Martin Meister, Kay Schröter, Diemo Urbig, Eric Lettkemann, Hans-Dieter Burkhard and Werner Rammert (2007)

# Construction and Evaluation of Social Agents in Hybrid Settings: Approach and Experimental Results of the INKA Project

*Journal of Artificial Societies and Social Simulation* vol. 10, no. 1
<http://jasss.soc.surrey.ac.uk/10/1/4.html>

For information about citing this article, click here

Received: 20-Jan-2006 Accepted: 18-Jul-2006 Published: 31-Jan-2007

## Abstract

We present an integrated approach to the modelling, implementation and examination of social agents as consecutive steps in an interdisciplinary research process. The multi-agent system developed is inspired by sociological role concepts to provide agents with the capability to negotiate on the basis of social expectations. The overall goal of our system is to allow direct interaction between agents and humans. In order to examine these hybrid constellations we developed an experimental approach, termed 'Interactivity Experiment'. An initial experiment showed that human settings, agent settings and mixed settings produce very different results, and that heterogeneous settings are superior to homogeneous settings.

**Keywords:**
Socionics, Negotiation, Roles, Mixed Human-Agent Systems

## Introduction

**1.1**

From the very beginning, the objective of the INKA[1] project (see http://www2.informatik.hu-berlin.de/ki/inka/index.html) has been to build computer agents that are capable of coping with the complexities, and even inconsistencies, of real-world organisations. This overall goal can be divided into two different sub-goals, which build on one another. The first sub-goal is to build a multi-agent system (MAS) that makes use of sociological insights into the practical side of human organisations, thus constituting a special version of artificial sociality. This is mainly a modelling task, which addresses especially the development of an agent architecture. Because this is done by the implementation of a social mechanism for agent coordination, we speak of 'social agents'. The second sub-goal is to examine the consequences of the re-entry of these social agents into real human organisations, thus constituting hybrid sociality. In our view, this is mainly a question of methodology because the existing methods do not directly cover the interaction of agents and humans.

**1.2**

In the simulations community and in computer supported cooperative work there seems to be growing interest in constellations which are inhabited by humans as well as by agents. While this recent attention stems from very different strands of research, it can be seen as narrowing the 'great divide' (Bowker et al. 1997) from two different angles.

**1.3**

Recently, 'participatory simulation' has become a widely accepted approach in the simulations community (see the overview in Ramanath and Gilbert 2004). This is mainly because simulation techniques and results are more likely to be accepted and useful in practice if the end-users are involved in the design of the simulation setting at an early stage of development. Thus, methods like scenario or role-playing games are used as an addition to the simulation itself (Barreteau et al. 2001; Guyot and Drogoul 2004). However, some researchers move beyond the question of acceptability by integrating 'stakeholders' into MAS simulations in order to achieve such aims as 'validating social interaction models', 'fostering reflection', 'teaching complexity' or 'learning of participants' (Guyot and Drogoul 2004). These approaches extend principles of software design to a concept of 'heterogeneous simulations' (ibid. 33) where the 'human is inserted in the loop' of iterative research steps (ibid.).

**1.4**

In computer supported cooperative work the new possibilities and risks of intelligent computer agents have recently been discussed under headings like ASCW ('agent supported cooperative work', Sierhuis et al. 2001; Bradshaw et al. 2003) in the sense of 'human-computer symbiosis'. Kortenkamp and Freed (2003) summarised the following issues as being central to 'human interaction with autonomous systems': agents' acceptance of human tasks, modelling of humans, and facilitating of human understanding of agents' tasks and actions. The most challenging issue comes to the fore when the two sides are seen together: the 'mutual adjustment of degrees of autonomy' of agents and humans in settings that are termed 'mixed initiative systems'. This version of an integrated view is backed by an understanding of 'complementary asymmetries in human-agent interaction' (Bradshaw et al. 2003). This means that there are gradual strengths and limitations in the nature of agents and humans which should complement one another for the achievement of specific tasks.

**1.5**

Though originating from a totally different standpoint, these two developments seem to constitute a common ground on which our concern with hybrid communities is also located. As part of the Socionics programme, our approach differs from the ones mentioned above in two respects. First, we address questions of agent (and human) coordination as a theoretical issue in both artificial intelligence and sociology. We have built the INKA-MAS based on these considerations. Second, in contrast to participatory simulation, we re-enter the MAS built into a human context. We term this process 'hybridisation'. Our MAS was, from the very beginning, designed to be an instrument for the examination of human-agent interaction, not only a stand-alone system.

**1.6**

From a computer scientist's point of view, the examination of hybrid constellations is relevant for the development of distributed open multi-agent systems. In these systems, increasingly, humans and agents interact beyond delegation from humans to machines: interactions occur on an equitable level in a multilateral coordination context. Due to the medium-specific communication restrictions - for example, no visual face-to-face communication - and due to an increase in agents' competence and autonomy it may become difficult to decide whether the interaction partner is human or artificial. Differing behaviours of artificial agents that act on behalf of humans may perturb the mutual expectations established between humans.

**1.7**

There are two ways of addressing these new challenges: First, we could force the humans to base their decisions on machine-like reasoning only and thus exclude expectations and other social aspects. In this way, problems of acceptability are likely to occur. People may feel threatened and finally may not restrict their behaviour as requested. This also calls into question the performance of the system. Second, we could enable the agents to cope adequately with social expectations. This would allow humans to maintain their common decision processes. These expectations may be learned during runtime (see Nickles et al. 2007). However, learning can take a long time, which is frequently not available. The alternative is to derive humans' mutual expectations empirically from the domain. The agent architecture, then, has to enable inclusion of this empirically derived data.

**1.8**

Of course, both ways are extremes. For practical application, an approach somewhere in-between is preferable. We have to trade off the complexity of the agents' social competence and the effort required for realisation. As far as our system is concerned, we base the agents' social competencies on generalised expectations (social types or practical roles; see 2.3). But we also limit the interaction possibilities of humans to a degree that can be handled by agents (see 3.4).

**1.9**

Overall, we contribute to computer science by developing a prototypical MAS that is based on empirically derived constructs from the level of organisational coordination. Furthermore, we contribute by developing software for a case study that investigates the behaviour of human participants in artificial societies that do not deal with artificial but with real-life problems, namely negotiating exchanges of work shifts.[2]

**1.10**

From a sociological point of view, the collaboration with computer scientists is beneficial because of the need to transform sociological concepts into numerical models and to evaluate them in computer simulation. But hybrid constellations have recently been recognised as a major sociological challenge in themselves. The relevant issues are the degrees of social agency in human-machine interaction and the nature of socio-technical constellations (Rammert 2003; Rammert and Schulz-Schaeffer 2002). However, in the social sciences these questions have largely remained on a purely conceptual level. A concrete examination of the integration of computer agents into complex human contexts has not been done. Thus, the major challenge is a methodological one and the development of an experimental approach addresses the sociological issues at hand.

**1.11**

The interdisciplinary work in our project was conducted in four phases described by our *socionic development cycle* (see Figure 1). The remainder of this article describes these four phases. Section 2 summarises the sociological conceptualisation of practical roles for the modelling of the MAS, and their application within the C-IPS approach, which we developed for structuring the reasoning of negotiating agents. Section 3 briefly describes the practical steps towards the system and its application: The implementation of the system, the methodological background and the set-up of our *interactivity experiment*, are presented. Section 4 sums up the major results from our initial experiment and some major lessons learned.

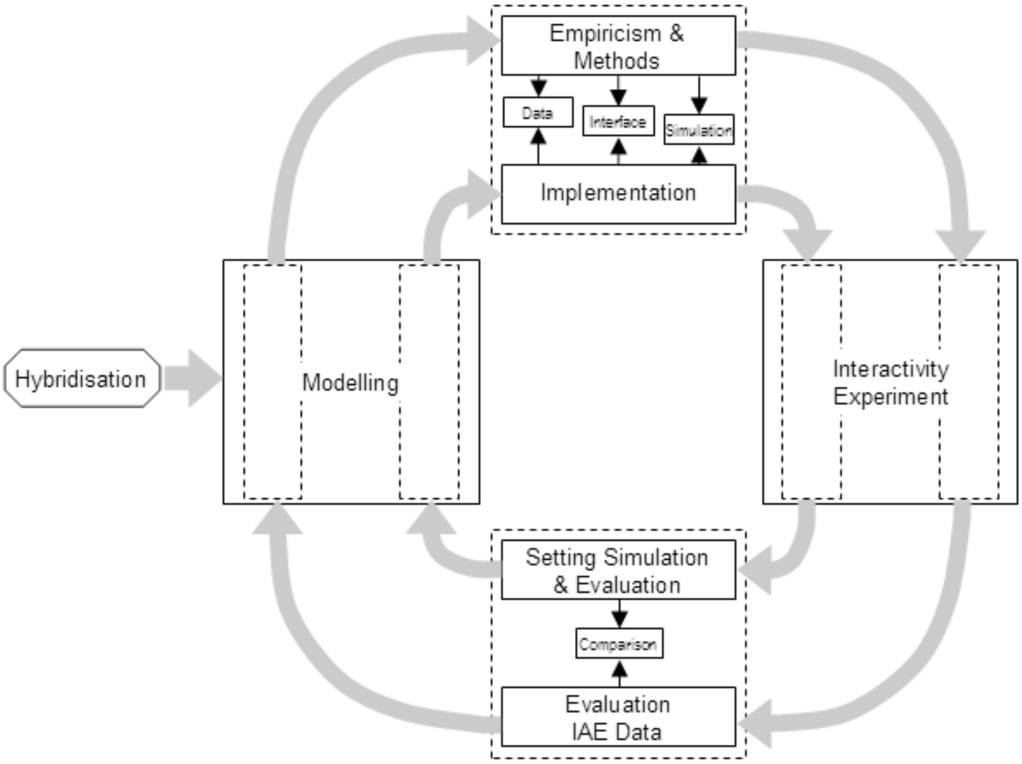



**Figure 1**. The socionics development cycle structures the interdisciplinary work in our project

## Agents enacting practical roles - the modelling approach

**2.1**

Every socionics approach starts with modelling: the production of a condensed and explicit model of the social coordination mechanism to be applied to an MAS, followed by the translation of this model into an appropriate agent architecture. We address the meso-level of organisational coordination because human organisations offer an effective way of coordinating complex individual behaviours, while at the same time remaining capable of flexible adaptation to changing environments (cf. Aldrich 1999). Concepts from the sociology of organisations are thus perceived as a promising blueprint for the design of MAS. However, in the social sciences two opposing ways of conceptualising organisations have been elaborated in recent decades: the first focuses on explicit and formalised ascriptions, the other on the multiplicity of local practices. The same dichotomy between a macro-view (formal roles) and an interactionist view characterises sociological role theory. Therefore, the first modelling challenge was to deal with this dichotomy, which in the MAS design debate was recently described as the challenge of 'formalising the informal' (Castelfranchi 2003). This challenge is to develop an MAS architecture between the extremes of fully formal social control and fully spontaneous social order: 'What is needed is some attempt to "incorporate" part of these layers and issues in the technology itself ... [The agents] should be able to manage - and thus partially understand - for example permissions, obligations, power, roles, commitments, trust' (ibid.).

**2.2**

Drawing on these considerations we focus on the practical (in Castelfranchi's terms: the 'informal') side of coordination by 'incorporating' practical roles in our agents' architecture. In this sense, once we get into the agent architecture we deal with the implementation of practical roles by means of social types as generalized expectations regarding peers. The INKA-agents are able to utilize knowledge about such social types but at the same time can enact those within an organisation's formal constraints.

### Sociological background and modelling decisions

**2.3**

Human organisations, unlike many other social entities, tend to work out an explicit description of their own principles of coordination: rules for membership, planning schedules, hierarchies, job descriptions adjusted to the internal division of labour, etc. These descriptions, often provided by management units, typically present a formally coherent and encompassing picture of the organisation's functioning. Sociological conceptualisations of organisational coordination, too, have elaborated this formal, explicit and hierarchical body of regulations as the basis for functionality and rational decision-making (cf. March and Simon 1958). However, the findings of empirical investigations into the 'real life of organisations' pointed to a picture of 'organised anarchies' (Cohen, March and Olsen 1972), since formalised descriptions turned out to be inconsistent and often mutually conflicting, thus creating individual frustration and large-scale inflexibility.

**2.4**

In sociology there are at least three different ways of dealing with these findings, which in turn lead to different consequences for the design of information systems. One way is to uncover misleading or conflicting rules, especially 'concurrencies' (ibid.). By this means, a redesign of formal structure can be initiated. A second direction of sociological research claims that the Weberian picture of the 'iron cage' does not describe any organisational reality at all, but 'dramatically reflects the myths of their institutionalised environments instead of the demands of their work activities' (Meyer and Rowan 1977, p. 431). Following this description, it is obvious that formal structures can by no means serve to guide the design of MAS. A third way focuses solely on local practices or 'situated action' (Suchman 1987), a stance mostly adopted in groupware research and especially in CSCW, and leading directly to a strong scepticism about whether modelling makes sense at all (cf. Berg and Toussaint 2003).

**2.5**

We do not claim that we have found a comprehensive way out of this seemingly paradoxical situation. We have developed a solution that regards both the formal and the practical side of organisational coordination as real (see Meister et al. 2005 for details). From organisational theory and empirical investigations, we know that in the 'real world' one way of counteracting formal regulations and especially their incoherences are negotiations by means of which the organisation's members themselves create a flow of problem solutions for their daily work practices. In our approach, negotiations are defined as a situated mechanism for problem solving in situations in which actors have a high degree of autonomy and at the same time a high degree of mutual dependency.

**2.6**

In order to model agent behaviour on the informal side of organisational coordination we draw on a specific type of role theory. We especially draw on the tradition of symbolic interactionism which focuses on the ongoing processes of interactively (re-)building social order. With respect to roles, this is especially important for the role-taking process which is defined in this tradition as follows: 'Grouping behavior into "consistent" units which correspond to generalizable types of actors' (Turner 1962: 32). In contrast to a broader definition of roles as pre-given norms, the consistency of roles in this approach is an achievement from 'the process of organizing behavior vis-à-vis relevant others' (ibid.). We termed these 'consistent units' practical roles, in contrast to the notion of formal roles, which are described as the force which the formalised regulations of the division of labour exert on the organisations' members (as in Dahrendorf's well known account of 'homo sociologicus' 1973; see also the critique in Esser 1999: 82ff).

**2.7**

Role differentiation (Turner 1962: 28), then, is a result of a mutual stabilisation of interpretations as part of the interaction process. This means that in practice a given fit of roles may make interaction easy or complicated. Because this fit of roles works against the one-dimensionality of formal role prescriptions (and 'repairs' its incoherences), it can be assumed that a high degree of diversity is the prerequisite for effective coordination (see our hypotheses in 3.3).

**2.8**

In our MAS these practical roles are 'incorporated' in the agents. At the present stage of our system development it is not possible for practical roles to emerge themselves, and formal roles are modelled as constraints on the agents' behaviour. We are aware of the fact that this is a very restrictive approach but it enabled us to focus our research and to keep complexity for modelling as well as for the empirical investigation at a manageable level. Due to this focus we do not consider specific relational or organizational structures as they are considered in related research on organisational structures in MAS, which usually has put a strong focus on what we term formal roles (see for instance Ferber 1995, Grossi et al. 2005, and also the MABS workshop series and the CoOrg workshop[3]).

**2.9**

Practical roles are patterned perceptions of the interaction partners. We modelled them according to the concept of 'group figures' (Popitz 1967), which can be observed among lasting informal groups (for example, juvenile gangs) and also within organisational units. Because the term 'figure' captures ascriptions, which can be assumed to differ only slightly between different social entities, we termed these patterned expectations 'social types' (see Almog 1998).

**2.10**

In order to achieve role-specific behaviour we had to equip our agents with distinctive expectations and interests. This can be described quantitatively in terms of an expected increase in capital as their 'motivation'. We modelled interest drawing roughly on Bourdieu's theory of capital sorts (1986) and distinguished economic, social, cultural and organisational capital, the latter transformed into symbolic capital on the practical side of coordination (for details see Meister et al. 2002).

**2.11**

We worked with a set of six social types. The names and characteristics of the types have been constructed on the basis of empirical evidence. We do not go into detail here (see Meister et al. 2002; the social types have been refined by a second empirical study), but illustrate the different types by means of a schema (see Table 1) which we presented to the human participants in the interactivity experiment. The transformation of this scheme into numerical values and a data structure used by our agents is explained in the Appendix (see Table 5).

**Table 1:** Profiles of the social types

| Social type | Willingness to compromise | Willingness to negotiate | Success effects relationship to partner | Preferred shift types | Preferred form of capital | Irrelevant forms of capital |
|---|---|---|---|---|---|---|
| **Family type** | Low | Average | No | Early shift, Night shift | Money | Reputation |
| **Team type** | High | High | Yes | All shift types | Reputation, relationships to colleagues | Money |
| **Uncooperative type** | Low | Low | No | All shift types | Knowledge acquisition | Positive relationships to colleagues |
| **Self-confident** | Low | Average | No | All shift | Knowledge | Positive |

| | | | | | | |
|---|---|---|---|---|---|---|
| **type** | | | | types | acquisition | relationships to colleagues |
| **Agreement-orientated type** | High | High | Yes | All shift types | -- | Money, reputation |
| **Pleasure type** | Average | Average | No | Late shift | Money | Reputation, knowledge acquisition |

*Note*: Columns 2-4 specify typical behavioural characteristics; columns 5-7 give typical preferences.

2.12
Our concrete domain is negotiations on shift exchanges in a hospital. This domain exhibits all the characteristics of complex organisations mentioned above. The official rosters do not adequately consider individual leisure-time interests. Therefore, the employees negotiate and trade single work shifts, thus making use of the scope deliberately provided by the management. These shift negotiations are a daily requirement under rigid time restrictions. Backed by empirical evidence, our theoretical assumption is that these negotiations are an effective means of coordination because they are carried out using a limited set of practical roles. Again, we have to mention that these practical roles do not show a complex interaction with the formal roles.

**Modelling negotiating agents using the C-IPS approach**

2.13
If an MAS is to be applied in a hybrid scenario, in which agents and humans have to interact equitably by means of life-like negotiations, it must meet a number of specific requirements:

- Most important, our agents have to incorporate social features. This allows them to estimate the complex behaviour of human partners and to act reasonably from the humans' point of view. In our approach, these features are social types (see 2.3), which define typical preferences and behavioural parameters. Social types serve as a guideline for the agents' negotiating behaviour and provide an estimation of the partners' attitudes. They are complemented by individual experiences. Experiences with individual partners, as well as with social types, are based on the success or failure of previous negotiations. This makes it possible to develop a tendency to prefer particular negotiation partners.
- The negotiation protocol has to enable sequences of interactions that humans are used to.
- The infrastructure of the system has to ensure equivalent interactions of humans and agents. All interactions will be done via this infrastructure. All required information is provided by this infrastructure.
- We have to ensure that people in physically distributed locations can interact in parallel.

2.14
In every negotiation, both humans and agents have to consider three main aspects: negotiation issues, negotiation partners and negotiation steps, that is, the next communicative act in the negotiation process. One influential approach conceptualises an agent's decision making in a negotiation by distinguishing the negotiation object, the negotiation protocol and the internal reasoning process (see Jennings et al. 2001; Huhns and Stephens 1999). This approach does not address the negotiation partners explicitly, and the reasoning process itself is unstructured. While some work has been done on the selection of issues and partners (Fatima et al. 2003; Kurbel and Loutchko 2002), these aspects have not been considered generally important to negotiating agents.

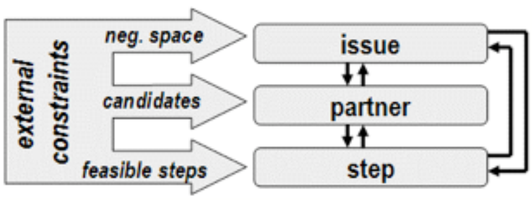



**Figure 2**. The C-IPS approach to negotiating agents

2.15
We require agents that are fully autonomous in all decisions regarding the negotiation, and we need a clearly structured decision process to localise the influence of social aspects. Hence, we extended and restructured the mentioned approach to negotiation modelling and developed the C-IPS approach (Urbig et al. 2003; Schröter and Urbig 2004). C-IPS (see Figure 2) distinguishes between external constraints (C) that restrict the agent's decisions and the internal reasoning process (IPS). The reasoning process is structured by three separate, but mutually dependent modules for the selection of issues (I), partners (P) and steps (S).

2.16
The interdependencies between these modules are specified explicitly. They are influenced by external constraints. Static interdependencies define allowed combinations of decisions by the different modules and have to be considered within the separate modules. Dynamic interdependencies specify the activation of modules due to decisions within other modules or due to perception of interactions with the environment (including partners).

2.17
C-IPS provides an approach to modelling negotiating agents which is on a level above protocol and strategy design or game theory. It can be combined with different approaches to decision making in each module. We only require that the modules follow the constraints and consider the interdependencies. In fact, this is also the basis for the integration of

humans into our multi-agent system (see Urbig and Schröter 2005): We replace our agents' automatic decision modules following the BDI approach (Rao and Georgeff 1991; Fischer et al. 1998) by modules that present a graphical user interface to get the decision from a human. We can even switch the modules independently and at runtime between the automatic and the manually controlled mode. Furthermore, when applying social concepts the modularisation makes it possible to separate how the concepts affect issue, partner and step selection (Urbig et al. 2003 give an example of this).

**Specifying the constraints, interdependencies and decision modules of our agents**

The agents must heed the following *constraints* [4] and *static interdependencies*:

- The agents' decisions are restricted to negotiations on shift exchanges.
- The need for a negotiation arises if there is an individual leisure time interest in relation to a shift the agent is scheduled for. We term this situation a conflict. The shift in question defines the issue of a negotiation.
- The negotiation concerns the shift the agent has to take in exchange for the conflicting shift.
- We allow only bilateral negotiations.
- Our negotiation protocol defines a negotiation by iterated offers. Besides concrete proposals it is also possible ultimately to propose an exchange, to ask the partner for a proposal, to agree on an exchange or to cancel the negotiation.
- Taking multiple shifts at the same time is not allowed.
- After an exchange the individual rosters have to meet a number of legal requirements concerning minimum rest periods. Our agents will not consider forbidden exchanges. The administration must confirm each exchange.

**2.18**

We define the following *dynamic interdependencies* between the agents' modules:

- Only one IPS decision module is active at any given time. We follow a sequential approach regarding decision making, starting with issue selection, then choosing an appropriate partner and finally selecting a negotiation step.
- Our agents are only engaged in one negotiation at a time.
- The information distribution in our system is asymmetric, that is, the agents do not know about each other's conflicts. Agents with conflicts have to ask others for a negotiation. We term the asking agent *initiator* and the other agent *responder*. Requests will be accepted, unless the responder is already negotiating. Once the responder has accepted the request, it sets the negotiation issue and partner accordingly. If the negotiation space is empty - for example, because there are no allowed exchanges according to the agent's local information -the responder cancels the negotiation.
- On realizing that no successful negotiations are possible on the selected issue, or with the selected partner on that issue, the agent marks those options as impossible and will not consider them during the selection processes for a certain time.
- In an open distributed system, it may happen that partners disappear, or conflicts are resolved by external events. Furthermore, timeouts may occur if partners do not want to wait any longer.

**2.19**

During the *issue selection*, an agent chooses a conflicting shift as the issue of a negotiation. From all the shifts assigned to the agent, it considers only those shifts *s* with a leisure-time interest above the significance threshold and that are not known to be currently impossible to negotiate. From those shifts the agent selects one with a minimum *shift utility* $u_{shift}$:

$$u_{shift}(s) = w_{LTI} \cdot lti(s) + w_{CA} \cdot \sum_{i=1 \in \{1...4\}} ci_i \cdot ca_i(s) \qquad (1)$$

**2.20**

The weights $w_{LTI}$ and $w_{CA}$, given by the agent's social type, balance the agent's goals in relation to fulfilling leisure time interests and in relation to accumulating different forms of capital by working a shift. The function *lti(s)* yields the agent's leisure time interest for shift *s*. The weights $ci_i$ define the agent's interest in capital form *i* as defined in its social type. The function $ca_i(s)$ yields the amount of capital of form *i* typically accumulated during shift *s*. If the agent realises that it cannot find a partner for the conflicting shift selected as the negotiation issue, it revises the selection.

**2.21**

The partner selection module decides on a promising negotiation partner for the selected issue. All available partners *p* that are not marked as impossible and that can take the issue shift are considered. For each of them the agent calculates the *partner utility* $u_{partner}$:

$$u_{partner}(p) = 0.5 \cdot (exp_{personal}(p) + exp_{typified}(p)) \cdot \sum_{(s,s') \in AE} max(0, u_{exchange}(s,s'))^2 \qquad (2)$$

**2.22**

The function $exp_{personal}(p)$ yields the experience in relation to partner *p*, while $exp_{typified}(p)$ yields experiences in relation to partner *p*'s social type.[5] *AE* denotes the set of all allowed exchanges with partner *p* regarding the selected issue *s*. The *exchange utility* $u_{exchange}(s,s')$ defines the utility of an exchange where the agent delivers shift *s* and gets shift *s'* from p in return:

$$u_{exchange}(s,s') = u_{shift}(s) - u_{shift}(s') \qquad (3)$$

**2.23**

If available, only partners without a social type-dependent leisure time interest for the issue are further considered. Now a partner with maximum partner utility is chosen. If the agent realises that the partner is already negotiating, or that a negotiation on the same issue with that partner has recently been unsuccessful, it revises the selection.

2.24
The *step selection* module has to find an appropriate next negotiation step. All possible exchanges are ranked by exchange utility. Each agent has a threshold for accepting exchanges and a threshold for cancelling the negotiation. These thresholds change over time. The 'accept' threshold linearly decreases and the 'cancel' threshold linearly increases. The actual initial values of the thresholds and the change over time depend on social type parameters and on the evaluation of currently available exchanges. Agents propose only acceptable exchanges, that is, exchanges that are above both thresholds. An agent agrees upon a proposal if the proposed exchange is acceptable. The negotiation is cancelled if the utility of an exchange proposed by the partner is below the cancel threshold. For ultimatums, the thresholds are advanced in time until they meet. Here any proposal is either acceptable or leads to cancellation of the negotiation.

2.25
After the negotiation is finished, the agents update their experiences according to formula (4) if the negotiation was successful or according to formula (5) if the negotiation failed due to cancels or timeouts. In the following formula $step_{exp}$ denotes a constant that specifies the magnitude of change and $p$ denotes the partner of the finished negotiation:

$$exp'_{\{personal \mid typified\}}(p) = (1-step_{exp})\cdot exp_{\{personal \mid typified\}}(p) + step_{exp} \quad (4)$$
$$exp'_{\{personal \mid typified\}}(p) = (1-step_{exp})\cdot exp_{\{personal \mid typified\}}(p) \quad (5)$$

Concrete numerical values on all constants, thresholds, weights and functions mentioned in this section are presented in the Appendix.

## Intermediate steps towards an examination of hybridisation

3.1
Having jointly modelled the MAS the two disciplines had to tackle different tasks: For sociology, the task was to generate concrete data by empirical investigations and to develop a methodological approach for the examination of hybridisation; for computer science, the task was to implement the system and to enable its distributed use.

### Implementation of the MAS

3.2
Since experiment participants access the system from different computers and even different rooms, we based our prototypical system on the JADE platform (see jade.tilab.com) that supports physically distributed agents with individual interfaces, complex agent architectures, and the modelling of agent communications using ontologies.[6]

3.3
Our system architecture (see Figure 3) distinguishes three types of agent: INKA agents, environment agent and statistic agent. INKA agents negotiate on shift exchanges. They comprise two components: (i) the negotiation component may either be automatic or manually controlled according to the C-IPS approach; (ii) the system component is responsible for the management of the agent's internal knowledge, the processing of requests and directives from non-INKA agents, and the logging of negotiation events. The environment agent provides the INKA agents with all required information and makes it possible to configure and control the system, including our step-based time model. The time model determines the speed of agents' decision-making. The environment agent also implements the administration. Finally, the statistic agent is responsible for continuously collecting, processing and visualising numerical data from the INKA agents.

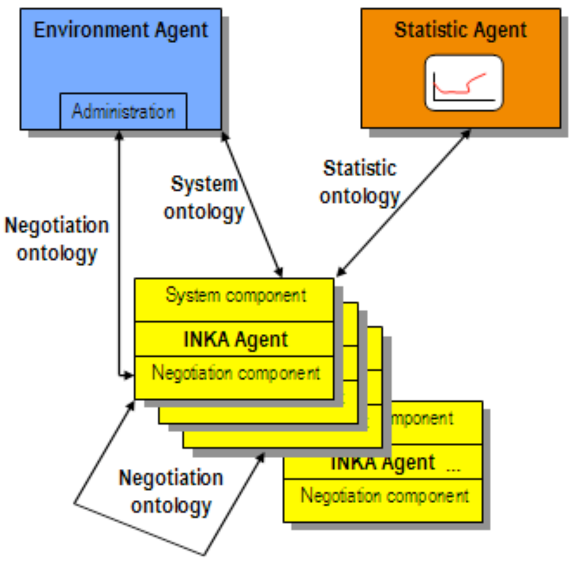



**Figure 3**. The architecture of the INKA system

### Observation of agent-human interactivity

**3.4**

The re-entry of a social MAS into a human environment constitutes what we call a hybrid constellation. Contrasting to philosophical considerations about the 'true nature' of interaction we define the term 'hybrid' as follows: If the mutual impact of technical entities, as well as their relations to human actors, can be described as social interaction, we will talk of 'hybrid communities'. Human-agent relations are termed 'interactivity' in distinction from inter-human interactions (Rammert 1998: 122).

**3.5**

In contrast to most approaches in participatory simulation and human-agent teamwork (see section 1) we address the direct interactivity of humans and agents. A mere combination of agent-based simulation and heuristics from the social sciences (like role-playing games) does not address this issue. Observation by empirical field studies, on the other hand, typically allows only a few general statements because faced with the complexity of social reality it is nearly impossible to decide which of the variables is only of local importance and which is not. Hence, we developed a new methodological approach which we termed 'interactivity experiments'.

**3.6**

From a sociological point of view this is a new type of observation because all data are generated by the automatic documentation of every single negotiation step. The design of an adequate interface has to meet two requirements. First, the scope of human behaviour has to be restricted in a way that it can be handled by our socially enhanced agents. The interface ensures that humans base their decisions on the same local information as the agents and that they take into account the same constraints and interdependencies. Second, usability has to be assured. Following ISO-NORM9241-10[7] we took the following aspects into consideration: The design of the interface should allow easy use, humans should be able to get quickly acquainted with the system and the interface should not influence the humans' decisions by arranging input and output elements suggestively. Our resulting interface is shown in Figure 4.

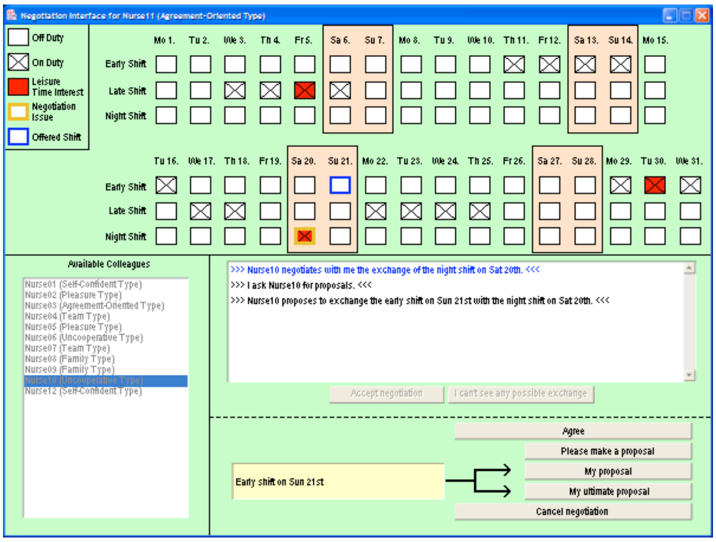



**Figure 4**. The interface for manual control of an INKA agent (click to enlarge the figure)

**3.7**

An important requirement of our experiment is the 'black boxing' of the nature of the interaction partners, that is, the assurance that nobody can recognise whether the negotiation partner is a human or an agent. To meet this requirement we had to do the following:

- The speed of the agents' activities was reduced.
- In situations during a negotiation where no new exchange becomes acceptable and the agent has to repeat the same proposal again and again, the number of such consecutively repeated proposals is restricted. If the maximum number of repetitions is reached, the thresholds are advanced in time until a new proposal becomes acceptable. This prevents the 'more impatient' humans from recognising the 'more patient' agents by their repeated proposals. This does not change the agents' negotiation strategy, they are only speeded up a little.
- After finishing a negotiation, the agents were forced to wait some time before initiating the next negotiation, to give humans a fair opportunity to contact agents outside a negotiation.

### Experimental approach to hybridisation

**3.8**

If human interaction with other humans or with agents can be logged automatically, and moreover the scope of human behaviour can be restricted in an adequate laboratory setting, it should be possible to directly observe and measure human-agent interactivity. Thus, an experimental approach offers an adequate methodological path. Initially we adopted a strong (or narrow) understanding of 'laboratory experiment' in social research methodology (Campbell 1988; Campbell and Stanley 1963), as, for example, in socio-psychological experiments about group behaviour (Aronson, Wilson and Brewer 1998), in which the examination is restricted to assumed effects of independent variables on the single variable under examination.

**3.9**

It is crucial for an experimental approach to choose meaningful hypotheses. In our domain, the results of interaction and interactivity can be measured as the degree to which the shift plan is optimised as a result of the negotiations. These negotiations work against the restricting tendency of formal prescriptions (see 2.3) and can be described as enabling a productive use of differences. In our scenario, we are able to address three kinds of difference between the interaction partners, two of them social, one ontological. For each of them we formulated one distinctive hypothesis.

**3.10**

The first difference is between heterogeneous settings, involving many social types, and scenarios with a small number of types or even homogeneous settings. Drawing on sociological role theory, we assume that resisting the tendency towards homogenisation makes poor quality outcomes more unlikely because there is larger scope for activities. The corresponding hypothesis is:

> A non-restrictive set of social types achieves better results with regard to shift-plan quality than a restrictive set of social types.

**3.11**

The second difference is between the concrete social types and the patterns of distribution between the types. This difference draws on the assumption that in practice a smooth fit of social types evolves over time, e.g. settings with outsiders and settings with sub-groups. The corresponding hypothesis can be seen as a specification of the first one:

> The pattern of distribution is the essential factor for the quality of shift-plans, independent of the individual social types involved.

**3.12**

A third difference is obvious in our approach: the one between humans and agents. Even if agents are modelled on social mechanisms they are still technical creatures and thus differ essentially from humans. Therefore, it has to be assumed that different ratios of humans and agents produce different collective results. The corresponding hypothesis is:

> Different qualities of shift plan depend on the nature of the negotiation partners - whether they are exclusively agents, exclusively humans or a mixture.

**3.13**

In experimental runs these hypotheses can be proved by the quality of the new shift-plans that result from the negotiation process. We developed the following indicators to measure this collective quality.

*Collective satisfaction with the negotiated shift-plan* (CS): Individual satisfaction can be expressed as the percentage of leisure-time interest each individual can realise. As a collective measure we define the average of the individual satisfaction values. This value is scaled on a range from 0 to 10. A higher collective satisfaction indicates a shift-plan of a higher quality.

*Collective frustration caused by negotiating the shift-plan* (CF): The individual interest in efficient shift negotiations is defined in negative terms. The measure of individual frustration is the ratio between unsuccessful negotiations - that is, negotiations that have been cancelled - and all the negotiations in which an individual has been involved. The collective value is the average of the individual values. Again, this value is scaled on a range from 0 to 10.

**3.14**

These two indicators are combined in an *overall score indicator* (OS) in the following way:

$$OS = (1 - w) \cdot CS + w \cdot (10 - CF) \qquad (6)$$

The weight $w$ is set to 0.2, so the influence of satisfaction is higher than the influence of frustration.

**3.15**

Finally, all social-science experiments call for internal validation criteria. We adopted several proposals from the literature (cf. Zimmermann 1972; Campbell and Russo 1999). A pre-test had to ensure that the measured effects were not merely an outcome of the instrument (or the experimental setting) itself; therefore running the whole procedure without agents was a good test. Furthermore, we explicitly divided the role of the experimenter into a number of different tasks: technical monitoring, technical advice, observation and formulation of exercises lay under the responsibility of different members of the research team. Thus we tried to avoid the problem of 'experimenter's bias'. In order to reduce well-known negative effects on the side of the probands, such as boredom or learning which can distort experimental results, we provided the test persons with different starting conditions and/or different exercises for every run.

### Patterns of negotiation results from simulation runs

**3.16**

Simulations are not part of our core approach. Nonetheless, they are important for different evaluation tasks within the socionics development cycle. Before the conduct of the experiment itself the task is to check whether different settings do in fact produce distinct outcomes and to contribute to a deeper understanding of patterns emerging from collective negotiations. We will focus on that point in this section. The task of checking experimental data against complementary data from the experiments will be discussed in 4.16.

**3.17**

For extensive simulation runs a system is required that is optimised for speed, can run experiments automatically and offers extended means for evaluation (for example, aggregated statistics for parallel simulation runs). The INKA-MAS cannot handle this, as it was designed for a different purpose.[8] Therefore, we implemented a second system that applies the same decision processes, but is relieved of dispensable or time-consuming components such as the agent platform enabling a distributed system, and supporting network communication and interfaces. Contrary to the INKA-MAS, the simulation system allows for full control over random processes inherent in distributed systems.

**3.18**

A significant quantity of simulation data makes possible the exploration of some of the basic patterns displayed by our system, especially those inherent in every negotiation process in contrast to those characteristic of specific settings. In order to illustrate this, we present the differences and similarities of two settings characterised by very different configurations of social types. Figure 5 shows the simulation results of a heterogeneous setting that consists of all six social types, while Figure 6 presents a homogenous setting that consists of family types only. Each graph represents the mean score of 500 runs over a time span of 170 virtual time steps.

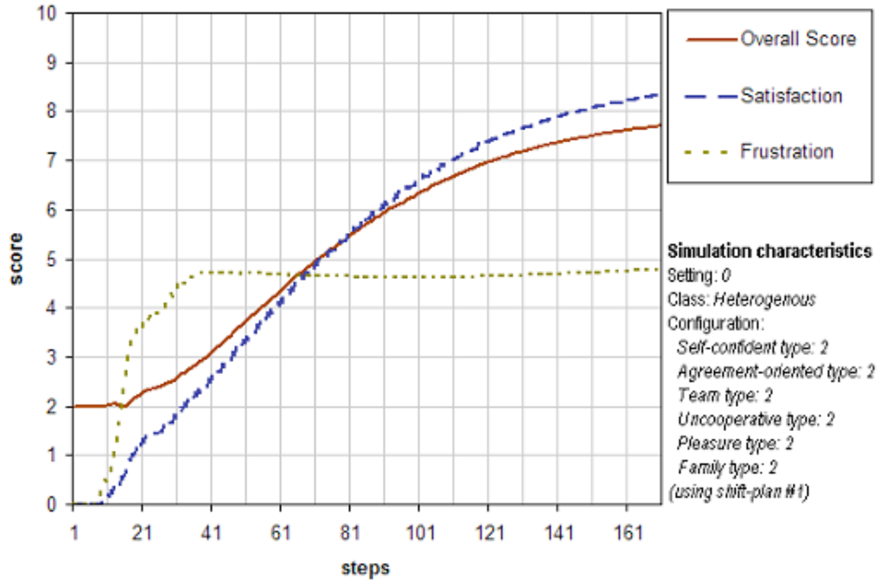



**Figure 5**. Simulation of a heterogeneous setting

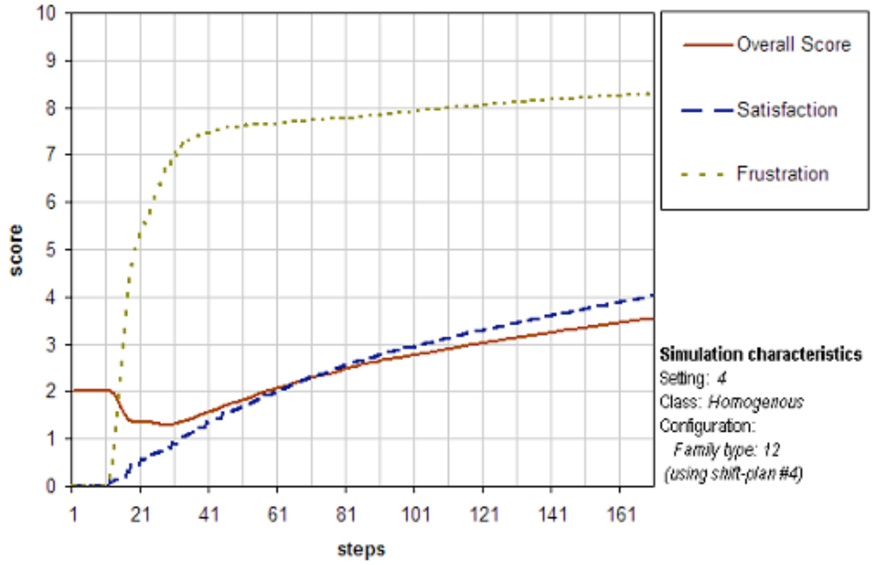



**Figure 6**. Simulation of a homogenous setting

**3.19**

Both figures show an abrupt rise in frustration at the beginning, a pattern inherent in every negotiation process. This is due to the fact that a few failures at the beginning have a stronger impact because they are not counter-balanced by successful negotiations. However, the score and the flattening rate of the frustration ratio differ strongly between the two settings. Throughout the simulations one can see that specific settings have specific frustration levels. The differences between the two figures can easily be interpreted if the difference between the two settings is taken into account: the lower frustration in Figure 5 characterises a heterogeneous setting where it can be assumed that successful negotiations are more likely. This also leads to a relatively significant rise of the satisfaction score which starts to flatten only towards the end, when most tasks are fulfilled. The lower level and slower rise of the satisfaction score in Figure 6 reflects the reduced possibilities for successful negotiations in a setting with homogenous social types, in which a mutual blockade of negotiation possibilities due to similar interests is likely to occur.

**3.20**

As regards overall score, the difference between the two figures is obvious. In Figure 6, in which the frustration score rises rapidly, the overall score starts to deteriorate before a slight but constant rise occurs. In contrast, Figure 5 presents a more rapid and steady rise that only flattens towards the end because satisfaction flattens too. This illustrates the strong difference in the negotiating capabilities of the two settings.

## Interactivity experiment: results and lessons learned

4.1

With the INKA-MAS and an appropriate method, we were able to enter the next stages of the socionics development cycle: the conduct and analysis of a first interactivity experiment.

**Experimental set-up**

4.2

In order to conduct the interactivity experiment we had to build the laboratory environment carefully because '[experimental] science is as opportunistic as a bacteria culture and grows only where growth is possible. One basic necessity for such growth is the machinery for selecting among alternative hypotheses, no matter how limited those hypotheses may have to be' (Campbell 1988: 165). This 'machinery' in our case consisted of the following parts:[9]

- Twelve probands from the domain that could be assumed to be familiar with the social types and shift negotiations from their daily practice.
- Separated PCs with an interface to accomplish the given negotiation tasks for each of the probands. As they did not know whether the other negotiation partners were humans or agents, this was a black box situation for them.
- A separate room with a PC running the system (the 'centre of command') where the different settings and the tasks for the probands were initialised and all the data collected.
- Negotiation settings including a shift-plan and three conflicting shifts which each proband had to exchange. The duration of each setting was 170 time steps (approx. fifteen minutes). Nineteen of these negotiation settings were run in the two-day experiment, each consisting of twelve players that could be agents as well as humans.

Because the probands were instructed at the beginning of each setting to play a specific social type and to take into account the other social types involved, the whole procedure had the character of a role playing game.

4.3

For observation data we drew exclusively on two computational data sources. First, the aggregated data for satisfaction and frustration collected by the statistic agent. Second, the individual and collective negotiation logs in which all relevant negotiation events are listed with their time stamp. It turned out that these log-files were very helpful (indeed, necessary) for the correction and interpretation of the raw data generated.

**General results**

4.4

The first thing we would like to emphasise about the interactivity experiment, from a general methodological perspective, is that the approach worked. Moreover, the direct interplay of technical specifications, probands in a role-playing game, indicators and data aggregation equations produced distinctive and meaningful results.

4.5

With respect to the three concrete hypotheses presented above the results differ significantly. The following tables show the mean of the overall scores achieved by the individual settings in the experiment. Note that these values necessarily contain different numbers of settings played - for human settings in particular, the number is low.

4.6

The first hypothesis was: A non-restrictive set of social types achieves better results with regard to shift-plan quality than a restrictive set of social types. Table 3 reveals at least a strong tendency towards confirmation of this hypothesis, which means that there is a productive use of social difference in the interaction. This holds true for the overall mean (in green), as well as for most of the possible types of settings.

**Table 3:** Test of hypotheses 1 and 3

| | **Agents** | **Hybrid** | **Humans** | **All** |
|---|---|---|---|---|
| **Homogeneous setting** | 5.22 | 7.20 | 8.53 | 5.97 |
| **Group setting** | 6.80 | 8.02 | 8.45 | 7.26 |
| **Outsider setting** | 6.87 | 8.19 | 8.29 | 7.29 |
| **Heterogeneous setting** | 7.54 | 8.27 | 8.20 | 7.81 |
| **All settings** | 6.59 | 7.94 | 8.37 | |

4.7

However, some of the homogeneous settings produce very good results, and the human settings are not in line with the overall tendency. These exceptional cases can at least partially be explained by the results of the test of hypotheses two and three (see below). The good performance of some of the homogeneous settings points to the fact that the variety of social types involved in heterogeneous settings enables a lot of opportunities for successful exchange, but this does not implicate that homogenous settings inhibit exchange per se. A few social types seem to perform better while others tend to block each other. This is presumably an effect of the more or less restrictive preferences of the different social types. In the case of the few human settings just those social types were involved, which later turned out to perform good in negotiations. Our second hypothesis strengthens this suspicion.

4.8

As a specification of the first, the second hypothesis was formulated as follows: The pattern of distribution is the essential factor in shift-plan quality, independent of the individual social types involved. The experimental results clearly disprove this hypothesis, as Table 4 shows. The values in this table are the mean of all settings in which one of our social types was involved (excluded are heterogeneous settings because these by definition consist of all social types). There seems to be a ranking with respect to the individual social types involved, and especially for the group settings the values differ significantly. This means that the impact of the individual features of social types is stronger than the influence of patterns of types, as the hypothesis would demand. This can be interpreted to mean that we were in some sense too successful in designing and implementing role-based agent behaviour.

**Table 4:** Test of hypothesis 2

| | **Homogeneous setting** | **Group setting** | **Outsider setting** | **All** |
|---|---|---|---|---|
| **Self-confident type** | - | 5.51 | 7.29 | 7.02 |
| **Uncooperative type** | - | 6.41 | 7.16 | 7.02 |
| **Agreement-oriented type** | - | 6.41 | 7.16 | 7.02 |
| **Family type** | 5.20 | 8.86 | 7.10 | 7.24 |
| **Pleasure type** | 6.65 | 8.19 | 7.79 | 7.73 |
| **Team type** | - | 8.86 | 7.50 | 7.94 |

**4.9**

The third hypothesis directly addresses the question of hybrid communities: Shift-plan quality differs depending on the nature of the negotiation partners - that is, if they are negotiated by only agents, only humans or are mixed. The experimental results show a clear tendency towards proof of this hypothesis (see, again, Table 3, especially the figures marked in green). However, we must confess that the ranking is not as we expected it to be because human settings produce the best values. This is an irritating result because from the underlying assumption of a productive use of difference it follows that hybrid settings should be the most productive.

**4.10**

There are some possible explanations for this finding. From the observation of the probands we have some hints for three specifically human properties that enabled them to perform better than the agents. First, it was obvious that humans are capable of a flexible interpretation of the roles prescribed to them. For humans, the situation at hand includes some scope for role-taking, whereas the agents are forced to deterministic role behaviour by the programming of the system. Second, humans are capable of a flexible way of decision making, which is not possible for the agents. For example, it was obvious that there is no equivalent to the numerically determined acceptance- and cancel-lines of the agents for the behaviour of the probands. And the humans are capable to develop and test different negotiation tactics and to deploy them as routines. For example, the probands tried to negotiate with specific partners despite of the issues, or they simply waited for promising offers. Tactics like these are not possible for the agents because IPS forces them to a specific succession of calculation steps. Third, humans are capable of systematically exploring patterns of behaviour of their negotiation partners and of typical reactions of the technical system, and they systematically try to learn from these experiences, especially in the long run (spanning over many settings). At the present stage of development of the INKA-system, exploration behaviour and learning are not possible for the agents. The implementation of routine behaviour and learning are the next working step on our project agenda.

**4.11**

These advantages of the humans can partly explain the irritating result that human settings performed best. But this interpretation would be one-sided without remembering that we had to restrict one of the biggest advantages of the agents for experimental reasons. We had to artificially slow the agents down in order to keep them in step with the humans. Otherwise it would have been easy for the probands to uncover the agents. It remains future work to find ways for the construction (and examination) of hybrid communities in which the specific strengths of humans and agents can unfold their potential advantages in a complementary way.

**Methodological lessons learned**

**4.12**

These overall results from the first experiments should be interpreted as more or less strong tendencies, and not as total proof or disproof of the hypotheses. This 'weakening' of the status of the results is not only necessary from the perspective of the concrete values produced (see discussion above), but also a consequence of some of the methodological problems we faced. These problems are severe in the context of a strong experimental claim because they indicate that there are other sources of influence on the dependent variable (the overall score of the shift-plan) than the independent variable (the distribution of social types) which we altered in the different settings. However, as the first experiment was an initial trial, these problems point to fundamental issues in human-agent interactivity. Moreover, we think that these experiences can be beneficial for related methodological approaches, such as participatory simulation or human-agent teamwork.

**4.13**

The first problem arises from the humans involved in the role-playing game. Even within the very restricted scope of proband behaviour (experimental set-up and interface) the conduct of the experiment depends on the goodwill of the probands and their ability to keep up the social roles given them by the experimenters. It turned out that in fact they did not always follow the prescribed roles; from this experience we must acknowledge that role-playing scenarios are not an easy complement to simulations. On the other hand, in our approach to directly protocolling the interactivity we are able to reconstruct single events of role-playing errors and to eliminate defective settings from the data used.

4.14 The second problem arises from the statistics used to compute the overall score of the settings. This of course must be taken into account whenever interactions are to be measured or compared. However, in our case it turned out that every step of the aggregation procedure strongly influences the results. For example, at the beginning we thought that it would be a good idea to introduce a variation equation into the collective aggregates in order to make the aggregate sensitive to extreme deviation of individual values. However, this formalism produced purely statistical effects which have nothing to do with the reality of the events observed. Statistical procedures of this kind generally bear this danger, and moreover tend to make interpretation more difficult. One way of facing this problem is to keep statistics simple (at least to keep the possibility of interpretation open); another way would be to apply and compare different aggregation principles.

4.15 The third and most serious problem is statistical spread. This means, that exactly the same settings produce different values in different runs. This holds true for human, hybrid and – as we analysed in simulation runs – agent settings. The dispersion can be damped by strictly avoiding variation equations (statistical effects) to some degree, but it surely weakens the significance of experimental proofs or disproofs of hypotheses. The problem of statistical spread raises strong doubts about methodological claims like the following: 'MAS models make it possible to conduct experiments with fully repeatable and controllable scenarios' (Barreteau et al. 2001). On the contrary, spread and path-dependency from early interactions seem to be an inherent feature of interaction systems like MAS and have to be seen as advantageous in terms of coordination capability and adaptability. Nonetheless, the strong experimental claim makes it possible to draw conclusions about tendencies (see above), and additional simulations can serve as a tool for analysing statistical spread in different types of setting.

### Checking the experimental data against simulations

4.16 The simulation system (see 3.16) makes it possible to produce negotiation results for a large number of runs of each setting. A detailed overview of our simulation's results can be found in the Appendix (see Table 8). In order to relate these data to experimental results, we had to ensure that the simulation system produce data comparable to data from interactivity experiments. We tested this functional equivalence of both systems by comparing multiple step-by-step simulations. Remaining minor differences, which do not affect the comparability of data produced by both systems, were due to factors beyond our control, such as the message distribution over the network or the activation sequence of the behaviours incorporated in our agents.

4.17 For us, the most interesting result of the simulations is the range of statistical spread of the overall score. Checking experimental data against this spread offers a heuristic tool to evaluate experimental data. To emphasise the possibilities of such a heuristic we again choose as examples a homogeneous (Figure 7) and a heterogeneous (Figure 8) setting. Both figures show the course of the overall score for a human and a hybrid experimental run of the same setting. Additionally, the spread of the overall score of 500 simulated runs is shown, depicting maximum, minimum and average scores over a running time of 170 steps.[10]

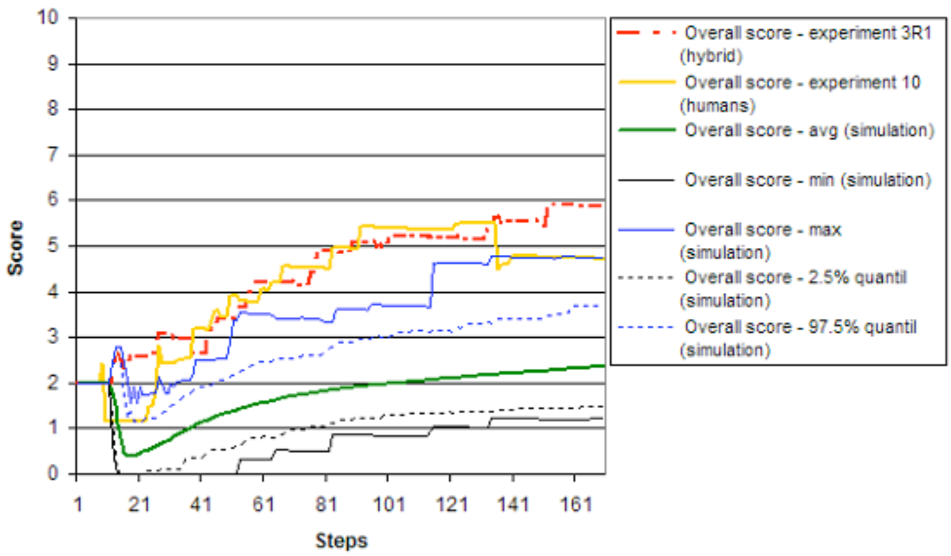



**Figure 7**. Overall score and statistical spread of homogenous settings (family type)

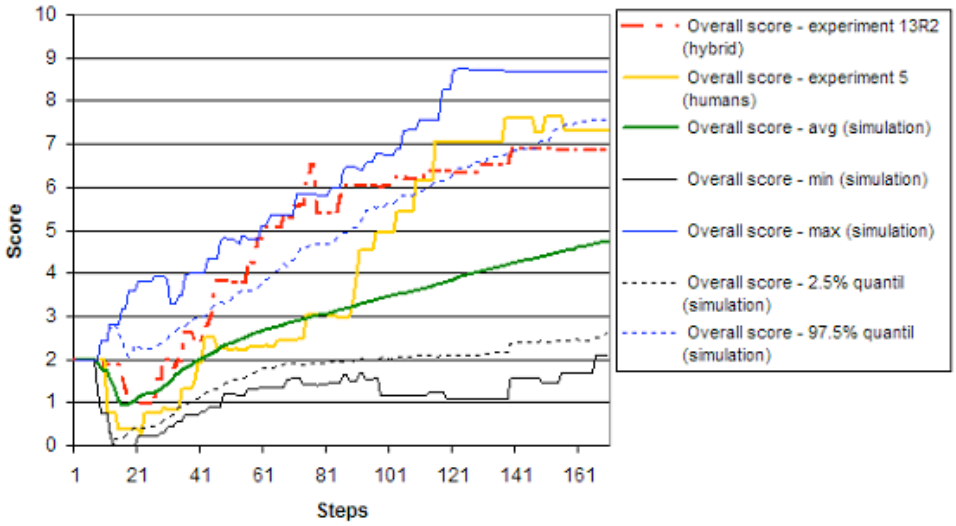



**Figure 8**. Overall score and statistical spread of heterogeneous settings

**4.18**

The more erratic course of the experimental graphs is a characteristic of single runs, in contrast to the statistically flattened simulation graphs. All graphs show the typical features of all negotiation settings (see 3.16): an early descent, followed by a more or less rapid rise, and a levelling off towards the end. Despite this similarity, all experimental graphs are located near the simulated runs' high scores. In the case of the homogenous setting, experimental runs even exceed the maximum scores of the simulations. As argued in 4.16, this effect can be explained by the negotiation capabilities of humans. Therefore we interpret the divergence of experimental and simulative results as additional confirmation of our third hypothesis. Of course, this is a 'soft' interpretation that should not be considered valid in an elaborated statistical sense. Nevertheless, simulations can give strong hints about the status of experimental runs.

## Conclusions

**5.1**

We have presented the consecutive steps of an integrated approach which 'incorporate' human sociality into computer agents, and then examined the re-entry of these social agents into human social contexts. We have argued that modelling agents on social roles not only enhances the coordination capacities of MAS, but, considering collective expectations, can also foster interaction of humans and agents. We introduced the interactivity experiment as a method for examining human-agent interaction and discussed some of the results of the initial experiment. We also pointed to some of the methodological problems - such as statistical spread and path dependency - that in our view are inevitable when interaction systems like MAS are examined. Nonetheless, we showed that it is possible to draw conclusions about distinctive tendencies. Moreover, we hope that this methodological approach can be used as a generic instrument for examining those mixed realities that will become more common if distributed systems leave the laboratories, or other kinds of direct interaction between humans and intelligent machines.

---

## Notes

[1] The acronym INKA stands for 'Integration of Cooperating Agents in Complex Organisations'. The project is carried out by the Artificial Intelligence Group of the Department of Computer Science at Humboldt-Universität zu Berlin and the Institute for Sociology, Technology Studies, at the Berlin Technical University.

[2] We do not develop a better tool for roster construction, but the problem of negotiating the exchange of work shifts is taken as a real-life problem for investigating systems composed of artificial and human actors.

[3] See http://www.pcs.usp.br/~mabs/ and http://boid.info/CoOrg06.

[4] These are especially defined by restrictions of the formal role.

[5] The implementation of experiences and how they are considered in selecting future interaction partners is in fact a two-layered reinforcement mechanisms, one layer is directed at the individual agent while the second one is directed at the more general social type. This approach is strongly related to the work on trust (see for instance Gambetta, 1999). Because the specific partner selection method is not in the focus in the article we do not discuss this in detail here. Our two-layered approach to reinforcement learning and trust implies that similar agents, e.g. same social type, are more likely to be correlated regarding trust. This view is supported by Ziegler and Lausen 2004.

[6] Compared to multi-agent-based simulation tools, for example, SWARM (http://www.swarm.org), the JADE platform is not suited for large reproducible simulations but for really concurrent distributed MAS.

[7] ISO-NORM 9241-10: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 10: Dialogue principles: International Organization for Standardization.

[8] The INKA System can be used for small simulations by running artificial setting (consisting of agents only). Such

simulations have been used for checking potential problems and pitfalls before the experiment with humans, and in order to fine-tune the experimental set-up.

[9] The set-up described was evaluated in a pre-test on a smaller scale.

[10] Exploratory simulations have shown that 500 runs per setting produce a good significance level for estimating means and variances of simulation results.

## Appendix

**Table 5:** Application of social types within the agents' decision processes

| Social type | Weight of leisure time interests $w_{LTI}$ | Default leisure time preferences | | | Weight of capital accumulation $w_{CA}$ | Capital interests | | | | Acceptance line | | Cancel line | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Early shift | Late shift | Night shift | | Economic $c_1$ | Cultural $c_2$ | Social $c_3$ | Symbolic $c_4$ | Start below best in % | Slope in % | Start below worst in % | Slope in % |
| **Family type** | -132,8 | 0 | 70 | 0 | 1 | 80 | 50 | 50 | 20 | 0 | -5 | 30 | 10 |
| **Team type** | -132,8 | 0 | 0 | 0 | 1 | 20 | 50 | 80 | 80 | 25 | -20 | 30 | 5 |
| **Uncooperative type** | -132,8 | 0 | 0 | 0 | 1 | 50 | 80 | 20 | 50 | 0 | -5 | 5 | 20 |
| **Self-confident type** | -132,8 | 0 | 0 | 0 | 1 | 50 | 80 | 20 | 50 | 0 | -5 | 30 | 20 |
| **Agreement-oriented type** | -132,8 | 0 | 0 | 0 | 1 | 20 | 50 | 50 | 20 | 50 | -20 | 55 | 5 |
| **Pleasure type** | -132,8 | 70 | 0 | 70 | 1 | 80 | 20 | 50 | 20 | 25 | -10 | 30 | 10 |

Note: For the application of social types within the agents' decision processes we had to transform the profiles of the social types given in Table 1 into concrete parameters and numerical values, as presented in this table. The *preference* and hence the *non-preference of certain shift types* is transformed into default leisure-time interests (0 resp. 70). Individual leisure-time interests override the default values. Notice that default leisure-time interests will not cause a negotiation, as they are not considered significant (see Table **7**). However, they may change the utility of an exchange. The capital interests result from the *preferred* (set to 80) and *irrelevant forms of capital* (set to 20) of the social type profile. Not explicitly mentioned forms of capital are set to 50. The weights are currently equal for all social types. They have been set so that a shift with maximum leisure-time interest (100) results in any case in a negative shift utility. We derive the parameters that define the strategic lines from the *willingness for compromise* (agree line) and from the *willingness to negotiate* (cancel line). All parameters are percentages of the difference between the utility of the best and the worst exchange. The greater the willingness to compromise, the lower the agree line starts and the steeper is its decrease. The smaller the willingness to negotiate, the higher the agree line starts and the steeper is its increase.

**Table 6:** Typical capital accumulation of the different shift types

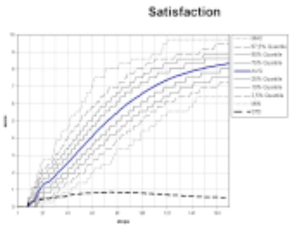
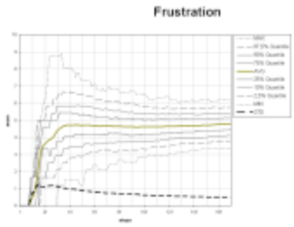
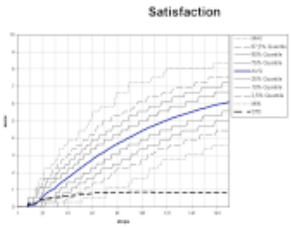
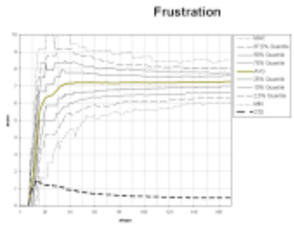
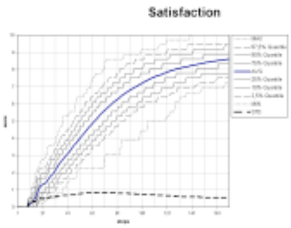
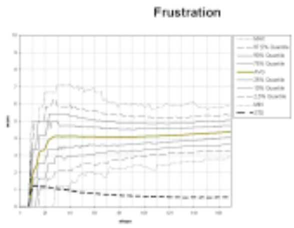
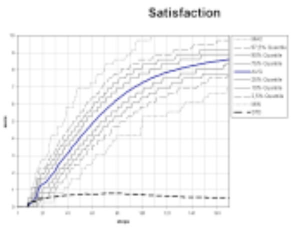
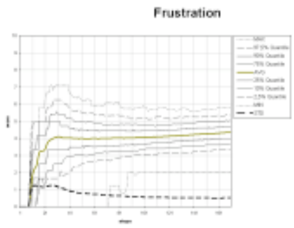
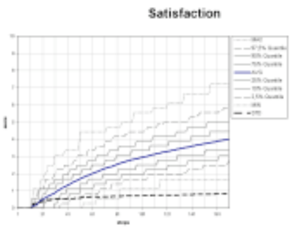
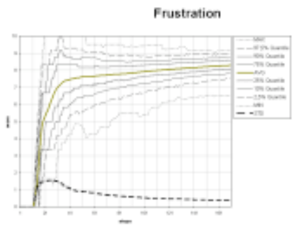
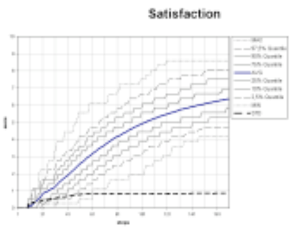
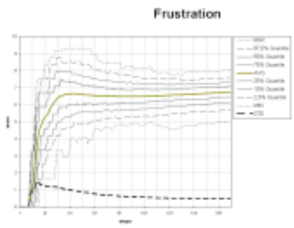
| Shift type | Economic capital | Cultural capital | Social capital | Symbolic capital |
|---|---|---|---|---|
| **Early shift weekday** | 0 | 40 | 20 | 20 |
| **Late shift weekday** | 10 | 40 | 10 | 22 |
| **Night shift weekday** | 20 | 20 | 0 | 16 |
| **Early shift weekend** | 50 | 40 | 13 | 39 |
| **Late shift weekend** | 50 | 20 | 10 | 30 |
| **Night shift weekend** | 100 | 20 | 0 | 48 |

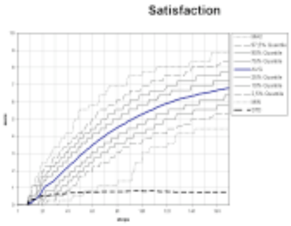
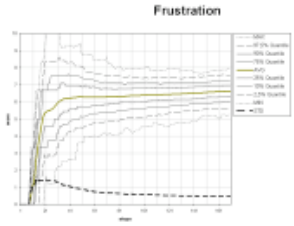
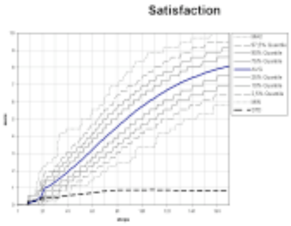
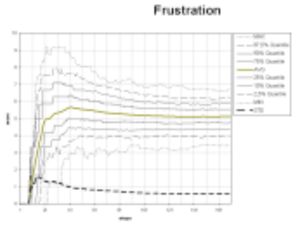
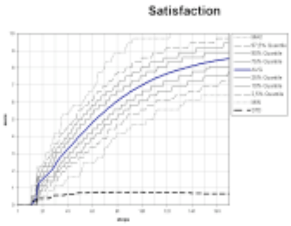
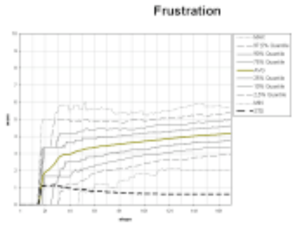
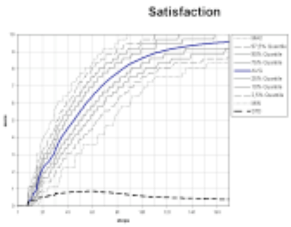
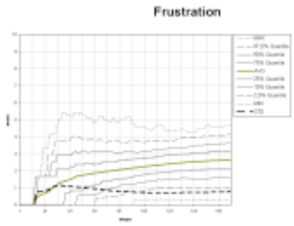
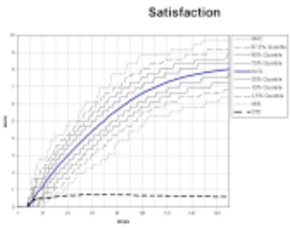
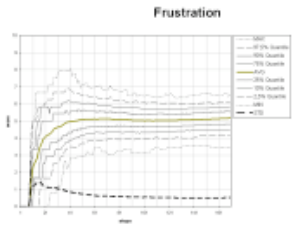
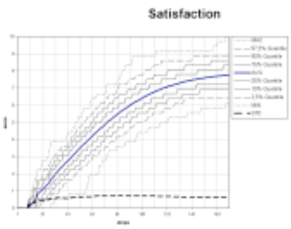
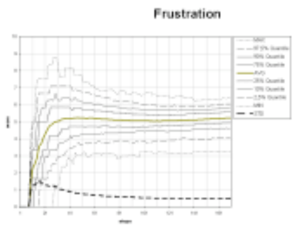
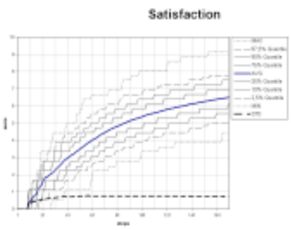
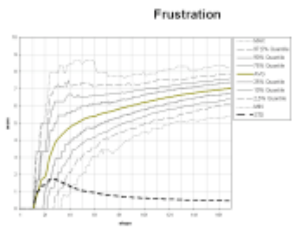
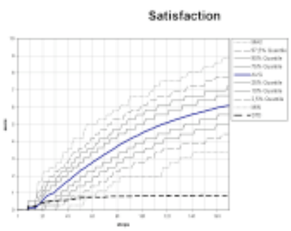
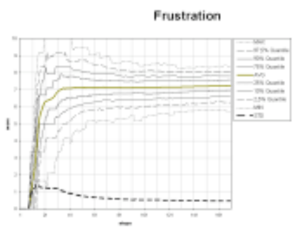
**Table 7:** Global parameters

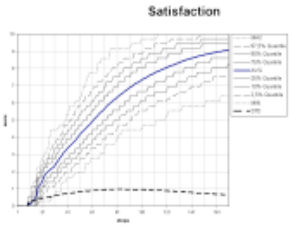
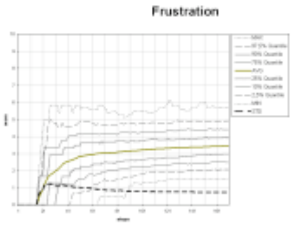
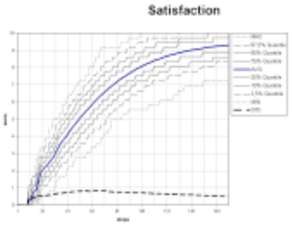
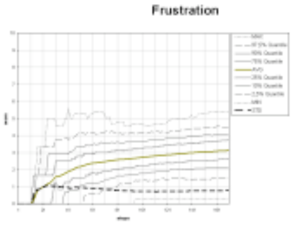
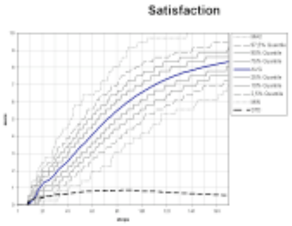
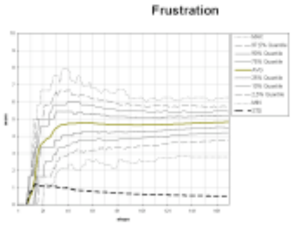
| Parameter | value / range |
|---|---|
| Significance thresholds for LTI | 75 |
| Magnitude of experience change $step_{exp}$ | 0.05 |
| Initial experience values | 0.5 |
| Time steps for which a certain issue with a certain partner is impossible after a | 30-50 |

| | |
|---|---|
| negotiation | |
| Time steps for which an issue is impossible if no partner can be found or any issue is impossible if no issue can be found at all | 3–7 |
| Time steps to wait before the next negotiation starts | 3–7 |
| Time steps to wait for the partner's reaction before a timeout is sent | 15 |
| Number of consecutive identical proposals before the booster is activated | 2–4 |

**Table 8:** Evaluation of the Simulation Data

| Setting | Setting class | Configuration (self-confident type, uncooperative type, agreement-oriented type, pleasure type, pleasure type, family type) | Data (click on figures to enlarge them) | |
|---|---|---|---|---|
| 0 | heterogenous | (2,2,2,2,2,2) using shiftplan #1 |  |  |
| 1 | group | (4,4,4,0,0,0) using shiftplan #2 |  |  |
| 2 | group | (0,4,4,4,0,0) using shiftplan #3 |  |  |
| 3 | outsider | (1,3,4,4,0,0) using shiftplan #3 |  |  |
| 4 | homogenous | (0,0,0,0,0,12) using shiftplan #4 |  |  |
| 5 | outsider | (4,3,4,0,0,1) using shiftplan #1 |  |  |

| | | | | |
|---|---|---|---|---|
| 6 | outsider | (4,3,4,0,0,1) using shiftplan #1 *(similar to setting 5, but individual positions in the shiftplan differ)* |   |   |
| 7 | heterogenous | (2,2,2,2,2,2) using shiftplan #2 |   |   |
| 8 | outsider | (1,0,0,3,4,4) using shiftplan #3 |   |   |
| 9 | group | (0,0,0,4,4,4) using shiftplan #4 |   |   |
| 10 | outsider | (0,4,4,0,3,1) using shiftplan #1 |   |   |
| 11 | outsider | (0,4,4,0,3,1) using shiftplan #1 *(similar to setting 10, but individual positions in the shiftplan differ)* |   |   |
| 12 | homogenous | (0,0,0,12,0,0) using shiftplan #2 |   |   |
| 13 | group | (4,4,4,0,0,0) using shiftplan #2 |   |   |

| | | | | |
|---|---|---|---|---|
| 14 | group | (0,0,0,4,4,4) using shiftplan #4 *(similar to setting 9, but individual positions in the shiftplan differ)* |  |  |
| 15 | outsider | (1,0,0,3,4,4) using shiftplan #3 *(similar to setting 8, but individual positions in the shiftplan differ)* |  |  |
| 16 | heterogenous | (2,2,2,2,2,2) using shiftplan #1 |  |  |

Note: The data has been retrieved from simulations of settings according to the original composition of the Interactivity experiment. Some configurations seem to be equal (regarding shift plan and composition of social types), though there are differences in the individual positions in the shiftplan in order to eliminate coincidences. Every data set in this summary evaluation is based upon 500 independently simulated models of the specified configuration, showing the impact of chance on the score results and its underlying indicators.

## References


ALDRICH H (1999), *Organizations Evolving*, London: Sage Publications.

ALMOG O (1998), The problem of social type: A review, *Electronic Journal of Sociology,* vol. 3, no. 4 http://www.sociology.org/content/vol003.004/almog.html (12.11.2003).

ARONSON E, WILSON TD, and BREWER MB (1998), Experimentation in Social Psychology, in Gilbert DT, Fiske ST, and Lindzey G (eds), *The Handbook of Social Psychology*, 4th edition, Boston, MA: McGraw-Hill, 99-142.

BARRETEAU O, BOUSQUET F, and ATTONATY J-M (2001), Role-Playing Games for Opening the Black Box of Multi-Agent Systems: Method and Lessons of its Application to Senegal River Valley Irrigated Systems, *Journal of Artificial Societies and Social Simulation,* vol. 4, no. 2 http://jasss.soc.surrey.ac.uk/4/2/5.html.

BERG M, and TOUSSAINT P (2003), The mantra of modelling and the forgotten powers of paper: A sociotechnical view on the development of process-oriented ICT in health care, *International Journal of Medical Informatics*, vol. 66, no. 2-3, 223-234.

BOURDIEU P (1986), The (three) Forms of Capital, in Richardson JG (ed.), *Handbook of Theory and Research in the Sociology of Education*, New York and London: Greenwood Press, 241-58.

BOWKER GC, STAR SL, TURNER W, and GASSER L (eds.) (1997), *Social Science, Technical Systems, and Cooperative Work. Beyond the Great Divide*, Hillsdale, NY: Lawrence Earlbaum.

BRADSHAW JM, SIERHUIS M, ACQUISTI A, FELTOVICH P, HOFFMAN R, JEFFERS R, PRESCOTT D, SURI N, USZO A, and VAN HOOF R (2003), Adjustable Autonomy and Human-Agent Teamwork in Practice: an Interim Report on Space Applications, in Hexmoor H, Castelfranchi C, and Falcone R (eds), *Agent Autonomy*, Kluwer, 243-80.

CAMPBELL DT (1988), Factors Relevant to the Validity of Experiments in Social Settings, in Overman ES (ed.), *Methodology and Epistemology for Social Science. Selected Papers*, Chicago and London: University of Chicago Press, 151-66.

CAMPBELL DT and RUSSO MJ (1999), *Social Experimentation*, London: Sage.

CAMPBELL DT and STANLEY J (1963), *Experimental and Quasi-Experimental Design for Research*, Chicago: Rand McNally.

CASTELFRANCHI C (2003), Formalising the informal? Dynamic social order, bottom-up social control, and spontaneous normative relations. *Journal of Applied Logic*, vol. 1, no. 1-2, 47-92.

COHEN MD, MARCH JG, and OLSEN JP (1972), A Garbage Can Model of Organizational Choice, *Administrative Science Quarterly,* vol. 17, no. 1, 1-25.

DAHRENDORF R (1973), *Homo Sociologicus*; London: Routledge and Kegan Paul.

ESSER H (1999), *Soziologie – Spezielle Grundlagen. Band 1: Situationslogik und Handeln*, Frankfurt a.M.: Campus.

FATIMA SS, WOOLDRIDGE M, and JENNINGS NR (2003), Optimal Agendas for Multi-Issue Negotiation, in *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'03)*, ACM, 129-36.

FERBER J (1999), *Multi-Agent Systems. Introduction to Distributed Artificial Intelligence*; Reading, MA: Addison Wesley.

FISCHER K, MÜLLER JP, and PISCHEL M (1998), A Pragmatic BDI architecture, in Huhns MN and Singh MP (eds), *Readings in Agents*, Morgan Kaufmann, 217-24.

GAMBETTA D (1998), Can we Trust Trust?, in Gambetta D (ed.), *Trust: Making and Breaking Cooperative Relations*; Oxford: Blackwell, 213-37.

GROSSI D, DIGNUM F, ROYAKKERS L, and DASTANI M (2005), Foundations of Organizational Structures in Multiagent Systems, in *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'05)*, ACM, 690-698.

GUYOT P and DROGOUL A (2004), Designing Multi-Agent based Participatory Simulations, in Coelho H and Epinasse B (eds), *5th Workshop on Agent-Based Simulations*, Lisbon: Publishing House, 32-37.

HUHNS MN and STEPHENS LM (1999), "Multiagent Systems and Societies of Agents". In Weiss G (ed.), *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, MIT Press, 79-120.

JENNINGS NR, FARATIN P, LOMUSCIO AR, PARSONS S, WOOLDRIDGE M, and SIERRA C (2001), Automated Negotiations: Prospects, Methods and Challenges, *Group Decision and Negotiation*, 10(2), 199-215.

KORTENKAMP D and FREED M (eds) (2003), *Human Interaction with Autonomous Systems in Complex Environments*, AAAI Spring Symposium Technical Report SS-03-04, Palo Alto: AAAI Press.

KURBEL K and LOUTCHKO I (2002), Multi-Agent Negotiation under Time Constraints on an Agent-Based Marketplace for Personnel Acquisition, in *Proceedings of the 3rd International Symposium on Multi-Agent Systems, Large Complex Systems, and E-Business (MALCEB 2002)*, 566-79.

MARCH JG and SIMON HA (1958), *Organizations*, New York: Wiley.

MEISTER M, URBIG D, GERSTL R, LETTKEMANN E, OSHERENKO A, and SCHRÖTER K (2002), *Die Modellierung praktischer Rollen für Verhandlungssysteme in Organisationen. Wie die Komplexität von Multiagentensystemen durch Rollenkonzeptionen erhöht werden kann*, Berlin: Technische Universität Berlin, Technology Studies TUTS-WP-6-2002, http://www.tu-berlin.de/~soziologie/Tuts/Wp/TUTS_WP_6_2002.pdf.

MEISTER M, URBIG D, SCHRÖTER K, and GERSTL R (2005), Agents Enacting Social Roles. Balancing Formal Structure and Practical Rationality in MAS design, in Fischer K and Florian M (eds), *Socionics*, Berlin, Springer: 104-131.

MEYER JW and ROWAN B (1977), Institutionalized Organizations: Formal Structure as Myth and Ceremony, *American Journal of Sociology,* 83, 340-63.

NICKLES M, ROVATSOS M, SCHMITT M, BRAUER W, FISCHER F, MALSCH T, PAETOW K, and WEISS G (2007), The Empirical Semantics Approach to Communication Structure Learning and Usage: Individualistic vs. Systemic Views. *Journal of Artificial Societies and Social Simulation*, vol. 10, no. 1, http://jasss.soc.surrey.ac.uk/10/1/5.html.

POPITZ H (1967), *Der Begriff der sozialen Rolle als Element der soziologischen Theorie*, Tübingen: Mohr (Paul Siebeck).

RAMANATH AM and GILBERT N (2004), The Design of Participatory Agent-Based Social Simulations, *Journal of Artificial Societies and Social Simulation,*, vol. 7, no. 4, http://jasss.soc.surrey.ac.uk/7/4/1.html.

RAMMERT W (1998), Giddens und die Gesellschaft der Heinzelmännchen. Zur Soziologie technischer Agenten und Systemen verteilter künstlicher Intelligenz, in Malsch T (ed.), *Sozionik. Soziologische Ansichten über künstliche Sozialität*, Berlin: Sigma, 91-128.

RAMMERT W (2003), Technik in Aktion: Verteiltes Handeln in soziotechnischen Konstellationen, in Christaller T and Wehner J (eds), *Autonome Maschinen*, Frankfurt a.M.: Campus, 289-315.

RAMMERT W and SCHULZ-SCHAEFFER I (eds) (2002), *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik* Frankfurt a.M.: Campus.

RAO AS and GEORGEFF MP (1991), Modeling Rational Agents within a BDI Architecture, in Fikes R and Sandewall E (eds), *Proceedings of Knowledge Representation and Reasoning (KRR-91)*, Kaufmann, 473-84.

SCHRÖTER K and URBIG D (2004), C-IPS: Specifying Decision Interdependencies in Negotiations, in Lindeman G et al. (eds), *Proceedings of the Second German Conference on Multiagent System Technologies (MATES 2004)*, Springer, 114-25.

SIERHUIS M, VAN HOOF R, CLANCEY WJ, and SCOTT M (2001), From Work Practice Models and Simulation to Implementation of Human-Centered Agent Systems in *Proceedings of ASCW01, Workshop on Agent Supported Cooperative Work,* Montreal, Canada, May 2001, http://www.agentisolutions.com/documentation/papers/Workshop_on_%20AgentSupportedCooperativeWork.pdf, accessed 02.09.2002.

SUCHMAN L (1987), *Plans and Situated Actions. The Problem of Man-Machine Communication*, Cambridge, UK: Cambridge University Press.

TURNER RH (1962), Role-Taking: Process versus Conformity, in Rose AM (ed.), *Human Behavior and Social Process. An Interactionist Perspective*, London: Routledge, 20-40.

URBIG D, MONETT DIAZ D, and SCHRÖTER K (2003), The C-IPS Architecture for Modelling Negotiating Social Agents, in Schillo M et al. (eds), *Proceedings of the First German Conference on Multiagent System Technologies (MATES 2003)*, Springer, 217-28.

URBIG D and SCHRÖTER K (2005), Negotiating Agents: from full Autonomy to Degrees of Delegation, to appear in *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS 2005)*, Utrecht, 25-29 July 2005.

ZIEGLER C-N and LAUSEN G (2004), Analyzing Correlation between Trust and User Similarity in Online Communities. In *Proceedings of the 2nd International Conference on Trust Management*, Volume 2995 of LNCS, 251-265.

ZIMMERMANN E (1972), *Das Experiment in den Sozialwissenschaften*, Stuttgart: Teubner.