SCALABILITY AND THE EVOLUTION OF NORMATIVE BEHAVIOR

JÖRG WELLNER[†], SIGMAR PAPENDICK[‡], AND WERNER DILGER[†]
[†] *Chemnitz University of Technology, Computer Science*
D-09107 Chemnitz, Germany
{jwe, wdi}@informatik.tu-chemnitz.de

[‡] *University of Konstanz, Department of Sociology*
D-78464 Konstanz, Germany
Sigmar.Papendick@uni-konstanz.de

We present an evolutionary approach for developing an agent system consisting of a large and varying number of agents. We start off by describing in short a sociological theory to problems of coordination in societies with many members. Elaborated cognitive explanations of handling individual information are rejected and the concept of symbolically generalized communication media is suggested instead. In a first attempt we have modeled an agent system based on this concept. Simulation results show that agents may coordinate their actions even though they have no individual representations of each other. Simulation starts with a small group of agents and evolves a system of several hundred agents which base their actions mainly on exchanged messages.

1 Introduction

Many coordination approaches for agent systems rely on mechanisms which include detailed knowledge of an agent architecture. For one agent it is essentially knowledge of another agent in order to cooperate with it. Agents personalized this knowledge in the way that they know goals, skills, or beliefs[1] for different opponents. To simplify the usual situation one can state that the more potential partners for an interaction an agent has the more agent specific knowledge it has to cope with. This is one reason why current logic based agent approaches scale so badly, because keeping track of information about other agents is an expensive matter. In the next section we consider in more detail a concept, developed by sociologists, to answer questions concerning the coordination of individuals in a society of a huge number of members, namely symbolically generalized communication media. A first approach of modeling the proposed mechanisms for one media is presented in Section 3, focusing on the ability of the agents to acquire a shared symbol system. Simulation results are discussed in Section 4. Section 5 concludes the paper, indicating that one can reasonably base multi-agent systems on the proposed sociological concepts in order to achieve a good scaling.

2 The concept of symbolically generalized communication media

Humans faced a scaling problem during the development from small groups to modern societies. In small groups it is possible for each individual to keep in mind relevant facts about other members of the group. Different strategies were developed to keep one's knowledge about each other up-to-date, e. g. gossip[2]. As groups became larger, personalized coordination mechanisms became less efficient, due to the necessary increase of cognitive capabilities which are – however – constrained.

In these situations, generalized media simplify communication and the representation of situations. The concept of generalized media has been introduced in sociological theory by Talcott Parsons[3] who used the term "generalized media of interchange". In the context of constructivistic sociological systems theory, the concept has been adopted as "symbolically generalized communication media" (SGCM) by Niklas Luhmann[4]. They offer a mechanism to allow coordinated behavior among individuals that have few or no representations about each other's individual goals, beliefs, intentions or restrictions – which used to be regarded as indispensable for behavioral selections in most of the dominating microsociological models. SGCM simplify the predictability of behavior because they offer a universalistic mechanism of generating strong motivations as a prerequisite for further cooperation. They symbolize the expectability of getting rewarded by others in situations of requested cooperation. A typical example of such an symbolic representation is money: Its possession symbolizes the expectability of having the option to instrumentalize cooperational behavior of others, for instance in case of spending money and getting goods or services, regardless of the time or situation this option is needed. If money is transferred, this option is transferred also and has to be represented and evaluated only as an option which can be used by the recipient. It can be coded, communicated and represented by a simple binary distinction of having or not having money. Therefore, it works as an generalizable and reliable communication mechanism of initiating coordinated behavior without ponderous and cognitive complex procedures of making others adopt the own goals in order to cooperate. Thus, the use of symbolically generalized media is an efficient way to reduce social complexity by symbolizing expectability.

Another important example of a SGCM is the symbolization of power, which is the main objective of our model. Like money, power is used as a mechanism to symbolize the instrumentability of coordinated behavior of others. A typical example of symbolizing power are policemen dressed in uniforms. By wearing a police-uniform, the option to apply expectable superior sanctions of criminal prosecution in case of normative deviance can be communicated very efficiently. It is also based on expectations, but – in contrast to money – these are predictions of getting sanctioned if requested cooperation is refused. As money, power as a symbolizing mechanism can only be established and preserved if its function can be proofed on demand. If the inability to apply sanctions being supported by others gets observable, the auto-catalytic mechanism of reinforcing a symbol by referencing networks of cooperation breaks down in dynamics comparable to inflationary processes of currency in economic markets.

Some concepts related to SGCM are already dealt with on a large scale in multi-agent research, especially norms[5,6,7] and market based coordination mechanisms[8,9,10]. All the mechanisms based on these concepts, first of all norms, have the goal of enabling interactions between agents which do not know much about each other, but do know something in general. Norms are condensed expectation structures. A population wide norm makes actions of agents predictable.

It is obvious that coordination mechanisms based on SGCM may play an important role in scaling huge agent systems. The main benefit of such media is that they reduce the amount of knowledge that agents need to interact. Interactions, controlled by a SGCM, are structured in a straightforward way: They do not ensure that an interaction always succeeds, but they ensure that agents know in advance on what aspect negotiation should be limited. Agents need not to know each other. Agents can be black boxes to each other, and indeed they cannot look inside their head. Coordination may still succeed and agents know in what stage an interaction currently is, and when it should be stopped. Every agent is only concerned with its own beliefs or goals, there is no need to take into account elaborated reasoning mechanisms about beliefs or goals of other agents, since they become immediately apparent to each other during an interaction to some extent. Whatever an agent wants or believes will be conveyed by a medium to another agent. A medium does not reveal an agent's goal or its beliefs, but it offers a way to achieve a goal or to verify or to strengthen its own beliefs.

In the remainder of this paper we concentrate on the SGCM power as proposed by Luhmann. We do not model predefined rules of power that allow an agent to interact with another one. We rather focus on the evolution of a shared symbol system and the meaningful use of a sanction mechanism that represents power, both, with respect to efficiently coordinated actions.

3 A first approach to the evolution of the power medium

A simulation consists of a large number of trials of a cooperation game which we called the "Planter-and-Harvester-Game" for simplicity. We introduce two different types of agents, with respect to their ability to change the environment. There are also two types of actions that change the state of the environment in an effective way, namely "planting" and "harvesting", which complement each other. Plant agents, called *Planter* can perform only plant actions effectively, harvest agents, called *Harvester* can perform only harvest actions effectively. At the beginning of a game the environment U is always in state $U_s = 0$. A plant action $Plant_I$ – performed by a *Planter* – transforms the environment into state $U_t = 1$, a harvest action $Harvest_I$ – performed by a *Harvester* – transforms it into the final state $U_e = 2$. In more complicated games the final state may be $U_e > 2$ assuming action sequences $Plant_I$, $Harvest_I$, $Plant_{II}$, and so on. Action $Plant_I$ in state $U = 1$ has no effect with regard to the state of the environment, similarly action $Harvest_I$ in state $U = 0$. Furthermore, a *Planter* might successfully apply in state $U = 2$ only action $Plant_{II}$, not actions $Plant_I$ or $Plant_{III}$.

At the beginning of a game two agents are randomly selected from the population, one of them is the start agent. This agent begins by sending a message M_0. The other agent receives this message and performs an action a_1 and sends another message M_1 to the start agent. Then, the first agent performs an action a_2 and sends a message M_2 to the second agent, and so on. A round is defined by a successive sequence of performing one action and generating a message for each of the two agents. Both types of agents have the same repertoire of actions regardless of the efficiency: apart from plant and harvest actions they have a *Null*-action without any effect, a *Sanction* action, an action *Exit*, and an action *Replace*. The later action affects the opponent agent in the way, that it gets replaced by another agent, randomly selected from the population. This may increase the general possibility for a successful coordination. A game may end by three different outcomes: an agent performed the *Exit* action, the environment reached the final state U_e, or the number of rounds in the games exceeded the defined threshold *rounds*.

There is a predefined set of symbols $\mathcal{S} = \{0, 1, 2, ..., S_{max}\}$ for message generation. A message consists exactly of one of these symbols. A symbol itself has no meaning to an agent, there is no predefined semantics at all.

A game ends successfully if the environment was transformed into the final state U_e. In this case, the last two agents, participating in the game, get a certain amount $E^\star$ of "energy". In other cases there is no energy payoff. Every action that an agent performs consumes a specified amount of energy of the agent. There are low cost actions (*Null*, *Exit*, and *Replace*) and high cost actions ($Plant_x$, $Harvest_x$). For a low cost action the agent consumes energy $E_l > 0$, for a high cost action $E_l + E_h$, $E_h > 0$. The cost of the action *Sanction* is $E_l + E_b$, $E_b > 0$. This action affects the other agent in the way that the sanctioned agent looses pain energy $E_p > 0$. At the beginning of an agent's life time its energy is set to $E = E_s > 0$, its start energy. If E ever falls below 0, the agent dies, that is, the agent is removed from the population.

An agent does not know its own type nor perceives the type of another agent. They are black-boxes to each other. An agent perceives the message of another agent, the state of the environment, and the fact of being sanctioned. In any case not all relevant aspects of the environment are known in the same way to all the participants, for instance the direct result of an action. Agents must test different actions at different times and the only hint to whether an action or message was appropriate is given by a reward signal. This signal is always generated by the agent itself, based on the energy difference between two consecutive actions. A sigmoid function generates the reward signal r based on the energy difference e_d; a positive energy difference results in a positive reward, a negative difference results in a negative reward. Thus, individual agents employ reinforcement learning. This definition of a reward signal is a weak one, since it does not assume any intelligent observer (outside the agent) who generates a reward signal based on its knowledge about correct actions.

Beside an energy value agents have an age A, which at the beginning of an agent's life time is set to 0. Any time an agent gets selected to play the game, its age will be incremented by 1. If the age reaches an individual maximum, A_{max}, the agent will be removed immediately from the population. At the start of the simulation, the population P consists of a certain number of agents P_s. The number of agents during the simulation may shrink or grow, depending on the fitness of the agents. An agent may enter the population if there are at least two agents, whose age is above value A_{sex} and whose energy value is above a value E_{sex}. The two "parents" are selected by a "Roulette wheel"[11] from all possible parent agents based on their energy value. Once a successful breeding occurred, the two parent agents are prevented from reproduction for a certain period of time t_{pause}. Whenever the number of agents in the population P_t falls below P_s, agents are randomly added to the population until $P_t = P_s$.

We focused explicitly on one certain aspect of media, namely the relevance of expectations in choosing an appropriate answer to a received message. Thus, we combine an internal state with the expectation of a received message. This results in a frame-like structure which will be executed on two levels. In a first step a set F_i of frame structures is chosen based on the state of the environment. This step will be performed without any learning by the agent and is totally determined by the environment. In a second step the agent chooses one frame structure from the previously chosen set F_i. The selected frame will be executed resulting in an action a_{t+1} and a new message M_{t+1}. A frame F is defined with respect to a received message $M_r = M_t$ in the following way:

$$\begin{array}{llllll} \text{if} & M_r = M_{e1} & \text{then} & a := act_1 & \text{and} & M := mes_1 \\ \text{elseif} & M_r = M_{e2} & \text{then} & a := act_2 & \text{and} & M := mes_2 \\ \text{else} & & \multicolumn{4}{l}{\text{execute a trouble frame in} \quad F^T ,} \end{array}$$

where $a_{t+1} = a$ and $M_{t+1} = M$. A "trouble frame" F_i^T will be executed in the case that the received message was neither M_{e1} nor M_{e2}. This frame has a special structure, because it does not check the occurrence of a certain message, rather it checks whether the agent has been sanctioned or not in order to determine the new action and message:

$$\begin{array}{llllll} \text{if} & \mathit{sanctioned} = \mathit{true} & \text{then} & a := act_{T1} & \text{and} & M := mes_{T1} \\ \text{else} & & \text{then} & a := act_{T2} & \text{and} & M := mes_{T2}\ . \end{array}$$

For every state of the environment the agent has two frames. The selection of a frame at time t will be guided by a Q-value Q_F, that is, reinforcement learning[12] takes place in order to choose an appropriate frame in a given (environmental) situation. The entire collection of frames for an agent by a given final state U_e of the environment is: $F_k = \{F_{(k,0)}, F_{(k,1)}\}$, for $k = 0, \ldots, U_e$. An additional frame set is employed by an agent when the agent starts the communication by generating the start message M_0. For the trouble state U^T the agent can choose also between two (trouble) frames $F^T = \{F_0^T, F_1^T\}$.

Evolution is based on frames, agents do not change frames during their life time, they are just able to change the Q-value of a frame with respect to the other frame inside the same frame set. At the beginning of the simulation, all frames of all agents are initialized randomly. In particular, variables M_{e1}, M_{e2}, mes_1, mes_2, mes_{T1}, and mes_{T2} get randomly chosen values from $\mathcal{S} = \{0, 1, 2, \ldots, S_{max}\}$, and variables act_1, act_2, act_{T1}, and act_{T2} get randomly chosen values from $\mathcal{A} = \{Null, Sanction, Exit, Replace, Plant_I, Harvest_I, Plant_{II}, \ldots\}$. Inheritance happens on the frame level, that is, cross-over takes place *between* frames, not inside a frame (but inside a frame set). Individual parts of a frame are subjected to mutation. Therefore, e. g. part M_{e1} or act_2 may get a new random value during mutation process. Q-values are not passed on to offspring, and are set to a small random value at the beginning of an agent's life time.

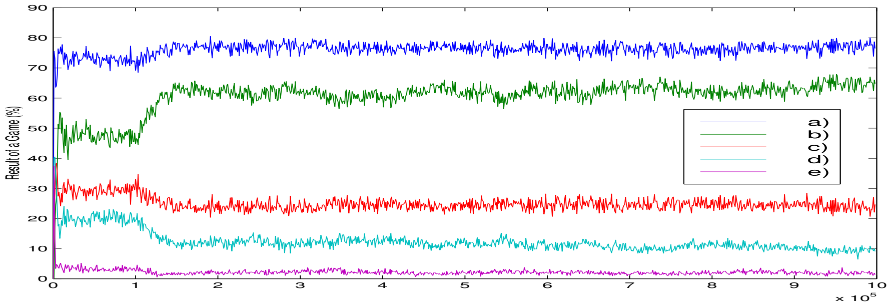

Figure 1. Simulation of 1000000 games (results averages 1000 games). Result of the simulation: *a*) maximum possible success (counting the occurrence of a "correct" pairing of the agents); *b*) the actually achieved success; *c*) correctly performed *Exit*; *d*) *Exit* in wrong situation; *e*) stopped, because maximum rounds exceeded. For example: after around 500000 games, the average result of 1000 games was 60% successful games, out of a maximum of 75% possible successful games, 25% were correctly and 10% were incorrectly exited by an agent, and 5% were stopped by the system (values approximated). $U_e = 4$, $S_{max} = 3$, $rounds = 10$, $E^\star = 10.0$, $E_l = 0.5$, $E_h = 2.5$, $E_b = E_p = 0.1$, $E_s = 50.0$, $A_{max} \in \{550, \ldots, 800\}$, $A_{sex} = 20$, $t_{pause} = 20$, $a = 5.0$, $b = 1.0$.

4 Simulation results

Figure 1 shows the general outcome of a simulation, and Figure 2 shows the statistics of the number of sanctions in 1000 games, the number of living agents, and the average energy of the agents. The maximum number of agents was set to 1024. The simulation started with 3 agents and as long as the number of agents was below 15 a higher energy pay off $E^\star$ was given for success than indicated in the caption of Figure 1 (to support an onset of evolution). The number of agents grew rapidly until the limit was reached. Later, evolution still took place optimizing the frame structures. This may result, for example, in changing cooperation sequences, or in a "competition" of different sequences as indicated in Figure 3. A sequence was defined by M_0, M_1 a_1 M_2 a_2 $\ldots$, that is, M_0 is the start message of the first agent, M_1 the answer message and a_1 the action of the other agent and so on. The coding of actions is: 0 - $Null$, 1 - $Sanction$, 2 - $Exit$, 3 - $Replace$, 4 - $Plant_1$, 5 - $Harvest_I$,.... Because we analyzed only sequences which did not contain a *Replace*-action, and which were successful, all these sequences end with action 7 ($Harvest_{II}$). Figure 3 shows the eight most frequent sequences of the entire simulation. The sequence 1 occurred 160877 times, out of 346727 successful sequences, without a *Replace*-action. In detail, sequences are shown

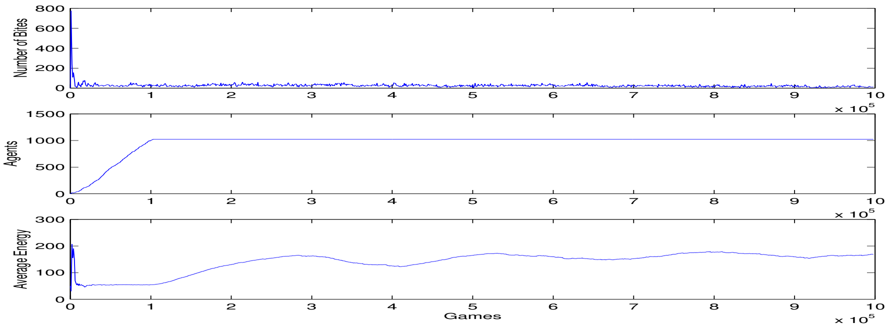

Figure 2. From top to bottom: Number of sanctions ("Bites", not averaged), number of living agents, and average energy of the agents. The number of agents was restricted to 1024. When this number was reached, agents did increase their amount of energy on the average.

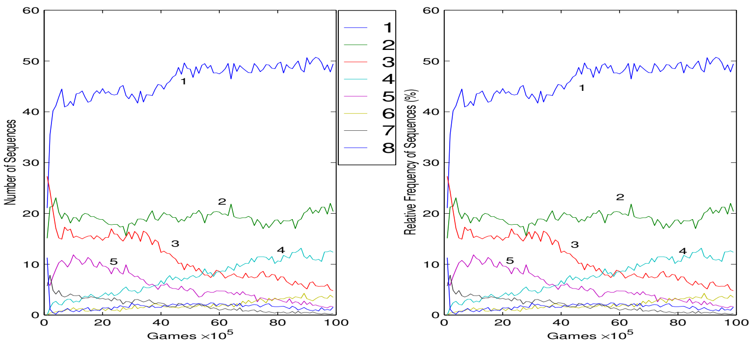

Figure 3. The eight main sequences of the frame based evolution. Left: Absolute occurrence of sequences (average of 1000 games), right: relative occurrence of the sequences (in relation to 346727 successful sequences). The eight sequences occurred 329895 times.

in Figure 4.

The communicative behavior of agents became more and more regular. Because there were two frames for each environmental situation it is obvious that a frame set is assumed to contain exactly one appropriate frame for *Planters* and one for *Harvester*. An individual only has to explore which one is better suited. A detailed analysis of the communicative behavior reveals indeed that communication controls the behavior of agents. As the results indicate, the agents were able to set up a population wide semantics for the

number (see Fig. 3)	number of occur.	seq. M_0 M_1 a_1 M_2 a_2 ...
1	160877	1 0 4 0 5 1 6 2 7
2	66551	2 0 4 0 4 0 5 1 6 2 7
3	37402	0 0 4 0 5 1 6 2 7
4	26721	0 1 5 0 4 0 5 1 6 2 7
5	19039	2 1 5 0 4 0 5 1 6 2 7
6	7118	0 0 4 0 4 0 5 1 6 2 7
7	6453	2 0 5 0 4 0 5 1 6 2 7
8	5734	2 1 7 0 4 0 5 1 6 2 7

Figure 4. Eight most sequences in detail.

exchanged symbols. The meaning of a symbol depends – of course – on the environmental state, however symbols became functional for the agent's choice of the next message or action. Sanctions became less important as the behavior became more normative. Although not shown here, simulations are easily adapted to cases where several thousand agents may evolve, still acting in a coordinated manner.

5 Conclusion

We have shown that a growing population of agents may act in a coordinated manner even in the case when the cognitive capabilities of the agents are limited and, moreover, when agents do not know anything about each other (apart from received messages). From an observer's point of view agents reveal a normative behavior, although we did not predefine any norms. We started by questioning what kind of mechanisms human society evolved in order to cope with a growing number of individuals. We found an interesting answer in the work of sociologists, especially SGCMs proposed by Luhmann. We have modeled one SGCM (power) in a first approach. However, our simulation is still too simple to establish all aspects of a symbolic medium.

Nevertheless, Luhmann's suggestions regarding SGCM, especially the aspect of structuring a situation by expectations, turned out to be useful. We modeled some aspects of his theory, mainly aspects of a closed communication system, but found interesting approach to answer well-known problems in multi-agent research, namely problems of scalability and the definition of norms. In subsequent work, we will deal with a more elaborated model of a symbolic medium. Further, the impact of more than one medium has to be analyzed, especially their potential for a more heterogeneous agent society and more complex problems to be solved by the agents.

Acknowledgement

We are grateful to three anonymous reviewers for their comments. This work is supported by the *Deutsche Forschungsgemeinschaft* under grant number DI 452/10-1 and part of a research project headed by Werner Dilger and Bernhard Giesen.

References

1. A. S. Rao and M. P. Georgeff. Modeling Rational Agents within a BDI-Architecture. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, pages 473–484, Cambridge, Mass., 1991.
2. R. Dunbar. *Grooming, Gossip, and the Evolution of Language*. Harvard University Press, Cambridge, Mass., 1996.
3. T. Parsons. *The Structure of Social Action*. Free Press, New York, 1968.
4. N. Luhmann. *Social Systems*. Stanford University Press, Stanford, Ca., 1995.
5. M. Paolucci and R. Conte. Reproduction of Normative Agents: A Simulation Study. *Adaptive Behavior*, 7(3/4):307–322, 1999.
6. K. Binmore. *Game Theory and the social contract*, volume 1: Playing fair. Cambridge, Mass.: MIT Press, 1994.
7. Y. Shoham and M. Tennenholtz. Social Laws for Artificial Agent Societies: Off-line Design. *Artificial Intelligence*, 73, 1995.
8. M. P. Wellman. A Market-Oriented Programming Environment and its Application to Distributed Multicommodity Flow Problems. *Journal of Artificial Intelligence Research*, **1**:1–23, 1993.
9. S. Park, E. H. Durfee, and W. P Birmingham. Emergent Properties of a Market-based Digital Library with Strategic Agents. In Y. Demazeau, editor, *Third Int. Conf. on Multi-Agent Systems (ICMAS98)*, pages 230 – 237, Los Alamitos, Cal., 1998. IEEE Computer Society.
10. G. Ballot and E. Taymaz. Technological change, learning and macro-economic coordination: An evolutionary model. *Journal of Artificial Societies and Social Simulation <http://www.soc.surrey.ac.uk/JASSS/2/2/3.html>*, 2(2), 1999.
11. D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Inc., Reading, Mass., 1989.
12. R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, Cambridge, Mass., 1998.