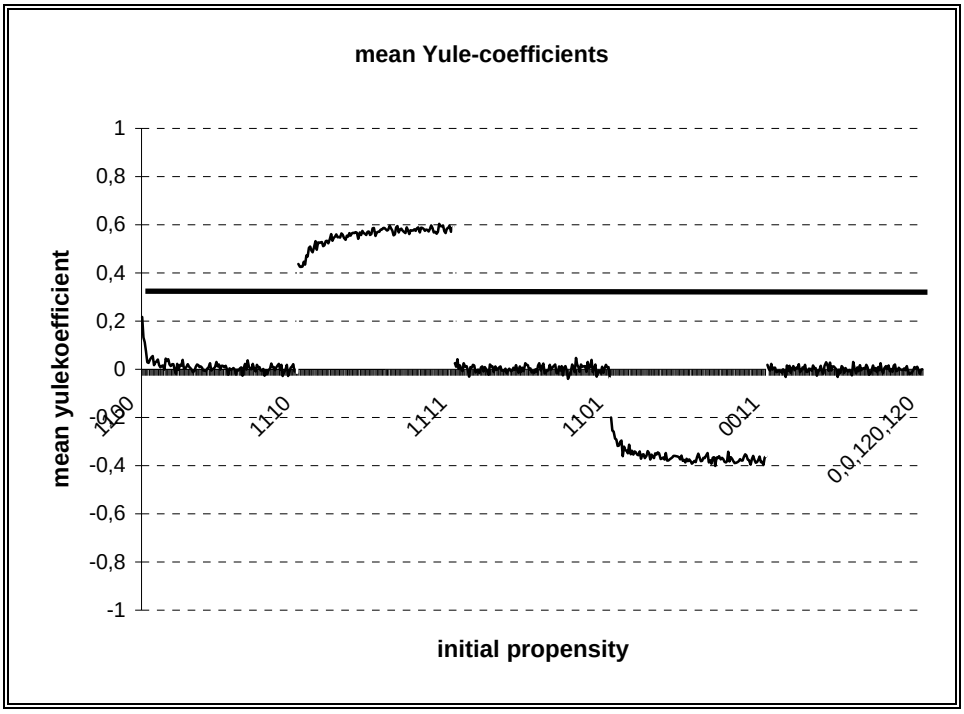


**Figure 6**: Mean Yule-coefficients.

Similar results could get by investigations of the initial propensities for simulations of CII.

## 4. Experimental Statistics and Simulation Results

### 4.1. CI with 9 Players

For each propensity vector $[q_1,...,q_4] \in \{1,...,10\}^2 \times \{0,...,10\}^2$ we run 1000 simulations according to the experiments with 100 periods. The numbers of the propensity vector refer to the strategies *place A, place B, direct* and *contrarian* in this order. We compared the mean values of each of the 1000 simulations of 6 statistical variables which are listed in table 3 with minimum and maximum values of the experimental data.

There were three parameter combinations which satisfied the requirement of yielding means for the six variables between the minimal and maximal experimentally observed values. This was the parameter combination *(1,1,2,1)* and *(2,2,1,1)* and *(3,3,4,2).*

| CI | Experiment | Simulations | | | Experiment |
|---|---|---|---|---|---|
| | **Minimum** | {1,1,2,1} | {2,2,1,1} | {3,3,4,2} | Maximum |
| **Player on A [mean]** | 4,19 | 4.48 | 4.50 | 4.54 | 4.74 |
| **Player on A [standard deviation]** | 0.67 | 1.45 | 1.48 | 1.50 | 1.50 |
| **Changes [mean]** | 0.59 | 4.32 | 4.18 | 4.51 | 5.17 |
| **Period of last Change** | 54.44 | 96.11 | 97.67 | 97.44 | 98.11 |
| **Yule Q [mean]** | -0.01 | 0.10 | 0.04 | 0.14 | 0.87 |
| **Yule Q [standard deviation]** | 0.33 | 0.50 | 0.40 | 0.35 | 0.76 |

**Table 2:** CI – 9 Players - Experimental Minima & Maxima vs. Simulation Means.

The parameter combinations seem to be reasonable vectors of initial propensities. There is no difference between place A and place B, so it is reasonable to have the same propensities for both places. In two of the three vectors the propensity of the direct mode is greater than the others propensities. It is remarkable that no initial propensities which contains only pure strategies does fit the experimental data.

### 4.2.2. CII with 18 Players

In set-up CII with 18 player we got only one parameter combination from the set $\{1,...,10\}^2 \times \{0,...,10\}^2$ which satisfied the requirement of yielding means for the six variables between the minimal and maximal experimentally observed values. This was the parameter combination *(4,3,3,2).* In table 3 compared the mean values of each of the 1000 simulations of 6 statistical variables which are listed in table 3 with minimum and maximum values of the experimental data.

| CII | Experiment Minimum | Simulations {4,3,3,2} | Experiment Minimum |
|---|---|---|---|
| **Player on B [mean]** | 5,85 | 5,95 | 6,17 |
| **Player on B [standard deviation]** | 1,59 | 1,65 | 1,99 |
| **Changes [mean]** | 4,62 | 5,17 | 5,38 |
| **Period of last Change** | 64,78 | 83,73 | 90,39 |
| **Yule Q [mean]** | 0,11 | 0,14 | 0,39 |
| **Yule Q [standard deviation]** | 0,53 | 0,61 | 0,75 |

**Table 3:** CII – 18 Players - Experimental Minima & Maxima vs. Simulation Means.

In the beginning of the game the players know that the capacity of A is greater than B. It seems to be reasonable to suppose that at least in the beginning the pure strategies A and B have a greater propensity sum than *direct* and *contrarian*. Like in CI no initial propensitie which contains only pure strategies fits the experimental data.

### 4.2.3. Simulations of CII with 18, 36, 54, 72, and 90 Players

Finally we compared the mean of six statistical variables of 1000 simulations with 18, 36, 54, 72, and 90 players with experiments of the same number of players. For simulations with 18$\lambda$ players we used the initial propensities $\lambda \cdot (4,3,3,2)$, $\lambda$=2,…,5. The vector *(4,3,3,2)* has been determined in section 4.2.2.

| Statistical Data CII | Data Sorce | Number of Players | | | | |
|---|---|---|---|---|---|---|
| | | 18 | 36 | 54 | 72 | 90 |
| Mean (# players on B) | E | 5,98 | 12,21 | 17,98 | 24,2 | 30,02 |
| | S | 5,95 | 11,91 | 17,9 | 23,83 | 29,02 |
| st. Dev. (# players on B) | E | 1,78 | 2,64 | 3,24 | 4,54 | 5,02 |
| | S | 1,65 | 2,39 | 3,04 | 3,78 | 4,58 |
| Mean (# of place changes) | E | 4,82 | 11,35 | 15,57 | 22,76 | 26,02 |
| | S | 5,17 | 10,07 | 15,98 | 21,32 | 23,04 |
| Mean (last place change | E | 81 | 82 | 86 | 89 | 88 |
| | S | 84 | 89 | 84 | 88 | 90 |
| Mean (Yule-coefficient) | E | 0,28 | 0,14 | 0,22 | 0,2 | 0,24 |
| | S | 0,14 | 0,16 | 0,15 | 0,15 | 0,16 |
| st. Dev. (Yule-coefficient) | E | 0,58 | 0,58 | 0,58 | 0,57 | 0,6 |
| | S | 0,61 | 0,54 | 0,54 | 0,52 | 0,56 |

**Table 5:** CII –Experimental Means (E) vs. Simulation Means (S).

## 5. Conclusion

We have run simulations based on a payoff sum reinforcement model. We applied this model on two similar experimental set ups CI and CII. Simulated mean values of six variables have been compared with the experimentally observed minimal and maximal of these variables. The simulated means were always in this range. Only four parameters of the simulation model, the initial propensities, were estimated from the data. In view of the simplicity of the model it is surprising that one obtains a quite close fit to the experimental data. With a linear transformation of the initial propensity the simulations fit experiments results with a higher number of players.

Two response modes can be found in the experimental data, a *direct* one in which changes follow bad payoffs and a *contrarian* one in which changes follow good payoffs. One can understand these response modes as due to different views of the causal structure of the situation. If one expects that A is crowded in period t A is likely to be crowded in period t+1 one will be in the direct response mode but if one thinks that many people will change in the next periode because it was crowded today one has reason to be in the contrarian response mode.

The strategies *direct* and *contrarian* are necessary to be represented in the simulations for fitting the experimental data. They appear in the simulations as the result of an endogenous learning behaviour by which initially homogeneous subjects become differentiated over time.

It is surprising that a very straightforward reinforcement model reproduces the experimental data as well as shown by table 5. Even the mean Yule coefficient is in the experimentally observed range. In spite of the fact that at the beginning of the simulation the behaviour of all simulated players is exactly the same. It is not assumed that there are different types of players.

## 6. Appendix

Figures 9-14 illustrate the experimental means in comparison to the simulated means of table 5. Black boxes represent the simulated values and the white boxes the empirical data.

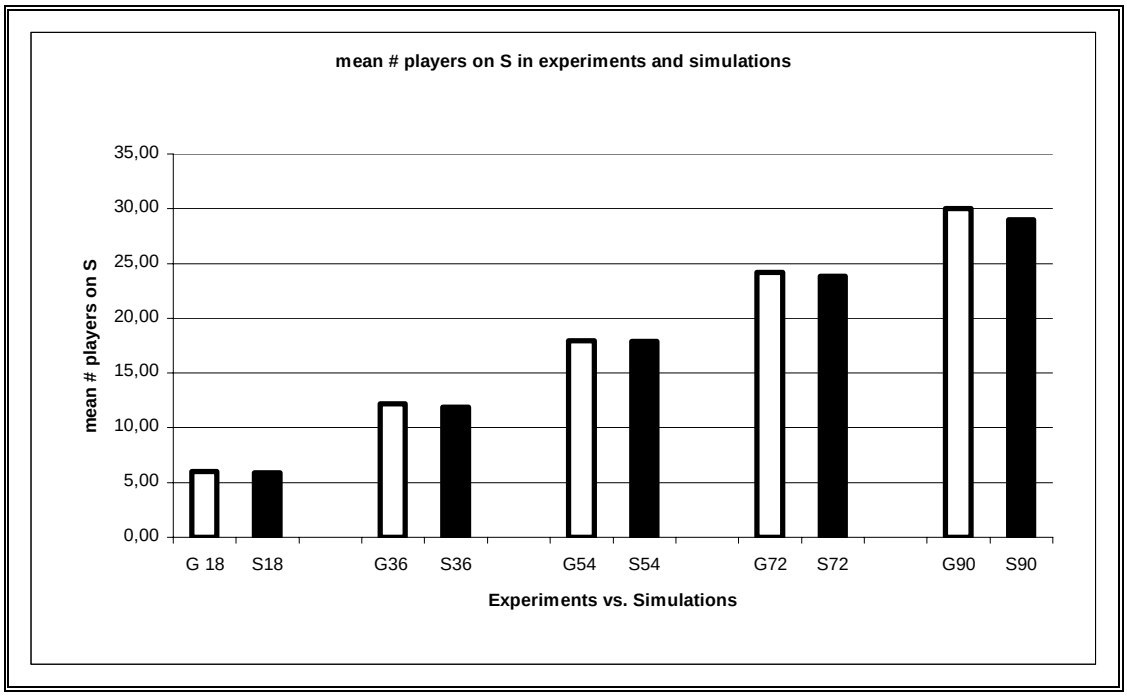


**Figure 9**: Mean Number of Players on S in Experiments and Simulations.

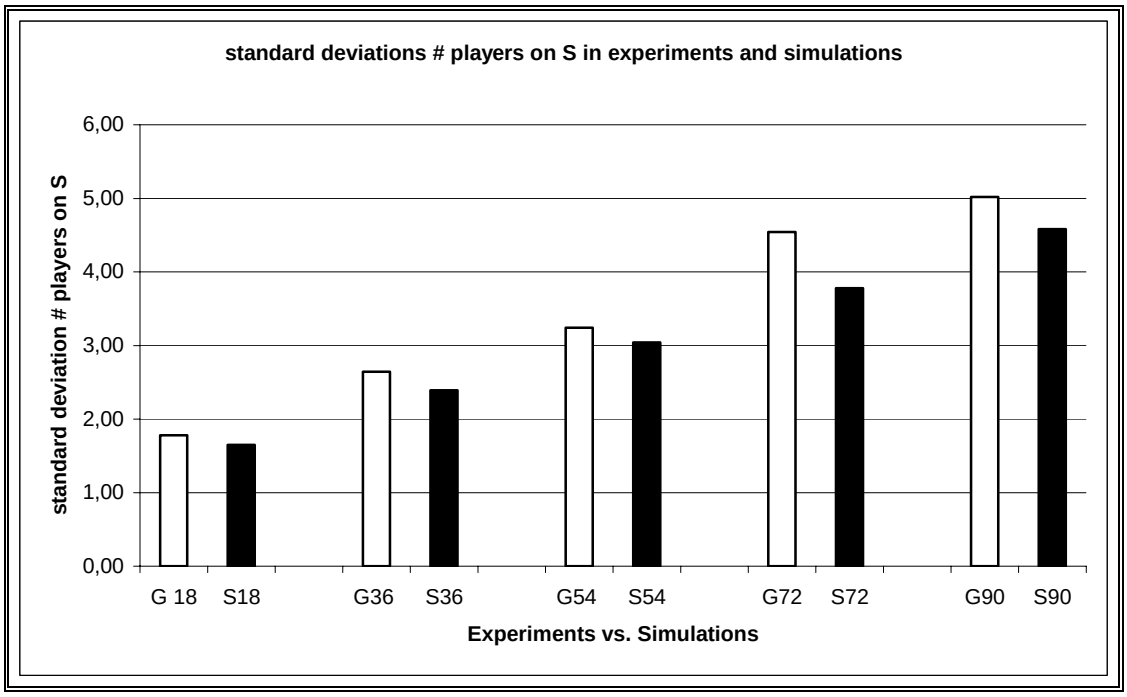


**Figure 10**: Standard Deviation number of Players on S in Experiments and Simulations.

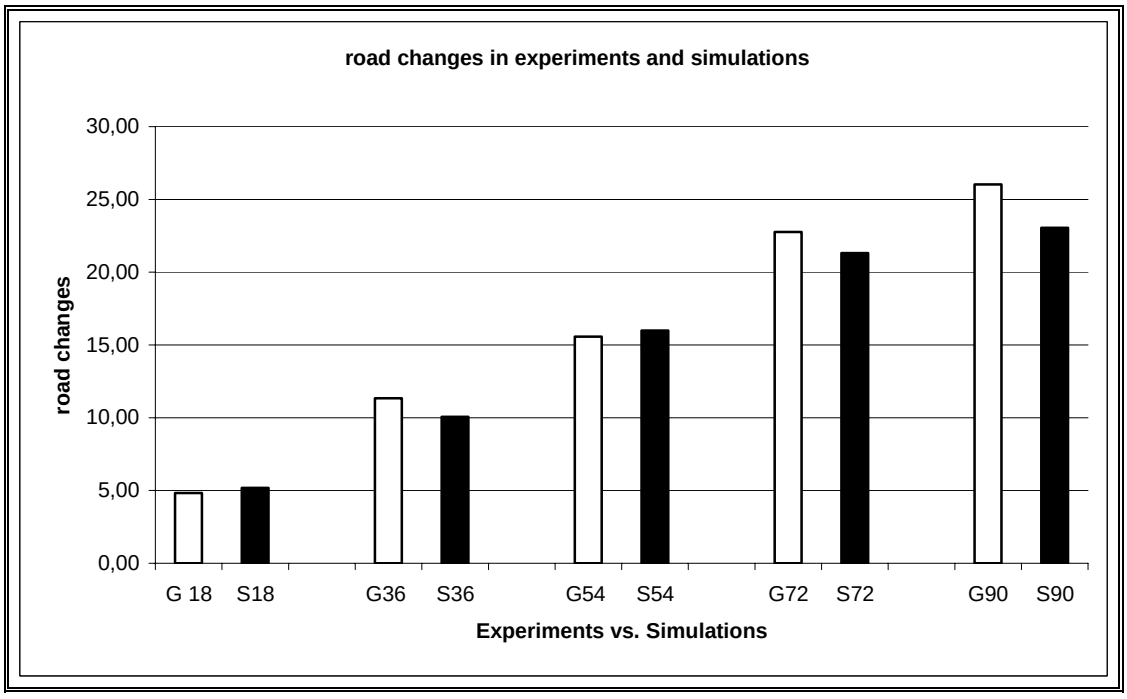


**Figure 11**: Number of Changes in Experiments and Simulations.

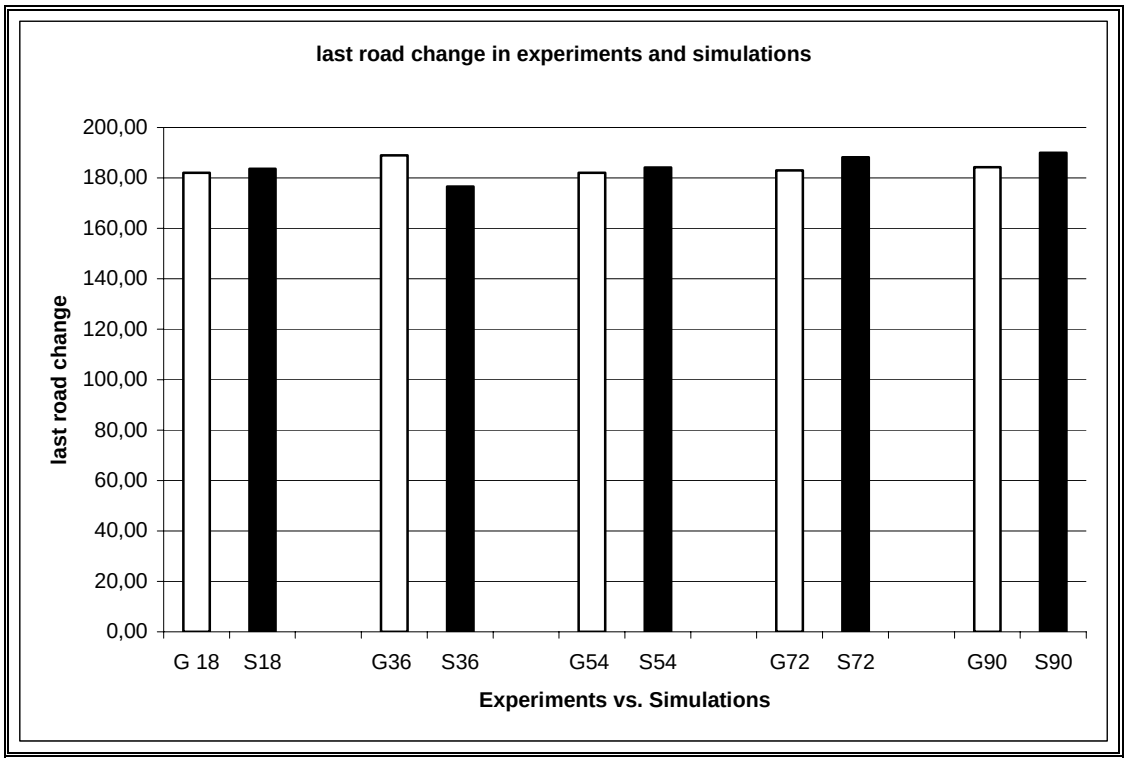


**Figure 12**: Last change in experiments and simulations.

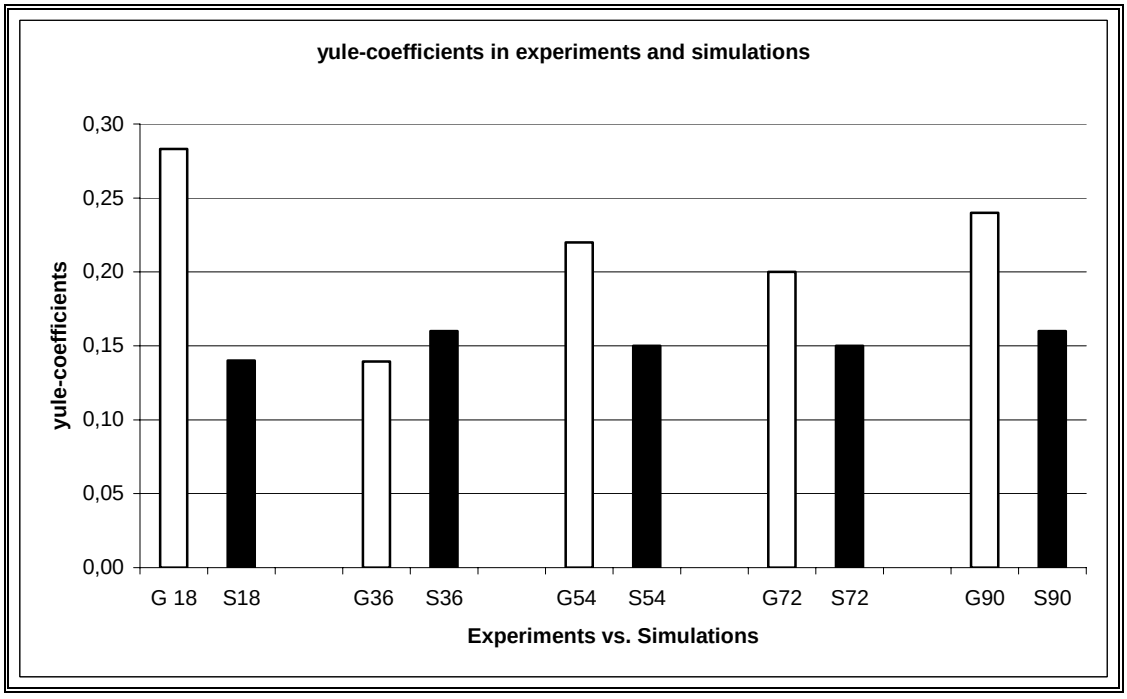


**Figure 13**: Mean yule-coefficients in experiments and simulations.

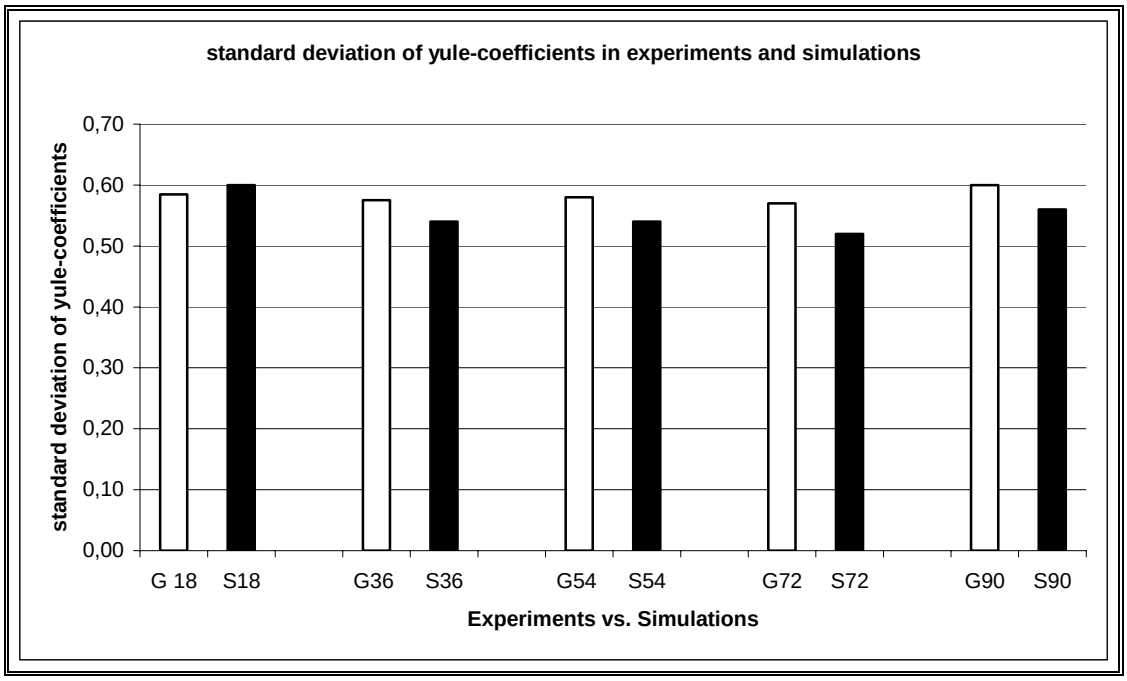


**Figure 14**: Standard Deviation of Yule-coefficients in Experiments and Simulations.

# References

ARTHUR W., B. 1991: Designing economic agents that act like human agents: A behavioural approach to bounded rationality. Amer. Econ. Rev. Papers Proc. 81 May, 353-359.

GIGERENZER, G., TODD, P.M. , AND ABC RESEARCH GROUP (eds.), 1999: Simple heuristics that make us smart. Oxford University Press.

HALL, R.: 1996, Route choice and advanced traveler information systems on a capacitated and dynamic network, Transpn. Res. C 4, 289-306.

HARLEY, C. B.: 1981, Learning in Evolutionary Stable Strategie, J. Teoret. Biol. 89, 611-633.

ROTH, A.E., EREV, I.: 1995, Learning in extensive form games: Experimental data and simple dynamic models in the intermediate term, Games and economic Behavior 8, 164 – 212.

ROTH, A.E., EREV, I.: 1998, Predicting how people play games: Reinforcement learning in games with unique mixed strategy equilibrium, American economic review 88, 848 – 881.

SCHRECKENBERG, M., SELTEN, R., PITZ, T., CHMURA, 2003: Experiments on Route Choice Behaviour, in: Ed. H. Emmerich, B. Nestler, M. Lecture Notes in Computers Science 32, Schreckenberg, Interface and Transport Dynamcis, Springer