Cognitive architectures of agent systems and social mechanisms of emergence and immergence

Martin Neumann[1]

Abstract. This paper develops a framework of a theory for the emergence of social reality that turns out by relating three different models or agent architectures, respectively: a model of the emergence of social hierarchies, an architecture of a normative system and an architecture of the delegation of social control. This reflects a complex feedback loop of the emergence of an autonomous sphere of social positions and immergence of social norms and execution of social control. It is shown how social structure emerges from and recursively affects the agents' cognitive structure.

1 INTRODUCTION

The notion of emergence is well known in agent-based simulation as well as in Social Theory. The aim of this paper is to demonstrate that agent-based methodology can provide a tool for the formulation of an emergentist Social Theory. An emergentist theory of society is in need to express social micro and macro phenomena in terms of one another. It will be demonstrated that the cognitive architecture of software agents can provide a framework to investigate this question.

Within Social Theory emergentist theories promise to bridge the gap between the micro-macro distinction, or dichotomy of action versus structure [1 – 5]. Traditionally, so-called holistic and individualistic theories are opposing approaches [6, 7]. The former claim the existence of a social reality with laws and theories in their own right, not reducible to theories of lower level domains. This has been suspected as the error of *reification* [8]. The latter emphasises that this social reality does not exist. Every social phenomenon should be explained in terms of individual action. This has been suspected as the error of *voluntarism* [8]. The problem in analytically distinguishing micro- and macro-phenomena is that they do not appear separated but rather are inherent in one another [9, 10]. Theories of social emergence postulate the autonomy of social reality without denying that this reality is constituted by individual actors [5, 7, 11]. However, the ontological status of emergent social reality is often left unexplained [12].

Agent-Based simulation techniques promise to shed new light on this old problem by generating macro phenomena in the course of individual interaction. In fact, it is claimed that agent-based simulation provides a tool for studying emergent processes in society [13]. This promises to allow for an understanding how individual actors produce and are in the same time a product of social reality [14]. Thus, the methodology of agent-based modelling proposes an integrated view on the theoretical problem: While it is an individualistic assertion that actors produce social reality, it is a holistic position that actors are products of social reality. Hence, agent-based modelling techniques suggest to built-up a framework to understand micro- and macro-phenomena in terms of one another. This paper argues that agent-based simulation provides a tool to investigate the constitution of social reality not only by its generative capacity but also by the design of the agents' cognitive capacities: since agent-based modelling allows for an explicit modelling of the agents' cognitive capacities, it allows to study how the emergent level is constituted by the agents' cognitive design.[3]

The paper concentrates on *ontological* questions related to the constitution of social reality rather then on the epistemological question if and how emergence is possible (comp. [15]). Hints to the application of concepts of emergence in complexity research can be found in [16 – 19]. The relation to evolutionary processes is stressed in [20, 21]. The relation to psychology and philosophy of mind is investigated in [22 – 25].

The paper proceeds as follows: the first section deals shortly with the conceptual framework of emergence. The following sections relate models (or architectures) to processes in human societies. Since Society is not a physical object it will be asked how it can be realised by the agents' cognitive design. Two ontological dimensions are differentiated: First, the process of emergence of social positions is investigated. This is an evolutionary process of differentiation of social reality and individual actors. A sloppy phrase would be to denote this as externalisation of society.[4] Secondly, the reverse process of immergence of society by social norms is analysed. This is the causal effectiveness of social reality in the minds of individual actors. Finally, it is shown how both processes are recursively related by social control.

2 CONCEPTS OF EMERGENCE AND IMMERGENCE

According to Bedau [27], the basic assumption of emergentist theories is that reality is constituted out of a hierarchy of levels of reality, for which hold that:

[1] Institute of Philosophy, Bayreuth University, Universitätsstr. 30, 95447 Bayreuth, Germany. Email: martin.neumann@uni-bayreuth.de.

[3] Obviously, the agent architecture is not a realistic representation of human cognitive capacities. Insofar it provides only an investigation the *possibility* of the emergence of a new level of reality by up- and downward causation through cognition. However, evidence exist that human consciousness is processed by downward causal chains [22].

[4] The term is borrowed from activity theory [comp. 26]. In activity theory, however, the deployment of this term is not identical to the way it is used here.

a) emergent phenomena are somehow *constituted* by and generated from underlying processes and

b) emergent phenomena are somehow *autonomous* from underlying processes.

Bedau defines a phenomenon as emergent when a macrostate P of a system S with microdynamic D can be derived from D and S's external conditions but *only* by simulation [27]. This is a comparably weak definition of emergence [comp 15]. However, it is a tellingly characterisation of the processes at work in agent-based simulation. One of the features of Multi-Agent Systems is that they enable to built up patterns on the macro-level by local interaction of individual agents. The famous Schelling model [comp. 28] is one of the most prominent examples: Initially randomly distributed groups of agents produce patterns of segregation. The patterns of segregation are a newly generated emergent property of the social macro-level.

However, only recently attention has been paid to a related, but somewhat different problem: the way back from emergent macro-social properties generating effects on the micro-level. This complementary process has been denoted as *immergence* [29, 30], in Philosophy of Science also known as Downward Causation [31]. In a recent paper, Conte et al. [32] distinguish two main ways in which Downward Causation occurs in human and Multi-Agent Societies:

- a *simple loop*, in which the emergent effects produce new properties on the generating micro-level. Examples are dependence networks. Here the emergent macrostructure creates a distribution of negotiation power among individual agents at the generating micro-level [33]. This is a structural property of the network, independent of the individual consciousness of this structure.
- a *complex loop*, in which the emergent effect determines new properties on the micro-level by means of which the effect is reproduced again. Hence, a recursive interaction between both levels is established in such a complex feedback loop [32]. A particular interesting case occurs, when the emergent effect is recognised by the producing system. This has been denoted as 2nd order emergence [28, 34]. Stressing the crucial role of language, Goldspink and Kay [35] introduce the notion of reflexive emergence; an effect that as a matter of fact exist in Human Societies. For instance, people recognise norms and act (sometimes) accordingly.

This paper will investigate the cognitive capacities needed to generate reflexive emergence. In terms of Social Theory this would contribute to the above mentioned micro-macro problem. It would be a fundamental contribution to Social Theory to represent 2nd order emergence in Multi-Agent Systems.[5] In particular, representing agents' conscious awareness of an emergent social reality would be a building block to develop a theory of Sociality, which would be able to explain recursively micro and macro phenomena in terms of one another. Such an investigation demands an examination of the agents' cognitive capacities. However, concurrently a framework will be outlined of what social reality actually consist.

[5] For reasons of terminological simplicity, in the following the paper will simply refer to the term immergence to denote the process of downward causation.

3 EMERGENT SOCIETY

Translating the philosophical concept of emergence into material conditions of human societies needs to identify the emergent object. In case of human societies it is not so obvious what are emergent levels of social reality, since different levels are interweaved in individual action. Actors and society exist parallel. Even though this need not be an exhaustive enumeration, it will be proposed that it is possible to distinguish at least two forms in which emergent and immergent social reality can be identified. Emergence is specified as a process of differentiation of social positions and individual actors. Immergence is specified as the causal power of social reality on individual behaviour by norms.[6] Thereby society finds its way back into the minds of the individuals. First the process of emergence is investigated.

Following Peter Blau [36], the emergence of social structure will be conceptualised as the distribution of a population among social *positions*. This is because social structure "nearly always includes the concept that there are differences in social positions, and that there are social relations among these positions" [36, p. 27]. Undoubtedly, social positions influence people's social relations, but they have to be distinguished from mere interaction. For instance, at different times the same position can be inhabited by different people. Therefore positions gain an autonomous reality. Hence, by the establishment of social positions a differentiation between social reality and individual actors take place.

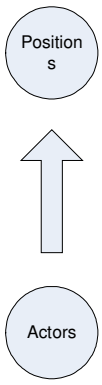

Figure 1. emergence of social positions

In fact, evidence exists that the emergence of social positions has been a concrete historical process which took place in space and time. Archaeological findings indicate that the process of differentiation between individual actors and social positions is the process of the emergence of social stratification, presumably located in the Palaeolithic period [38 – 42]. No archaeological indicators for social stratification can be found in earlier societies. In course of cultural evolution, however, "egalitarian principles of burials were violated when extraordinary items of gold started to be placed with certain individuals, presumably to

[6] There might exist more instances of social reality than social norms and positions. For instance, some sociological theories [e.g. 37] stress the crucial role of communication and thus language. Presumably language is a precondition for the generation of cognitive capacities that enable the emergence of social norms and positions. It is not denied that there might exist a hierarchy of several levels of social reality.

mark their social difference." [42, p. 112]. It is highly plausible that the process of social stratification goes hand in hand with the emergence of social positions. This is a process of the emergence of an autonomous element of social reality. The emergence of cultural symbols denote the emergence of a new form of reality, consisting of role differentiation and social stratification. However, since positions are no physical object, the emergence of positions involves some kind of transformation the cognitive structure of individual actors. In the following a model is analysed with regard to the question what cognitive capacities are needed for this innovation in the organisation of human relations.

agent models of social emergence

There exist archaeological models of the emergence of stratification. One – quite old, but in this respect still outstanding model will be considered briefly: the EOS model of the emergence of organised society.[7]

The target of the EOS model is to develop an agent-based model of a theory [38] of the growth of social complexity in the Upper Palaeolithic period of South–West Europe, that is 15 000 to 30 000 years ago [45]. In contrast to egalitarian societies, complexity is defined as containing centralised decision-making, ranking, role differentiation, and territoriality [46]. Hence, among other features, the evolution of social stratification and role differentiation is denoted by the notion of social complexity. The main features of the model are [45]:

a) a two-dimensional simulated environment providing clusters of resources that can be gathered by the agents. The resources have a specific regeneration cycle and complexity, which is defined as the number of agents necessary to acquire them.

b) a population of 32 to 50 agents. The agents are able to collect sensory data and move around in the environment. In particular, they form plans for resource acquisition and communicate about these plans. Agents need to gain resources.

To examine the question of how this model is capable to represent emergent sociality, it is of particular interest to investigate the agents' cognitive capacities: The working memory of the agents contains a resource model, where the agents keep their beliefs about the resources, and a *social model*, where an agent stores its beliefs about itself and other agents.

In course of the simulation the agents start without any knowledge of groups or other agents. They collect information about their environment and, if they are able to collect resources individually, they do so. If there are resources that need co-ordinated activity, then they develop plans for collective resource gathering and attempt to recruit others for the execution of the plan. Therefore they send out information about the resource and the others evaluate this information to decide whether or not to follow. Agents that are able to recruit others become group leaders. The agents, whose plans are selected, gain 'prestige'. This leads to a "semi-permanent leader-follower relationship" [45, p. 106]. This group structure becomes part of the social model of the involved agents. This process may be iterated, thus leading to a situation where a group leader together with its group members becomes a participant of another group. Hence, by iterations a social hierarchy is formed.

These hierarchies have the ability to persist, but may also break down after a while. This is affected by how easily agents decide to operate independent of their leader and how long they believe to be part of a group when they are not in contact with it [45, p. 214]. Hence, the EOS model allows the study of mechanisms of the emergence of social stratification out of egalitarian groups of agents.

Agent cognitive capacities

It thus has been concluded that the hierarchy is an "implicit property of the agents' social model" [47, p. 154]. However, regarded from the perspective of duality of structure and action, the way back from emergent structure (namely: hierarchies) into the agents' social model is a crucial property of the agents' cognitive structure. Namely, in this model social reality is expressed in terms of individual agents. The restriction of this model is not primarily that the hierarchy is a property of the agents' social model but that the agents' social model is restricted to knowledge about *individual actors*. It does not include the notion of social positions, where actors in such positions are responsible to form plans. In terms of George Herbert Mead [48], they lack the notion of the generalised other. This restriction of the agents' cognitive capacities leads to the result that the model is incapable of generating a social sphere of its own. It shall not be question whether the model is a correct representation of the social reality at around 20 000 BC.. However, the differentiation of social positions and individual actors is a qualitative switch in human prehistory that agent-based modelling technology has not represented so far. The cognitive capacities of the agents do not include a kind of *abstraction* process from individual leaders to the abstract notion of a leader and thus social positions. Eventually, the current New-Ties Project might provide new inside into such processes.[8]

4 IMMERGENT SOCIETY

The process of differentiation between actors and positions is a process of the emergence of a new level of social reality. It is represented both in material symbols of social status and in the mind of the actors, able to identify the symbolic meaning of the material items. Classical role theory [49 – 51] emphasises that positions are characterised by social *norms*.[9] Hence the emergence of positions includes the dual process social *immergence*. For instance, in the EOS model social hierarchies are a representation of the agents' social model. This is a complex feedback loop, in which the emergent effect is recognised and reproduced by the producing actors. This is a social norm: Norms are means to regulate individual behaviour in a way prescribed by society (for instance, the norm to respect

7 This model has been selected since more recent models such as those concerned with the decline of the Anasazi culture [43, 44] do not capture the process of social differentiation.

8 `http://www.cs.vu.nl/~gusz/newties/newties.html`

9 The immergent effect of norms is essential to establish social positions. However, norms may have existed before the emergence of social positions took place. Presumably, the emergence of norms went along with the emergence of language (compare also footnote 6). However, positions enable the establishment of a new kind of norms (such as money) for large and anonymous societies. Hume [52] denoted this as artificial virtues.

the authority of a group leader). Social norms are essential features in the coordination of populations of actors. Hence, by the means of social norms, social reality is causally effective in the minds of the individual actors [29]. In the following, a closer examination of agent-based models of norms is undertaken.

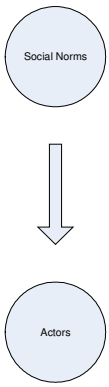

Figure 2. immergence of social norms

Agent models of social immergence

The current development of normative agent systems mostly follows the Belief-Desire-Intention Architecture [53], by extending this approach to a Belief-Obligations-Intention-Desire (BOID) Architecture [54]. Examples for such an approach are [55 – 60]. Obligations are introduced to constrain individual intentions and desires on the one hand, while preserving individual autonomy on the other [54]. Agents are able to violate normative obligations. This implies that agents dispose of the capacity of normative reasoning. A sophisticated agent design following a similar logic, however without an explicit notion of obligations can be found in [61].

In this paper, a closely related, but nevertheless somewhat different account is examined in more detail: Boella/van der Torre's 'An architecture of a normative system' [62]. This architecture is selected, since it relies on John Searle's theory of constructing social reality [63]. It is thus an approach to explicitly represent social reality in the agents' cognitive capacities. The primary technical terminus in Searle's theory are the so-called *counts-as conditionals*. Searle's theory of social reality distinguishes between brute and institutional facts. Institutional facts are build upon social norms. Two types of norms are distinguished: some norms regulate pre-existing forms of behaviour, while other norms create the possibility of that activity. For instance, playing chess is constituted by the rules of the game. Following Searle, the general form of a normative ontology is 'X counts-as Y in context C'. Boella/van der Torre's architecture intends to represent this theoretical approach in the agent's design.

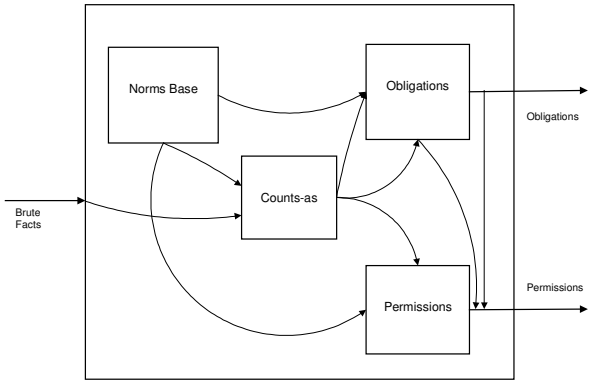

Figure 3. simplified Architecture of a Normative System

In the architecture the inputs of the normative system are brute facts. The outputs are obligations and permissions. The normative architecture has four components: counts-as conditionals, obligations, permissions and a norms base. Thus, if a proposition p is flowing over a permission channel it is interpreted as a permission P(p). The *norms base* has no input, it is assumed as fixed. It includes background knowledge of institutional constraints. The inputs of the *counts-as component* are brute facts, counts-as conditionals and institutional constraints. Counts-as conditionals and institutional constraints come from the norms base. Brute facts are external inputs. The outputs are brute and institutional facts. They are sent to the obligation and permission components. The input of the *obligation component* contains conditional obligations (coming from the norms base), a context (a mixed description of brute and institutional facts coming from the counts-as component) and constraints. Constraints are added to avoid inconsistencies. The outputs are obligations. The design of the *permission component* is similar, but without constraints: the inputs are conditional permissions (coming from the norms base) and brute and institutional facts (coming from the counts-as component). The outputs are permissions. Thus, the whole architecture transforms brute facts into obligations and permissions.

Agent cognitive capacities

In analysing this architecture it is striking that first, the input of the whole architecture is restricted to brute facts and secondly, that the norms base has no input.[10] At first sight this seems to be straightforward. However, this leaves a problem unresolved: the norms base can only be updated off-line. Norms are thus *not* emergent features of the system; there exist no feedback loop between the environment and the norms-base.

It can be presumed that this is due to the fact that sensory data consist of brute facts. Based on ethnological evidence, Emil Durkheim [64] proposed an alternative view in his 'elementary form of religious life'. First, he claimed that religion has been the very first representation of a norm setting authority in the

[10] This is not specific to this example. In fact, also for the BDI inspired architectures it holds that belief updating is realised by sensory data. Moreover, in attempts to resolve conflicts between the different components it is a commonplace that beliefs override the other components. This is the architecture of so-called realistic agents [54]. In fact, the update of the obligation component is an open problem for the design of normative agents.

evolution of human culture. Religion entails belief formation affecting the norms base. For instance, if you convert to a muslim you are prohibited to drink alcohol. This is uncontroversial. However, Durkheim also claimed that religion always entails a cosmology. Religious belief is not only about supernatural entities but about the structure of the world. Hence, brute and institutional facts cannot be separated. An obvious objection is that here two meanings of belief are confused. It is a difference to belief that it is raining just now or to belief in the existence of a supernatural god. However, this is exactly what is contested by Durkheim's analysis: "Religious representations are collective representations which express collective realities" [64, p. 22].

Moreover, it is worth noting that they are *collective* representations: by analysing totemism Durkheim concluded that primitive religion had been the very first representation of society by its members. Totemism established the identity of the clan by the shared name of the totem and thus a relation between the individual and the society. The symbolic unity, generated by this classification scheme, was the source of this belief system to exert moral authority in much the same way as the belief that it rains might trigger the intention to open an umbrella.

Obviously, this is difficult to implement computationally. However, as the notion of a 'shared name of the totem' indicates, the problem is closely related to the emergence of *language* [comp. 65]. The advent of language opens up the possibility of a symbolic representation of the world in a consensual linguistic domain [35]. Language is thus a precondition for religious collective representations. However, *this* is a computationally traceable problem. For instance, the emergence of a commonly shared lexicon is an implemented feature of agent models [66, 67]. Eventually, the current New-Ties Project might provide new insights "which system components carry the knowledge structures that make up world models".[11] Anyway, to represent the *human* social dynamic in agent systems it is necessary to close the feedback loop between generating and emergent phenomena. Since the only environmental input in the agent's design is directed to the belief component (here the counts-as conditionals), this problem is in some way related to the relation between the update of beliefs and obligations. In the following we will see that beliefs can exert social control (similar to obligations).

5 SOCIAL CONTROL

While the former section was concerned with the question of how social norms work *in* the minds of individual actors, this section is concerned with the question of how they get *into* the minds. This is closely related to the problem of social control. In particular the invention of 'Artificial Virtues' [52] in large and anonymous societies demands for social control.

Sociologically, the problem of social control refers to the problem discussed above: the emergence of social positions. It has already been remarked that, presumably, this process was driven by population concentration. Hierarchical organisation of society allows for larger populations than prehistoric bands. Evolutionary psychology estimates that ancestral hominids lived in groups of 20 to 100 persons [68]. In small groups social control can be exerted in direct peer-to-peer interaction. In larger societies, however, this becomes precarious because actors can preserve anonymity.[12] A possibility to assert social order are new organisational features such as hierarchies and role differentiation: namely, by vesting hierarchical positions with norm setting authority. As a matter of fact, this process occurred in course of cultural evolution. In large societies, social control is executed by specialised institutions, providing specific social positions. [13]

Agent models of social control

In the following, an architecture of such a normative control system is investigated, described in the paper 'Norm governed multiagent systems: the delegation of control to autonomous agents' [69]. It has been stressed by several authors that normative obligations involve both – a norm addressee and someone wanting the norm to be fulfilled [70 – 73]. However, this architecture is selected for closer inspection, since it explicitly introduces agents with specialised social *roles*.

In the paper it is distinguished between 1) agents whose behaviour is governed by norms, 2) so-called defender agents that monitor violations and 3) a normative system that issues norms and monitors the defender agents. This reflects that in modern states the government is separated into a legislative system, responsible for the norm setting and a judicial system, responsible for the control of norm compliance. Thus, there exist three classes of autonomous agents: (class of) agent 1 is subject of obligations, (class of) agent 2 is responsible for norm control and sanctioning of violations. Agent 2 is the defender agent. (class of) agent 3 is the central authority that imposes obligations and permissions and monitors the defender agents. All agents make their decisions autonomously, i.e. based on their interests and states of belief.

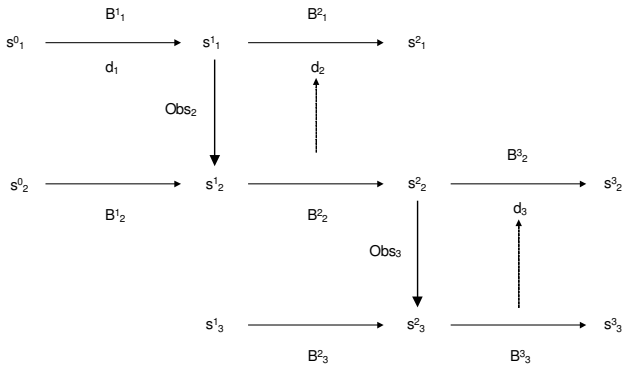

Figure 4. (Three agent) Control System

The sequence of actions in a three agent scenario is the following (from the perspective of agent 1):

Agent 1 makes its decision d_1 at time 0. It beliefs to be in a state s^0_1 (subscripts denote the agents, superscripts denote time).

[11] **http://www.cs.vu.nl/~gusz/newties/newties.html**

[12] This refers to the well known problem of cooperation among strangers. It shall not be questioned that this problem also can be resolved by decentralised interaction [74].

[13] The problem that the generation of a certain surplus is a presupposition to provide specialised positions will not be investigated in this paper [comp. 75].

The expected consequences of the decision at time 1 are calculated by a belief rule B^1_1. The expected consequences are denoted as the epistemic state s^1_1 of the agent 1. In making its decision, agent 1 tries to take the decision of agent 2 into account. Therefore it has a representation of what it beliefs to be agent 2's initial state s^0_2. Agent 1 beliefs that its decision has the consequence that agent 2 then beliefs to be in the state s^1_2, since agent 2 observes agent 1. Then agent 2 makes its decision d_2. The decision is based on whether agent 2 counts the action of agent 1 as a norm violation or not. Thus, (next to goals and desires) agent 1 builds its decision on expectations about the belief system of agent 2.

However, the same holds for agent 2: At time 1 (when agent 2 observes agent 1), agent 2 beliefs agent 3 to be in the epistemic state s^1_3. Moreover, it beliefs that the epistemic state of agent 3 changes to s^2_3 as a consequence of its decision d_2. This in turn will cause a decision of agent 3. Agent 2 beliefs that this will lead to the epistemic state s^3_2 for itself and to s^3_3 for agent 3. Thus, (next to goals and desires) agent 2 builds its decision on expectations about the belief system of agent 3.[14]

The decision process is characterised as follows: the agents are assumed to be of a selfish stable agent type. That is, it is not implemented that agents automatically obey norms, but calculate an optimal decision. An optimal decision maximises expected utility.

Obligations, i.e. norms, are characterised as follows: Agent i beliefs that it is obliged to do x with sanction s under condition q if it beliefs that agent i+1 (the defender agent) desires and has the goal x. Moreover, agent i beliefs that agent i+1 has the desire not to sanction and agent i itself has the desire not to be sanctioned. However, agent i beliefs that agent i+1 has the desire that there is no norm violation and has the goal and desire to recognise a norm violation. Finally, agent i beliefs that agent i+1 desires to sanction a norm violation if it recognises it.

Note, that defender agents do not intrinsically desire to sanction. Defender agents desire to sanction a norm violation because they are monitored by the norm setting authority. The norm setting authority is represented by agent i+2. This means (from the perspective of agent i) that agent i beliefs that it is obliged to do x, if it beliefs that agent i+2 desires and has the goal that x and that there is no norm violation. Moreover, agent i beliefs that (agent i+2 beliefs that) agent i+1 is conditionally obliged by agent i+2 to sanction a violation by agent i. The obligation for the defender agent to sanction norm violation is again represented in the same way: namely by the possibility that the norm setting authority, agent i+2, sanctions violations of the obligation to sanction norm violation. Hence, defender agents sanction because of fear of sanctions.[15]

[14] In principle, this can be iterated: for example, the central authority (agent 3) can delegate the control of the defender agent (agent 2) to another defender agent. This leads to a hierarchy of agents in which each agents considers the reaction of the agent in the subsequent hierarchy level when choosing an action. Conversely each agent observes the agents on the next lower hierarchy level.

[15] In principle, this is the same structure than in Robert Axelrod's so-called meta-norms game [76]. In this game agents can sanction defecting agents and agents not sanctioning defections. The difference is, that role differentiation is introduced here. Sanctioning norm violations and monitoring sanctions is ascribed to specific types of agents.

Orders of emergence

In this architecture norm enforcement is ascribed to specific types of agents. It thus explicitly represents social *role differentiation*. This implies the existence of social positions assigned to specific tasks. It is worth noting, that first, defender agents are obliged to sanction because of their professional duties and secondly, that this role differentiation is a feature of the agents' *belief* structure. Social positions are represented as a feature of the agents' cognitive capacities.

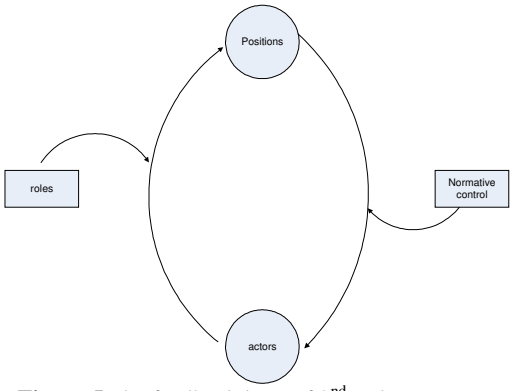

Figure 5. the feedback loop of 2nd order emergence

The stabilisation of the process of norm *immergence* by means of social control (in large societies) can be enforced by the *emergence* of social positions, enabling the delegation of control. The notion of positions refers to the emergence of hierarchical organisation of society as analysed in the EOS model [45]. Hence, both processes of emergence and immergence refer to each other. The delegation of social control to specific agent roles closes the feedback loop between emergent and immergent processes. It enables the possibility of the invention of 'Artificial Virtues'. However, currently the feedback loop is not closed: the existence of positions is a pre-given feature of this architecture. Only an integrated perspective on all three cases enables a representation of a complex feedback loop of social emergence.

6 CONCLUSION & FUTURE WORK

The task of this paper was to argue that the cognitive structure of software agents opens-up a new perspective to an explanation of social emergence. Agent-based models and architectures have been investigated with regard to the question of how individual agents are able to produce and to be in the same time a product of social reality. A three-stage scenario of the evolution of social reality has been developed. Particular attention has been paid to the question how society gets effective by the agents' cognitive structure. This enables to develop a theoretical framework that is capable to represent simultaneously the social micro- and macro-level.

The process of differentiation of social positions and individual actors has been identified as the *emergence* of social reality. The EOS model [45] demonstrates how social structure can be represented in the agents' social model. However, it still lacks a process of abstraction of social positions from individual

agents. The representation of counts-as conditionals in the architecture developed by Boella/van der Torre [62] opens-up a perspective of how emergent social reality can be *causally effective* in the individual agents' minds. By the generation of permissions and obligations, society is reproduced again. Also here, society is a component of the agents' cognitive structure. However, the shortcoming of this architecture is that the norms base can only be updated off-line. If agents would be able to develop conjointly a symbolic representation of the world in a consensual linguistic domain, it could be possible that a norms base could be developed in this process. The processes of emergence of social positions and immergence of social norms are related by norm enforcement exerted by actors ascribed to *specific roles* (that is, social positions), *transferring* social norms into individuals. The principles of role differentiation are described in Boella/van der Torre's paper [69]. Note, that these roles are features of the agents' belief structure. In this model, however, positions (of defender agents and normative authority) are pre-given.

So far the models stand in isolation. It is an open problem to link the models recursively. The task would be to integrate a model of the emergence of positions, responsible for the execution of social control, with a model of norm internalisation: if agents would be able to recognise the emergent norm setting authority, a complex feedback-loop of 2nd order emergence would be established. It is thus left for future work to close the feedback-loop and to develop a more comprehensive model of emergence in the loop.

ACKNOWLEDGEMENT

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".

REFERENCES

[1] K. Knorr-Cetina, and A. V. Cicourel (Eds.), *Advances in social theory and methodology*. London: Routledge (1981).

[2] M. Archer, *Realist social theory: the morphogenetic approach.* Cambridge: Cambridge University Press (1995).

[3] R. Mayntz, Individuelles Handeln und gesellschaftliche Ereignisse: Zur Mikro-Makro-Problematik in den Sozialwissenschaften. MPIfG Working Paper, 99/5 (1999).

[4] K. Sawyer, Artificial Societies: Multiagent Systems and the Micro-Macro Link in Sociological Theory. *Sociological Methods and Research,* 31/3 (2003).

[5] B. Heintz, Emergenz und Reduktion. Neue Perspektiven auf das Mikro-Makro-Problem. *Kölner Zeitschrift für Soziologie und Sozialpsychologie,* 56/1 (2004).

[6] M. Brodbeck (Ed.), *Readings in the Philosophy of the Social Sciences.* New York: MacMillan (1971).

[7] J. O'Neill (Ed.), *Modes of Individualism and Collectivism.* London: Heinemann (1973).

[8] R. Bhaskar, *The Possibility of naturalism: A philosophical critique of the contemporary human sciences.* (Second Edition) Hemel Hempstead: Harvester (1989).

[9] R. Collins, On the Microfoundations of Macrosociology. *American Journal of Sociology,* 86 (1981).

[10] A. Giddens, *The constitution of Society: Outline of the Theory of Structuration.* Cambridge: Polity Press (1984).

[11] K. Sawyer, *Social Emergence: Societies as Complex Systems.* Cambridge: Cambridge University Press (2005).

[12] M. Neumann, Emergence as an Explanatory Principle in Artificial Societies, in: *Proceedings of the Conference on Epistemological Perspectives on Simulation II,* K. G. Troitzsch, F. Squazzoni (Eds.). Sythese Library, Berlin: Springer (forthcoming).

[13] A. Drogul, J. Ferber, Multi-Agentsimulation as a tool for studying emergent processes in societies. In: *Simulating Societies. The computer simulation of social phenomena.* N. Gilbert, J. Doran (Eds.). London, UCL Press (1994).

[14] G. Deffuant, S. Moss, W. Jager, Dialogues Concerning a (possibly) new Science. *Journal of Artificial Societies and Social Simulation* 9/1 (2006). `http://jasss.soc.surrey.ac.uk/9/1/1.html`

[15] A. Beckermann, H. Flor, J. Kim, (Eds.) *Emergence or Reduction?* Berlin, New York: De Gruyter (1992).

[16] J. Holland, *Emergence - From Chaos to Order.* Oxford, Oxford University Press (1998).

[17] J. A. Goldstein, Emergence as a construct: History and Issues, *Emergence* 1 (1999).

[18] V. Darley, Emergent Phenomena and Complexity, In: R. Brooks, P. Maes (Eds.). *Artificial Life IV: Proceedings of the Fourth International Workshop on the Sythesis and Simulation of Living Systems.* Cambridge, Mass.: MIT Press. (1994).

[19] K. Richardson, On the Limits of bottom-up computer simulation: Towards a nonlinear modelling culture, in: *Proceedings of the 36th Hawaiian International Conference on System Science,* IEEE, California (2003).

[20] C. Emmeche et al., Explaining Emergence: Towards an Ontology of Levels, *Journal for General Philosophy of Science* 28 (1997).

[21] E. Mayr, *Eine neue Philosophie der Biologie,* München, Piper (1991).

[22] E. Thomson, F. Varela, Radical Embodyment: neural dynamics and consciousness, *Trends in Cognitive Science* 5 (2001).

[23] W. Sullis, Archetypical dynamical systems and semantic frames in vertical and horizontal emergence, *Emergence: Complexity and Organisation,* 6/3 (2004).

[24] M. Pauen, G. Roth (Eds.), *Neurowissenschaften und Philosophie.* Wilhelm Fink Verlag, München (2001).

[25] R. Oerter Entwicklungskrisen im Jugendalter: Eine Systemtheoretische Perspektive, *Psychotherapie Heft 1* (2002).

[26] Y. Engeström, R. Mietinen, R.L. Punamäki (Eds.) *Perspectives on activity theory.* Cambridge: Cambridge University Press (1999).

[27] M. Bedau, Weak Emergence. In: *Philosophical Perspectives: Mind, Causation, and World*, Vol. 11. J. Tomberlin (Ed.). Malden MA: Blackwell (1997).

[28] N. Gilbert, Varieties of emergence. Paper presented at the social Agents: Ecology, Exchange and Evolution Conference on Social Agents: ecology, exchange and evolution. Chicago (2002).

[29] C. Castelfranchi, Through the Minds of the Agents. *Journal of Artificial Societies and Social Simulation,* 1/1 (1998). `http://www.soc.surrey.ac.uk/JASSS/1/1/5.html`

[30] G. Andrighetto, M. Campennì, R. Conte, M. Paolucci, On the Immergence of Norms: a Normative Agent Architecture. In: *Proceedings of AAAI Symposium, Social and Organizational Aspects of Intelligence,* Washington DC (2007).

[31] D. Campbell, Downward Causation in Hierarchially Organised Biological Systems. In: *Studies in the Philosophy of Biology,* F. Ayala, T. Dobzhansky, (Eds.). London: MacMillan (1974).

[32] R. Conte, G. Andrighetto, M. Campennì, M. Paolucci, Emergent and Immergent Effects in Complex Social Systems. In: *Proceedings of the AAAI ,07,* (2007).

[33] J.S. Sichman, R. Conte, C. Castelfranchi, Y. Demazeau, A Social Reasoning Mechanism Based on Dependence Networks. In: *Proceedings of the 11th European Conference on Artificial Intelligence*, A.G. Cohn (Ed.) Baffin Lane, England: John Wiley and Sons (1994).

[34] D. Dennet, *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon and Schuster; London: Allen Lane (1995).

[35] C. Goldspink, R. Kay, Emergence in Social Systems: Distinguishing Reflexive and Non-reflexive modes. In: *AAAI Fall Symposium: Emergent Agents and Socialities: Social and Organizational Aspects of Intelligence*. Washington (2007).
[36] P. Blau, A Macrosociological Theory of Social Structure. *American Journal of Sociology,* 83/1 (1977).
[37] N. Luhmann, *Soziale Systeme: Grundriss einer allgemeinen Theorie,* Frankfurt a.M.: Surkamp (1984).
[38] P. Mellars, The ecological basis of social complexity in the Upper Palaeolithic of southwestern France. In: *Prehistoric hunter-gatherers: the emergence of cultural complexity.* T. Price, J. Brown (Eds.). New York: Academic Press, (1985).
[39] J. Aigner 1989, Frühe Siedlungen im arktischen Nordamerika. In: *Siedlungen der Steinzeit,* Heidelberg: Spektrum, (1989).
[40] M. Kolb, Monumental Grandeur and the rise and fall of Religious Authority in Precontact Hawaii. *Current Anthropology,* 34 (1994).
[41] C. Renfrew, Symbol before Concept: Material Engagement and the Early Development of Society. In: *Arachaeological Theory Today,* I. Hodder (Ed.). Cambridge, Polity Press, (2001).
[42] T. Earle, Culture Matters in the Neolithic Transition and Emergence of Hierarchy in Thy, Denmark: Distinguished Lecture. *American Anthropologist,* 106, (2004).
[43] C. Kresl, E. Van West, Carr, R. Wilshusen, Be There Then: A Modeling Approach to Settlement Determinant and Spatial Efficiency among late Ancestral Pueblo Populations of the Mesa Verde Region, U.S. Southwest. In: *Dynamics in Human and Primate Societies: Agent-Based Modeling of Social and Spatial Processes.* G. J. Gumerman, T. A. Kohler (Eds.). Oxford: Santa Fe Institute and Oxford University Press (2000).
[44] A. Gumerman, J. Swedlund, S. Harburger, R. Chakravarty, J. Hammond, J. Parker, M. Parker, Population Growth and Collapse in a Multi-Agent Model of the Kayenta Anasazi in Long House Valley. In: *Proceedings of the National Academy of Sciences of the United States of America* 99, 3 (2002).
[45] J. Doran, M. Palmer, N. Gilbert, P. Mellars, The EOS Project: modelling Upper Palaeolithic social change. In: *Simulating Societies.* N. Gilbert, J. Doran (Eds.). London, UCL Press, (1994).
[46] M. Cohen, Prehistoric Hunter-Gatherers: The Meaning of Social Complexity. In: *Prehistoric hunter-gatherers: the emergence of cultural complexity,* T. Price, J. Brown (Ed.). New York: Academic Press, (1985).
[47] N. Gilbert, Emergence in social simulation. In: *Artificial Societies: The computer simulation of social life,* N. Gilbert, R. Conte (Eds.). London: UCL Press, (1995).
[48] G.H. Mead, *Mind, Self, and Society.* Edited by C.W. Morris. Chicago: University Press (1934).
[49] T. Parsons, *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, (1968 [1937]).
[50] T. Parsons, E.A. Shils, *Towards a General Theory of Action.* Harvard: Harvard University Press, (1951).
[51] R. Darendorf, *Homo Sociologicus. Ein Versuch zu Geschichte, Bedeutung und Kritik der Kategorie der sozialen Rolle.* Opladen: Westdeutscher Verlag (1956).
[52] D. Hume, *A Treatise in Human Nature,* Vol 2, Edinburg Edition, (1826).
[53] A. Rao, M. Georgeff, Modelling rational agents within a BDI architecture. In: *Proceedings of the KR 91,* (1991).
[54] J. Broersen, M. Dastani, J. Hulstijn, Z. Huang, L. van der Torre, The BOID Architecture: Conflicts between Beliefs, Obligations, Intentions, and Desires. In: *Proceedings of the 5th International Conference on autonomous agents,* (2001).
[55] J. Broersen, M. Dastani, L. van der Torre, Beliefs, Obligations, Intentions, and Desires as Components in an Agent Architecture. *International Journal of Intelligent Systems,* 20 (2005).
[56] G. Boella, Deliberate normative Agents. Basic Instructions. in: *Social Order in Multiagent Systems.* R. Conte, C. Dellarocas (Eds.). Norwell: Kluwer (2001).
[57] A. Garcia-Cumino, A. Rodriguez-Aguilar, C. Sierra, W. Vasconcelas, Norm-oriented programming of electronic institutions. *AAMAS '06,* (2006).
[58] F. Dignum, D. Kinny, L. Sonenberg, From desires, obligations and norms to goals. *Cognitive Science Quarterly,* Vol. 2 (2002).
[59] F. Lopez y Lopez, M. Luck, M. d'Inverno, Constraining autonomy through norms. *AAMAS '02,* (2002).
[60] J. Vazquez-Salceda, H. Aldewereld, F. Dignum, Norms in Multi-Agent Systems: From theory to practice. *International Journal of Computer Systems and Engineering,* 20 (2005).
[61] C. Castelfranchi, F. Dignum, C. Jonker, J. Treur, Deliberative normative agents: Principles and Architecture. In: *Intelligent Agents: Theories, Architectures, Languages.* Springer, Berlin (2000).
[62] G. Boella, L. van der Torre, An Architecture of a Normative System. *AAMAS '06*, ACM Press (2003).
[63] J. Searle, *The construction of social reality.* London: Penguin Press, (1995).
[64] E. Durkheim, *The Elementary Forms of Religious Life.* New York: Free Press (1965[1912]).
[65] H. Whitehouse, Modes of religiosity: towards a cognitive explanation of the sociopolitical dynamics of religion. *Method and Theory in the study of religion,* 14 (2002).
[66] L. Steels, Constructing and Sharing Perceptual Distinctions. Paper presented at the European Conference on Machine Learning: Berlin, (1997).
[67] T. Gong, J. Ke, J. Minett, W. Wang, A computational Framework to simulate the Co-evolution of language and Social Structure. In: *Artificial Life IX.* Boston, (2004).
[68] J. Tooby, L. Cosmides, Conceptual Foundations of Evolutionary Psychology. In: *Handbook of evolutionary psychology.* D. Buss (Ed.). Hoboken: Wiley, (2005).
[69] G. Boella, L. van der Torre, Norm Governed multiagent systems: the delegation of control to autonomous agents. In: *Proceedings of the IEEE/WIC IAT Conference,* IEEE Press (2003).
[70] R. Conte, C. Castelfranchi, From Conventions to prescriptions: Towards an integrated view of norms. *Artificial Intelligence and Law,* 7 (1999).
[71] G. Boella, L. Lesmo, Deliberate Normative Agents. In: *Social Order in Multiagent Systems.* R. Conte, C. Dellarocas (Eds.) Norwell: Kluwer, (2001).
[72] R. Conte, F. Dignum, From Social Monitoring to Normative Influence. *Journal for Artificial Societies and Social Simulation,* 4/2 (2001). `http://www.soc.surrey.ac.uk/JASSS/4/2/7.html`
[73] F. Lopez y Lopez, A. A. Marquez, An Architecture for Autonomous Normative Agents. In: *Proceedings of the Fifth Mexican International Conference in Computer Science ENC '04,* IEEE (2004)
[74] R. Schüssler, *Kooperation unter Egoisten: 4 Dilemmata.* München: Oldenbourg, (1990).
[75] N. Müller, *Civilization Dynamics, Vol. 1: Fundamentals of a model oriented description.* Aldershot: Avebury, (1989).
[76] R. Axelrod, An evolutionary approach to norms. *American Political Science Review,* 80 (1986).