Can Simulation Models Support Social Research? A Critical Discussion.

G. Rohwer

Version 1 August 2000

Summary

Fakultät für Sozialwissenschaft
Ruhr-Universität Bochum, GB 1
44780 Bochum

goetz.rohwer @ ruhr-uni-bochum.de

dsm.tex August 2000
Homepage: http://www.stat.ruhr-uni-bochum.de [papers]

Contents

1 Introduction

1. The word 'simulation model'[1] covers a broad range of different modeling approaches (see, e.g., Troitzsch 1990, Troitzsch et al. 1996, Gilbert and Troitzsch 1999). Very often, the attribute 'simulation' is just added to indicate that one intends to construct a version of some model that can be programmed and executed on a computer. If one thinks of possible uses of simulation models in the social sciences, it might be helpful to make the following distinction.

a) Simulation models can be used to represent and combine statistical knowledge about social processes. Early developments of this approach can be found in Orcutt et al., 1961. The main goal of this kind of model building is to use available statistical data for prediction of social processes and for exploring possibilities and possible consequences of policy interventions. In order to refer to this kind of model building, one may speak of 'pragmatic simulation'. The idea is that these models should be useful for some practical purpose; there is no claim that the model can contribute to an understanding of social life.

b) On the other hand, there is a growing literature that proposes to use simulation models as tools for a discussion of theoretical questions concerning human interaction and its social conditions and consequences; see, e.g., the introduction titled "Computer simulation for social theory", by R. Conte and N. Gilbert (1995). Most remarkable are approaches claiming that simulation models can provide a suitable conceptual (and perhaps computational) framework for reasoning about social actors in social settings. A recent example is Epstein and Axtell's "Growing Artificial Societies" (1996).

This paper intends to discuss some problems connected with the second kind of approaches. It is not questioned that pragmatic simulation might be a valuable tool for making sense of statistical data about social processes.[2] The paper is concerned, however, with the pretension that simulation models, as they have been developed mainly by importing ideas from computer science and other technologically oriented disciplines,[3] might provide a suitable conceptual framework for theory building in the social sciences.

[1] Single quotes are used to refer to expressions, double quotes are used for citations and to indicate metaphorical, ambiguous and opaque use of expressions.

[2] For introduction and further references, see Gilbert and Troitzsch 1999, ch. 4.

[3] "Simulation modeling in the social sciences is an adaptation of approaches that arose in the physical and life sciences. Many social science modelers track developments in fields as diverse as population ecology, meteorology, physics, and computer science, and seek to import ideas from these fields." (Hanneman and Patrick, 1997, sec. 5.1)

2. I want to stress that, in this paper, my interest concerns theoretical claims since this is not necessarily the main concern in the simulation literature. The following passage from a recent introduction to "Simulation for the Social Scientist" by N. Gilbert and K. G. Troitzsch (1999, p. 2) provides an illustration.

> "Like statistical models, simulation have 'inputs' entered by the researcher and 'outputs' which are observed as the simulation runs. Often, the inputs are the attributes needed to make the model match up with some specific social setting and the outputs are the behaviour of the model through time. An example – based loosely on the work of Todd (1997) – may make this clearer. Suppose that we are interested in how people choose a marriage partner. Do you (perhaps, did you?) keep looking and dating until you found someone who meets all your romantic ideals, or do you stop as soon as you find someone 'good enough'? Do people use a sufficiently rigorous search procedure or, as Frey and Eichenberger (1996) suggest, should they search longer, possibly reducing the divorce rate as a result?
>
> Asking people about their searching behavior is unlikely to be very helpful: they may not be following any conscious strategy and may not reveal it even if they did have one. Instead, we might set up a model (in this case, a computer program) which embodies some plausible assumptions and see what happens, comparing the behaviour of the program with the observed patterns of searching for a partner."

Obviously, in this example, the creation of a simulation model is not intended to provide any insights into the nature and characteristics of the processes by which people actually find partners. There is no intention to investigate these processes empirically; *instead*, one invents a model that is based on "some plausible assumptions". Consequently, the model cannot serve an understanding of people's behavior. For what purpose, then, might such a model be useful? The authors go on to tell the reader:

> "Todd (1997) explores a number of possible strategies, including those which have been proved analytically to be optimal in terms of finding the best partner, but which require unrealistic amounts of search, and some other strategies which are much simpler and have better results when one takes into account that search is expensive in time and effort. He also begins to investigate the

implications for search strategies when there is a possibility that you might want to settle down with a partner, but the partner may still be wanting to continue to search for someone else. Even in this much more complex situation, simple strategies seem to suffice."

This suggests that the model should serve to explore mating strategies for someone (maybe, you) who has decided to consider finding a partner as a problem that should be solved by using some "optimal" strategy and, one may add, *regardless* of whether other people follow the same view (remember that, at the beginning, it has been explicitly questioned whether people use strategies at all). We may call this a 'strategic use of simulation models': to explore strategies from the point of view of the model builder who provides the problem definition and criteria for an assessment of solution strategies. As an implication, it suffices to have a purely instrumentalistic, even definitely wrong view of the social context where to apply the model builder's strategies. However, I do not here intend to criticize this understanding of model building but simply want to stress that I am concerned with a different question: whether simulation can also serve a theoretical understanding of social actors and their interaction.

3. In order to follow this question I proceed as follows. In sections 2 – 4 I sketch a general framework for an understanding of simulation models. I then consider the question of what can be represented by using this conceptual framework. What I intend to show is that the conceptual tools provided by simulation models are quite limited and, finally, insufficient to understand, and reason about, the behavior of human actors.

2 A Conceptual Framework

1. In this section I sketch a general conceptual framework that can provide a starting point for an understanding of a broad range of currently developed simulation models. There are three basic ingredients: (a) a set of objects, (b) a space where the objects have a location and, in some models, are allowed to move, and (c) a time axis to provide a framework for thinking of changes and movements. The word 'object' is used here in a very general sense, including what is sometimes called 'agents' in the simulation literature.

2. Almost all models require a time axis in order to think of objects that may change their attributes over time. Here one has the choice between a

continuous and a discrete time axis. Most object-based simulation models that are intended to be implemented on a computer assume a discrete time axis. I follow this idea and assume a time axis

$$\boldsymbol{\mathcal{T}} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

I refer to the elements of $\boldsymbol{\mathcal{T}}$ as 'points in time', but of course, each of these time points has an intrinsic duration. However, being concerned only with a conceptual framework, it is not necessary to fix a specific interpretation for the durations. The only requirement is an order relation defined for the time points in the usual sense.

3. Analogously, one can introduce a space that allows to speak of locations. I shall use the notation $\boldsymbol{\mathcal{R}}$ to refer to a set of locations. The idea is that these locations can be used to locate objects in a space. How to interpret the locations depends on the purpose of a model, and thereof also depends whether one needs to introduce a topology or metric. This is required when a model shall be used to think of objects that can move in a space; as an example, one can think of segregation models that have been developed by using the conceptual framework of cellular automata. On the other hand, there are many models that are not explicitly interested in spatial movements and therefore do not need an explicit representation of locations to place the objects. One can then simply ignore the requirement of a space or, equivalently, assume a space that contains just one location.

4. Finally, one needs a representation of objects. I will use the notation Ω to refer to a finite set of names for objects. It is not required that objects are all of the same kind. It is the task of introducing specific models to explicate what kinds of object shall be considered. It also depends on the specific model whether it suffices to consider a timeless set of objects, say

$$\Omega = \{\omega_1, \ldots, \omega_n\}$$

or whether the model needs to consider that objects do only exist for a limited duration. For example, most demographic models would need to take into account that humans do not live forever, but are born at some point in time and eventually die at some other point in time. It is then necessary to index the object sets by elements from $\boldsymbol{\mathcal{T}}$, that is, to conceive of Ω_t as the set of objects that exist at time t.

5. It is not enough to have names for objects of whatever kind. A name only allows to refer to an object. In addition one needs conceptual tools to represent objects, that is, to think of attributes that can be used to characterize objects. This requires a conception of property spaces. In general, a property space is simply a collection of attributes that can be used to characterize the elements of the object set, Ω. I will use the notation

$$X_t : \Omega \longrightarrow \tilde{X}$$

In this notation $\tilde{X}$ is the property space, and the only requirement is that it allows to define a mapping, X_t, that associates with each object $\omega \in \Omega$ a specific attribute $X_t(\omega) \in \tilde{X}$ that characterizes the object at $t \in \boldsymbol{\mathcal{T}}$. Since X_t is here used to represent properties of objects, it is called a 'representational variable'. As an example, one can use $\boldsymbol{\mathcal{R}}$ as a property space for the definition of location variables

$$L_t : \Omega \longrightarrow \boldsymbol{\mathcal{R}}$$

that represent the locations of the objects at time t. One can specify many other property spaces. They can be used, in particular, to distinguish different kinds of object in Ω. Furthermore, they can be combined into multidimensional property spaces that allow to define multidimensional representational variables having the form

$$(X_t, Y_t, Z_t, \ldots) : \Omega \longrightarrow \tilde{X} \times \tilde{Y} \times \tilde{Z} \times \cdots$$

It is possible, however, to refer to this multidimensional variable again by a single letter, and so it is mainly a question of notational practicality whether to use an explicit representation of different property spaces.

6. The conceptual framework so far introduced can be used for many different purposes. In particular, it is a useful framework for a definition of statistical concepts. Assuming that the main purpose of statistics, as it is traditionally conceived, is to provide representations of sets of objects which actually exist in the human environment by using observations of some of their attributes, immediately leads to think of the representational variables defined above as statistical variables that can be characterized by frequency distributions. All further statistical concepts can then be definitionally derived from this starting point. However, the intention in developing simulation models is somewhat different. Their main purpose is not to provide a representation of "what there is", but to provide insights into the development of processes and their

"mechanisms"; simulation "aims to explicate the mechanisms of social processes." (Gilbert 1996, p. 449) This is also the main concern of this paper: whether simulation models can sensibly contribute to this task. The conceptual framework introduced above does not provide any specific hints. While it certainly can be used to define several notions of 'social process', simply by using sequences of representational variables, there is so far no idea of what drives such processes and how to think of "mechanisms".

3 Statistical and Behavioral Rules

1. The basic notion used in the construction of simulation models is that of 'rule'. As said by J. L. Casti (1992, p. vii): "Mathematical modeling is about rules – the rules of reality." However, 'rule' is a very general notion which is used in several contexts with different meanings, so one needs to find an understanding of how this notion is used in the construction of simulation models. In object-based simulation models that provide the examples for our discussion, rules refer to the behavior of objects; one may therefore speak of 'behavioral rules'. As it seems, this notion is closely connected to the talk of "mechanisms". When talking of "mechanisms underlying a social process", many authors seem to have in mind some idea of behavioral rules that "govern" the behavior of the objects whose attributes are used to define the process.

2. Before trying to get an understanding of behavioral rules it is worthwhile to briefly look at the statistical approach to conceive of rules in social processes. In the simplest case, the starting point is a two-dimensional statistical variable, say

$$(X, Y) : \Omega \longrightarrow \tilde{X} \times \tilde{Y}$$

The statistician is not, however, interested in the behavior of individual objects (members of Ω), but in the distribution of variables. The focus is on how the distributions of X and Y are related. A notion of rule might then be introduced by considering the distribution of one variable, say Y, conditional on values of the other variable, X. This is the basic idea of statistical regression models. The statistician tries to find a mapping that associates with each value of X (each element or subset of $\tilde{X}$) a conditional distribution of Y. This may be written in the following way:

$$\tilde{x} \longrightarrow \Big\{ \tilde{y} \longrightarrow \mathrm{P}(Y \in \tilde{y} \,|\, X \in \tilde{x}) \Big\}$$

where $\tilde{x}$ and $\tilde{y}$ are subsets of the property spaces $\tilde{X}$ and $\tilde{Y}$, respectively. Since the values of functions of this kind are itself functions (conditional distributions), statisticians have invented a lot of simplifying representations, often called 'regression models'. Most often the conditional distributions are represented by some characteristics as, for example, conditional means, quantiles or frequencies. The technical details need not concern us here, but two points should be emphasized. First, while it is quite possible to interpret statistical regression functions as rules that can be used for conditional predictions, they are clearly not behavioral rules; they simply do not refer to whatever might be attributed to individual objects in Ω as behavior.[4] A second point concerns the epistemic status of statistical rules (regression functions). These rules are derived from data in the sense of values of statistical variables. These are representational variables that reflect aspects of a historical process for some limited region in space and time. This does not exclude the possibility to use the derived rules for the formation of expectations and predictions (nothing is implied for being successful or not); but it creates a question: In which way can one think of statistically derived rules not just as representing the data used to derive the rules but in some way providing "empirical regularities"? I shall not discuss this question here but only wish to make clear that a central idea behind the construction of simulation models is, not to begin with assumptions about rules that connect statistical distributions, but to assume that rules can be formulated with respect to the behavior of individual objects. This often leads to the additional idea that statistical regularities can finally be derived from more basic rules formulated in terms of behavior.

3. What is meant by the idea that objects "can behave", or "show some behavior"? One can try to begin with two basic requirements: (a) that objects can change some of their properties over time, and (b) that objects can interact with an environment. The formulations are obviously very unspecific. In particular, nothing is said about the possible meanings of 'can' and 'interaction'; for example, in which sense might it make sense to say that a stone *can* kill a man? However, while it is certainly necessary to consider such questions seriously, it is remarkable that object-based simulation models employ a notion of 'behavioral unit' that is completely unspecific and therefore, it seems, can be interpreted

[4]It should be stressed that the notion of statistical distribution can only be sensibly used when referring to a set of objects or, more precisely, to a statistical variable defined for a set of objects.

in almost any way one likes. The notion is imported from mathematics and known as an 'automaton with input and output channels' (see, e.g., Weisbuch, 1991). In order to explain the idea I refer to an element $\omega \in \Omega$. To conceive of ω as an automaton[5] requires two considerations. First, a suitable representation consisting of three variables.

a) A state variable intended to represent the state of ω for each point in time. I use the notation $s_t(\omega)$ to refer to ω's state at $t \in \mathcal{T}$. Of course, this needs a specification of a set of possible values, say $\tilde{S}$; this is a property space as introduced in the previous section, often called a 'state space'.

b) In addition, it is assumed that ω can get inputs from some environment. This can be represented by an input variable, $x_t(\omega)$, that can take values in an input state space $\tilde{X}$.

c) Correspondingly, it is assumed that ω can create values of an output variable. I use the notation $y_t(\omega)$ and assume that this variable can take values in an output state space, $\tilde{Y}$.

These notations finally allow to define an automaton by a behavioral rule that prescribes how the automaton changes its states and creates its outputs. There are actually two rules:

$$\begin{aligned} s_{t+1}(\omega) &= r_s(s_t(\omega), x_t(\omega)) \\ y_{t+1}(\omega) &= r_y(s_t(\omega), x_t(\omega)) \end{aligned}$$

The first rule, r_s, specifies how the automaton changes its states (or remains in the same state), depending on its previous state and input; the second rule, r_y, specifies the output, again depending on the automaton's previous state and input. Together, both rules specify the behavior of the automaton.[6]

4 Automata-based Simulation Models

1. The previously introduced notions provide a starting point to define a broad class of simulation models. The general idea should already be

[5]Here I always use this word as implying input and output channels.

[6]Note that the notations do not exclude the possibility to assume that automata have a memory. This can be achieved by using part of the object's state variables to keep track of previous states. In this way one can also make the behavioral rules, r_s and r_y, dependent on previous states.

visible: conceive of the objects in Ω as automata and assume behavioral rules that can effectively be computed. However, one thing is yet missing. In order to arrive at a computational model one also needs a rule that allows to calculate values for the input variables, $x_t(\omega)$. The idea is that these values depend not only on the object's states, $s_t(\omega)$, but also on attributes of their environment, including output from other objects in the environment. So one needs to find some notation for this idea.

2. What is the environment of an object? No simple answer can be given. But one might say that what counts is which other objects are contained in an environment, since otherwise there would be only empty space. This suggests to conceive of environments also as sets of objects. Then, given that the conceptual starting point has fixed just one basic object set, Ω, environments become subsets of Ω. This reasoning can be fixed by assuming that, for each $\omega \in \Omega$ and each time point $t \in \mathcal{T}$, there is an environment $U_t(\omega) \subseteq \Omega$. It will be called a local environment if $U_t(\omega) \neq \Omega$. For notational convenience I shall assume that $\omega \in U_t(\omega)$.

3. This then allows to formulate rules for the input states of the automata. The idea is to assume that $x_t(\omega)$ is determined by the outputs of all objects in ω's environment. The general form of a rule for the input states will then be

$$x_{t+1}(\omega) \;=\; r_x(\{y_t(\omega') \,|\, \omega' \in U_t(\omega)\})$$

The argument is written as a set since, in general, an environment is simply defined as an unordered set of objects. Given a space for the location of objects for which one has specified a topology,[7] this would induce some structure in the environments, too. In any case, this can be treated as attributes of the environments and does not require a separate notation.

4. There remains, however, a quite important question. How are the environments for the objects generated? Two possibilities come easily into mind. One is that environments are fixed at the beginning and cannot change through the objects' behavior. As an example, one can think of cellular automata models where objects are identified with spatial locations and environments are defined by a fixed set of adjoining locations. Another possibility would be that each object is allowed to define its own

[7] It would also be possible to explicitly introduce "social relations" for Ω that induce corresponding relations in environments.

environment. But as long as there is no way of talking about "choices" that might be taken by the objects, this simply means to make $U_t(\omega)$ a function of $s_t(\omega)$. Assuming a space with a topology this would allow, for example, to make environments a function of the spatial location of objects.

5. Taken together, the previously introduced ideas and notations can be used to explain the construction of automata-based simulation models. There are three steps involved.

a) The first step is to create names for a set of objects, Ω, and specify a space, $\mathcal{R}$, and time axis, $\mathcal{T}$, to locate the objects.

b) The second step is to define the objects as automata. This requires a specification of the state spaces, $\tilde{S}$, $\tilde{X}$ and $\tilde{Y}$, and the rules, r_s, r_x and r_y. These must be computational rules, meaning that they are specified in such a way that one can effectively calculate values from their arguments.[8]

c) The final step is to fix some initial state for the variables $s_t(\omega)$ and $x_t(\omega)$, for all objects in Ω. One can then calculate the values of all variables for all further points in time.

These three steps then define a general notion of 'automata-based simulation model'. The definition covers virtually all simulation models which refer to objects and their behavior and can be realized on a computer by using traditional or object-oriented programming languages.

5 What Can be Represented?

1. There is an obvious relationship between automata-based simulation models and statistical models. Given an automata-based simulation

[8] In fact, there are two further requirements. First, in order to realize a simulation, it must be possible to apply the rules sequentially (but not necessarily in a predetermined order) to all objects. Whether this is possible depends on the specification of mutual dependencies. This condition has been called 'simulatability' and is equivalent to the existence of a computational mapping from the states of *all* objects in t to their next states in $t+1$; for a discussion see Rasmussen and Barrett, 1995. A second requirement, in a sense already implied in the first one, is that "simultaneously" (defined by reference to one time step in the simulation) applying the rules to all objects must not lead to a violation of logical and physical constraints; for example, that two objects occupy the same spatial location at the same time. A discussion of the implied "coordination problem" will be postponed to a later section.

model one can define statistical variables

$$(S,X,Y)_t : \Omega \longrightarrow \tilde{S} \times \tilde{X} \times \tilde{Y}$$

Of course, their values depend on the initial state assumed for the simulation. But given an initial state, the statistical variables have a fixed distribution and can be used to provide a statistical description of the simulation results. This also motivates to think of the behavioral rules specified for the objects as providing a characterization of the "mechanism" underlying a statistical description of the results of the simulation.[9] On the other hand, beginning with a process defined in statistical terms, it is not generally possible to find a unique automata-based simulation model that might have created the statistical distributions.

2. But one also needs to ask what can be represented by using the conceptual framework of an automata-based simulation model. The intention is to represent the behavior of objects and, in particular, their interaction and what might result from their interaction. An important question, therefore, is whether the conceptual framework offered by automata-based models allows to sensibly represent human actors and their environments. In the simulation literature this is often taken for granted without (much) discussion. Consider, for example, the following statement about simulation methodology by J. Doran (1998, sec. 1.2):

> "Some explanation and justification of methodology is appropriate. By working in the computational domain we are able to study multiple agent systems *independently* of any particular modeling interpretation. We can establish rigorously and objectively what consequences, including non-obvious consequences, flow from what assumptions. We can therefore hope to develop an abstract theory of multiple agent systems and then to transfer its insights to human social systems, *without* an *a priori* commitment to existing social theory. Of course, at a certain level assumptions must be built into whatever systems we create and experiment with, but the assumptions may be relatively low level, and their consequences (including emergent properties) may be discovered, not guessed."

In fact, since the general notion of an automata-based model is purely formal (mathematical), one can study such models independently of interpretations. However, it is not at all evident that these formal models allow a sensible representation of social reality.

[9] Actually, it does not suffice to refer to the behavioral rules, r_s, r_x and r_y, since results also depend, in an often crucial way, on the specification of rules to generate environments. For a discussion of examples see, e.g., Kephart, 1994.

3. Questions concerning interaction will be postponed to a later section. With respect to behavior, one sees that there is nothing in the conceptual framework of automata-based models that can directly be linked with a common-sense understanding of 'behavior'. The model just provides variables taking values that, at the beginning, are set by an external actor, the creator of the model or whoever "runs" the simulation, and then go on to take values according to rules which also are fixed at the beginning.[10] It is certainly possible to talk metaphorically of behavior by referring to changing values of variables in the model, but there remains a conceptual conflict. Ordinary talk of behavior gets its *meaning* from the view, or assumption, that behavior is some kind of activity as opposed to simply being in one or another state.

4. The intuition behind the ordinary understanding of behavior becomes clearer when one recognizes that it is intimately linked to the idea of "causal agency".[11] The idea is that objects can exhibit some behavior and, by this behavior, can cause changes in their environment. However, while such intuitions are quite fundamental, not only for ordinary talk of behavior but also for an interpretation of simulation models, they do not find any conceptual representation in the model formulation.[12] In fact, automata-based simulation models conceptually derive from an opposite idea. The process that becomes realized when a simulation model is run on a computer does not result from the behavior of the objects

[10] Some authors have remarked that one can construct simulation models where rules can change while the model is executing. This, however, is an unwarranted play with words. Rules that define how the model works belong to the definition of a simulation model and consequently cannot change while one creates one of its possible realizations. Of course, one can formulate a rule, say R, by a formulation like: Being in state $\tilde{s}$ apply rule r_1, otherwise apply rule r_2. So one can construct a model where the frequencies for the application of r_1 and r_2 change over time. Nevertheless, the behavior of the objects remains determined by the rule R which, by definition, cannot change.

[11] To provide such an idea is also one of the main functions of the metaphorical language that most often associates the construction of simulation models. As an example, consider the following statement from the authors of the Swarm approach to simulation: "The basic unit of a Swarm simulation is the agent. An agent is any actor in a system, any entity that can generate events that affect itself and other agents." However, "A typical agent is modeled as a set of rules, responses to stimuli." (Minar et al. 1996, p. 3)

[12] Using a distinction proposed by Goldspink (2000), one might say that automata-based models can only represent "passive agents", but not "active agents". Unfortunately, Goldspink takes for granted, without providing any discussion, that objects in simulation models can also be viewed as "active agents".

that the model pretends to represent, but originates from the creator of the model and the rules he has prescribed for the process. Consider the question, What "drives" a process? With respect to social processes many people would agree that these processes consist of actors, and in particular human actors, whose behavior brings about the processes. But applying the same view to processes simulated on a computer leads to the conclusion that the actors that drive the process are not the objects, represented by the model, but just two quite different actors: the human creator of the model and its implementation, and the computer's clock that drives the computation to become a process.

5. The conceptual conflict becomes most dramatical when a simulation model is intended to simulate the behavior of objects that we assume can choose between different behaviors.[13] An automaton, by definition, cannot choose between different ways of behavior. The values of its state and output variables are determined by its previous states and inputs.[14] One should note that the idea that some objects can choose between different behaviors is quite distinct from ideas concerning "goals" or "desires".[15] It is quite easy to think of objects having goals or desires, and so it

[13]The assumption has been denied, for example, by B. F. Skinner (1973, p. 101): "In what we may call the pre-scientific view [...] a person's behavior is at least to some extent his own achievement. He is free to deliberate, decide, and act, possibly in original ways, and he is to be given credit for his successes and blamed for his failures. In the scientific view [...] a person's behavior is determined by a genetic endowment traceable to the evolutionary history of the species and by the environmental circumstances to which as an individual he has been exposed." I shall not discuss here whether this version of a scientific view can be formulated in a self-consistent way.

[14]Here I refer to an automaton as a concept, as distinct from some realization. Any artificially created object to be used as the realization of an automaton can, of course, deviate from its prescribed behavior and, depending on the kind of realization, it might be possible to say that the realized automaton has chosen to deviate from what it is designed to do.

[15]This is not to deny that the notions are related; it seems quite plausible that one cannot seriously explain what choice is without reference to goals. Nevertheless, being able to make choices is semantically different from "being governed by goals". This should be stressed because some authors seem to suggest that a notion of 'cognitive agent' can solve the conceptual conflict. The following passage from A. Drogoul and J. Ferber (1994, p. 132) may serve as an example: "Cognitive agents have a symbolic and explicit representation of their environment on which they can reason and from which they can predict future events. Cognitive agents are driven by intentions, i.e. by explicit goals that affect their behavior and make them able to choose between possible actions." Common sense would have it the other way around: that to the extent that someone is driven by goals his ability to choose decreases. The important question concerns in which way one can say that people "have" goals and whether one can sensibly assume that goals exist logically in advance of actions. For a discussion

is easy to simulate such ideas.[16] But it seems not possible to represent the notion of 'choice' in the conceptual framework of an automata-based simulation model. The reason is that 'choice' is not a representational category; it cannot be defined in purely descriptive terms. To say that an object has chosen some behavior implies, by the semantics of the word 'to choose', the assumption that the object might have chosen some other behavior. Therefore, the notion of 'choice' not only is different from, but is in a clear conceptual conflict with the assumption that the behavior of objects is governed by rules. In particular, conceiving of the objects in a simulation model as automata, by definition, implies that one cannot sensibly talk about choices.

6. One can, nevertheless, try to find a conceptual framework for a representation of objects that are assumed to be able to make choices. The notion of automaton, as introduced in section 3, can be used as a starting point. What would be necessary to make sense of saying, for example, that an object ω can choose one of the values of $\tilde{Y}$ for its output $y_t(\omega)$? A sufficient condition would be simply to drop the rule r_y, that is, no longer to assume that the value of $y_t(\omega)$ is in some way determined by the object's previous states and inputs. This then can be called a 'non-deterministic automaton'.[17] Exactly the abandonment of the assumption that the automaton's output is determined by a rule allows to say that the automaton has chosen whatever value will become realized.[18] Of course, given that the automaton has chosen some behavior one can, afterwards, try to find an explanation; but in order to stay

of these questions see Hawkins, 1968.

[16]This was done, for example, by Epstein and Axtell (1996, p. 25) with the following rule: "Agent movement rule M: • Look out as far as vision permits in the four principal lattice directions and identify the unoccupied site(s) having the most sugar; • If the greatest sugar value appears on multiple sites then select the nearest one; • Move to this site; • Collect all the sugar at this new position." Observing an object following this rule might then be described in terms of a desire to consume sugar, as much as possible.

[17]In the following, when using the word 'automaton' without a qualifying attribute, I always mean a deterministic automaton as introduced in section 3. Note also that there are other definitions of 'non-deterministic automata' in the literature. For example, Ginzburg (1968, p. 56) calls an automaton 'nondeterministic' if its initial state is not fixed but only given by a set of possible states.

[18]Non-deterministic should not be confused with stochastic automata. Like deterministic automata, also stochastic automata are governed by rules, the only difference is that in the latter case rules are formulated in terms of probability distributions (for state spaces). Stochastic automata therefore do not offer any help for a reconciliation of the conceptual conflict between actors and automata.

consistent with the supposition of a non-deterministic automaton, one cannot understand such post-factum explanations as providing rules that "govern" the automaton's behavior. One might add that presupposing a non-deterministic automaton also does not exclude the possibility to use knowledge about the automaton's past behavior to conceive of *epistemic* rules for predicting its future behavior. But such rules belong to the behavior of an observer and should not be confused with rules that are assumed to "govern" the behavior of observed objects.

7. The question how to think of choice is intimately related to another one concerning the possible meanings of the word 'can'. In defining a model, the designer needs to fix the sets of possible values for the model's variables. He might say, for example: I fix the set $\tilde{S}$ in the following way ..., meaning that the automaton's state variable *can* take values in $\tilde{S}$. But which values of the state variable are really possible is not fixed by such a definition but depends on the automaton's initial state and the transition rules. In any case, talking of variables that *can* take values in a set of possible values is unspecific since, without further explanation, it does not provide any idea about the processes that bring about specific values of the variables. In the realm of pure mathematics it is not necessary to say anything about such processes because one deals with logical possibilities. However, being concerned with real objects, the possible sense of models for their behavior crucially depends on that they provide at least some hints about the processes by which the model's variables get their values. And this, in turn, implies that a sensible model should provide hints for answering the question, Which objects, and by what kind of actions, *can* change (in the sense of being causal agents) at least some values of a model's variables? But then we are back where the discussion began. In an automata-based simulation model there simply are no causal agents which *can* change anything. Only the designer of the model can. But one then needs to refer to the designer as an actor who really *can* do something.[19]

8. The simple conclusion is that an automata-based simulation model cannot provide a conceptual framework that would allow to think of the

[19]It is interesting to note how J. McCarthy and P. J. Hayes (1969) have tried to provide a meaning for the view that an automaton, say A, being part of a system of two or more automata, *can* influence the state of the system. The proposed definition simply consists in the following: substitute A by an external agent, say A', having the same state space and output channel as A, and consider what options are available, for A', to bring about changes in the system.

model's objects as being actors who can choose between different behaviors. This would require to introduce a notion of non-deterministic automata. But the idea of a system consisting of two or more non-deterministic automata would not provide a useful starting point for a simulation model. One certainly could use a computer as a tool to find out all possible states for such a system. But the automata in such a system would do nothing because the model's designer, who "runs" the model, would have no means to provoke any specific behavior (simply as a consequence of the fact that he is the only real actor in the given context). Nevertheless, the notion of a non-deterministic automaton is useful as a conceptual tool and will be used again in a later section for a discussion of different kinds of interaction.

6 Why Using Simulation Models?

1. How to view an automata-based simulation model strongly depends on the language one has chosen to talk about its objects and corresponding state spaces. Without any further interpretation, such a model is simply an algorithm that allows to compute values of $(S, X, Y)_t$ for a number of time points, assuming that some initial assignment of values has been fixed. The basic motivation for the development of a simulation model is, however, that it might provide a tool for representation of, and reasoning about, objects and their behavior that we know, at least to some extent, by experience. This distinguishes the construction of a simulation model from using a computer simply as a computational tool, for example, to solve a set of equations. So it becomes essential that one can use basically the same language, both for communicating knowledge that derives from experiences with the objects to be modeled, and for talking about a simulation model. This creates the question whether the formal notations used in the definition of a simulation model can be given an interpretation that is consistent with the experiences that motivate the simulation. The question is particularly important when simulation models are intended to provide tools for reasoning about human actors in social settings since in this case we not only have a lot of experiences, but we also have a natural language that provides the corresponding meanings.

2. As already indicated in the previous section, the problem most importantly concerns the notion of 'actor'. The view that people are actors is fundamental not only for our understanding of specific human individuals that we personally know, but also for our understanding of social processes that we think of are "driven" by the behavior and interaction of people being actors. Consequently, if simulation models should be able to support reasoning about social processes, they should allow a representation of human individuals that is consistent with viewing them as actors. But then obviously arises a problem when one thinks of the framework for simulation models provided by the notion of automata. In this framework the objects are defined as automata and this clearly contradicts our common-sense view of actors, and in particular, human actors.

3. How is this problem discussed in the literature that develops simulation models? As far as I can see there is no serious discussion at all. Instead one finds a thoughtless use of metaphorical talk that, while actually talking about automata, uses a language that only makes sense when referring to (human) actors. That such metaphorical talk might be misleading has been recognized by several authors. Gilbert and Troitzsch (1999, p. 159) make the following remarks:

"Applied to people, the concept of agency is usually used to convey the purposive nature of human activity. It is thus related to concepts such as intentionality, free will, and the power to achieve one's goals. When applied to agents as computer programs, the scope of agency is generally rather weaker. Wooldridge and Jennings (1995) note that computer agents typically have the following properties:

- *autonomy* – agents operate without others having direct control of their actions and internal state;
- *social ability* – agents interact with other agents through some kind of 'language' (a computer language, rather than natural language);
- *reactivity* – agents are able to perceive their environment (which may be the physical world, a virtual world of electronic networks, or a simulated world including other agents) and respond to it;
- *proactivity* – as well as reacting to their environment, agents are also able to take the initiative, engaging in goal-directed behavior.

In addition, agents are often attributed a degree of intentionality. That is, their behaviour is interpreted in terms of a metaphorical vocabulary of belief, desires, motives, and even emotions, concepts which are more usually applied to people rather than to computer programs. For example, we might say that an agent built to collect relevant items from a supply of news articles was 'trying' to find something appropriate for the user, 'wanted' to get the most relevant article, and 'believed' that articles on a related topic would also be interesting. The habit of attributing intentionality to software agents in this way is liable to cause a great deal of philosophical confusion to the unwary (Shoham 1990). For our purposes, it is only necessary to view the ascription of intentionality to agents as a matter of modelling: a computer agent does not have intentionality, but is constructed to simulate some (much simplified) aspects of human intentions."

4. Before commenting on this passage, one should make a distinction concerning the different uses of automata-based models in computer science and social research. Originally developed in computer science and related technological disciplines, such models serve as tools to develop technical artifacts. The final goal is to develop automata, or networks of automata, that can substitute activities formerly done by humans or can support their activities in various ways. Concerning this context, metaphorical talk of automata as "agents" having properties similar to people is intended to be metaphorical and can be justified as providing intuitions for the design of automata.[20] As an example, consider the following statement from Y. Shoham (1994, p. 271-2):

"I will use the term '(artificial) agents' to denote entities possessing formal versions of mental state, and in particular formal versions of beliefs, capabilities, choices, commitments, and possibly a few other mentalistic-sounding qualities. What will make any hardware or software component an agent is precisely the fact that one has chosen to analyze and control it in these mental terms."

However, the situation is quite different when simulation models are proposed as tools to study the behavior of people and their social interaction. The models then no longer refer to automata that eventually should be realized as technical artifacts, but they refer to people as they really live and interact in a historical context. The focus is no longer on automata, and how to design automata for some technical purpose; but the model should allow to reason about real people and their social interaction in a sensible way. So it becomes of fundamental importance that the model is based on concepts that allow to talk about human actors without denying the essential properties that distinguish human actors from other kinds of object. But exactly this requirement is missed when, in a simulation model, actors are represented by automata and the resulting conceptual conflict is not taken seriously but ignored by using

[20]This also seems to be the main line of argumentation in the introductory contributions to Malsch, 1998.

metaphorical and, on the conceptual level, wrong characterizations of automata.[21]

5. Unfortunately, this important distinction concerning the aims of simulation models is not generally recognized in the literature. Consider, for example, the following statement by C. Castelfranchi and E. Werner (1994, p. x):

> "Like in Cognitive Modelling, so in Social Modelling AI is interested in describing and explaining (modelling) real social phenomena (such as negotiation, persuasion, alliances, conflicts, social hierarchies, etc.)."

This statement confuses two quite different goals. One is to design, and eventually build, automata that may be useful in some way; this is actually the main goal of AI and related technological developments.[22] The other is to develop models, that might or might not be simulation models to be described in a computational language, that can help in "describing and explaining real social phenomena". Given the latter goal, already the word 'simulation' becomes somewhat misleading since the real task is not so simulate in some way the behavior and social interaction of human actors, but to adequately describe and explain what they are doing, how they interact and what results from their interaction. For the same reason I also find the last sentence in the above quoted passage from Gilbert and Troitzsch misleading. As I understand this sentence, it says that, in the construction of simulation models, it is not necessary to use adequate conceptual tools *because* one only wants to simulate some aspects of their behavior. But this contradicts the claim that the simulation model provides a conceptual framework for a better understanding

[21] This is not always obvious because many authors use terminology that derives from human interaction in a completely *unspecific* way. Here is an example: "One of the most stated characteristics of agents refers to their autonomy, in a sense, that an agent's action does not need continuous human guidance or intervention [Shoham 1993]. Autonomy requires the capability of an agent to maintain its boundary, the distinction between itself and its environment [Heylighen 1990]." (Uhrmacher 1996, p. 433) Taken literally, this implies that virtually every object is autonomous. On the other hand, it is quite understandable that engineers are interested in constructing "autonomous" artifacts in the sense that they do not need continuous human guidance or intervention.

[22] One needs to add, however, that these technological projects are surrounded by a discussion that aims to make distinctions between 'artificial' and 'natural' obsolete. This becomes particularly visible in what has been called the "artificial life" project. As said by Langton, a primary promoter of this project: "We would like to build models that are so lifelike that they would cease to be models of life and become examples of life themselves." (Langton 1987, p. ??)

of human action and interaction.

6. One reason why there is almost no serious discussion of human action and interaction in the simulation literature might be that the focus is on a seemingly different question. In much of the literature the main questions do not concern actors and interaction, but the "emergence of system-level properties" out of the interaction of "simple agents following simple rules". In particular, most authors involved in constructing versions of artificial society models show a special interest in this problem. Epstein and Axtell (1996, p. 6) even proposed to make a study of this problem the defining characteristic of artificial society modeling:

> "Typically, we release an initial population of agent-objects into the simulated environment (a lattice of site-objects) and watch for organization into recognizable macroscopic social patterns. The formation of tribes or the emergence of certain stable wealth distributions would be examples. Indeed, the defining feature of an artificial society model is precisely that *fundamental social structures and group behaviors emerge from the interaction of individual agents operating on artificial environments under rules that place only bounded demands on each agent's information and computational capacity.*"

Before continuing with a discussion of problems that result from the notion of 'actor', in contrast with the notion of 'automaton', we shall try to find out what simulation models might contribute to an understanding of "emergent phenomena".

7 How to Think of "Emergence"?

1. As has often been noted, while ideas about "emergence" are quite central to much of the simulation literature, the notion is still obscure. In fact, no simple and consistent definition has yet emerged.[23] In order to approach the discussion it seems sensible, therefore, not to deal directly with the notion of 'emergence', but follow the complementary question, What might be possible results of interaction? As will be shown, this requires to distinguish between different kinds of interaction and also between different ways of describing possible results.

2. One reason why the discussion of 'emergence' in the simulation literature is difficult to understand, is due to confusing questions concerning

[23] This might also be due to the fact that the notion of 'emergence' already has a long and controversial history; see, e.g., Hoyningen-Huene, 1994.

results of interaction with the idea that one can sensibly distinguish two or more "levels" of reality. For example, Gilbert and Troitzsch (1999, p. 10) say:

"Emergence occurs when interactions among objects at one level give rise to different types of objects at another level. More precisely, a phenomenon is emergent if it requires new categories to describe it which are not required to describe the behaviour of the underlying components."

Confusion is likely to arise because the notion of 'level' is not only obscure but, if at all, refers to an idea conceptually different from 'emergence'. Think, for example, of a chemical reaction by which out of two different substances a new one results. This is often used as an example for emergence but obviously does not require any talk about levels. On the other hand, when sociologists make a distinction between a micro and a macro level, there is no obvious sense in which one can think of macro level phenomena as emerging from behavior defined on the micro level.

3. A widespread approach to explain ideas about different levels uses set-theoretic notions.[24] The basic idea is that a set of objects should be distinguished from its individual members. Correspondingly, there are two levels of description. One can describe the individual members of a set, and one also can describe the set, for example by saying that it contains a certain number of elements. In particular the whole conceptual apparatus of statistics is on the level of sets, beginning with the basic notion of statistical distribution. In fact, many approaches to look for "emergent phenomena" simply proceed in terms of statistical distributions (or notions derived thereof). However, it seems quite misleading to think of a statistical distribution as "emerging" from the properties of a set of objects that is used to define a statistical variable which, in turn, allows to define a statistical distribution. Using statistical distributions is just another way to describe the properties of a set of objects. It is the statistician who creates specific ways to describe sets of objects, but this does not make the set of objects in any sense a new object that in some way emerges from its members. In addition, one can note that there

[24]Not always explicitly, but most often the view is that higher-level units can be conceived of as sets consisting of lower-level units. The following quotation from Coleman (1976, p. 85) may serve as an example. "Most scientific disciplines must deal with the problem of shifting between levels of organization. In sociology, this shift manifests itself in the movement from persons as units of analysis to groups or organizations that have persons as members, to organizations or social systems that have groups or organizations themselves as members."

is no temporal relationship between a set of objects having properties and a statistical description of the distribution of properties in the set. The relationship is purely conceptual. On the other hand, the notion of 'emergence' as it is most often used refers to a process: there are two or more objects and out of their interaction "emerges" something new. This clearly implies a temporal view but not, in whatever sense, an idea of different levels.

4. I therefore propose to conceptually distinguish the notion of 'emergence' from ideas that one might develop regarding different levels; and furthermore, to think of emergence as implying that one can also think of a process that in some way brings about emergent phenomena, objects or situations. Linking then, as it is a central idea in many simulation models, conceptions of emergence to ideas about interacting objects, the heuristically leading question should be, What can result from interaction and how to find appropriate conceptual tools for representing those results?[25]

8 Can Different Levels be Simulated?

1. However, since the idea that one can sensibly distinguish different levels plays an important role in much of the simulation literature, one should also ask whether this idea can be expressed with the conceptual tools of automata-based models. Difficulties occur as soon as one tries to find a sensible definition. A first approach could be to conceive of a model that contains different kinds of object. Assume two kinds of object. The object set might then be written $\Omega = \Omega_1 \cup \Omega_2$, and the elements might be

[25]This proposal deliberately deviates from approaches that rely on some notion of 'system'. The following definition proposed by M. Bunge (1996, p. 20) may serve as an example: "I propose the following definition. P is an *emergent* property of a thing b if and only if either b is a complex thing (system) no component of which possesses P, or b is an individual that possesses P by virtue of being a component of a system (i.e., b would not possess P if it were independent or isolated)." There are two problems. A first one concerns the generality of the notion of 'system'. Most definitions imply that anything can be regarded as a system, that is, consisting of parts which are related, in some way. Bunge's definition then becomes virtually meaningless because there would be no properties which are not emergent. Another difficulty derives from that definitions in terms of systems tend to be purely static and therefore in conflict with temporal views of emergence. In contrast, the proposal made above closely follows the ordinary view that "something emerges". Following the rhetoric of systems, one would need to ask how systems come into being and develop. But given that anything can be viewed as a system the formulation would not be helpful.

called 'lower-level' and 'higher-level' objects, respectively. The rhetoric might be justified by specifying certain relations (via input and output channels) between the objects in the two sets. However, the distinction would be *purely* rhetorical. Without some externally given attributes it would not be possible to distinguish the elements in Ω with respect to membership in Ω_1 or Ω_2. Given only a description of the model, including names of the objects and a complete description of state spaces and rules, it would not be possible to distinguish between lower-level and higher-level objects. This is not to say that it would be impossible to design a model that contains different *kinds* of object that can be distinguished, for example, by reference to their state spaces and forms of embeddedness into local environments. But the argument is that one cannot formulate any essential distinction of levels.

2. The argument relies on the assumption that a notion of different levels requires to define some form of membership relation. In a way this can easily be done by defining subsets of the set of objects, Ω. However, these subsets cannot, simply by definition, be taken as representing new objects. In the given context, a minimal requirement would be to introduce rules for the behavior of new entities that shall be considered as objects. But in an automata-based model it seems not be possible to introduce rules for subsets of Ω in any meaningful way. Actually, the behavior of subsets of Ω is already completely determined by the rules prescribed for the elements of Ω. Any attempt to define behavior of the subsets would require to make use of statistical distributions that derive from the behavior (states) of the elements of the subset. In any case, it seems not possible to introduce a notion of higher-level objects which are not substitutable by a reference to their elements.[26] Any talk of different kinds of object, or even different kinds of actors, would be purely rhetorical without a conceptual foundations.

3. As an example to illustrate the difficulties I briefly refer to a model proposed by R. Axelrod (1995). The purpose and approach of the model is described as follows:

"How can new political actors emerge from an aggregation of smaller political actors? This chapter presents a simulation model that provides one answer. In its broadest perspective, the work can be seen as part of the study of

[26] This observation can be linked to a discussion of 'emergence' that focusses on possibilities of "reductionism". In automata-based models, properties of sets of objects can always be "reduced" to properties of their members.

emergent organizations through 'bottom-up' processes. In such 'bottom-up' processes, small units interact according to locally defined rules, and the result is emergent properties of the system such as the formation of new levels of organization." (Axelrod 1995, p. 19)

I shall not describe the complete model,[27] but concentrate on Axelrod's idea to define "collective actors". In his model, the basic object set is

$$\Omega \;=\; \{\omega_1, \ldots, \omega_{10}\}$$

with elements referred to as "political actors". The time axis $\boldsymbol{\mathcal{T}}$ is discrete with units called 'years'. The space, $\boldsymbol{\mathcal{R}}$, is taken as a set of fixed locations for the elements of Ω, arranged on a circle to allow to speak of neighbors. The idea then is that the objects can interact. In Axelrod's interpretation, this means that they can threaten, and eventually fight, each other in order to accumulate wealth. The state space, $\tilde{S}$, consists of two parts:

$$\tilde{S} \;=\; \tilde{W} \times \tilde{D}^n$$

$\tilde{W}$ is used as a property space for 'wealth', the numerical representation used by Axelrod is the set of natural numbers. $\tilde{D}$ is a property space for 'degree of commitment'. The idea is that the objects in Ω can have "commitments" to each other that can be numerically represented and change over time; Axelrod uses the numerical representation $\tilde{D} = \{0, \ldots, 10\}$.[28] For each point in time t, the state of an object $\omega \in \Omega$ may then be written in the following form:

$$s_t(\omega) \;=\; (w_t(\omega), d_t(\omega, \omega_1), \ldots, d_t(\omega, \omega_n))$$

Finally, Axelrod introduces rules that allow to sequentially calculate the states for all objects, beginning with an externally given initial configuration (so his model is a special case of a one-dimensional cellular automaton). Application of the rules implies that the mutual commitments change over time. This can be represented by a relation, say

$$\delta_t : \Omega \times \Omega \;\longrightarrow\; \tilde{D}$$

[27] A description and comments can also be found in Gilbert and Troitzsch 1999, pp. 133-6.

[28] "The model assumes that increases and decreases of commitment are in constant amounts, namely increments of 10 per cent. In addition, one actor's commitment to another can never be more than 100 per cent nor less than 0 per cent." (p. 25)

By definition of the rules, this is a symmetrical relation and the starting point for Axelrod's definition of "collective actors", also called "alliances". The definition consists in introducing subsets of Ω, that I will denote by $V_t(\omega) \subset \Omega$, consisting of all objects that have positive commitment to ω and do not have higher commitment to another object.

4. Is there any possibility to sensibly interpret the sets $V_t(\omega)$ as a new kind of object or even as "higher-level objects"? Already the notation makes this questionable. There is exactly one set $V_t(\omega)$ for each element $\omega \in \Omega$. This might be restricted by requiring that mutual commitment in "alliances" must exceed some minimum level. In any case, each set $V_t(\omega)$ can only be defined by referring to a specific object in Ω. In a sense they are comparable with local environments that dynamically change as implied by the model's rules. Correspondingly, they can easily overlap. This might not be an objection because individual objects may well belong to several higher-level objects at the same time. The main point is, however, that the subsets $V_t(\omega)$ are not required to formulate the model. In fact, they only serve as computational devices to facilitate the application of rules which are defined in terms of the individual elements of Ω. In particular, there are no rules defined on the level of the subsets $V_t(\omega)$, and, as has been remarked above, this would not be possible in the given modeling context. As a consequence, Axelrod's talk of behavior, with respect to the "alliances" $V_t(\omega)$, remains completely rhetorical. Furthermore, since these subsets cannot be sensibly interpreted as objects it also would not make sense to view them as emerging from the behavior of the objects in Ω.

9 What Might Result from Interaction?

1. Given the difficulties to find any sensible representation of different levels in automata-based simulation models, I return to the proposal made at the end of section 7: to think about the question, What might result from interaction and how to describe results? It is helpful to look at examples. Assume you take a stone and throw it against a window-pane. This will result in some kind of interaction between the stone and the window-pane and one can think of, or actually observe, certain results. Those results can be described by referring separately to the stone and to the window-pane. Alternatively, one can speak of an event and give a description of the event. While the notion of 'event' plays a fundamental role in our language and is often used to talk about results

from interaction, it is an open question whether events should be viewed as entities sui generis, different from objects. In any case, although it would be quite difficult to make the distinction precise, we are normally able, at least in the human environment, to distinguish between objects and events. And so this provides two ways to describe the results from an interaction of objects.

2. The example just given exemplifies only one kind of interaction. A different kind of interaction occurs when the interaction actually results in a new object. Standard examples refer to chemical reactions.[29] But there are other kinds of examples too, e.g. mixing substances or building a table by using raw material from different trees. The general idea is that two or more objects not just become parts of a new object, but by the same time loose there previous way of existing as separately identifiable objects.

3. So far there are two kinds of interaction. With respect to interaction of the first kind it normally does not make sense to view the outcome as a new object but is more natural to describe the result of the interaction as changes in the properties of the interacting objects. On the other hand, interactions of the second kind clearly result in new objects. Following J. St. Mill, one may call the first kind of interaction 'mechanical'; the second one will be called 'substantial interaction'. Now, social interaction among human actors most often does not fit into either of these two kinds. While there are clearly examples of mechanical interaction, an account of such examples does not require any specific reference to human actors. What remains? It should be easy to think of a lot of examples; people talk together, people live together, people work together. As a matter of fact, as the word 'interaction' is used in the literature, it not always refers to situations where two or more people do something together; so one needs again to distinguish different kinds of human interaction. This will be done in a later section. Here I shall follow the leading question, What might result from human interaction? More specifically, the question concerns how to describe what might result from human interaction and whether it might make sense to think of those results as new objects.

[29] The first systematic discussion of the differences between the two kinds of interaction can be found in J. St. Mill's "Logic" (book III, ch. 6). Mill's discussion also played an important role for the development of the notion of 'emergence', see Hoyningen-Huene, 1994.

4. It seems undeniable that humans, by their actions and interactions, can create new objects. In fact, much of the human environment consists of artifacts that have been created by humans. When interaction leads to the creation of new objects one can sensibly describe the new objects as having emerged from the interaction. With respect to the objects that are used as raw materials this may involve mechanical or substantial interaction. However, thinking of human actors as being the creators of new objects, *their* interaction most often cannot sensibly be described as mechanical or substantial. I shall speak of 'creative interaction' to refer to processes by which actors create new objects that are conceptually distinct from the actors.

5. One should note that a characterization of objects that emerge from creative interaction as *new* has two possible meanings that should be distinguished. One is temporally new, the other is historically new. Whenever human action, or interaction, creates an object it is temporally new in the sense that there is some temporal location where the object comes into being. To speak of historically new objects, in the ordinary understanding, requires a previous conception of kinds of object to allow for saying that two or more objects are of the same kind. Then, but only in retrospect, might it be possible to think of a temporal location where an object of some kind has been created for the first time.[30] In any case, both meanings of 'new' should be distinguished from characterizations in terms of 'being surprising' or 'unexpected' which have quite different connotations.

6. Creative interaction can result in many different kinds of new objects. These need not be technical artifacts in a narrow sense of the word. There are clear differences, for example, between building a computer and breeding new kinds of plant or animal. Concerning social research, two kinds of creative interaction have special importance. One is the creation of new human beings. This is also an example where any distinction between 'artifacts' and 'natural kinds' becomes misleading. The other is

[30]The idea that sometimes historically new objects come into being has an interesting implication: such objects cannot be generated according to a rule. This is because it must be possible to think of, and actually formulate, a rule that is assumed to govern some aspects of a process, logically in advance of any realization of the process. However, in order to formulate a rule for creative interaction one must already know the kind of object that might become created. But by implication of 'historically new', this knowledge is only available after at least one example of the new kind of object has been created. This reasoning exhibits an essential epistemological limit to using rule-based models for an understanding of historical processes.

that people are able to create organizations that can behave as corporate actors.[31] How to make a clear distinction between corporate actors and other forms of social interaction is an important, but also difficult question and I shall not here try to provide a discussion. The essential point is that a corporate actor can be viewed as an object sui generis, existing separately from its individual members. It therefore becomes sensible to think of processes by which corporate actors emerge.

7. Is there anything else that can result from social interaction? Our language offers many possibilities to refer to forms of social interaction, for example, 'talking together' or 'joint work on some project'. But it seems obvious that these are not new objects that emerge from social interaction. The expressions simply refer to some way in which social interaction takes place. Things become a little bit more complicated because our language not only allows to abstract forms of interaction from actual realizations but to use such abstractions as grammatical subjects in sentences, for example, to talk about a partnership of two people. But think of two persons who become engaged in a partnership. It would seem odd to say that, by this engagement, a new object comes into being, such that there would then be three objects, the two partners and, in addition, their partnership. The main point is that the notion of emergence only has a clear meaning when one can refer to some object that allows to think of a process by which the object actually came or, regarding future possibilities, might come into being. But most expressions that language offers to talk about social relations do not refer to objects but to *forms* of social interaction; and there simply is still no clear meaning in the idea that forms of social interaction might result (emerge) from social interaction.

8. One should not become confused by the fact that it is quite possible to make groups of people the "unit of analysis", for example, households. It is possible to refer to a collection of households and use statistical variables to represent properties of the households. But this does not make households a new kind of object that exist separately of, and in addition to, the individual persons living together in such a way that allows to view them as members of a household. The definition of statistical variables only requires something that can be used for attributing properties, and these things might well be defined as sets without implying an empirical existence of corresponding objects. This is not to deny

[31]The term is adapted from Coleman, 1974.

that one can sensibly speak, for example, of households, but to suggest that the word refers to a form of social interaction and therefore does not immediately allows to formulate questions in terms of emergence. It would, nevertheless, be possible to ask for how people become members of households; for example, two people might decide to live together or a child becomes born into an already existing household. But if two people decide to found a common household they do not, by this decision, create a new object. They simply plan certain forms of cooperation for the future, and again, this planning can take place in a broad variety of forms, for example, to set up an explicitly formulated contract or simply doing it. In any case, only the future behavior of the involved individuals can show whether they can sensibly be considered as being members of a common household, and this again testifies that the word, 'household', does not refer to an object but to (a broad variety of) *forms* of social interaction.

9. Unfortunately, a distinction between objects and forms of interaction becomes easily neglected when beginning with an abstract notion of 'system'. But this notion only has a clear meaning when it is possible, in the first place, to refer to an object. Then it becomes possible to think of the object as a system, meaning that it consists of parts which in some way interact and contribute to characteristic features of the object. This can be called an *analytical* use of the notion of 'system'; viewing an object as a system serves as a starting point for an analysis. On the other hand, when social scientists sometimes speak of "social systems", they cannot make any analytical use of this phrase simply because they cannot, then, refer to any kind of object. In this context the notion of 'system' is just a linguistic construct used to refer to some *set* of people who in some way interact. But most often such a set cannot sensibly be conceived of as an object, or even some kind of collective actor; instead, the primary analytical task concerns the *forms* of interaction, how to find appropriate conceptual tools for describing different forms of interaction and their results. Here the phrase 'results' might, or might not, refer to objects that result from human interaction. It seems obvious that, by actually performing interactions, the involved actors themselves undergo changes, learn and develop views and habits. On the other hand, it is in no way obvious how, and under what conditions, one can sensibly think of systems emerging (resulting) from human interaction.

10. Finally, questions concerning "emergent phenomena" (in the sense of new objects) should also be distinguished from questions concerning the

emergence of ways of describing social interaction and social relations. It is certainly possible to investigate the history of human language as it refers to historically changing forms of social interaction; and it seems quite plausible that social interaction also depends on how people describe and understand what they are doing. Nevertheless, if one wishes to develop a theory of social interaction, ordinary language does not already provide a sufficient conceptual framework; for the simple reason that there are different problem definitions. Ordinary language has its primary use in serving the practice of social interaction. In contrast, a theory of social interaction aims to understand how it works.

10 Interaction in Automata-based Models

1. One can think of interaction in two different ways. One is thinking of two or more objects as being dependent on each other, in some way; alternatively, one can follow the idea that actors can do something together. Since automata-based models do not allow a representation of actors, they can only be used to formalize some notion of interaction in the sense of mutual dependence. This, of course, is a further serious limitation of these models as tools for understanding social interaction. We shall consider therefore, in a later section, whether non-deterministic automata can provide a better framework for thinking about different forms of interaction. However, since the simulation literature actually uses automata-based models, it is important to understand the implied view of interaction.

2. Definitions are straightforward when using the notational framework introduced in section 3. An automaton, ω, is directly dependent on another automaton, ω', if an output channel from ω' is connected with an input channel of ω and, one may add, if the behavioral rule prescribed for ω's behavior uses the output from ω' as an argument in a nontrivial sense. One might then say that ω' can influence the behavior of ω (but remember the remarks about different meanings of 'can' made in section 5). This then leads to a corresponding definition of 'direct interaction': ω and ω' directly interact if ω is directly dependent on ω' and vice versa. It is obvious how, in the same way, one can also define non-direct dependencies and interactions.

3. This notion of interaction is essential for almost all simulation models proposed in the literature. In particular, it provides the starting point for the study of "emergent phenomena". It is important, however, in what sense one can speak of "emergent phenomena" when using an automata-based simulation model. The question, What might result from interaction? becomes identical, then, with the question, What processes can be generated with an automata-based simulation model? Of course, it would be possible to enlarge the conceptual framework and allow for new objects to be generated by already existing ones, simply by adding rules for birth (and death) events. But this is not meant when authors of simulation models speak of "emergent phenomena". They rather mean something that can be identified as properties of processes that can be generated by a simulation model. As said by N. Gilbert (1995, p. 148):

"Because complex systems, whether adaptive or not, consist of many agents, their behavior can be described either in terms of the actions of the individual agents or at the level of the system as a whole. In some system states the global description may be very simple (e.g. if the agents are either not interacting or interacting in repetitive cycles, the global description might be that 'nothing is happening'), or exceedingly complex (e.g. if the agents are in complete disequilibrium). In some circumstances, however, it may be possible to discover a concise description of the global state of the system. It is in these latter circumstances that it becomes possible to talk about the 'emergence' of regularities at the global level."

4. As a preliminary to follow these intuitions, one needs a representation of the process generated by a simulation model and then find ways to describe the process. The representation immediately follows from the conception of an automata-based simulation model. As already discussed in section 5, the process can be represented by a sequence of representational variables

$$(S, X, Y)_t : \Omega \longrightarrow \tilde{S} \times \tilde{X} \times \tilde{Y}$$

This immediately also sets the course for a description of the process: to interpret the variables as statistical variables and proceed in terms of their distributions. This allows to apply all the conceptual tools provided by statistics to describe a process that has been generated by a simulation model.[32]

[32]There are actually two somewhat different possibilities to approach a statistical description of processes. One, most often used in the current simulation literature, views a process as a sequence of cross-sectional variables. Another one derives from the possibility to identify objects over time. This allows to formally define "life courses" of the objects and use statistical tools for a description of these life courses.

5. However, while this is, at least in principle, straightforward it does not lead to any specific notion of "emergent phenomena". The passage quoted above from Gilbert is actually not very helpful. The author obviously tries to avoid the view that whatever might result from the automatons' interaction should be said to have "emerged". The idea is to link a notion of 'emergent phenomena' to specific *possibilities of describing* a process. But this creates several difficulties. First, what might be sensible criteria for distinguishing "emergent" phenomena from other ones? Next, any notion of 'emergent phenomena' becomes then dependent on an observer whose description of the process defines what he is willing to recognize as "emergent phenomena". And finally, as already remarked in section 7, it is no longer possible, then, to think of "emergent phenomena" as *resulting* from interaction. The reason is that the meaning of 'resulting' implies that one can think of a process that has brought about a new object, or situation; but properties of statistical distributions do not result, in this sense of the word, from properties of objects that a statistical variable is intended to represent. Otherwise one would simply confuse the different meanings of 'resulting' (which implies a temporal relationship) and 'representing' (which does not).

6. Some authors have proposed that it might be possible to define "emergent phenomena" by relying on some notion of 'predictability'. A version of this idea has also been proposed by Gilbert (1995, p. 149-50):

"Some criterion is required which will distinguish emergent behaviour from behaviour which is predictable from the individual characteristics of the agents. The description of complex systems suggests that a candidate criterion is that it should not be possible to derive analytically the global emergent behaviour solely from considerations of the properties of agents. In other words, emergent behaviour is that which cannot be predicted from knowledge of the properties of the agents, except as a result of simulation."

However, this proposal is difficult to make sense of. I do not deny that it might be possible to give a notion of 'unpredictability' (except by simulation) some meaning with respect to automata-based models (see, e.g., Darley, 1996). But obvious problems arise when referring to predictability in real-life situations. It is a fact that we, to some extent, are able to successfully predict future events, or "state of affairs". Whether this is possible depends on the time horizon and what it is to be predicted. In any case, predictions might turn out to be wrong. Should we say, therefore, that whatever happens in the future is emergent because it cannot be predicted with absolute certainty? But the notion of 'emergence' would then loose any specific meaning. On the other hand, should we say that some happening turned out to be emergent because a prediction failed? For example, you intend to visit another city and predict that, by using a car and performing appropriate actions, you will arrive there, after some while. Then, if you actually arrive, this was not an emergent phenomenon because successfully predicted; but if you don't arrive and instead wake up in a hospital, this was an emergent state of affairs. But talk of "emergent phenomena" would then become purely subjective, and completely dependent on realized outcomes of predictions. Furthermore, the same phenomena can both be emergent and not emergent simply by referring to different predictions.

7. Actually, most authors who claim to have created simulation models that can generate "emergent phenomena" do not provide any clear definition of what, in their view, distinguishes "emergent" phenomena from whatever else. They often simply rely on an observer's visual imagination as a means to "detect" whatever might then be called "emergent phenomena".[33] This approach should not only be criticized because it makes "detection of order and structure" to rely on preconceived intuitions which cannot be explicated in terms of well-defined concepts. There is a deeper problem concerning the claim that simulation models may help in theory construction. Visualization of "emergent phenomena" hides the theoretical deficit that we actually don't know what we are searching for when trying to find "order", or "structure", or what you like. Brian Smith (1996, p. 19) has pointed to this deficit with the following remarks concerning simulation models based on cellular automata.

> "It is relatively easy to program cellular automata with various kinds of rule, and to see – palpably, in front of one's own eyes – little 'organisms' and other organized or patterned entities emerge – dynamically, like worms or clusters or hives of activity. But when one 'sees' such a creature emerging, one is relying on one's perceptual apparatus. No one yet has a theory that, given a field of cellular activity, can reliably identify such 'objects' or 'emergent entities'. And *identification* is not really the problem, anyway. If the underlying theory – of selection, say, or organization, or behavioral emergence, or evolution – is to be worth is salt, it should be defined in terms of a theory of such things."[34]

[33]This is often seen as a particularly useful feature of computer simulations. For example, in the introduction to Liebrand, Nowak and Hegselmann (1998), the editors say: "An additional major contribution of computers to our understanding of social dynamics is that often, only through their use, visualization of social processes and their properties becomes possible. Intelligent color coding of numbers may enable us to use our most powerful sensory organ, namely our eyes, to recognize patterns we never would have recognized in looking at a list of thousands of numbers." (p. 16)

However, the main problem does not result from relying on visual imagination but from the belief that "emergent phenomena" can sensibly be defined in terms of statistical distributions. This is not to deny that statistical concepts can be useful tools for a description of sets of objects, or processes defined in terms of their attributes. But using such tools does not create new objects, only new ways of describing a given set of objects.

11 Technical and Political Models

1. Actually, what can be done with automata-based models is not an investigation of the emergence of new objects but something else. One can investigate how a process defined in terms of representational variables, $(S, X, Y)_t$, depends on rules that the model's creator has prescribed for the behavior of its objects. In order to follow this question it is certainly helpful to describe the process in terms of statistical distributions or concepts derived thereof. For example, one can create a model that simulates the behavior of pedestrians in certain environments by prescribing specific rules for their behavior, and then investigate how some measures of overall flow depend on the rules and the structure of the environment. It would be misleading, however, to assume that such a model could be used to detect how some kind of order emerges from the behavior of the pedestrians (as suggested, e.g., by P. Molnár, 1996). The mistake would be to believe that order can be "detected". 'Order' is not a descriptive category but essentially depends on a purpose that must be established logically in advance of any talk of order. Without being able to refer to some purpose, whatever process results from pedestrians' behavior may be called "ordered" or "unordered". And the same can be said, for example, regarding segregation models. Whatever spatial distribution of objects turns out to become realized might be described as "ordered" or "unordered", but without any meaning unless one has specified some criteria that make a distinction between different distributions sensible.

[34]The author then adds: "It has even been speculated that the entire field of nonlinear dynamics, popularly called 'chaos theory', could not have happened without the development of such displays. No one would have 'seen' the patterns in textual lists of numbers."

2. But then arises the question who can establish such criteria. Thinking of human actors in social environments it might be worthwhile to investigate *their* views, whether they actually have some ideas about "social order" and whether and in which way such ideas influence their behavior. But when using the conceptual framework of automata-based models *their* ideas cannot be represented; the model builder must establish his own criteria if he intends to distinguish between possible outcomes. Of course, this is exactly what the model builder wants to do if the model is intended to serve a practical purpose. He is then interested in how certain performance measures, derivable from observations of the simulated process, depend on specifications of the model and, in particular, the rules prescribed for the behavior of the model's objects. This is obvious when the model is intended to explore possibilities to design and build machines of whatever type. In a similar way models can serve to think about possible designs of organizations for human behavior. For example, one can think of models to explore implications of different designs of traffic rules. But obviously, the question then is not how "order" can emerge from local interactions, but how to *design* an environment, and rules, for the behavior of objects or actors.

3. It seems appropriate to speak of *technical models* if they are intended to explore possible designs of machines, and of *political models* if they are intended to serve a study of possibilities to organize human behavior. An obvious difference concerns the possibilities of prescribing behavior. In any case, although political models might use the conceptual framework of automata-based models, one would need a different interpretation of the model's rules. It would become wrong to think of behavioral rules, as this term was introduced in section 3, simply because nobody could guarantee that people behave as prescribed by the rules. Instead, the rules should be viewed as *norms*. The difference is significant. The notion of 'norm' implies that people can deviate, for whatever reasons. From a logical point of view, norms are suggestions for behavior. In fact, the very idea of norms presupposes that humans are not, or at least not completely, "governed" by behavioral rules.[35] It is this presupposition that distinguishes political from technical models.

[35] I here stress the difference because many authors in the tradition of behavioral science seem to believe that norms can be defined in terms of observable behavior. For example, R. Axelrod (1986, p. 1097) has proposed the following definition: "*Definition:* A norm exists in a given social setting to the extent that individuals usually act in a certain way and are often punished when seen not to be acting in this way."

4. Neither technical nor political models can sensibly be called simulation models. Their aim is not to simulate (whatever) but to serve as tools to think of possible designs. In fact, authors who propose to use simulation models in social research most often are not interested in political models but think of their models as conceptual tools to represent social processes as they actually develop in historical time. For example, Gilbert and Troitzsch (1999, p. 14) say:

> "We shall assume that there is some 'real world' phenomenon which you, the researcher, are interested in. This we call the *target* (Doran and Gilbert 1994, Zeigler 1985). The aim is to create a *model* of this target which is simpler to study than the target itself. We hope that conclusions drawn about the model will also apply to the target because the two are sufficiently similar. [...] In the social sciences, the target is always a dynamic entity, changing over time and reacting to its environment."

This implies that a model, in this sense, and in particular a simulation model, is not a conceptual tool to explore possible designs, but involves a *representational claim*.[36] In my understanding, it is exactly this representational claim that makes simulation models potentially interesting for social research, and so this claim should be taken seriously.

12 How to Think of Social Interaction?

1. The main limitation of automata-based models derives from the fact that they do not allow for a sensible representation of actors. This implies that using such models to simulate social processes requires a metaphorical interpretation that contradicts the conceptual framework for the model. One possible alternative, already pointed to in section 5, is to begin with the idea of non-deterministic automata. We shall continue with the notations already introduced for deterministic automata but

[36] Actually, such a claim is sometimes explicitly denied. For example, Conte and Gilbert (1995) made the following remarks: "Once the process of modelling has been accomplished, the model achieves a substantial degree of autonomy. It is an entity in the world and, as much as any other entity, it is worthy of investigation. Models are not only necessary instruments for research, they are themselves also legitimate objects of enquiry." (p. 2) The authors then add: "In this mode of research, the target is no longer a natural society, but an artificial one, existing only in the mind of the researcher." (p. 3) This obviously implies that the model has no longer any representational claim but simply serves to explore what can be done with computers; see also Gilbert and Troitzsch 1999, p. 19.

assume that there is no rule, r_s, that completely determines the automaton's state, $s_t(\omega)$, given values for its previous state and input, but that the automaton (ω) is free to choose some part of its state. One can assume, for example, that $s_t(\omega)$ consists of two parts, say

$$s_t(\omega) \;=\; (b_t(\omega), c_t(\omega))$$

in such a way that the first part, $b_t(\omega)$, is determined by a deterministic rule but ω is free to choose values for the second part, $c_t(\omega)$. The set of possible choices will be denoted by $\tilde{C}_t(\omega)$ and, as indicated by this notation, may depend on $b_t(\omega)$ and values in the input state space.

2. A first question concerns how to conceive of interdependencies among the objects. One possibility comes easily into mind. One can think that ω's choice set, $\tilde{C}_t(\omega)$, depends on some previous behavior of another object, say ω'. This is consistent with interpreting $\tilde{C}_t(\omega)$ as a set of alternatives from which ω can choose. However, using this interpretation, it would be somewhat misleading to say that ω's *choice* (in the sense of selecting one alternative) depends on the behavior of other objects. One should say, instead, that the *choice situation* of ω depends on the behavior of other objects. This kind of dependence which can, of course, be mutual will be called *c-dependence*; formally: ω is c-dependent on ω' in t if $\tilde{C}_t(\omega)$ depends on previous output from ω'.

3. While the notion of c-dependence seems quite obvious, it actually leads into conceptual difficulties. One of these difficulties can be explained as follows. Given a choice set $\tilde{C}_t(\omega)$, for an object ω, it should consist of *feasible* options. Thinking of several objects, each should be able to select one element of its individual choice set. However, one can easily imagine situations where the choice sets of two (or more) objects are, in some sense, incompatible. Think of spatial moves. The choice sets provide coordinates for possible locations. Now, in which way can we sensibly reason about a situation where, for two objects ω and ω', the intersection of their choice sets is not empty? One might say that, nevertheless, both can *try* to make an individual choice and finally they will be successful or not. However, the conceptual framework developed so far provides no means for talking about actions that objects can try to perform. It simply allows to speak of choices with the implied meaning that, when ω selects an element $\tilde{c} \in \tilde{C}_t(\omega)$, this will become realized.

4. The example shows that an approach to interaction that begins with a notion of choice sets immediately leads to a coordination problem. This is

well known in the literature dealing with distributed AI (see, e.g., Müller 1993, p. 12). However, one needs again to distinguish between different kinds of problem definition. In technologically oriented disciplines, like distributed AI, the problem is how to find solutions for the coordination problem that can effectively be implemented for a cooperation of artifacts (and, possibly, also human actors). This kind of problem definition depends on two major presuppositions. First, that their is someone who can design appropriate rules; and secondly, that one can design, or influence, the objects in such a way that they actually will follow the rules. A similar problem definition with respect to interaction of human actors would lead to political models as defined in section 11. However, the primary task of social research is different, not to construct political models but to understand how interaction among people really works, and to find a suitable conceptual representation.[37] Given this task it is, in fact, questionable whether a conceptual apparatus that begins with a notion of choice sets will enable us to understand social interaction.

5. Before trying to show this I briefly consider game theory as being the most developed approach to social interaction in terms of choice sets. It is remarkable how this approach has managed to avoid what was called above the coordination problem. The basic idea is to conceptually separate choices from outcomes. In a sense, following game theory, one can think of the elements of the choice sets as being actions, but actions with an undefined outcome. More precisely, each element of a choice set is defined by a set of possible outcomes, and it is assumed that the realized outcome depends on the actions selected by all players simultaneously. When referring to just two players, say ω and ω', the assumption is that one can rely on the existence of a mapping,

$$\pi : \tilde{C}_t(\omega) \times \tilde{C}_t(\omega') \longrightarrow \tilde{O}$$

that associates with each pair of selections from the choice sets a specific outcome, an element in an outcome space $\tilde{O}$ (which may then be linked to further variables in a model). By assuming that such a mapping actually exists as a (computational) rule, and consequently independent of what

[37] It is surprising that the coordination problem is almost never considered as a *theoretical* problem in the simulation literature. While the problem regularly shows up in the construction of models for interacting objects, it is most often treated just as a technical problem to be solved by the model's designer, for example, by introducing priority rules. From a theoretical point of view one should ask how the objects (actors) might become able to solve the coordination problem.

the actors actually might do, the game theoretic approach avoids the coordination problem, in so far as such a problem might result from incompatible choice sets. It is exactly the existence of the mapping, π, that makes the choice sets compatible. However, as a consequence of this approach to avoid the coordination problem, one looses the possibility of thinking about interaction in terms of cooperation. Not only are possible actions conceived of as selections of elements from choice sets which, by definition, must be done by each player separately. Moreover, the conceptual approach excludes the possibility to think of joint actions as means to achieve specific results.

6. We shall not discuss whether the game theoretic approach can provide useful models at least for some kinds of social interaction. In my understanding, its primary purpose should be viewed as providing a conceptual framework for strategic reasoning in situations where, for whatever reasons, actors do not want or are not able to cooperate, but not to provide a conceptual framework for a general theory of social interaction. In fact, most occurrences of social interaction take place as cooperation, meaning that two or more people are doing something together. This will be called 'cooperating interaction' or simply 'cooperation'. The question is how to find a suitable conceptual framework for describing, and reason about, this kind of interaction. The essential point is that cooperation cannot sensibly be described in a framework that conceives of actions as selecting alternatives from a choice set. One can easily imagine examples where this would lead to misleading descriptions. As an example, adapted from R. Tuomela (1984), one can think of two persons carrying a table upstairs. A proper description would require to assume that both persons not only have an understanding of what they want to achieve, but also are able to coordinate their individual actions in order to become part of a joint action. But this, in turn, implies that they have an understanding of what a joint action is, a way of doing something together. Only by dispose of this understanding they become able to make their individual actions to become a part of the joint action. Of course, where understanding is required, misunderstandings may take place.

7. The example just given, and many similar ones, support the view that also non-deterministic automata, when defined in terms of choice sets, cannot provide a suitable conceptual framework for a theory of social interaction. In my view, the problems do not primarily, or particularly, concern interaction, as opposed to individual action. It is a simple fact that humans can act and cooperate in several distinguishable ways.

It seems odd that some authors aim to develop theories intended to show how cooperation might be possible. Learning to act and learning to cooperate can only artificially be separated. This is not to deny that cooperation can take place in several different forms, including some extreme forms of strategic interaction that might give some plausibility to a game theoretic approach. If the goal is to develop models that allow to reason about the behavior of human actors, in whatever form this behavior may take place, the primary obstacle concerns the notion of 'actor'. I have tried to show that the conceptual framework provided by automata-based models contradicts attempts to think of the model's objects as actors. So there remains the fundamental question how to find an appropriate conceptual representation of actors. The idea that objects are actors to the extent that they are *not* "governed" by rules, as suggested by the notion of 'non-deterministic automata', might provide a starting point. However, it is difficult to see how this might lead to an understanding.

References

Axelrod, R. 1986. An Evolutionary Approach to Norms. American Political Science Review 80, 1095–1110.

Axelrod, R. 1995. A Model of the Emergence of New Political Actors. In: N. Gilbert, R. Conte, eds. Artificial Societies. The Computer Simulation of Social Life, 19–39. London: UCL Press.

Bunge, M. 1996. Finding Philosophy in Social Science. New Haven: Yale University Press.

Castelfranchi, C., Werner, E. (eds.) 1994. Artificial Social Systems. Berlin: Springer-Verlag.

Casti, J. L. 1992. Reality Rules: I. Picturing the World in Mathematics: The Fundamentals. New York: Wiley.

Coleman, J. S. 1974. Power and the Structure of Society. New York: Norton & Comp. [German translation: Macht und Gesellschaftsstruktur. Tübingen: Mohr 1979.]

Coleman, J. S. 1976. Social Structure and a Theory of Action. In: P. M. Blau, ed. Approaches to the Study of Social Structure, 76–93. London: Open Books.

Conte, R., Gilbert, N. 1995. Computer Simulation for Social Theorie. In: N. Gilbert, R. Conte, eds. Artificial Societies. The Computer Simulation of Social Life, 1–15. London: UCL Press.

Darley, V. 1996. Emergent Phenomena and Complexity. In: R. A. Brooks, P. Maes, eds. Artificial Life IV. Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems, 411–416. Cambridge: MIT Press.

Doran, J. 1998. Simulating Collective Misbelief. Journal of Artificial Societies and Social Simulation 1, 1. [http://www.soc.surrey.ac.uk/JASS/1/1/3.html]

Doran, J., Gilbert, N. 1994. Simulating Societies: An Introduction. In: N. Gilbert, J. Doran, eds. Simulating Societies. The Computer Simulation of Social Phenomena, 1–18. London: UCL Press.

Drogoul, A., Ferber, J. 1994. Multi-agent Simulation as a Tool for Studying Emergent Processes in Societies. In: N. Gilbert, J. Doran, eds. Simulating Societies. The Computer Simulation of Social Phenomena, 127–142. London: UCL Press.

Epstein, J. M., Axtell, R. 1996. Growing Artificial Societies. Social Science from the Bottom Up. Cambridge: MIT Press.

Gilbert, N. 1995. Emergence in Social Simulation. In: N. Gilbert, R. Conte, eds. Artificial Societies. The Computer Simulation of Social Life, 144–156. London: UCL Press.

Gilbert, N. 1995a. Simulation: An Emergent Perspective. University of Surrey: Department of Sociology. [http://www.soc.surrey.ac.uk/research/simsoc/tutorial.html]

Gilbert, N. 1996. Simulation as a Research Strategy. In: Troitzsch et al., eds., Social Science Microsimulation, 448–454. Berlin: Springer-Verlag.

Gilbert, N., Conte, R. (eds.) 1995. Artificial Societies. The Computer Simulation of Social Life. London: UCL Press.

Gilbert, N., Doran, J. (eds.) 1994. Simulating Societies. The Computer Simulation of Social Phenomena. London: UCL Press.

Gilbert, N., Troitzsch, K. G. 1999. Simulation for the Social Scientist. Buckingham: Open University Press.

Ginzburg, A. 1968. Algebraic Theory of Automata, New York: Academic Press.

Goldspink, C. 2000. Modelling Social Systems as Complex: Towards a Social Simulation Meta-Model. Journal of Artificial Societies and Social Simulation 3, 2. [http://www.soc.surrey.ac.uk/JASS/3/2/1.html]

Hanneman, R., Patrick, S. 1997. On the Uses of Computer-Assisted Simulation Modeling in the Social Sciences. Sociological Research Online, Vol. 2, no. 2.

Hawkins, D. 1968. The Natur of Purpose. In: H. v. Foerster, J. D. White, L. J. Peterson, J. K. Russell, eds. Purposive Systems. Proceedings of the First Annual Symposium of the American Society for Cybernetics, 163–179. New York: Spartan Books.

Hoyningen-Huene, P. 1994. Zu Emergenz, Mikro- und Makrodetermination. In: W. Lübbe, ed. Kausalität und Zurechnung, 165–195, Berlin: de Gruyter.

Kephart, J. O. 1994. How Topology Affects Population Dynamics. In: C. Langton, ed. Artificial Life III, 447–463. Reading: Addison-Wesley.

Langton, C. Studying Artificial Life with Cellular Automata. Physica 22D, 120–149.

Liebrand, B. G., Nowak, A., Hegselmann, R. (eds.) 1998. Computer Modeling of Social Processes. London: Sage.

Malsch, T. (ed.) 1998. Sozionik. Soziologische Ansichten über künstliche Sozialität. Berlin: Edition Sigma.

McCarthy, J., Hayes, P. J. 1969. Some Philosophical Problems from the Standpoint of Artificial Intelligence. In: B. Meltzer, D. Michie, eds. Machine Intelligence 4, 463–502. Edinburgh: Edinburgh University Press.

Minar, N., Burkhart, R., Langton, C., Askenazi, M. 1996. The Swarm Simulation System: A Toolkit for Building Multi-Agent Simulations. Santa Fe Institute. [http://www.santafe.edu/projects/swarm/]

Molnár, P. 1996. A Microsimulation Tool for Social Force Models. In: Troitzsch et al., eds. Social Science Microsimulation, 155–170. Berlin: Springer-Verlag.

Müller, J. (ed.) 1993. Verteilte künstliche Intelligenz. Methoden und Anwendungen. Mannheim: BI-Wissenschaftsverlag.

Orcutt, G. H. et al. 1961. Microanalysis of Socioeconomic Systems: A Simulation Study. New York: Harper and Row.

Rasmussen, S., Barrett, C. L. 1995. Elements of a Theory of Simulation. In: F. Morán, A. Moreno, J. J. Merelo, P. Chacón, eds. Advances in Artificial Life, 515–529. Berlin: Springer-Verlag.

Shoham, Y. 1994. Multi-Agent Research in the Knobotics Group. In: C. Castelfranchi, E. Werner, eds. Artificial Social Systems, 271–278. Berlin: Springer-Verlag.

Skinner, B. F. 1973. Beyond Freedom and Dignity. Harmondsworth: Penguin.

Smith, B. C. 1996. On the Origin of Objects. Cambridge: MIT Press.

Troitzsch, K. G. 1990. Modellbildung und Simulation in den Sozialwissenschaften. Opladen: Westdeutscher Verlag.

Troitzsch, K. G., Mueller, U., Gilbert, G. N., Doran, J. E. (eds.) 1996. Social Science Microsimulation. New York: Springer-Verlag.

Tuomela, R. 1984. A Theory of Social Action. Dordrecht: Reidel.

Uhrmacher, A. 1996. Object-Oriented and Agent-Oriented Simulation: Implications for Social Science Application. In: Troitzsch et al., eds. Social Science Microsimulation, 432–447. Berlin: Springer-Verlag.

Weisbuch, G. 1991. Complex Systems Dynamics. An Introduction to Automata Networks. New York: Addison-Wesley.