Developments in Applied Statistics
Anuška Ferligoj and Andrej Mrvar (Editors)
Metodološki zvezki, 19, Ljubljana: FDV, 2003

Can Simulation Techniques Contribute to Microsociological Theory? The Case of Learning Matrices

Anselm Eder[1] and Walter J. Gutjahr[2]

Abstract

Nearly all simulation techniques have one big disadvantage in common: oversimplification, leading to unrealistic results. And nearly all simulation techniques have one big advantage in common: oversimplification, which is the only way to express a theoretical model in clear terms. The oversimplification in our simulation model consists in assuming two interacting partners whose actions/reactions are determined by only two sets of parameters: 1. a matrix of reaction probabilities which is updated according to the subjective evaluation of each reaction of the partner to the actor's behaviour at each interaction, and 2. a payoff matrix, reflecting each partner's subjective evaluation for each pair of action/reaction, which remains stable over a longer sequence of interactions. Applications of this model to several problem areas, such as socialization agents, game theory approaches, and Ant Colony Optimization in previous publications of the authors, have shown some practical results. Here, we want to focus on three paradigms which we believe can be deducted from our work: 1. The law of sociodynamics: Social systems whose organization is similar to the model conditions formulated above, tend to a decrease in entropy as they get older, in contrast to physical systems, which seem to do the opposite. 2. Cultural values prevail over individual behavioural dispositions: We believe to have found an argument that social systems with properties similar to our model assumptions are likely to provide individuals with behavioural dispositions which depend a lot more on the (culturally more stable) values attributed to behaviour than on initial behavioural dispositions, reflected in reaction probabilities.

And 3. The slower the learning process, the better the results. Optimal solutions to typical dilemma situations such as the iterated prisoner's dilemma are found more frequently when individuals show "silly"

---

[1] Institute of Sociology, University of Vienna, Rooseveltplatz 2, 1090 Wien, Austria; anselm.eder@univie.ac.at.

[2] Institute of Statistics and Decision Support Systems, University of Vienna, Universitätsstraße 5, 1010 Wien, Austria; walter.gutjahr@univie.ac.at.

behaviour, i.e. the rate at which they change their behavioural dispositions is low, whereas a quick learning rate results more often in sub-optimal results for both partners.

1 How it all started: Peter's story

In a therapeutic institution we came across the following situation: A 33-year old man, whom we will call "Peter", and who is called "the boy" by his parents, is brought to the institution by the police. The police had been called by the 72-year old father of "the boy", after a violent attack of this "boy" against his mother. "The boy" had taken drugs and had attacked his mother for her attempt of taking away the drugs from him. In the therapeutic setting, we see "the boy" together with his father, his mother and two therapists. He shows an extremely cooperative behaviour, repeating again and again how much he regrets having caused his beloved parents so much trouble. He does not really understand how all this could have happened. And he is willing to work very hard, with the help of the therapists, to improve his behaviour, so that the otherwise so harmonious family life can be re-established. His father says that he does not understand either how these terrible things can have happened. He has spent all his life trying to provide a solid basis for a well-functioning family. But, naturally, when the son attacks the mother violently, and when the father, with his 72 years of age, does not have the physical strength to protect his wife against their son's attacks, there just isn't anything else he could possibly do than call the police. At this moment, the mother takes a deep breath and gets ready to say something. But the father bends forward and puts his right hand on her left hand, which transforms her attempt to intervene into nothing but a deep sigh.

In the therapist's protocol, we read a very optimistic statement about the prognosis of this case: the whole family is committed to find a cooperative solution of the problem. The compliance of the client is very high, and during a supervision session, the two therapists express high motivation to work with this client and his family. At this point, the supervisor of the two therapists, who had observed the scene just described through a one-way mirror and also videotaped it, got in contact with the authors. He told us about a funny feeling he had during the whole situation. Somehow, he could not get rid of the idea that "the boy" played some kind of game with his parents, and the two therapists joined in, without being very aware of it. They also did not want to look at it from such a perspective, even after a closer look at the client's records revealed that almost exactly the same chain of events had happened three months before: a violent attack on his mother, police intervention, ending by a transfer to (another) therapeutic institution. And an investigation at this other institution revealed another previous episode of almost exactly the same kind, another three months before that. But still, the therapists were reluctant to accept the idea of a game, following a certain set of rules, being played here. Therefore, the question with which we were confronted, was: could there be any way to make the implicit rules of a game plausible to the persons involved in it?

2 The model: Its basic elements

There are three basic elements in our model which need to be explained:

1. A matrix of values which actors attribute to behavioural sequences. Following the terminology of game theory, we call this matrix the Payoff Matrix. Its rows represent the strategies (behaviours) of the actor, or the ego, and its columns represent the strategies (behaviours) of his partner, opponent, or the alter. The entries in the cells represent the amount of benefit which a person receives (or interprets) when the partner chooses the column strategy after the ego had chosen the row strategy.
2. A matrix of probabilities with which the actor reacts to the partner's strategy. Following the terminology of Markov chains, we call this matrix the Transition Matrix. Its rows represent the strategies (behaviours) of the partner, or the alter, and its columns represent the strategies of the ego. The entries in the cells represent the probabilities with which the actor, or the ego, will react to the strategy of the alter, as represented in a row of the matrix, by his own strategy, as represented in a column of the matrix. This means that the cell entries will be numbers between 0 and 1 and add up to 1 in each row. But unlike (homogeneous) Markov processes, which assume this transition matrix to remain stable over time, we assume that at each interaction, the probabilities in this matrix change, depending on ego's evaluation of how the partner has reacted. The amount of change is determined by the entries in the payoff matrix. The details of this change will be described later.
3. The final distribution, showing the percentages with which, in a large number of simulations (interaction stories), every combination of strategies (action-reaction pair) has remained after a long series of interactions. In other words, this distribution gives, for each action-reaction pair, the probability that it occurs in the stable end state resulting from a long interaction process (as to the stability of this end state, cf. the remarks below).

4 What the model does: About the nature of virtual interaction partners

Modelling human interactions has led to relatively complex simulation algorithms in the past, as outlined, e.g., by authors like Schmidt (2000), Mosler (2000), and many others. But in our model, we are trying to look at the possible outcomes of a two-person interaction in which two partners have a limited spectrum of actions to choose from, information about the spectrum of actions of the other, and a given matrix of preferences for each action-reaction-pair. They have no information about what the other partner will do next, and they have no direct influence on the other. But they will change the probabilities for their own behaviour according to their preferences: i.e. according to how much they like the partner's reaction to their own preceding action. This procedure basically follows the paradigm of psychological learning theory: In psychological learning theory, it is either behaviours (in our terminology: "actions", or "strategies") that are reinforced, i.e. whose probabilities increase, or, as in discriminative learning, action-reaction pairs. In our approach, the more flexible model of discriminative learning is used, thus leading to increases in conditional probabilities rather than unconditional probabilities. In Eder, Gutjahr, and Neuwirth (2001), we have tried to show that even a model as simple as this one can make sense in terms of counselling: it can help us understand the process of mutual reinforcement of two partners, leading to sometimes unexpected results. The example discussed there was the situation of a smoking child, and an attempt to figure out in which way even a slight change in preferences of the actors will be likely to produce quite remarkable effects on the probability of smoking. In Gutjahr and Eder (2001), we have tried to show that our model can also be used to understand some properties of Ant Colony Optimization: a heuristic optimization technique for computer-based solutions of decision support problems in several areas of business, management and technology; see, e.g., Dorigo and di Caro (1999). This optimization paradigm ist derived from insights into the behaviour of "society-building insects", obtained by biologists, and is therefore closely related to learning in social contexts. In this paper, we would like to point out some properties of our model which we believe can be used to make a few general statements about the organization of dyadic interaction systems.

A more careful description of the model has been given in the contributions cited above. Here we want give only a short review of how the model is functioning (Eder, Gutjahr, and Neuwirth; 2001).

In our model we have two interacting partners who know absolutely nothing about each other and have no such thing as a "personality", a "character", or a "morality" of any kind. And just like the two prisoners in the well-known prisoner's dilemma, they have no idea what the other one will do. And their decisions are not influenced by any idea of what's good or what's bad. The only thing they know is their own part of their payoff-matrix. *A* knows how much he gets when *B* reacts to *A*'s strategy *i* with *B*'s strategy *j*, for all pairs of possible strategies. In other words: For each possible action of *A,* and for each possible reaction of *B, A* knows how much he will like or dislike it. And the same goes vice versa for *B*.

A starts an activity. And since he has not the slightest idea what he should do, he throws a coin. *B* does the same. *A* again. Now, *B* has the possibility of evaluating what has happened. He interprets his own strategy as a reaction to the preceding strategy of *A,* and *A*'s response to his own strategy as a sign of success or failure of his own strategy.

The fact that our two actors had started with bland personalities is represented in their vectors of reaction probabilities, which we can call transition probabilities in analogy to the theory of Markov chains, although in our case, not transitions between states, but rather between actions are considered: For every possible strategy of *B, A* has a vector which contains the transition probabilities of reacting with each of his possible strategies, given the strategy of *B*. In the beginning, all these probabilities can be the same. In the simplest possible case, when we have only two possible actions for each actor, they are all 0.5 (the coin). When we list these vectors in rows, we get the transition matrix as described above.

Now, as we said, *B* interprets his strategy as a reaction to what *A* had done previously. If *A* had chosen strategy *i* first (by throwing a coin), then *B* will find the probability of reacting to *A*'s strategy *i* by choosing strategy *j* with the probability in row *i* and column *j* of his matrix of transition probabilities.

Furthermore, *B* interprets *A*'s strategy *k* which follows his own strategy *j* as a reaction to this strategy *j* which he just chose. The extent to which he likes or dislikes this reaction is represented in his payoff matrix, in line *j* and column *k*. If he likes the reaction – a positive value in cell (*j,k*) of his payoff matrix – his transition probability of reacting to strategy *i* of *A* with strategy *j* will increase, as a function of the value in column *j* and row *k* in the payoff matrix: First, the probability in row *i* and column *j* is increased by a fixed learning rate, multiplied by the achieved payoff; after that, the whole row *i* is re-normalized to a sum of 1 by division by (1 + (learning rate)*payoff). Thus, *B* will be conditioned to those reactions that have shown to be successful for him.

Exactly the same goes for *A*. In the terminology of learning theory, both actors mutually take the role of environment, conditioning the other one to those actions that rank high in the other's payoff matrix. In this way, we avoid the drawback of psychological learning theory criticized by Bateson and Jackson (1964): an arbitrary determination of the roles of the person providing the stimuli and of the person responding by reactions makes their roles appear as fixed, whereas in reality, they alternate, resulting in a cyclic process.

Apparently, a lot of things can happen to the transition probabilities during a simulation run. Every single simulation, consisting of a number of interaction steps during which transition probabilities are changed, can lead to another transition matrix at the end of the process. – Can it really? A little later we shall see that this is not quite the case.

For the moment, let us just distinguish between two types of transition matrices: first, a transition matrix which allows no prognosis whatsoever about what the most likely reaction of an actor will be. We chose to call such a matrix an "entropic" matrix, and it is characterized by the same entries in each cell, each being 1 divided by the number of columns. And second, a matrix which gives us some information on what an actor will do, given a preceding action of his partner. We call such a matrix a "non-entropic" matrix. The least entropic transition matrix that we can think of is a deterministic one, containing a 1 in each row, all the other entries in each row being 0. Figure 1.1 gives an example of entropic matrices of two actors *A* and *B*, where there are 4 strategies available for *A* and 3 for *B*.

Now the big question is: Assuming that the updating process of an initially entropic matrix through discriminative learning has worked for some time: is there anything that we can say about how an initially entropic matrix will look in the end, or not? Is everything possible, or can we expect an initially entropic transition matrix to converge to something predictable?

It can be shown both theoretically and by repeated simulation that, no matter how entropic a transition matrix is in the beginning, at the end of the interaction story the behaviour of *A* and *B* will always be deterministic. To be precise, this is true if all the values in the payoff matrix are at least marginally greater than zero. The proof for this is straightforward.3)

By what type of transition matrices can a deterministic behavior of both actors be achieved? The two matrices must have a specific relation to each other, effecting that each actor's decision is completely determined by the other actor's previous decision. We call a pair of matrices of this type a **cyclic matrix pair**.4) A cyclic matrix pair forces both actors to run again and again through a deterministic cycle of actions. This cycle can also be the trivial one, consisting of only one fixed action per partner.

One might argue that the matrices in such a pair must be deterministic, i.e., contain only entries 0 or 1, but this is not true: Figure 1.2 shows an example of a cyclic matrix pair. The last row of the right matrix contains noninteger values. Nevertheless, the resulting behavior is still deterministic, because the last row will never have a chance to "play": *A* will never play strategy *4*, and therefore, *B* will never get a chance of using the transition probabilities in his fourth row.

Let us summarize: No, we cannot say which combination of strategies will remain at the end of an interaction story. But we can say that whatever is left from an initially entropic pair of transition matrices at the end of an interaction story, will be deterministic, i.e. non-entropic, and cyclic.

1.1: Entropic state

$$\begin{array}{c|cccc}
\multicolumn{1}{c}{} & \multicolumn{4}{l}{A \longrightarrow} \\
\multicolumn{1}{c}{B} & & & & \\
\multicolumn{1}{c}{\downarrow} & 1 & 2 & 3 & 4 \\ \cline{2-5}
1 & .25 & .25 & .25 & .25 \\
2 & .25 & .25 & .25 & .25 \\
3 & .25 & .25 & .25 & .25
\end{array}
\qquad
\begin{array}{c|ccc}
\multicolumn{1}{c}{} & \multicolumn{3}{l}{B \longrightarrow} \\
\multicolumn{1}{c}{A} & & & \\
\multicolumn{1}{c}{\downarrow} & 1 & 2 & 3 \\ \cline{2-4}
1 & .33 & .33 & .33 \\
2 & .33 & .33 & .33 \\
3 & .33 & .33 & .33 \\
4 & .33 & .33 & .33
\end{array}$$

1.2: Cyclic state

$$\begin{array}{c|cccc}
\multicolumn{1}{c}{} & \multicolumn{4}{l}{A \longrightarrow} \\
\multicolumn{1}{c}{B} & & & & \\
\multicolumn{1}{c}{\downarrow} & 1 & 2 & 3 & 4 \\ \cline{2-5}
1 & 0 & 0 & 1 & 0 \\
2 & 1 & 0 & 0 & 0 \\
3 & 0 & 1 & 0 & 0
\end{array}
\qquad
\begin{array}{c|ccc}
\multicolumn{1}{c}{} & \multicolumn{3}{l}{B \longrightarrow} \\
\multicolumn{1}{c}{A} & & & \\
\multicolumn{1}{c}{\downarrow} & 1 & 2 & 3 \\ \cline{2-4}
1 & 0 & 0 & 1 \\
2 & 1 & 0 & 0 \\
3 & 0 & 1 & 0 \\
\multicolumn{1}{c}{4} & 0.3^{*)} & 0.5^{*)} & 0.2^{*)}
\end{array}$$

Cycle:

$A{:}1 \rightarrow B{:}3 \rightarrow A{:}2 \rightarrow B{:}1 \rightarrow A{:}3 \rightarrow B{:}2 \rightarrow A{:}1$

$^{*)}$: or any other non-integer value between 0 and 1, adding up to 1 in row 4.

Figure 1: Two corresponding transition matrices in an entropic state (1.1) and a cyclic state (1.2).

Intuitively, the importance of the notion “cyclic matrix pair” can be illustrated by our introductory case report: The story of Peter gave us the impression that there are certain behavioural cycles which the actors “play”, without being able to stop the “game”. And each actor perceived himself as being completely determined by somebody else’s preceding action. Nobody had a choice. But not only that: The behaviour repeated, leading to cycles, just as in the above example.

5 The law of sociodynamics

Social systems whose organization is similar to the model conditions formulated above, tend to a decrease in entropy as they get “older”, i.e. the longer the sequence of their interactions, in contrast to physical systems, which seem to do the opposite. In the example in Figure 2, we look at a simulation with two interacting partners with three strategies each. As we have stated above, the process will develop uneven probability distributions in the transition matrices. If we use entropy as a measure for equality of distribution, we get a pattern as in Figure 2 (the symbols p_{ij} in the formula denote the probabilities in the transition matrix). Different simulations may show different slopes and different amounts of decrease, but a decrease can always be observed, and it mostly has the same shape as in Figure 2.

In other words: the more interactive steps we see, the less balanced the probability distributions become: less and less reactions become more and more likely, and after a while the behaviour is no more stochastic at all: to a given action, we have a probability of 1 for one specific reaction, and of 0 for all the others. When this stage has been reached ,–the reaction is completely predictable. We also might say: The entropy of the reaction to this considered action has decreased to its minimum value. This occurs for all actions the considered actor eventually chooses, which reduces the entropy of the overall system. We might express this observation as follows: Whereas thermodynamics tell us that physical systems (where other laws are at work) tend towards a state of **increased** entropy (i.e. unpredictability of their states, or equal distribution of all possible states), when they are uninfluenced by forces outside the system, social systems do the opposite: "Sociodynamics" tell us that social systems tend towards a state of **decreased** entropy, predictability of their states, unequal distribution of all possible states, if no other influences than the mechanism of operant conditioning are at work. We even might call this state of decreased entropy "social order", depending on how generous we want to be with using that term. This social order is manifested by a determinism in behavioural sequences, which, however, can look a little neurotic at times.

We believe to have here a formal representation of a process which is one of the basic questions often asked in social theory and social research: By which processes is it that small social systems establish social order, where one important aspect of social order is predictability of behaviour. The importance of our model, in our view, lies in the fact that it seems to show how implicit rules of behaviour can emerge by a mere functioning of mutual reinforcement in dyadic interactions, even before considering the roles of institutions, socializing agents, values, asf. Later in this paper, in the epilogue, we will try to show how such a general observation can have quite concrete consequences for interventions in dyadic systems.

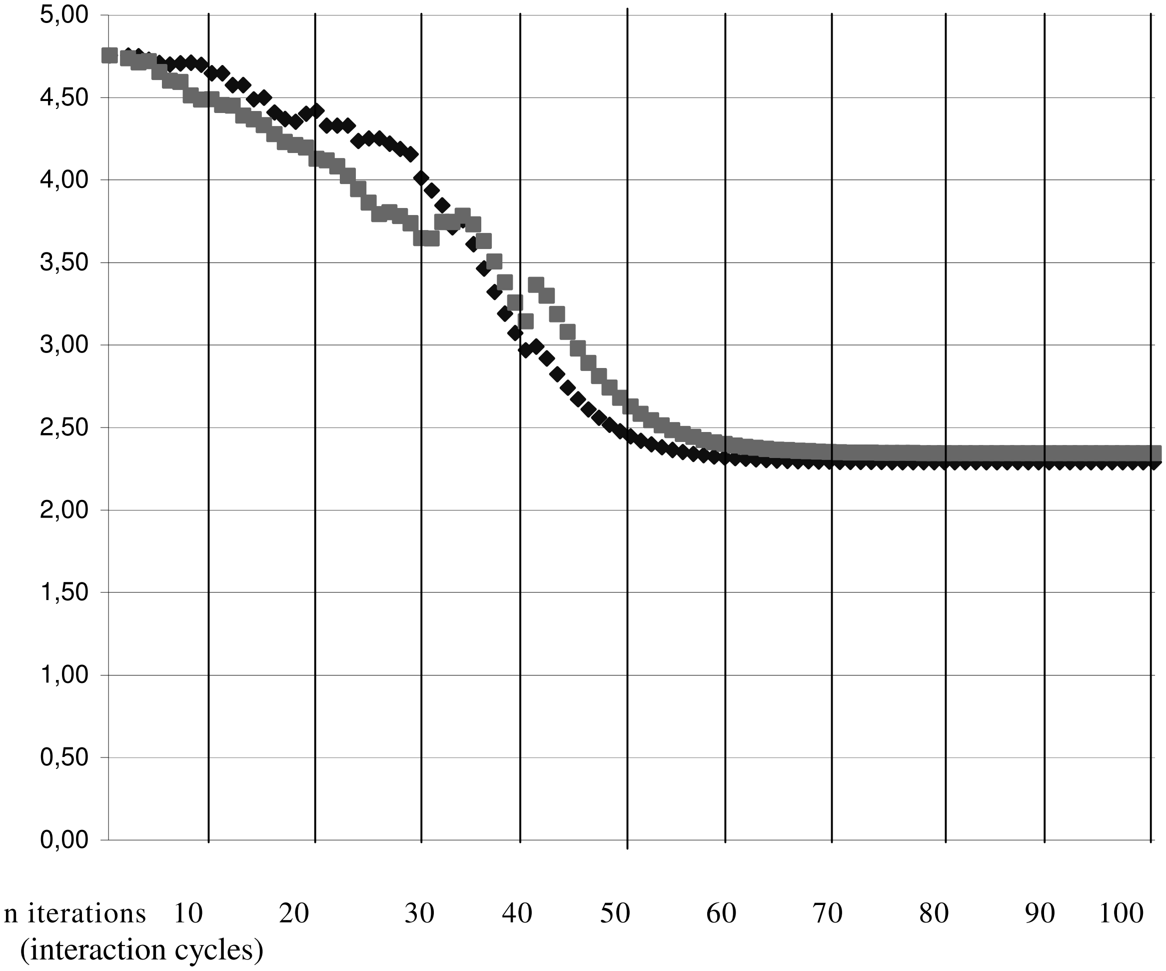


Figure 2: Entropy of reaction probabilities for 2 actors with 3 strategies each and 100 interactions, where entropy$=-\sum_i \sum_j p_{ij} \log_2 p_{ij}$.

What could this mean in terms of a sociological argument?

Social order is one of the basic issues of authors like, e.g., Soeffner (1992). He states the necessity of rituals to constitute social order. But how are rituals created in social systems? Maybe our little simulation program provides us with one answer: rituals can be looked at as those behavioural routines at the end of an interaction story which are left over, after all other behavioural possibilities have been sorted out by a process of discriminative learning; a process which we have called the "law of sociodynamics": Each interaction story can be seen as a process during which two actors mutually condition each other to a subset of behavioural possibilities ("strategies", to use game theory terminology), which is contained in the range of possible behaviours with which the actors have originally started their common history. The longer the story, the less entropic, i.e. the more deterministic, the interaction becomes. Every actually chosen strategy of partner A will cause with a probability close to 1 one specific strategy of partner B, which will in turn cause another strategy of A, asf. Therefore, the result of each interaction story will be a more or less deterministic sequence of actions, or strategies, where each partner will see the cause of his own behaviour in the partner´s preceding behaviour. And the longer it takes, the more automatic it will become. Such a quasi-automatic sequence of behaviours will look very much like what Sociologists tend to call rituals.

6 The impact of values as opposed to behavioural dispositions

We believe to have found an argument that social systems with properties similar to our model assumptions are likely to provide individuals with behavioural dispositions which depend a lot more on values attributed to behaviour than on initial behavioural dispositions, reflected in reaction probabilities.

In order to illustrate that, we refer again to the example cited above. In Eder, Gutjahr, and Neuwirth (2001), we have shown that a small change in the payoff matrix of two interacting partners can have strong effects on the distribution of strategies with which the partners end up after a longer interaction story (in our case; 100 interaction cycles). We call this distribution the “final distribution”. In the following (Figure 3), we compare 4 scenarios. Scenarios 1 and 2 differ by a rather big change in the initial transition matrix, scenarios 1 and 3 differ by a comparably small change in the payoff matrix. The same goes for scenarios 3 and 4, and scenarios 2 and 4, respectively.

In the schematic representation of Figure 4, we just look at the distances (boxes) between the distributions of the final action-reaction sequences in the four scenarios (circles). The degree of similarity between two distributions is measured by computing the Euclidean distance of the two 4-dimensional frequency vectors (in percent). The comparison between the left two circles and the right two circles is a comparison between entropic and non-entropic initial transition matrices, and the comparison between the upper two circles and the lower two circles is a comparison between a pair of payoff matrices with no Nash-equilibrium5) in dominant strategies and another one with an equilibrium in dominant strategies. (In terms of the Nash solution concept of game theory, the upper case refers to a game where the Nash equilibrum is not unique, while it is unique in the lower case.)

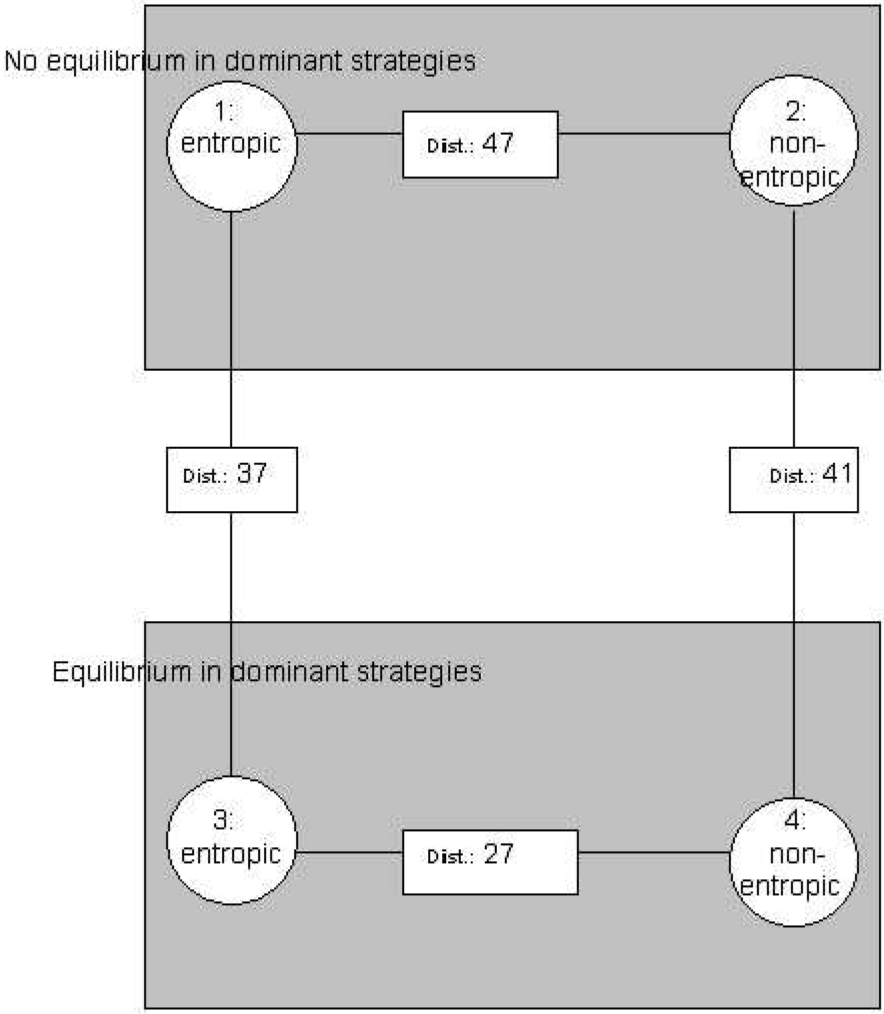
This set of comparisons shows quite clearly: When the payoff matrices do not have an equilibrium in dominant strategies, then the initial transition matrices make a lot of difference for the final distributions. If we try to translate this into an interaction situation, we might say: when the payoff situation does not make a clear distinction between different strategies, then the outcome of an interaction story will depend quite strongly on behavioural dispositions, as reflected in initial transition matrices. But we also can see in Figure 4: as soon as there is an equilibrium in dominant strategies in our payoff matrices (lower circles), then the difference between an entropic and a non-entropic initial transition matrix has a much smaller effect on the final distributions than a difference between a situation with an equilibrium vs. a situation without. (Again: we use the term "equilibrium" here in the sense in which NASH described rational solutions for N-person games, and not in the sense in which physical systems are described5), the crucial difference being that equilibrium in physical systems means that the current state *must* remain invariant according to physical laws, whereas equilibrium in game theory means that no actor has an *incentive* to change the current state.)

Trying to translate that into a situation of social interaction, we might be tempted to say: As soon as the evaluation system of outcomes gives us a clear distinction as to which situation is preferable, such a distinction, even when small, will have a much higher impact on the outcome of an interaction story, measured by final distributions, than individual behavioural dispositions, even when quite accentuated.

scenario	1		2		3		4	
	A	B	A	B	A	B	A	B
Payoff	2 0	0 2	2 0	0 2	1 0	0 2	1 0	0 2
Matrices	2 0	0 2	2 0	0 2	2 0	0 1	2 0	0 1
	A and B		A and B		A and B		A and B	
Initial transition matrices	0,5	0,5	0,8	0,2	0,5	0,5	0,8	0,2
	0,5	0,5	0,8	0,2	0,5	0,5	0,8	0,2
Final	9	31	49	19	4	12	26	9
Distributions	49	12	31	1	79	5	64	0

Euclidean	dist 1-2	Dist 1-3	dist 1-4	dist 2-3	dist 2-4	dist 3-4
Distances	46,79	36,54	33,79	66,29	41,46	27,26

Figure 3: 4 scenarios to observe the impact of differences in payoff matrices vs. differences in initial transition matrices.



1. entropic initial transition matrix and no equilibrium in payoff matrix,
2. non-entropic initial transition matrix and no equilibrium in payoff matrix,
3. entropic initial transition matrix and equilibrium in payoff matrix, and
4. non entropic initial transition matrix and equilibrium in payoff matrix.

Figure 4: Euclidean distances between the final distributions of four different scenarios.

In other words: it takes an enormous change in initial probabilities to achieve a change in final distributions, but it takes a much smaller change in payoffs to achieve a similar effect.

What could this possibly mean in terms of a theoretical sociological argument?

If we interpret our payoff matrices as a more or less culturally stable pattern of evaluation of actions which does not change over time, and if we interpret the matrix of initial transition probabilities as behavioural dispositions, then the argument is clearly: the impact of values overweighs the impact of behavioural dispositions. Behavioural dispositions have an impact only in situations of unclear preferences.

7 The slower the learning process, the better the results

Pareto-efficient solutions to typical dilemma situations such as the iterated prisoner's dilemma are found more frequently when individuals show a behaviour which might superficially be rated as "silly", i.e. the rate at which they change their behavioural dispositions is low, whereas a quick learning rate results more often in sub-optimal results for both partners. Here, we want to focus on the parameter of learning speed: When the rate at which transition probabilities are changed is high, then the chance of finding the efficient solution (cooperation/cooperation in the prisoner's dilemma) is lower than in cases where the rate at which transition probabilities are changed is low. Therefore, slow learning, in the sense of keeping up a readiness to explore unsuccessful alternatives again and again, turns out to be more "intelligent" than fast learning, at least in the cases we observed.

In Figure 5, we compare solutions to a sequential version of the chicken game (a variation of the prisoner's dilemma) with fast vs. slow learning rates, for 3 different initial transition probability matrices. We can see that the payoff sum is always higher at a slow learning rate than it as at a fast learning rate.

What does this mean for sociological theory?

A slow learning rate means that success of an action is not at once transformed into a readiness to repeat that same action at each occasion, but to try out the unsuccessful action again and again. In psychological terms, we might interpret slow learning as an inclination towards presumably non-rewarded behaviour. In several settings, such as the classical prisoner's dilemma, such an inclination looks very much like a paradoxical intervention. And we know enough about situations requiring paradoxical interventions so that this result makes some sense. We could also interpret our observation as an argument for the necessity of what might at first glance look like "irrational behaviour", for optimal solutions, just as in the case of ACO (Ant Colony Optimization), where deviations of individuals from the "pheromone track" are a necessary condition for the whole colony to find the optimal (i.e. shortest) path to food. The track they actually follow is marked by individual ants, secreting a chemical substance called pheromone, as they search food. It increases the probability of other individuals to follow this track during their search for food, but not in a deterministic manner: Single individuals sometimes "break the rule" and take another way. And a certain percentage of these "disobedient" individuals eventually turns out to be successful, resulting in a new track.

Payoff matrix:

	Cooperation	competition
Cooperation	3	2
Competition	4	1

Initial transition probabilities:

0,5	0,5
0,5	0,5

learning rate:	final distribution:	payoff A:	payoff B:	sum of Payoff
slow (0,1)	15 41 44 0	3,03	2,97	**6,00**
fast (10,0)	8 42 38 13	2,70	2,78	**5,48**

Initial transition probabilities:

0,2	0,8
0,2	0,8

learning rate:	final distribution:	payoff A:	payoff B:	sum of Payoff
slow (0,1)	4 49 45 2	2,93	3,00	**5,93**
fast (10,0)	0 25 24 51	1,96	2,00	**3,96**

Initial transition probabilities:

0,9	0,1
0,1	0,9

learning rate:	final distribution:	payoff A:	payoff B:	sum of Payoff
slow (0,1)	39 25 35 1	3,08	2,89	**5,97**
fast (10,0)	33 21 8 38	2,11	2,36	**4,47**

Figure 5: The effect of learning rates on payoff in the chicken game (a variation of the prisoner's dilemma).

Experiments have demonstrated that this system, in the long run, produces better results in terms of path optimization than a system where individuals apply a deterministic "greedy" behaviour. (Compare Dorigo and di Caro, 1999), and

Gutjahr and Eder (2001). On certain conditions, it can even be shown mathematically to always converge to the optimal path (Gutjahr, 2002).

Sociological theory has often claimed that a certain amount of individuals "breaking the rule" is a necessary condition for the development of societies. For a specific set of conditions, as described above, this can even be shown mathematically.

8 Conclusions

We believe that a relatively simple process of mutual discriminative learning can contribute to an explanation of three phenomena whose explanations would otherwise require a lot more complicated assumptions.

First: Human behaviour needs a certain amount of predictability in dyadic interactions, in order to make societies "function". This necessity of predictability can be shown on a macrosocial level: anomia is one of the major issues in sociology, describing societal conditions in which predictability of behaviour, based on normative regulations, does not work. Anomia has both clinical consequences - a rise in psychosomatic disturbances and diseases - and political consequences: a rise in criminality. It also has a suicide aspect, as we know since Durkheim's early work. Sociological descriptions of how individuals try to make their behaviour predictable include role theory, the analysis of institutions, norms, values, asf. But perhaps predictability of behaviour can often be described as a very simple process by which individuals condition each other, without even reflecting any of the aforementioned stabilizing agents. Maybe there are a lot more aspects of "functioning" in societies which require a lot less sociological terminology than we have believed so far. Or, as Heinz von Foerster (2001) has put it: "The whole social structure can be seen as a closed operator which makes certain stable values und predictable forms of interaction emerge from an infinity of behavioural possibilities. They select themselves from the infinite plurality of possibilities and cannot be explained from an analytical point of view, but can be predicted from the perspective of experience. Eigenvalues and intrinsic behaviours emerge, stable forms of interaction." 6)

Second: In order to predict the pattern of interaction which will prevail after two individuals have conditioned each other, it may be more important to know how much the interaction partners benefit from each combination of actions, than to know their "personalities": i.e. with which probabilities they are likely to react.

And third: "Good" societies require a certain amount of "bad" people (at least if less adapted behavior is viewed as bad). Optimal solutions to dilemma situations are achieved more often when the trial-and-error-phase during which solutions are found is longer. This means that during this phase, some amount of non-conformist behaviour is required: Behaviour which does not seem to be goal-oriented, which does not look promising, which does not look "logical", which does not meet the other's expectations. We might say that on the long run, it is the outsiders in society who open up a chance for improvement.

9 Epilogue: and what does all this tell us about Peter's story?

The statements we wanted to make about the contribution of our work to microsociological theory are said by now. But some of the readers might feel that there is still something missing. We started with a case report on Peter's story and claimed our model would be an instrument to improve our understanding of how the behavioural cycles described in this story would "function". So why don't we try to apply our model to Peter's story and see what it tells us about Peter and his parents?

For the sake of simplicity, we skip the role of the therapists and just look at how Peter tends to react to his parents, and vice versa. We define two actors in this "game": 1. Peter, and 2., his parents, at whom we look as if they were one single actor. And we define two different strategies for each actor: Peter's strategy 1 is to "be good" and show the behaviour we have seen in the therapeutic setting, and his strategy 2 is to be a bad boy, hit his mother, and do all these nasty things. His parents' strategy 1 is to accept him as their beloved son, and their strategy 2 is to reject him (or his behaviour), and/or call the police. If we try to translate the behavioural cycles reported during the therapeutic setting, we get the following set of cyclic transition probability matrices (Figure 6):

Peter's strategies:

strategies:

Peter's parents' strategies:	1: be good	2: be a bad boy
1: accept	0	1
2: reject	1	0

Peter's parents'

Peter's strategies:	1: accept	2: reject
1: be good	1	0
2: be a bad boy	0	1

Figure 6: Hypothetical transition probability matrices for Peter's and his parents' behavioural cycles.

These matrices are easy to read: Whenever Peter feels acceptance by his parents, he sooner or later becomes nervous, and starts misbehaving. And as soon as he feels rejected, he starts to become a good boy again. His parents' behaviour looks a lot more like what might superficially be rated as “logical”: when Peter is good, they accept him, and when he is a bad boy, they reject him and call the police – what else should they do?

This is what we observe. We know not much about the preference system of Peter and his parents. All we know is that the parents declare that all they want from Peter is to be good, and Peter declares that all he wants from his parents is to take him back and accept and love him. The payoff matrices resulting from these declarations would also be easy to sketch, and they would look approximately as in Figure 7:

Peter’s parents’ preferences

Peter’s strategies after his parents’ strategies	1: be good	2: be a bad boy
1: accept	1	0
2: reject	1	0

Peter’s preferences:

Peter’s parents’ strat.: after Peter’s strategies	1: accept	2: reject
1: be good	1	0
2: be a bad boy	1	0

Figure 7: Hypothetical payoff matrices for Peter’s parents’ and his own preferences.

These payoff matrices, we believe, are also very easy to read: no matter what Peter’s parents have done before, they want Peter to be good. And no matter what Peter has done before, he wants his Parents to accept him. Everybody wants the other to just be good. Again: we do not know if this is what they *really* want, but this is what they say they want, and this is what they would like to get by the help of therapy.

Peter’s strategies:

strategies: Peter’s parents’ strategies:	1: be good	2: be a bad boy
1: accept	46	13
2: reject	11	30

Peter’s parents’

Peter’s strategies:	1: accept	2: reject
1: be good	46	11
2: be a bad boy	13	30

Figure 8: Final distributions for Peter’s and his parents’ behavioural cycles, based on the payoff matrices as in Figure 7.

Our model shows us that a process of mutual discriminative learning would lead to the following final distributions, starting with initial transition probabilities that were all 0.5.

The final distributions of Figure 8 do not show too much similarity with what we have observed. Based on the payoff system as in Figure 7, at least 46% of all cases should show a happy end: a well-behaving Peter, being accepted by his parents. Only 13% of the families with a value system as in Figure 7 would eventually produce behavioural rituals where acceptance of the parents induces Peter to his terrible behaviour. Could there be a payoff system producing final distributions that would show higher similarity with what we have observed in our case study?

We actually do not know what really happens in that family. All we know is their behaviour in a therapeutic setting. And if we reduce our payoff matrix more closely to what we actually could observe, then we might suspect that perhaps it is the outcome of therapy which the actors are after: They say that they want the therapy to make Peter be a good boy, and thereby enable the parents to accept him again. If we follow the hypothesis for a little while that well-behaviour of Peter in his family setting does not have any favourable consequences for him, and that it is only well-behaving in the therapeutic setting which he sees as rewarding, then we would have to hypothesize a payoff matrix as the following (Figure 9):

Peter's parents' preferences

Peter's strategies after his parents' strategies	1: be good	2: be a bad boy
1: accept	0	0
2: reject	1	0

Peter's preferences:

Peter's parents' strat.: after Peter's strategies	1: accept	2: reject
1: be good	0	0
2: be a bad boy	1	0

Figure 9: Another hypothetical payoff matrices for Peter's parents' and his own preferences.

In Figure 9, it is the therapy outcome that counts, and nothing else. As soon as Peter has been a bad boy, he highly emphasizes being accepted (again) by his parents. But without such a condition, he does not care too much. And as soon as his parents have rejected Peter, they are highly interested in his well-behaviour (again), but his well-behaviour does not make much of a difference for them otherwise. And the final distributions produced by such a system of preferences will look like Figure 10.

	Peter's strategies:	
strategies: Peter's parents' strategies:	1: be good	2: be a bad boy
1: accept	0	12
2: reject	10	79

	Peter's parents'	
Peter's strategies:	1: accept	2: reject
1: be good	0	10
2: be a bad boy	12	79

Figure 10: Final distributions for Peter's and his parents' behavioural cycles, based on the payoff matrices as in Figure 9.

We can see: although we have made only one minor change in the preference system as reflected in the payoff matrices, assuming that there is no experience of reward for well-behaving inside the family, the "logical" outcome, as produced by simple discriminative learning, is dramatically different from the first scenario: Whenever Peter's parents accept him, he will be a bad boy, an event which accounts for 12% of all behavioural combinations; this seems to match better with what we have observed in his biography than the final distribution in Figure 8. And whenever he behaves nicely, his parents will reject him.

The feedback to the therapists from this seems to be quite clear: Do not let the clients impress you too much by what they declare during the therapy session. Rather, try to find out what happens in that family during those phases that are rated as "normal", as everyday routine, by the clients. If it should turn out that their everyday routine is as dull and free of any rewards as it is in our hypothetical payoff matrices in Figure 9, implying that they need the therapists to provide a rewarding link between behaviour and outcome, then the symptoms presented would turn out to be not more than a logical consequence of the payoff system which this family uses. If the family feels that reconciliation is more exciting than routine, then this is what they will get: and problems become a necessary condition for later reconciliation.

References

[1] Bateson, G. and Jackson, D.D. (1964): Some variations of pathogenic organization. In D. Rioch (Ed.): *Disorders of Communication,* **42,** 273.

[2] Dorigo, M. and Di Caro, G. (1999): The ant colony optimization metaheuristic. In D. Corne, M. Dorigo, and F. Glover (Eds.): *New Ideas in Optimization.* New York: McGraw-Hill, 11-32.

[3] Durkheim, E. (1970, first published Paris 1897): *Suicide. A study in Sociology*. London: Routledge&Kegan.

[4] Eder, A., Gutjahr, W.J., and Neuwirth, E. (2001): Modelling social interactions by learning Markovian matrices. (or: Why we may not need early childhood to develop neuroses, although it helps a lot). In D. Elizur (Ed.): *Facet Theory: Integrating Theory Construction with Data Analysis.* Proceedings of the 8th international facet theory conference. MATFYZ Press, Praha.

[5] Foerster, H.V. and Pörksen, B. (2001): *Wahrheit ist die Erfindung eines Lügners*. Gespräche für Skeptiker. Bonn 2001, 4. edition, 61.

[6] Gutjahr, W.J. and Eder, A. (2001): A Markov model for dyadic interaction learning. In: M.A. Wiering (Ed.): Proceedings of the fifth european workshop on reinforcement learning. October 5-6, 2001, Utrecht University.

[7] Gutjahr, W.J. (2002): ACO algorithms with guaranteed convergence to the optimal solution*, Information Processing Letters,* **82**, 145-153.

[8] Mosler, H.-J. (2000): *Computersimulation Sozialpsychologischer Theorien. Studien zur Veränderung von Umwelteinstellung und Umweltverhalten.* Psychologie Verlags Union, Weinheim.

[9] Schmidt, B. (2000): *Die Modellierung menschlichen Verhaltens*. Soc. for Computer Simulation Int., Delft, Erlangen.

[10] Soeffner, H.G., (1992): *Die Ordnung der Rituale - Die Auslegung des Alltags 2*, Frankfurt a.M.

Footnotes

[3] If, in steady state (stable transition matrices), strategy i of the partner is actually responded by action j (thus implying that the corresponding $p_{ij} > 0$) and all elements of the payoff matrix are at least marginally greater than 0, then p_{ij} must be 1. This can be shown as follows:
We denote with p_{ij}^{t} the probability of an actor to react to the partner's strategy i by strategy j at time t, and p_{ij}^{t+1} the same probability at time $t+1$, and with a_{jk} the value which is attributed when the partner's reacts to ego's strategy j with strategy k. Then the condition for every nonzero element p_{ij} in a stable transition probability matrix, showing no more changes, can be formulated as:

$$p_{ij}^{t+1} = p_{ij}^{t}$$

Our updating process as described above, is defined as follows:

$$p_{ij}^{t+1} = \frac{p_{ij}^{t} + a_{jk}}{1 + a_{jk}}$$

Therefore, an updating process resulting in no changes in the transition probability matrix can be described as

$$p_{ij}^{t} = \frac{p_{ij}^{t} + a_{jk}}{1 + a_{jk}}$$

This is equivalent to $p_{ij}^{t} + p_{ij}^{t} a_{jk} = p_{ij}^{t} + a_{jk}$ or $p_{ij}^{t} a_{jk} = a_{jk}$

Apparently, there are three cases in which the above equation is true: when either p_{ij} equals 1, or when a_{jk} equals 0, or both. Therefore, since we have excluded the case that $a_{jk} = 0$, p_{ij} must be 1. The case in which we allow for $a_{jk} = 0$, admits both deterministic and non-deterministic limiting behaviour.

---

4) Formally, a cyclic matrix pair can be defined in the following way: There is a subset of actions of actor 1, and a subset of actions of actor 2, such that the transition matrix of actor 1 assigns to each action of the first subset a single action of the second subset with probability 1, and the second transition matrix assigns to each action of the second subset a single action of the first subset with probability 1 (see Gutjahr and Eder, 2001).

---

5) It is important to point out that the term “equilibrium” used here refers to a solution concept in game theory, introduced by NASH, and not to an equilibrium in physical systems. The term “NASH-equilibrium” is formally defined as follows: A combination (s_1*, …, s_n*) of the strategies of N players is a Nash equilibrum, if no player i can improve his/her payoff by changing from strategy s_i* to another strategy s_i, given that all other players choose their strategies s_j* as contained in the strategy combination (s_1*, …, s_n*).

---

6) Translation into English by the authors. The original German text is as follows:

„Die gesamte soziale Struktur kann als ein geschlossener Operator verstanden werden, der aus den unendlichen Möglichkeiten des Verhaltens gewisse stabile Werte und vorhersehbare Formen der Interaktion entstehen lässt, sie schälen sich – aus der unendlichen Vielfalt des Möglichen – heraus und sind von einem analytischen Standpunkt aus unerklärbar, aus der Perspektive des Erfahrbaren jedoch prognostizierbar. Es entstehen Eigenwerte bzw. Eigenverhalten, stabile Formen der Interaktion.“