Martin Neumann (2008)

Homo Socionicus: a Case Study of Simulation Models of Norms

Journal of Artificial Societies and Social Simulation vol. 11, no. 4 6
<http://jasss.soc.surrey.ac.uk/11/4/6.html>

For information about citing this article, click here

Received: 11-Dec-2007 Accepted: 17-Jun-2008 Published: 31-Oct-2008

Abstract

This paper describes a survey of normative agent-based social simulation models. These models are examined from the perspective of the foundations of social theory. Agent-based modelling contributes to the research program of methodological individualism. Norms are a central concept in the role theoretic concept of action in the tradition of Durkheim and Parsons. This paper investigates to what extend normative agent-based models are able to capture the role theoretic concept of norms. Three methodological core problems are identified: the question of norm transmission, normative transformation of agents and what kind of analysis the models contribute. It can be shown that initially the models appeared only to address some of these problems rather than all of them simultaneously. More recent developments, however, show progress in that direction. However, the degree of resolution of intra agent processes remains too low for a comprehensive understanding of normative behaviour regulation.

Keywords:
Norms, Normative Agent-Based Social Simulation, Role Theory, Methodological Individualism

Introduction

1.1

The past decade(s) have witnessed a growing interest in the inclusion of norms in multi-agent simulation models. Numerous factors are responsible for attention being paid to norms, ranging from technical problems with moving robots (Shoham and Tennenholtz 1992; Boman 1999) to philosophical interest in the foundation of morality (Axelrod 1986; Skyrms 1996, 2004). So far, the investigation of normative multi-agent systems is a highly diverse field of research. Therefore an overview of agent-based social simulation models that exist so far might provide insight into converging or diverging trends in the conceptual realisation of normative agents as well as shedding light on whether and how particular attempts might benefit from the findings of other research fields.

1.2

The questions directed towards normative agent-based social simulation models in this survey originate in a distinct point of view: namely, they are posed from the perspective of the foundations of sociology. The rationale for choosing this perspective is that the notion of norms is a central concept in classical role theory of mid-twentieth century sociology (Parsons 1937; Parsons and Shils 1951; Merton 1957). Role theory claims that action is guided by a normative orientation. This theory of action is often paraphrased as the 'homo sociologicus'. For a long time this remained the dominant stream of sociological theory. Thus, normative agent-based social simulation do in fact address the questions of this old approach to the foundations of sociology: the normative orientation of actors. For this reason it will be asked whether these models are able to capture the methodological principles of this account.

1.3

However, in the past 20 years, this paradigm has been heavily criticised[1]: Role theory built on the old claim that social phenomena should be explained with social factors (Durkheim 1895). Roles are a pre-given element in this theoretical architecture. Roles (and thus: norms) emanate from society. However, the origin of both roles and society is left unexplained. It has been suspected that this is a reification of society. Others have suggested building sociology on the foundations of individual actors. This is the programme of the so-called *methodological individualism* (Boudon 1981; Raub and Voss 1981; Coleman 1990; Esser 1993). A shift 'from factors to actors' (Macy and Willer 2002) can be observed in the foundations of sociology. Agent-based models contribute to this theory building strategy: in agent-based simulation models (Artificial Societies), structures emerge from individual interaction. The great advantage of this methodology is that it enables the investigation of the feedback loop between individual interaction and collective dynamics. This has led to a growing awareness of its potential for investigating the building-blocks of social structure. For instance, it is even claimed that agent-based simulation allows us to "discover the language in which the great book of social reality is written" (Deffuant et al. 2006).

1.4

Nevertheless, the problem how to explain normative orientation still remains. Hence, to include norms in agent-based models would enable to recover the findings of role theory in terms of individual actors; in fact, it would represent a great advance for the foundations of sociological theory.

1.5

To examine how far this project has been already developed, this paper proceeds as follows:

1.6

First, the *questions* that need to be addressed to normative agents from the perspective of sociological role theory will be developed. To focus the investigation on methodological issues, these questions concentrate on the methodological characteristics of the role theoretic norm conception. This section contains two parts: In the first, a brief outline of the role theoretical concept of norms will be given (Section 2). It will be briefly outlined (Section 3) why and how normative agent-based models might contribute to discovering the language of social reality from the perspective of sociological critics of classical role theory.

1.7

The main part of this paper consists of a review of existing simulation models, which is organised as follows: Two classical papers, representing two typical approaches, are considered in more detail. The strength and weakness of both approaches will be elaborated (Section 4 — 5). Then a short overview of further developments will be given (Section 6). Two exemplary models of both traditions will be examined in more detail, again with the aim of answering the question: can a convergence of both traditions be observed, and how, if at all, does this contribute to an actor-

oriented reconstruction of normatively framed social roles?

Development of questions to normative agent-based models

2.1

In this section the questions will be developed that need to be posed to normative muti-agent models from the perspective of sociological theory. To specify where norms are placed in the theoretical architecture, let us consider the example of the famous Mr Smith, introduced by Ralf Dahrendorf (1956, pp. 19 ff.) in his analysis of the 'homo sociologicus' to characterise key elements of sociological role theory. We first meet him at a cocktail party and want to learn more about him. What is there to find out?

2.2

Mr Smith is an adult male, circa 35 years old. He holds a PhD, and is an academic. Since he wears a wedding ring, we know that he is married. He lives in a middle-sized town in Germany and is a German citizen. Moreover, we discover that he is Protestant and that he arrived as a refugee after the 2nd World War in a town populated mostly by Catholics. We are told that this situation caused some difficulties for him. His is a Lecturer by profession and he has two kids. Finally, we learn that he is the third chairmen of the local section of a political party, Y, a passionate and skilful card player and a similarly passionate though not so good driver. This approximates to what his friends would tell us.

2.3

In fact, we may have the feeling that we know him rather better now. After all, we have some expectations as to how a lecturer is likely to behave. As a lecturer he stands in certain relations to colleagues and pupils. As a father he will love and care for his children, and card playing is also typically associated with certain habits. If we know the party Y, we will know a lot more about his political values. However, all that we have found out represent *social facts*. There are a lot more lecturers, fathers and German citizens beside Mr Smith. In fact, none of this information tell us anything about the unique identity of Mr Smith. We simply discovered information about social positions, which can, of course, be occupied by varying persons. However, social positions are associated with specific *social roles*. Roles are defined by specific attributes, behaviour and social relations. Demands of society determine—to a certain degree—individual behaviour. Individuals are faced with obligations and expectations. This social demand is transmitted to the individual by *norms*. Norms are the 'casting mould' (Durkheim 1895) of individual action. They regulate how lecturers, fathers and members of political parties should act to fulfil the role expectations of society. In particular, Talcott Parsons (1937) emphasised that the ends of individual actions are not arbitrary, but rather are prescribed by social norms. Thus, norms represent a key concept within sociological role theory.

2.4

Moreover, Dahrendorf presumes that Mr Smith will undertake a considerable part of the education of his pupils without recourse to the cane. Presumably, nowadays we will be unlikely to find references to such educational practices. Likewise, it is improbable that we will be told anything about driving competence. Thus, we have learned another lesson: Norms may *change* over the course of time.

2.5

To examine the explanatory account of this theoretical approach, the present investigation will abstract from a description of the content of concrete norms. We concentrate on the methodological characteristics of norms in general, rather than the content of specific norms. On closer inspection of this example, we find out that the concept of social norms is characterised by three key elements:

- First, norms show some degree of generality. They are regarded as the 'casting mould' of individual action (Durkheim 1895). The very idea of role theory is that a social role must not be restricted to a unique individual. For instance, the roles of lecturer or chairman of a political party can be performed by different individuals. It might not, of course, be arbitrary as to who will play this role. In fact, it is a major focus of the empirical counterpart of role theory, statistical analysis of variables, to investigate the distribution of roles. For instance, monetary background might determine an individual's chances of securing an academic position. Nevertheless, academic positions are a feature of society, not the individual. In the classical account, the generality of norms is simply a given. However, agent-based models start from individual agents. Thus, in the individualistic approach norms have to spread in some way from one agent to another to gain generality. The explanation of norm spreading is essential for a reconstruction of social norms in terms of individual actors.
- Secondly, the role set of father or lecturer encompasses a huge action repertoire. The choice of a concrete action cannot be determined only solely by an external force. The ends of an action have to be determined internally by the individual actor executing a specific role. For instance, the norms to love and take care of his children have to be an internalised property of a father. Thus, a role also possesses a subjective element. It has already been highlighted by Durkheim (1903 [1973]) that the internalisation of norms constitutes a crucial element of the education process.
- Thirdly, this approach is characterised by a certain type of analysis: the normative integration of society (Parsons 1937; Davies and Moore 1945; Merton 1957). Hence, the question is to a lesser extent concerned with the origin of norms than with the function of norms for a society. For instance, the role of the father is to educate his child. The role of the lecturer is crucial for the socialisation of pupils. Thus, both roles are functionally relevant for the reproduction of the society.

Criticism of role theory

3.1

However, classical role theory has been severely criticised in the past decade. It is not the purpose of this article to provide a comprehensive review of this debate.[2] However, two points shall be highlighted that are relevant for normative multi-agent systems:

- First, the norm conception of role theory has only a quite dubious epistemological basis. For Durkheim as well as for Parsons, society is a reality '*sui generis*' (Balog 2000). Both authors did not pay much attention to the question of the origins of society. Thus, the focus of this approach is on a functional rather than on a causal analysis. This has been criticised as a reification of society (compare Gellner 1971; Archer 1995).
- Secondly, role theory has been criticised for sketching an oversocialised picture of man (Wrong 1961). Already in the 1960s, Homans (1964) claimed to 'bring man back in'. In fact, individual actors in role theory are more or less social automata.[3] If they have properly internalised the norms, they execute the program prescribed by their roles (Balog 2000).

3.2

Agent-based social simulation contributes to a better comprehension of these two problems. First, with this methodology, it is possible to generate macrosocial structures through individual interaction. It enables a causal understanding of the processes at work. Hence, normative agent-based models can provide a sound basis for the origins of a macrosocial structure like social norms. Also the second problem can also be addressed with agent-based models. Since structures on the macro level are a product of individual interaction, it seems possible to fill the gap between individual action and social structure with agent-based models. Hence, agent-based modelling suggests to contribute to some essential questions that are related to the question of the origin of society. Yet the question remains whether these are also able to rediscover the findings and results

of the role theoretic approach. Therefore they should capture the above mentioned properties of the social roles; that is, some answer need to be given to the following questions:

1. Can they provide insights into the normative regulation of society; that is, do they also reproduce the findings of a functional analysis of norms (*focus of contribution*)?

Moreover, do they allow for a causal reconstruction of the mechanisms that generate the functional interconnectedness on the social level? This implies that two further questions have to be addressed:

2. What transforms the agents in such a way that they factually follow norms? That is, what are the causal mechanisms at work that enable an internalisation of norms (*transformation problem*)?
3. By what mechanisms in the model can norm-abiding behaviour spread to or decay from one agent to another (*transmission problem*)?

Literature review

4.1

These questions will be examined in this section. Again, it has to be emphasised that the investigation concentrates on methodology, not on the contents of norms governing concrete roles such as father or lecturer. However, existing models are clustered around various intuitions about norms, conventions or standards of behaviour. The concrete research question differs from model to model. Some models concentrate on the emergence or spreading of norms. Others concentrate on functional aspects or the feedback of norms on individual agent's behaviour. A multiplicity of concepts is at hand. Hence, a comprehensive review of all models and accounts that may be in some way related to the study of norms would go beyond the scope of this investigation.

4.2

Instead, the report concentrates on a more specific sample. It focus on agent-based simulation models possessing certain characteristics: in some way both intra -and inter-agents processes need to be involved in the model. While the notion of inter-agent processes refers to a population of interacting agents, the notion of intra-agent processes indicates that the agents cannot not be purely reactive: some internal processes should be at work.[4] These processes have to be of such a kind that—in a broad and liberal sense—they provide insights into normative behaviour regulation and/or transformation.

4.3

The overwhelming mass of the research so characterised can be traced back to (or is at least influenced by) two traditions in particular: first, game theory and secondly an architecture of cognitive agents with some roots in Artificial Intelligence. Tradition and theoretical background has a direct impact on the terminology used. Depending on their background, the models tend to be communicated in different scientific communities. Additionally, references in articles tend to depend on their authors' background. For instance, five out of seven papers attributed to the AI tradition mention a more technical paper of Shoham and Tennenholtz (1992). This acknowledgement is predominantly enacted in the introductory notes, where one usually indicates the research field to which a paper is intended to contribute. Under the perspective of content, the models in the AI tradition typically contain references to conceptual articles relating to agent architectures. Articles with models in a more game theoretical tradition typically refer to game theoretic literature for the characterisation of the interaction structures in which the authors are interested.

4.4

Of course, this tradition-influenced framing, publishing and referencing is a tendency. It does *not*

constitute a clear-cut disjunction without any intersection. It has to be emphasised that this is neither a very precise nor a disjunctive categorisation. To some degree, the distinction between game theory and DAI is a distinction in the mode of speech employed by the authors. Some problems of game theoretic models could also be formulated in a DAI language and vice versa. The categorisation of models as following the DAI tradition shall only indicate that the agents employed by these models are in some way cognitively richer than those in the so-called game theoretic models.

4.5

Nevertheless, this distinction gives a rough sketch of the line of thought followed by the models, and also of the kind of problems, the concepts for their investigation, and the mode of speech in which the paper is presented. Moreover, this categorisation provide hints to other areas of research that are closely related to the models considered in this article: For instance, simulation is only a small sub-discipline of game theory in general and the distinction between analytical and simulation results is only gradual[5]. Simulation models might describe problems in game theoretic terms, but the method of resolution is not that of analytical game theory (Binmore 1998). In fact, investigating norms with the means of analytical game theory is a highly active research field. Also research on norms in the Artificial Intelligence tradition is by far not restricted to simulation experiments. For instance, important contributions can be found in more conceptually oriented considerations, often closely related to deontic logic (Boella et al. 2007). A number of contributions focus on the proper design of normative architectures[6]. A comprehensive review of all these approaches would soon exceed the scope of a single article. Identifying these two traditions enables to characterise the target of this review as a small, but important intersection of these two traditions: it concentrates on such accounts that describe results of simulation experiments. The particular focus of this review is an investigation of the contribution of simulation experiments to sociological theory.

4.6

The following considerations will be centred around the outlined distinction. First a closer inspection of the classical accounts will be undertaken. These models are selected because of their high impact on further research. How far do they proceed with regard to the questions outlined in the previous section? Does the difference in the theoretical tradition affect the models?

4.7

Subsequently, the models following these lines of tradition are taken under investigation from a broader comparative point of view: it will turn out that, in fact, converging trends can be observed. Therefore, a closer inspection of two models of both traditions will be undertaken. These are not selected because of their impact but because they can be regarded as striking examples of more elaborated accounts to integrate the frame of the other research tradition into the model. This will enable an evaluation inasmuch normative agent-based models have so far reached the goal to discover 'the language in which social reality is written'.

The beginnings

5.1

First, a more detailed description of the classical articles of these modelling traditions is given, each followed by an analysis of their strength and weakness with respect to the 'homo sociologicus' conception of norms. It has to be emphasised that it need not be the intention of the author to reconstruct a specific sociological theory. Some authors explicitly address those questions, others not. Some authors are only interested in explaining the emergence or function of norms without reference to any social theory. However, it is an implicit property of simulation models of norms that they provide building-blocks to rediscover role theoretical arguments and insights from an actor oriented perspective.

5.2

The classical paper in the game theoretic tradition is Robert Axelrod's 'an evolutionary approach to norms' of 1986. It has been analysed and replicated several times (e.g. Deguchi 2001; Galan and Izquierdo 2005), and remains the point of reference for this line of tradition. The classical model in the tradition of models employing cognitive rich agents is the model described in Conte and Castelfranchi's 1995 paper on 'Understanding the functions of norms in social groups through simulation'. This model has also been replicated several times (Conte, Castelfranchi and Paolucci 1998; Saam and Harrer 1999; Staller and Petta 2001; Hales 2002). It has been extended in several ways and is still the reference point for authors in this tradition.

1) Robert Axelrod: An evolutionary approach to norms. *American Political Science Review*, Vol. 80, 1986

5.3

In this paper simulation models of a norms game and a meta-norms game are described. Axelrod defines the existence of a norm as the extent to which individuals usually act in a certain way and are often punished when seen not be acting in this way.

5.4

To analyse possible mechanisms of how norms develop and change over time an evolutionary approach is utilised. It does not rely on the assumption of rationality, but on the assumption that effective strategies are more likely to be retained than ineffective ones. This is interpreted as a form of social learning.

5.5

The norms game works in the following manner:

At the beginning, an individual player has the options to defect (e.g. by cheating in an exam) or not defect. This is accompanied by a certain chance of being observed by other players. The defector receives a certain payoff, while all other players are slightly hurt. Yet, if player j observes the defection of player i, player j can decide to punish (or not to punish) player i. In the case of punishment, player i gets a negative payoff. However, player j has to pay an enforcement cost. The choice of the strategies is dependent on two variables: the *boldness* B_i, which determines the probability that player i will defect, and the *vengefulness* V_i , which determines the probability that player i will punish defectors.

5.6

The simulation proceeds as follows:

The agent population consists of 20 players. Their initial strategy is set at random. Each individual gets four opportunities to defect with a randomly determined probability of being observed. Then the reproduction rate of the players is determined: an individual with a score one standard deviation above the average gets two offspring, an individual with average success gets one offspring, and an individual with a score one standard deviation below the average gets no offspring. These steps are repeated for 100 generations. 5 simulation runs are executed.

5.7

The result of the norms game is ambiguous:

One run resulted in a high degree of vengefulness and a low degree of boldness. However, two runs resulted in a moderate level of both boldness and vengefulness, while a further two runs arrived at a state of almost no vengefulness combined with a high degree of boldness.

5.8

As it is in fact verified by Galan and Izquierdo (2005), Axelrod assumes that these different outcomes are due to one common mechanism: at first, the boldness level starts to decrease because of the costs of being punished. Thus, the rate of defection decreases. However, this leads to a decrease in the level of vengefulness, because punishment is also costly. This in turn makes it attractive to defect again. Thus, *the final stable state is a state without any norms at all*.

5.9

This result raises the question of how the establishment of norms is possible. Axelrod proposes a second mechanism called meta-norms: punish those who do not punish defectors.

5.10

The game proceeds like the norms game, yet a further step is introduced: As in the norms game, agent i has the choice to defect or not-defect. If agent j observes a defection it has the choice of punishing or not punishing agent i. However, if agent j decides not to punish, it can be observed with a certain probability by agent z. Now a further step is introduced: agent z can decide to punish (or not to punish) agent j for not punishing agent i. The probability for punishment is determined by the degree of vengefulness of agent z.

5.11

The result of the simulation is unambiguous:[7]

5.12

All runs resulted in a high degree of vengefulness and a degree of boldness near to zero. Thus, a *norm against defection* has been established. The mechanism behind this result is that the players have a strong incentive to be vengeful, simply to avoid punishment.

Axelrod's contribution to answering the questions:

5.13

The agents in the model are very flexible in term of changing their behaviour. Thus, a transformation of the agents' behaviour can be observed. Norm compliance is transmitted by punishment, that is to say, an external force. This represents an advantage of Axelrod's modelling strategy: The model provides a sound *causal mechanism of norm spreading*. Agents' adopt a norm in fear of punishment. It provides an answer to the transformation and the transmission problem.

5.14

In fact, Axelrod's model has been the starting point for many models investigating *normative dynamics* by applying a game theoretic mode of the problem description. The great advantage of this account is to shed light on the process of norm change. As it has been become apparent in discussing Mr Smith, this process can also be observed in human societies. However norm change is only barely captured by the functional account of role theory.[8] It is, in fact, a theoretical progress compared to the role theoretical approach.

5.15

On the other hand, the model only includes a restricted functional perspective: On an individual level, the agents' choice of action is guided by the functional consideration of calculating the expected utility. However, a corresponding analysis on the social macro-level cannot be found in Axelrod's account. This first result can be summarised in the following table:

Table 1: Axelrod's contribution to answering the questions

Transmission	yes
Transformation	yes

Function	no

5.16

However, from the perspective of the role theory of action, a weakness of this approach becomes apparent, one immediately related to the game theoretic problem description. Agents are faced with a strategic (typically binary) decision situation. Thus, they have a fixed set of behaviour (Moss 2001). Faced with this situation, agents choose the alternative that maximises their expected utility. Behaviour change goes along not with goal change. Agents can do no more than react to different environmental conditions. An active element of normative orientation in the choice relating to the ends of action cannot be found in a game theoretic approach. This is simply due to the fact that agents do not possess any internal mechanism to reflect and eventually change their behaviour, other than the desire to maximise utility. This point has already been highlighted in Parsons' critique of 'utilitarian theories' of action (Parsons 1937): namely, that the ends of individual actions are in some way arbitrary. In Axelrod's model, this circumstance can be nicely illustrated by the difference between the results of the norms game and the meta-norms game. Even though agents quickly modify their behaviour, the ends of the action remain unchanged: the goal is to maximise utility. In this respect, the relation between the action and the ends of the action remains arbitrary.

5.17

However, the very idea of role theory is to provide an answer to the question: Where do ends come from? Parsons' (and Durkheim's) answer was the internalisation of norms. A corresponding answer to this problem is not supplied in Axelrod's model. This is due to the fact that agents do not act because they want to obey (or deviate from) a norm. They do not 'know' norms. Their behaviour can only be interpreted as normative from the perspective of an external observer. Thus, *transformation is not identical with internalisation*. While the model provides a mechanism for behaviour transformation, it cannot capture the process of internalisation. Compared to the classical role theory, this is a principle limitation of a game theoretical description of the problem situation.

2) Rosaria Conte, Cristiano Castelfranchi: Understanding the functions of norms in social groups through simulation. In: Nigel Gilbert, Rosaria Conte (Eds.) *Artificial Societies: the computer simulation of social life*. London, UCL Press, 1995

5.18

The starting point of this paper addresses considerations concerning what should be considered as a norm: The paper differentiates between norms of co-ordination, which might be purely conventional, and norms that include explicit prescriptions, directives, or commands. The latter form the focus of this model. It investigates the functions performed by norms, in particular for the control and reduction of aggression.

5.19

The simulations takes place on a 10 × 10 grid. This world contains 50 agents and 25 randomly scattered 'food resources'. The agents are equipped with an initial 'strength' value of 20. Every action is costly in terms of reducing the agents' strength value. The food items contain a nutritional value of 40. Once consumed, this value is added to the agents' strength value. If a resource is consumed, a new one appears at a random location.

5.20

The agents are able to perform the following actions: They can move one step per round. Each move reduces the strength of the agent by 1. Agents are equipped with a visual field. Additionally, they can observe their environment by so-called 'smelling'. This means that they can identify a food resource, but cannot detect whether there is another agent between them and the 'smelled'

food. If an agent occupies a cell which contains a food resource, it eats the food. This lasts for two rounds. Finally, the agents are aggressive: if a neighbouring agent is eating food, they are able to attack them. The result of the attack is dependent on the relative strength of the agents. An attack reduces the strength of both agents by 4 units, no matter who is winning. Yet, the winner gains the food.

5.21

Three kinds of experiments are undertaken:

1. *Blind aggression*: the initial set-up contains no means to control aggression. Agents always attack eaters when no other food is available. In particular, they do not take the agent's strength into consideration. They will attack, even if they are weaker and bound to lose the battle.
2. *Strategic aggression*: this is a first step in aggression control. Aggression is constrained by strategic reasoning. Strategic agents only attack those agents whose strength is not higher than their own.
3. *Normative agents*: in this setting, norm-based action control is introduced. Normative agents do not always avoid aggression. Yet they obey a finder-keeper norm: the agent that initially detects a resource is regarded as its possessor. Note, that multiple possession is allowed. Possession is maintained over time, even when agents move away from their possession. Normative agents do not attack agents eating their own food. Nevertheless, they can eat possessed food from unoccupied cells. In this case they can be attacked by other agents.

5.22

The units of analysis pertaining to these experimental conditions are as follows: First, the rate of aggression, that is, the number of attacks occurring during the simulation experiment, is stored. Secondly, the average strength of the agents at the end of the simulation is recorded. At this level of analysis, the strength of individual agents is interpreted as a measure of welfare. Finally, the variance of the individual strength of agents is documented. This is interpreted as a measure of equality. The greater the variance, the more unequal the agent society. The performance of the three experimental conditions is then compared according to these three units of measurement.

5.23

The results are as follows:

1. *Degree of aggression*: the experimental setting with agents blindly attacking each other shows the highest degree of aggression. The strategic scenario has a considerably lower rate of attacks. Thus, aggression can be controlled even with non-normative means. However, the experimental condition of a normative agent society results in the lowest number of attacks.
2. *Aggregated welfare*: inverse ordering can be observed for the welfare of societies: the 'blind attack' society is the poorest, followed by the strategic agents' society. The normative agents' society is the richest.
3. *Equality*: with regard to the variance of strength it can be observed that the strategic scenario exceeds the blind aggressive scenario to a small degree. Hence, the strategic agents' society is the most unequal. The experiment with normative agents results in by far the most equal distribution of welfare.

Conte and Castelfranchi's contribution to answering the questions:

5.24

This paper provides an example of how agent-based models are able to contribute to a functional analysis of norms. The evaluation of welfare, equality or aggression control concerns the social macro level. The paper shows that a norm of aggression control can be beneficial for a society as a whole. However, in contrast to the classical scheme of functional explanations in the social

sciences, this result is reached by interactions of individual agents.

5.25

The classical scheme of a functional explanation assumes a social phenomena P, whereby P has a (functional) effect n for the society. Individual actors have reasons to practise P independently of the functional effect n. Moreover there is a feedback loop so that in the case of a decrease of P, there is a cause for an amplification of P. Thus, society remains in equilibrium. It is claimed that this state of affairs is crucial for the 'survival' of the society.

5.26

However, this explanation has often been criticised (compare e.g. Homans 1964; Haller 1999) for failing to explain the primary existence of the social phenomena P and for the fact that the society is assumed to be an unexplained 'social organism'. For instance, the explanation of a phenomenon by reference to the survival of the society is regarded as a suspicious reification. The origin of the societies needs is unexplained. The mere existence of a phenomenon indicates its necessary character. However, the model proves that in principle it is also possible to undertake a functional analysis on the basis of individual actors without reference to the needs of a social unit. Obviously, the model does not include a process of self-regulation. The task of the model is much less demanding. However, this avoids the danger of teleological explanations. Compared to the classical role theory, this is in fact a theoretical progress.

5.27

Moreover, in this model a much stronger notion of norms is deployed than in Axelrod's model. Norms are not just reached by mutual agreement, but are an explicit action routine. This conception of norms allows for a wider field of applications that could cover the role theoretic norm conception: these can be interpreted as internalised properties of the agents. However, the process of internalisation is not investigated. Contrary to the concept of norms in Axelrod's model, not even a mechanism for a normative transformation of the agents is given.

5.28

The model also shares a weakness of the role theoretic account in the social sciences: In particular, norms are simply given in this model. The process of norm spreading is left unexplained. Neither a transformation nor a transmission process can be discerned in the model. In this respect, the criticism of functional explanations also holds for this model. These criticism go along with the fact that agents cannot deliberate about norms and eventually deviate from them. Agents have no individual freedom. As critics accuse the role theory, the action repertoire is also (depending on conditions) deterministic. Thus, even though the authors succeed in 'bringing man back in', the agents in the model are merely normative automata. Insofar as the norms are a pre-given element in the model, the approach can also be regarded as an 'oversocialised' conception of man. The result can be summarised in the following table:

Table 2: Conte and Castelfranchi's contribution to answering the questions

Transmission	no
Transformation	no
Function	yes

Summary

5.29

Thus, from a role theoretical perspective, the strength of both models is complementary:

On the one hand, Axelrod's model, employing a game theoretic mode of problem description, provides a causal explanation for norm spreading. This includes a designation of mechanisms of norm transmission and normative transformation. An investigation of the functional effect of norms is left aside. Even though the model provides a mechanism for the transformation of the agents, this is not identical with norm internalisation, which remains beyond the scope of this account.

On the other hand, Conte and Castelfranchi's model, utilising cognitive agents in the AI tradition, provides a causal explanation of how norms can have a functional effect on the social level. However, the process of norm spreading is left unanalysed. The result can be summarised in the following table:

Table 3: Comparison of both accounts

	Axelrod	Conte/Castelf.
Transmission	yes	no
Transformation	yes	no
Function	no	yes

5.30

The process of norm internalisation is beyond the scope of both models. While Axelrod's model contains a process but no internalisation, the norms in Conte and Castelfranchi's model can be interpreted as internalised but no mechanism is given. Hence, it is be plausible to assume that both lines of research could benefit from each other. Such convergence could represent a step towards the explanatory account of the classical role theory in terms of individual actors.

Further development

6.1

This section examines the extend to which research to date on normative agents has progressed towards convergence. To this end, a *sample* of game theoretic models and models employing cognitive agents will be investigated.

a) Models employing a game theoretic mode of speech:

- Coleman, J. (1987) The emergence of norms in varying social structures. *Angewandte Sozialforschung,* Vol. 14 (1) 17 - 30.

  Coleman investigates the effect of interaction structures on the evolution of co-operation in a prisoner's dilemma situation. Only small groups can prevent the exploitation of strangers.

- Macy, M.; Sato, Y. (2002) Trust, cooperation, and market formation in the U.S. and Japan. *PNAS,* Vol. 99, 7214 - 7220.

  Macy/Sato examine the effect of mobility on the emergence of trust among strangers in a trust game. While agents with low mobility trust only their neighbours, high mobility supports the evolution of trust among strangers.

- Vieth, M. (2003) Die Evolution von Fairnessnormen im Ultimatumspiel: eine spieltheoretische Modellierung. *Zeitschrift für Soziologie,* Vol. 32 (4) 346 - 367.

Vieth investigates the evolution of fair division of a commodity in an ultimatum game. Including the ability to signal emotions leads to a perfectly fair share. If detection of emotions is costly the proposals even exceed fair share.

- Bicchieri, C.; Duffy, J.; Tolle, G. (2003) Trust among strangers. *Philosophy of Science,* Vol. 71, 286 - 319.

  Bicchieri et al. present a model of a trust game. It demonstrates how a trust and reciprocate norm emerges in interactions among strangers. This is realised by several different conditional strategies.

- Savarimuthu, B.; Purvis, M.; Cranefield, S.; Purvis, M., (2007) How do norms emerge in Multi-Agent Societies? Mechanism Design. *The Information Science Discussion Paper Series* , 2007 (01).

  Savarimuthu et al. study the convergence of different norms in the interactions of two different societies. Both societies play an ultimatum game against each other. Two mechanisms are examined: a normative advisor and a role model agent.

- Sen, S.; Airiau, S. (2007) Emergence of norms through Social Learning. *IJCAI-07*, 1507 - 1512.

  In this model, a co-ordination and a social dilemma game are examined. Agents learn norms in repeated interactions with *different* agents. This is denoted as social learning to distinguish this interaction type from repeated games with the same player. The whole population converges to a consistent norm.

b) Models utilising cognitive agents:

- Castelfranci, C.; Conte, R.; Paolucci, M. (1998) Normative Reputation and the costs of compliance. *JASSS,* 1 (3) 3 http://www.soc.surrey.ac.uk/JASSS/1/3/3.html.

  This is an extension of the author's first model. The paper studies the interaction of different agent populations. The interaction leads to a breakdown of the beneficent effects of norms, which can only be preserved with the introduction of normative reputation and communication among agents.

- Saam, N.; Harrer, A. (1999) Simulating norms, social inequality, and functional change in Artificial Societies. *JASSS,* 2 (1) 2 http://www.soc.surrey.ac.uk/JASSS/2/1/2.html

  Saam and Harrer present an extension of Conte and Castelfranci's model. They investigate the influence of social inequality and power relations on the effectiveness of a 'finder-keeper' norm.

- Epstein, J. (2000) Learning to be thoughtless: Social norms and individual computation. *Center on Social and Economic Dynamics Working Papers,* No. 6.

  Epstein examines the effect of norms on both the social macro- and the individual micro level. On the macro level, the model generates patterns of local conformity and global diversity. At the level of the individual agents, norms have the effect of relieving agents from individual thinking.

- Flentge, F.; Polani, D.; Uthmann, T. (2001) Modelling the emergence of possession norms using memes. *JASSS,* 4 (4) 3 http://www.soc.surrey.ac.uk/JASSS/4/4/3.html

Flentge et al. study the emergence and effects of a possession norm by processes of memetic contagion. The norm is beneficent for the society, but has short-term disadvantages for individual agents. Hence, the norm can only be retained in the presence of a sanctioning norm.

- Verhagen, H. (2001) Simulation of the Learning of Norms. *Social Science Computer Review,* 19 (3) 296 - 306.

  Verhagen tries to obtain predictability of social systems while preserving autonomy on the agent level through the introduction of norms. In the model, the degree of norm spreading and internalisation is studied.

- Hales, D. (2002) Group Reputation supports beneficent norms. *JASSS,* 5 (4) 4 http://jasss.soc.surrey.ac.uk/5/4/4.html

  Hales extends the Conte and Castelfranchi model by introducing stereotyping agents. Reputation is projected not on individual agents but on whole groups. This works effectively only when stereotyping is based on correct information. Even slight noise causes the norms to breakdown.

- Burke, M.; Fournier, G.; Prasad, K. (2006) The Emergence of Local Norms in Networks. *Complexity,* 11 (5) 65 - 83.

  Burke et al. investigate the emergence of a spatial distribution of a binary norm. Patterns of local conformity and global diversity are generated by a decision process dependent on local interaction with neighbouring agents.

6.2

Obviously, all these models have been developed for differing concrete purposes. To examine the extend to which these models capture the explanatory problems of the contribution problem, transformation problem and transmission problem, the various accounts of the different models will be outlined in a table. Moreover, a short hint to the concrete implementation is provided.

Table 4: Tabular comparison

	Contribution	**Transformation**	**Transmission**	**Implementation**
Axelrod (GT)	norm dynamics (*norms* broadly conceived!)	sanctions	social learning; replicator dynamics	dynamical propensities
Colemann(GT)	norm dynamics	punishment by defections (memory restrictions for identifying defections as sanctions)	a) group size (acquaintance) b) additionally: replicator dynamics[1]	conditional strategies
Macy and Sato (GT)	norm dynamics	losses by exclusion from interaction	social learning	dynamical propensities
Vieth (GT)	norm dynamics	losses by rejection	social learning; replicator dynamics	dynamical propensities
Bicchieri et al. (GT)	norm dynamics	sanctions by retaliating super game strategies	strategy evolution; replicator dynamics	conditional strategies

Savarimuthu et al.(GT)	norm dynamics; functional analysis	losses by rejection; Advice	Advice updating based on collective experience	dynamical propensities
Sen and Airiau (GT)	norm dynamics	experience	social learning guiding behaviour convergence	dynamical propensities
Conte and Castelfranchi 95(AI)	functional analysis	–/–[2]	–/–[2]	conditional strategies
Castelfranchi, Paolucci and Conte 98(AI)	functional analysis	updating conditionals (of strategies) through knowledge	updating knowledge by experience (and communication[1])	conditional strategies
Saam and Harrer (AI)	functional analysis	a) –/–[2] b) internalisation[1]	a) –/–[2] b) obligation[1]	conditional strategies
Epstein (AI)	norm dynamics; functional analysis	observation	social learning	dynamical updating
Flentge et al. (AI)	functional analysis	memetic contagion	contact	conditional strategies
Verhagen (AI)	norm dynamics	internalisation	communication	decision tree
Hales (AI)	functional analysis	updating conditionals (of strategies) through knowledge	updating knowledge by experience (and communication[1])	conditional strategies
Burke (AI)	norm dynamics	signals	social learning guiding behaviour convergence	dynamical propensities (Threshold)

[1] only in a second experiment

[2] the agents are/are not already moral agents

6.3

The tabular description allows for a quick look at what progress has been made so far with regard to the explanatory problems:

1. Both traditions still predominantly contribute to a different focus of analysis. As can be seen from the table, while all models in the game theoretic tradition investigate some form of norm dynamics, most of the models in the AI tradition investigate functional aspects of norms. There are exceptions such as the functional analysis of Savarimuthu et al., but here degree of convergence is low.
2. However, with regard to the transformation problem, the cognitive agents have become more flexible than in the very first model of Conte and Castelfranchi. So far, a number of models have implemented some mechanisms of agent transformation. However, a key difference between both traditions still remains: while game theoretic models mainly concentrate on some form of sanctioning as the transformation mechanism (Sen's model is an exception), in models of cognitive agents a number of different accounts is used.[9]
3. Obviously, the transformation has to be transmitted in some way. Thus, the transmission problem is no longer a blind-spot of cognitive agents. By comparison, communication plays

a much more important role, and is much more explicitly modelled in models with the AI tradition. However, an investigation of the process of internalisation remains in the fledgling stages.

4. It is striking that social learning is implemented in many game theoretic models by a replicator dynamics. If applied in a context where no real natural selection, rather than some kind of learning, is at work, then using a replicator dynamics amounts to saying: Somehow the individuals learn in a way that—measured by the relative overall success of their type of behaviour—more successful types of behaviour become more frequent. As an effect this may be true. However, no mechanism is indicated. In this dimension, the models struggle with the same kind of problem as functional analysis which the individualistic program tries to resolve; namely, the lack of a causal explanation.

6.4

It has become apparent that with regard to the transformation and the transmission problem, the borderlines of both approaches are no longer so clear cut. To gain a more detailed insight into the convergence of models, an example from both traditions has been selected for closer inspection: The models of Verhagen (2001) and Savarimuthu et al. (2007). These models have been selected since they represent the most advanced examples of models that incorporate elements from the other tradition.

1) Harko Verhagen (2001) Simulation of the Learning of Norms. *Social Science Computer Review*, Vol. 19

6.5

Verhagen investigates the problem of how to obtain predictability on the social level while preserving autonomy at the level of individual agents. Therefore, norms are employed.

6.6

In the model, the agents are situated in a two dimensional grid landscape containing on each cell either the resource A, the resource B, or nothing at all. The agents are clustered to a group with one agent denoted as leader. Agents have the choice of either doing nothing, moving to another cell, or of consuming either resource A or B. To decide which action to perform, the agents calculate the expected pay-off in a decision tree. For every possible action, the utility of the action is multiplied with the probability of success. The action with the highest score is chosen.

6.7

To integrate autonomous individual agents into a group, the agents have two decision trees: a private decision tree, denoted as *self-model*, and a group decision tree, the *group-model*. The self-model contains personal evaluations and the group-model represents what the agent presumes to be group evaluations. The group-model is the agent's appreciation of group norms. Faced with a decision situation, an agent combines both decision trees. The self-model is weighted with an *autonomy value* a, the group-model is weighted with (1—a). The agent then announces its decision to the group and executes the decision. Group members send the agent their own self-model relating to the situation. After receiving n messages (n is the agent's memory length) the agent updates its group-model. The feedback from the leader is weighted more strongly than those of other agents. Also the self-model is updated in a manner that is dependent on the factual pay-off gained by the executed action. This is denoted as feedback from the environment.

6.8

This framework is used to calculate the *spreading and internalisation of norms*. These variables are used to investigate whether norms have the capability to induce predictability on the social level, while preserving the agent's autonomy. For this reason several hypotheses are tested with particular emphasis being placed on norm spreading and internalisation.[10] Norm internalisation is calculated according to the sum of the differences between the self-model and the group-model of

all agents with regard to all action alternatives. With the self-model denoted as s_i, the agent's group-model as g_i, n action alternatives, and m agents, this reads as follows:

$$\Sigma(\Sigma\, |s_i - g_i| / n)\ m \qquad (1)$$

6.9

Norm spreading is calculated according to the deviation of the sum of the personal group-models of the individual agents from the mean group-model of all agents with regard to the number of alternatives. With g_i as the personal group model, g as the mean group model, n alternatives, and m agents, this reads as follows:

$$\frac{\Sigma\sqrt{\Sigma\, g_i - g / n}}{m} \qquad (2)$$

Surprisingly, Verhagen claims, that a higher norm spreading indicates a higher variance of behaviour.[11] Conversely, a higher variance of behaviour is defined as a lower degree of norm internalisation.

Verhagen's contribution to answering the questions

6.10

The strength of the model lied in its high resolution of both inter- and intra-agent processes. Contrary to the forerunner of the tradition of cognitive agents, the model of Conte and Castelfranchi, it has a dynamic view on norms. Thus, it captures both the process of norm transmission and agents' transformation, which is left aside in the classical model. While Conte and Castelfranchi's agents are normative automata, a degree of autonomy is preserved in the agents of this model. Traditionally, this represents the strength of the game theoretic tradition. However, contrary to Axelrod's model—the forerunner of the game theoretic tradition—the model is exceptional with regard to the problem of internalisation: it explicitly specifies a causal mechanism of the process of internalisation.[12] It is thus, able to overcome the particular weak point of game theoretic models: namely, that agents do not 'know' anything about the concept of norms. The model constitutes a step towards convergence regarding the advantages of both approaches. Moreover, via the notion of a leadership value, it also includes the effect of social differentiation on normative dynamics.

6.11

In this respect, the model enables theoretical progress: it is highly advanced in constructing a feedback loop between individual and collective dynamics. By the combination of the self- and the group-model a representation of the (presumed) beliefs held in the society is integrated in the belief system of individual agents. Conceptually, this is quite close to Mr Smith.

6.12

However, the implementation of internalisation is conducted on a rather ad hoc basis. The formula calculates the difference between personal and (presumed) social values. This captures the idea of internalisation. Yet in the formula the difference can vary from one time step to another. The degree of internalisation can also diminish. This is in contrast to the socio-psychological assumption that it is difficult to change norms once they are internalised. Internalised norms are presumed to be tightly entrenched by processes in early childhood (Wallis and Poulton 2001). Thus, the model is able to generate an effect which might look similar (but not identical) to processes in human societies, but crucially it does not represent the processes at work. However, at this point, the model reproduces a comparable weakness of the sociological account: Also classical role theory paid more attention to the functional effects—that is, norm internalisation—than to

cognitive mechanisms of the socialisation process.

6.13

Moreover, it has to be emphasised that—compared to the roles performed by Mr Smith—in this model the norms are not particularly rich in content. In this respect, it is more a methodological exercise. It does not contribute to an analysis of the function of norms. These are too weak in content to reasonably examine their performance for the society. Therefore, and finally, another model, from the game theoretic tradition is examined.

2) Bastin Savarimuthu, Maryam Purvis, Stephen Cranefield, Martin Purvis (2007) How do Norms emerge in Multi-Agent Systems? Mechanism Design. ***The Information Science Discussion Paper Series*****, No. 2007 / 1**

6.14

This paper studies two mechanisms with regard how expectations converge. Shared expectations are defined as norms. The models simulate the interaction of two agent societies with two different norms. The assumption is that during the course of interaction the two different norms converge, and moreover that this convergence improves the performance of the society.

6.15

To verify this assumption an *Ultimatum game* is set in motion. One player receives a certain amount of a commodity under the condition that it proposes a part of it to another player. The proposal can vary between 0 and 100% of the whole amount. The other player has two options: to accept or reject the proposal. If it rejects the proposal, both players get nothing. The agents play the game against all members of the other society. The two societies differ in their principles of proposal and rejection. Society 1 is denoted as selfish, society 2 as benevolent. Agents of the selfish society try to maximise their utility. They propose the least amount of money while accepting every amount apart from 0. This is consistent with Rational Choice assumptions. Benevolent agents, on the other hand, propose more than a fair share, but reject every proposal above a fair share. The values of proposal and acceptance are the norms under investigation.

6.16

In the first mechanism proposed, the agents have two possibilities to obtain a specific norm: they can follow a group norm (G norm) or a personal norm (P norm). Both norms continuously evolve based on social learning. While the G norm is shared by all members of the society, the P norm is specific to individual agents. The probability that they choose either the G norm or the P norm is calculated according to an autonomy value (between 0 and 1) of the agents. The G norm of each society is calculated by a so-called normative advisor agent: at the end of each round of the game, all agents from each society submit their successful proposal and acceptance values to their normative advisor agent. The average successful value is then used by the normative advisor agent to update the G norm of the society. The P norm is updated by each agent itself, dependent on its success.

6.17

In the second mechanism, instead of a normative advisor agent, a role model agent is introduced. The agents dispose only of a P norm. However, agents can decide to ask the role model agent for advice. If they ask for advice, they can decide if they follow the advice or not. Both decisions are based on the autonomy value of the agent. The role model agent calculates advice in the same way as the normative advisor agent updates the G norms.

6.18

The questions under investigation relate to the average performance of the societies in different settings and ask whether the proposal norms stabilise on a commonly shared value. Three experiments are undertaken: an initial experiment with societies that do not change initial norms,

followed by experiments with both mechanisms.

6.19

Societies that resist change return the worst performance. Obviously, the norms of the two societies do not converge in this case. However, norms do converge if either mechanisms one or two is applied. In the long run, the average score in both cases is near to a utopian society of a perfectly fair share. However, convergence time is faster when mechanism one is applied. The authors ascribe this result to the fact that mechanism two does not possess the concept of a G norm.[13]

Savarimuthu's contribution to answering the questions:

6.20

As usual in game theoretic models, the model provides an answer to the transmission and transformation problem. Moreover, this model explicitly addresses the question of the performance of norms for the society. It contributes to a functional analysis. This is only rarely observed in models within the game theoretic tradition. Moreover, by the notion of a normative advisor or role model agent, social norms are explicitly represented in the model. It thus represents an important step to integrate the perspective of the cognitive tradition in a game theoretical framework.

6.21

However, the price the authors have to pay is that this model also shares a weakness of classical sociological accounts. Consider the normative advisor and role model agent: they are both crucial for the model's results, since they possess global knowledge of all successful proposal values in their society. Their knowledge exceeds that of any individual agent. In fact, a role model is a well known concept in social science, but the traditional concept of a role model is based on the notion of prestige and status in peer groups. This is very different from the implementation in this model.

6.22

What is missing is a justification of these agents: what is their epistemological status apart from the purely computational level? The notion of normative advisor or role model agent is in danger of intermixing a target system with a computational trick. Such effect generating modelling is particularly suspicious because it is not only a modelling shortcut to generate effects assumed to be generated by processes that are at work in reality, but are not explicitly modelled. In fact, no such processes are specified by the authors. Considered as a real world entity, the normative advisor or role model agent would equate to an omniscient central bureaucracy. This, of course, is very unlikely. They can be suspected of representing a reification of society. Hence, it turns out that the model is faced with the same objections that already have been raised against functional explanations—in particular from the perspective of methodological individualism.

Conclusion

7.1

An examination of normative agent-based social simulation models raises many questions, not all of which can be answered in a single paper. In particular, the following two questions have *not* been addressed:

1. The concrete content of specific norms has not been considered. The circumstances that the models examined in this paper originate in highly diverse research fields, is reflected in the different nature of the norms used in the models. Many of the game theoretic models investigate the emergence of norms centred on trust and reciprocation. However, it would call for ethnographic studies and participatory modelling methods to determine what the essential norms of human societies to be replicated in computer models actually are.
2. Moreover, an examination of the concrete mechanisms at work in the diverse models

generating specific effects has not been undertaken. Presumably, it would be possible to assemble typical model structures resulting in typical effects. For instance, in two models (Epstein's model and the model of Burke et al.) an effect of local conformity and global diversity has been generated. In both models, this result is caused by local interactions. However, it is left for future work to provide a comprehensive overview.

7.2

The focus of this article was first to detect the methodological requirements for agents to compare agent-based models with role theoretical arguments and findings. This approach has been guided by the claim of agent-based modelling to 'discovering the language in which the great book of social reality is written'. The finding that norms regulate a good deal of social interaction remains beyond dispute. Thus, to finally achieve this ambitious goal, it is essential to gain a proper understanding of the structure of normative influence by the means of the actor oriented approach of agent-based modelling. This report has accomplished an identification of three fundamental methodological tasks that normative multi-agent systems need to fulfil for this purpose:

- in some way a transmission of norms has to be ensured by the model. To regulate social interaction norms have to possess a certain degree of generality. If norms are not assumed to be pre-given, it has to be explained how they spread in a society.
- The process of norm transmission implies that in some way a transformation of the agents must be possible. The transformation problem can be further subdivided into strategic adaptation to, and the internalisation of norms.
- Finally, it is an advantage if an analysis of norms at work can include the function of norms in some way. Such an inclusion allows the investigation of a broader range of research questions.

7.3

Secondly, the extend to which these requirements are fulfilled by the existing agent-based simulation models of norms has been assessed. Existing models originate primarily in two different traditions: Game theory and Artificial Intelligence. The first models of Axelrod (game theoretic tradition) and Conte and Castelfranchi (AI tradition) could only partially fulfil these requirements, failing in other aspects:

7.4

Axelrod's model is effective at explaining the dynamics of norms, in particular the strategic adaptation of agents to changing environmental conditions. Namely, they react to punishment. Thus, there exist a mechanism for norm transmission and agent transformation. In contrast to classical sociological accounts it lacks of an active element of normative orientation in the choice of the ends of action. The agents do not 'know' norms. Thus, the models does not capture the process of norm internalisation. Also a functional analysis is not in the scope of the model.

7.5

On the other hand, Conte and Castelfranchi's model primarily demonstrates the effects of norms. Moreover, it includes norms that exceed strategic adaptation. Norms can be interpreted as an internalised property. However, this is also the weak point of the model: the agents are merely normative automata. No mechanisms for the transmission and transformation problem are given.

7.6

The further development shows a convergence of both traditions. The models of Verhagen and Savarimuthu et al. include elements of the other line of thought. This enables (partial) answers to the questions of transmission, transformation and contribution. However, the review shows also how difficult it is to implement the requirements: While Verhagen's model is able to generate the effect of internalisation, but does not represent the process of internalisation, in Savarimuthu's model there is not even a process indicated which could generate the effects of normative advisor or role model agents. This suggests a suspicion of reification. In this respect, both models also

replicate not only the findings, but also the shortcomings of classical role theory. There is, then, still a lot to do with regard to achieving a comprehensive understanding of how actors produce, and are at the same time a product of social reality.

7.7

In conclusion, future work could profit from a finer-grained resolution of internal processes of normative reasoning based on explicit representations of norms. While agent-based modelling has reached a substantial understanding of inter-agent processes, an investigation of the recursive impact of inter- and intra-agent processes is still in its fledgling stages.

Acknowledgements

This work has been undertaken as part of the Project 'Emergence in the Loop' (EmiL: IST-033841) funded by the Future and Emerging Technologies programme of the European Commission, in the framework of the initiative “Simulating Emergent Properties in Complex Systems”. The author would like to thank Rainer Hegselmann for helpful discussions and comments, two anonymous referees for helpful and encouraging recommendations and Teresa Gehrs for correcting English expressions. Their contribution is gratefully acknowledged.

Notes

[1]It is often denoted as 'variable sociology'—with a critical undertone. The empirical counterpart of role theory is the statistical analysis of relations between variables. A criticism has been levelled that already in the process of drawing a representative sample, individual actors are neglected. However, the relations between variables do not provide a causal explanation (Esser 1987; Faulbaum 1992). It is claimed that its partial success has been only accidental: since in the 20th century the ties between social origins (for instance Catholicism or working-class) and specific attitudes (e.g. towards the pope) had been very strong. Therefore such interconnections could appear in statistical samples, even if they do not provide a causal generative mechanism. However, it is claimed that in course of the process of individualisation since the 1980s these ties have been lost and these 'artificial' relations have vanished (Esser 1989, 1996).

[2]Beside the fundamental paradigm shift towards methodological individualism, Parsons has also been criticised for inherent inconsistencies (compare Balog 2000; Haller 1999; Oakes 1980; Warner 1978; Gouldner 1971; Black 1961). This judgement is emphasised also by evidence from attempts to model Parsons' theory (Jacobsen and Bronson 1997).

[3]This is in remarkable contrast to Parsons original intention: in fact, he criticised 'utilitarian' theories as deterministic. He claimed that the active role of an actor is “reduced to one of the understanding of his situation and forecasting of its future course of development” (Parsons 1937 [1968], p. 64). Thus, the actor is reduced to a situational automaton. Parsons claimed that a 'voluntaristic' theory of action has to include the active choice of the ends of action. However, in explaining these ends, he relied on the pre-existence of social norms and in so doing reduced the individual actor again to the status of an automaton. Normative orientation is identified with conformity with norms. This is in contradiction to his own approach.

[4] For this reason, the work of Brian Skyrms, for instance, is not included. Without a doubt, the evolution of the social contract (Skyrms 1996) is a highly relevant question for the foundation of social norms. However, Skyrms' models remain on the level of population dynamics. Intra agent processes are not taken into regard. The results of the models may be true, but the mechanisms

cannot be covered by such an approach. Here we will concentrate on models including in some way intra-agent processes.

[5]A Replication of a simulation model developed by Robert Axelrod undertaken by Galan and Izquierdo (2005) use analytical tools as well as simulation experiments.

[6] For an overview of the broad range of moral dynamics compare Hegselmann (2008). A representative sample of normative architectures is examined by Neumann (2008).

[7]Yet Galan and Izquierdo (2005) prove that the results are not unequivocal.

[8]The criticism has been levelled that too much attention has been paid to the concept of a an equilibrium (Merton 1957; Gouldner 1971). This shortcoming is closely related to functional explanations. If a deviation exist from the equilibrium, it is assumed that forces also exist that push the social system into a state of equilibrium again.

[9]Even though in the models in the AI tradition often punishment is possible, punishment does not induce a transformation of the punished agent.

[10]The hypotheses are first, that a higher degree of autonomy reduces the predictability of the behaviour. Secondly, that a higher leadership value induces a higher predictability of the behaviour. Thirdly, if the personal decision tree equals the initial group decision tree, it is assumed that the predictability will be higher compared to an initial random group decision tree. In fact, a higher autonomy value leads to a higher degree of norm spreading. Surprisingly, a higher autonomy does not lead to a lower norm internalisation—which indicates (according to Verhagen's assumptions) a higher variance of behaviour. Thus, the Hypothesis is rejected for norm internalisation. The same result holds for the effect of the leadership value: A higher leadership value leads to a higher norm spreading but not to a higher degree of norm internalisation. Only the final assumption is verified for both norm spreading and internalisation.

[11]Presumably, the assumption is that higher norm spreading indicates a higher difference between the agent's self- and the group-model. However, no reference to the self-model is given in the formula.

[12] A similar concept can also be found in the model of Saam and Harrer (1999). The authors deploy the notion of institutionalisation of norms in a society as a whole. Institutionalisation of a norm n means that n is saved in the memory of each agent of the society. Thus, a norm is institutionalised, if it is internalised in each agent of the society. While internalisation refers to the micro-level of individual agents, institutionalisation is a concept that operates on the social macro-level. However, in Saam and Harrer's model, institutionalisation is simply switched on (or off) by a so-called redistribution agent. There is no mechanism at work that could explain the transmission of a norm to individual agents.

[13]The slower convergence time might be due to the fact that in the case of mechanism two the autonomy value is applied twice: for determining the probability that agents ask for advice, and again for the probability that they accept it.

References

ARCHER, M. (1995) *Realist Social Theory: the Morphogenetic Approach*. Cambridge: Cambridge University Press.

AXELROD, R. (1986) "an evolutionary approach to norms". *American Political Science Review* 80 (4) 1095 - 1111.

BALOG, A. (2000) "Theorie als 'theoretisches System': Parsons Beitrag zur soziologischen Theorie". In Staubmann, H., Wenzel, H. (Eds.) *Talcott Parsons: Zur Aktualität eines Theorieprogramms,* Wiesbaden: Westdeutscher Verlag.

BOMAN, M. (1999) "Norms in artificial decision making". *Artificial Intelligence and Law* 7 (1) 17 - 35.

BINMORE, K. (1998) "Review of the book: The complexity of cooperation: agent-based models of Competition and Collaboration, by Axelrod, R. Princeton, Princeton University Press". *Journal of Artificial Societies and Social Simulation* 1 (1) http://jasss.soc.surrey.ac.uk/1/1/review1.html.

BLACK, M. (1961) "Some Questions about Parsons' Theories". In Black, M. (Ed.) *The Social Theories of Talcott Parsons*, New York: The Free Press.

BOELLA, G., van der Torre, L., Verhagen, H. (2007) "Introduction to Normative Multiagent Systems". *Dagstuhl Seminar Proceedings 07122*.

BOUDON, R. (1981) *The Logic of Social Action*. London: Routledge.

CASTELFRANCHI, C., Conte, R., Paolucci, M. (1998) "Normative Reputation and the Costs of Compliance". *Journal of Artificial Societies and Social Simulation* 1 (3) 3 http://jasss.soc.surrey.ac.uk/1/3/3.html.

COLEMAN, J. (1990) *Foundations of Social Theory*. Harvard: Belknap.

DAHRENDORF, R. (1956) *Homo Sociologicus. Ein Versuch zu Geschichte, Bedeutung und Kritik der Kategorie der sozialen Rolle*. Opladen: Westdeutscher Verlag.

DAVIES, K., Moore, W. (1945) "Some Principles of Stratification". *American Sociological Review* 10 (1) 242 - 249.

DEFFUANT, G., Moss, S., Jager, W. (2006) "Dialogues Concerning a (Possibly) new Science". *Journal of Artificial Societies and Social Simulation* 9 (1) 1 http://jasss.soc.surrey.ac.uk/9/1/1.html.

DEGUCHI, H. (2001) "Mutual Commitment, Norm Formation and indirect Regulation of Agent Society". *Graduate School of economics, Kyoto University Working paper* No. 53.

DURKHEIM, E. (1970 [1895]) *Regeln der Soziologischen Methode*. Neuwied: Luchterhand.

DURKHEIM, E. (1903 [1973]) *Erziehung, Moral und Gesellschaft. Vorlesungen an der Sorbonne 1902/1903*. Neuwied: Luchterhand.

ESSER, H. (1987) "Warum die Routine nicht weiterhilft—Bemerkungen zur Stagnation soziologischer Theoriebildung". In Mueller, N., Stachowiak, H. (Eds.) *Problemloesungsoperator Sozialwissenschaft Vol. 1*. Stuttgart: Enke

ESSER, H. (1989) "Verfaellt die 'soziologische Methode'?" *Soziale Welt* 40, 57 - 75.

ESSER, H. (1993) *Soziologie: allgemeine Grundlagen*. Frankfurt a. M.: Campus.

ESSER, H., (1996) "What is wrong with 'Variable Sociology'?" *European Sociological Review* 12 (2) 159 - 166.

FAULBAUM, F. (1992) Von der Variablenanalyse zur Evaluation von Handlungs- und Prozesszusammenhaengen. *ZUMA Arbeitsberichte,* May 1992.

GALAN, M., Izquierdo, L. (2005) "Appearances can be deceiving: Lessons learned Re-Implementing Axelrod's 'Evolutionary Approach to Norms'". *Journal of Artificial Societies and Social Simulation* 8 (3) 2 http://jasss.soc.surrey.ac.uk/8/3/2.html.

GELLNER, E. (1971) "Holism versus individualism". In Brodbeck, M. (Ed.) *Readings in the Philosophy of the Social Sciences*. New York: Macmillan.

GOULDNER, A. (1971) *The Coming Crisis of Western Sociology*. London: Heinemann.

HALES, D. (2002) "Group Reputation supports beneficent norms". *Journal of Artificial Societies and Social Simulation* 5 (4) 4 http://jasss.soc.surrey.ac.uk/5/4/4.html.

HALLER, M. (1999) *Soziologische Theorie im systematisch-kritischen Vergleich*. Leverkusen: Leske + Budrich.

HEGSELMANN, R. (2008) "Moral dynamics." In: *Encyclopedia of Complexity and System Science*. Berlin: Springer

HOMANS, G. (1964) "Bringing man back in". *American Sociological Review* 29 (5) 809 - 818.

JACOBSEN, C., Bronson, R. (1997) "Computer simulated empirical tests of Social Theory: Lessons from 15 Years' experience". In Conte, R., Hegselmann, R., Terna, P. (Eds.) *Simulating Social Phenomena*. Heidelberg: Springer.

MACY, M., Willer, R. (2002) "From Factors to Actors: Computational Sociology and Agent-Based Modelling". *American Review of Sociology* 28, 143 - 166.

MERTON, R. (1957) *Social Theory and Social Structure.2nd Edition*. Glencoe: The Free Press.

MOSS, S. (2001) "Game Theory: Limitations and an Alternative". *Journal of Artificial Societies and Social Simulation* 4 (2) 2 http://jasss.soc.surrey.ac.uk/4/2/2.html.

NEUMANN, M. (2008) "A classification of normative architectures". *Proceedings of the World Congress on Social Simulation 2008*.

OAKES, G. (1980) *Die Grenzen kulturwissenschaftlicher Begriffsbildung*. Frankfurt a. M.: Surkamp

PARSONS, T. (1968 [1937]) *The Structure of Social Action. A Study in Social Theory with Special Reference to a Group of Recent European Writers*. New York, London: Free Press.

PARSONS, T., Shils, E.A. (1951) *Towards a General Theory of Action*. Harvard: Harvard University Press.

RAUB, W., Voss, Th. (1981) *Individuelles Handeln und gesellschaftliche Folgen: Das individualistische Programm in den Sozialwissenschaften*. Darmstadt: Luchterhand.

SAAM, N., Harrer, A. (1999) "Simulating norms, social inequality, and functional change in Artificial Societies". *Journal of Artificial Societies and Social Simulation* 2 (1) 2 http://jasss.soc.surrey.ac.uk/2/1/2.html.

SAVARIMUTHU, B., Purvis, M., Cranefield, S., Purvis, M. (2007) "How do norms emerge in Multi-Agent Societies? Mechanism Design". *The Information Science Discussion Paper Series*.

SHOHAM, Y., Tenneholtz, M. (1992) "On the synthesis of useful social laws for artificial agent societies (preliminary report)". *Proceedings of the Tenth AAAI Conference*.

SKYRMS, B. (1996) *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

SKYRMS, B. (2004) *The Stage Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.

STALLER, A., Petta, P. (2001) "Introducing Emotions into the Computational Study of Social Norms: a First Evaluation". *Journal of Artificial Societies and Social Simulation* 4 (1) 2 http://jasss.soc.surrey.ac.uk/4/1/2.html.

VERHAGEN, H. (2001) "Simulation of the Learning of Norms". *Social Science Computer Review* 19 (3) 296 - 306.

WALLIS, K., Poulton, L. (2001) *Internalization: The Origins and Construction of Internal Reality*. Buckingham: Open University Press.

WARNER, S. (1978) "Towards a Redefinition of Action Theory: Paying the Cognitive Elements its Due". *American Journal of Sociology* 83 (6) 1317 - 1349.

WRONG, D. (1961) "The oversocialised conception of man". *American Sociological Review* 26 92) 183 - 193.

Return to Contents of this issue