Science, Technology & Innovation Studies
Vol. 10, No. 1, January 2014

ISSN: 1861-3675

STI Studies
www.sti-studies.de

Interaction of Human Actors and Non-Human Agents

A Sociological Simulation Model of Hybrid Systems

Robin D. Fink (TU Dortmund University, robin.fink@tu-dortmund.de)

Johannes Weyer (TU Dortmund University, johannes.weyer@tu-dortmund.de)

Abstract

Despite comprehensive research on sociable robotics in different disciplines, sociological theory of action so far has almost completely disregarded the issues of agency of technology and of human-machine interaction and left the field to human factors research or to novel approaches such as the Actor Network Theory (ANT). The following paper links research on human-machine interaction to sociological theory of action and proposes a method to investigate these issues experimentally.

First, it sketches a sociological sound model, which describes the "co-action" of technology in a way that allows investigating the question of non-human agency empirically. Bruno Latour's provocative argument of symmetry of humans and nonhumans is taken as a starting point to show that a sociological theory of action, based on Hartmut Esser's model of sociological explanation (MSE), is also capable to cope with non-human agency.

In order to better understand the interaction of human actors and non-human agents in highly automated systems, we therefore construct a model of sociological explanation of hybrid systems (HMSE), which treats both parts of the system as deciders, who act according to the principle of subjective expected utility (SEU). The overall behaviour of the hybrid system thus can be modelled as the aggregated result of the actions of both parts.

The data from experiments with an agent-based computer simulation, implemented on the basis of the HMSE, show that human test persons indeed attribute agency to the technical systems. Additionally, they describe the relation of human and machine as symmetrical. Finally, we discovered that test persons also tended to attribute responsibility for the achievement of certain goals to the technical system – although the experimental setup implied equally distributed responsibility among humans and nonhumans.

The HMSE can help to gain new insights into the interplay of humans and nonhumans and provide a deeper understanding of this kind of hybrid interaction, grounded on a sociological theory of action.

1 Introduction

Autonomous technical systems, such as software agents or robots, present a challenge to sociology, because they raise the issue of agency of technology (Rammert/Schulz-Schaeffer 2002). Most sociological theories, however, are not able to deal with this question, since they grant the status of an actor exclusively to humans. It is ascribed to human actors only to act intentionally and to interact with others. This way they produce effects that may be relevant for society as a whole (Parsons 1967, Coleman 1990).

Modern societies, however, are increasingly shaped by objects that perform actions, which formerly have been executed by humans. For example, the automatic spam filter deletes harmful mails without intervention of the user. The autopilot controls the aircraft precisely and safely from take-off to landing. Regarding the resulting effects, it is hard to distinguish whether these effects have been accomplished by smart systems or by humans. Smart, autonomous systems seem to be capable to act almost human-like. Modern planes or cars thus have to be regarded as hybrid systems, where agency is distributed among humans and nonhumans who act and interact in a way that is only partly understood in terms of sociological theory.

Additionally, new generations of robots will operate in environments shared with people, such as museums or hospitals (Breazeal 2004b). These robots will be equipped with advanced capabilities of social interaction (Breazeal 2004a), provoking questions of social intelligence and socially acceptable behaviour of robots (Huettenrauch et al. 2006, Turkle 2006).

Research on human-machine interaction has brought about important results for example on trust in automation, overreliance, and situational awareness especially in highly automated systems (Lee/See 2004, Sheridan 1999, Parasuraman et al. 2008, Grote 2009). Research on human-robot interaction has pointed to the fact that human-robot cooperation requires treating your counterpart as a partner – seen both from the perspective of the human and the robot (Breazeal 2004b). As the CASA approach (computers as social actors) argues, people interacting with computers "engage in the same kinds of social responses that they use with humans" (Takayama/Nass 2008: 174).

Although the practical use of this research cannot be disputed, from our point of view a *theoretical* foundation of interaction models, applied in automation research or research on sociable robots, is still missing. We suppose that a deeper understanding of the mechanisms of interaction between humans and autonomous technology from a sociological perspective may help to gain new insights about the functioning of smart systems.

In the paper at hand we will sketch a sociological model, which describes the co-operation of autonomous technology, and thus might allow us to analyse the issue of agency of technology empirically. This pragmatic approach frequently meets critique of people who argue that humans are unique and are exclusively able to act intentionally - contrary to animals, objects or even robots (Sturma 2001). In order to avoid fundamentalist debates on such ontological issues, we refer to Lucy Suchman, who in the second edition of "Plans and situated actions" – contrary to previous work – calls for a reorientation of the debate on "nonhuman agency", which should "be reframed from categorical debates to empirical investigations of the concrete practices" (Suchman 2007: 1). It is no longer important, "whether humans and machines are the same or different" (ibid.: 2), but how these categories and differences are used in practice. Additionally, experiments conducted by the CASA group have shown that human-computer interaction works “in much the same way” (Takayama/Nass 2008: 175) as human-human interaction (Reeves/Nass 1996).

In terms of this shifting perspective we have developed a model of sociological explanation of hybrid systems (HMSE) grounded on Hartmut Esser’s macro-micro-macro model of sociological explanation (MSE) which makes use of subjective expected utility (SEU) on the micro level (for further details on Esser’s approach see the excursus in section 3.1). We then implemented this model as a computer simulation that allows us to perform interactive experiments and to observe the issue of distributed agency empirically.

2 State-of-the-art

Despite the remarkable disinterest of sociological theoreticians there is a long tradition of sociological research on interaction of humans and technology.

Sherry Turkle: Computer Cultures

For example Sherry Turkle has analysed computer cultures by means of ethnographic methods. She studied real processes of interaction of younger people and computers and of elder people and pets such as the robot dog AIBO (Turkle 2005, 2006, Turkle et al. 2006). She didn't reflect that much about the issue of "whether", but took interaction as self-evident and concentrated on the repercussions of human-computer interaction on the respective persons. Even today her publications are a valuable source for psychoanalytic and cultural theoretic studies. However, her approach does not provide us with options for a deeper theoretical analysis of human-computer interaction.

Lucy Suchman: Workplace studies

Lucy Suchman, one of the founders of workplace studies, has analysed - also by means of ethnographic methods - "the ways people use technologies to accomplish and coordinate their day-to-day practical activities" (Luff et al. 2000a: 12). She focuses on "the contingent and situated character of practical action" (ibid.: 13). However, in her view machines are inferior to humans, since they have fundamental shortcomings. She states "radical asymmetries" (Suchman 2007: 5) of humans and machines, which are rooted in "severe limitations" (Suchman et al. 1999: 395) of the machine. Consequently she claims that "the analysis of everyday human conversation provides a baseline from which to assess the state of interactivity between people and machines" (Suchman 2007: 178), thus making human action the benchmark for assessing nonhuman action.

Although workplace studies have generated valuable insights into the everyday practices of dealing with technology, the thesis of lacking machine capabilities obstructs the view for an unbiased analysis of the interaction of men and autonomous technology.

Bruno Latour: Nonhuman Actors

The actor network theory, developed by Bruno Latour, Michel Callon and others, takes a very different perspective. In contrast to Suchman, Latour presents a radically symmetrical ontology, which does not accept any presupposed distinctions between human actors and nonhuman actants, since both of them are able to bring about changes (Latour 1988, 1996, 1998). A human may close the door, but the automatic door-closer can do this as well, thus translating the human who wants to enter the house. By means of different translations a network emerges, consisting of human actors and nonhuman actants. Latour thus tries to overcome the traditional divide between the technical and the social realm and to establish a symmetrical perspective, which allows to catch processes of hybridisation. For example by mutual translation of a human (e.g. a citizen) and a technical device (e.g. a handgun) a new hybrid actor emerges, the citizen-gun or the gun-citizen, who finally commits murder, which none of the singular parts could have done alone (Latour 1998: 34).

Although some of the instances chosen by Latour to present his new approach seem to be rather bizarre, the basic question has to be taken seriously, who makes a phone call (the user, the telephone or both of them together) or who sends an e-mail (the user, the computer or both of them together). There has been an intense debate in science and technology studies for years, heavily criticising or defending actor network theory (for an overview cf. Gad/Bruun Jensen 2010). Instead of summing up this debate, we want to point to the fact that most contributions were rather theoretical – and - empirical studies on the question of symmetry are still rare. Latour himself has only presented ad hoc cases e.g. on key fobs which do not meet methodical standards. Additionally, these cases are not related to smart technology but as a rule to conventional technology such as keys or door-closers.

Werner Rammert and Ingo Schulz-Schaeffer: Attribution Processes

In contrast to the ontological perspective of ANT, Werner Rammert and Ingo Schulz-Schaeffer propose to "treat the question of agency of technology as empirically open" (2002: 50). According to Rammert and Schulz-Schaeffer people attribute agency even to technical objects. They construct a model of "distributed agency" (ibid.: 21) which allows to determine a "stream of actions" (ibid.: 41) with activities distributed among humans and nonhumans. However, the attribution of agency or responsibility to human or nonhuman is constructed by the observer.

This model may help to better understand that activities in complex technological systems are distributed among humans and smart technology. However, despite of their call for an empirical approach, Rammert and Schulz-Schaeffer did neither refer to a specific theory of action nor operationalize their model in a way that enables empirical studies, e.g. with a quantitative focus.

Methods of Research on Hybrid Systems

Latour's provocative arguments serve us as a starting point to analyse if the processes of hybrid interaction of humans and technology can be integrated into the sociological theory of action. We want to analyse human-machine interaction empirically without losing contact to mainstream sociology. In the end our approach will not be able to answer fundamental questions about the ontological status of actors and actants, since we do not have empirical access to those subject matters. Empirically observable are only real interactions as well as processes, in which humans attribute agency to technology (insofar there is a structural asymmetry, since the opposite direction is not observable).

Recent research on hybrid systems has up to now used different methods to observe human-machine interaction, such as:

1. Observation and measurement of real interactions of human and technology, for example in smart cars (Stanton/Young 2005) or in control rooms of complex facilities (Moray et al. 2000, Cummings/Bruni 2009).

2. Ethnographic observation and thick description of human-machine interaction, for example encounters with robots or avatars, also in real settings of working environment (Brooks 2002, Turkle 2005, Braun-Thürmann 2003, Krummheuer 2010, Luff et al. 2000b), partly using auto-

matic recording of interactions (Hahne et al. 2006).

3. Case studies on advanced technical systems such as the Traffic Alert and Collision Avoidance System (TCAS) in aviation and on incidents and accidents that have been caused at least partly by the system (Brooker 2008, Grote 2009, Weyer 2006).

4. Surveys of experts or laymen concerning their experiences with and their attitudes towards smart technology (Graeser/Weyer 2010, Weyer et al. 2012).

5. Computer simulation of social processes by means of the method of agent-based modelling and simulation (ABMS), as e.g. applied in growing artificial societies (Epstein/Axtell 1996, Epstein 2007) and other projects (Macy 1998, Macy/Willer 2002).

Our approach combines methods 1, 4 and 5. In using computer simulation we refer to the model of sociological explanation (MSE), established by Hartmut Esser (1991, 2000) and others, who on their part refer to James Coleman (1990). MSE is a sociological theory of action, which has been elaborated in many details and has already been formalised by its founders, so that it is well suited for modelling and simulation (for details see the excursus in the following section).

Our model of sociological explanation of hybrid systems (HMSE) is a further development of the MSE, which only adds a new component: the agency of technology. We want to show that a sociological theory of action is capable to grasp the phenomenon of co-action of technology, without forcing us to give up basic assumptions such as the intentionality of action, as Latour suggests.

First, we developed a hybrid model of action (Chapter 2), implemented this model in a computer simulation (Chapter 3) and then performed experiments with real probands, who had to solve a driving task in a simple traffic simulation conjointly with autonomous technical systems (Chapter 4). During these experiments we measured the real distribution of agency by recording certain performance data. Besides, we documented the attribution of agency to technology by questioning the probands during and after the test runs.

Our hypotheses are:

(H1) The interaction of humans and autonomous technical systems can be modelled by means of the HMSE as a symmetrical interaction.

(H2) Human actors, which are part of the hybrid system, attribute agency to technical systems and perceive the relation of human and technology as a symmetrical one.

(H3) The concept of agency of technology can be operationalized and empirically investigated by experiments via computer simulation.

3 The model of sociological explanation of hybrid systems (HMSE)

In this chapter we introduce the model of sociological explanation of hybrid systems all, we start with a short excursus: The MSE and the SEU calculation of actions, the theoretical basis of the HMSE, are explained. Later on, we present a combination of MSE with ideas from Latour and Rammert/Schulz-Schaeffer that lead consequently to the HMSE.

3.1 Excursus: SEU theory and the model of sociological explanation

In general, sociology focuses on the explanation of macro phenomena. Sociologists try to determine, how the current state of a social system has dynamically emerged from a previous one. According to Esser (1993a) a sociological in-depth explanation consists of three explanatory steps: the logic of situation, the logic of selection and finally the logic of aggregation.

In the first step, the logic of situation, the researcher "has to reconstruct the [...] situation for typical actors in typical situations" (ibid.: 8) and has to formalize this perception.

In the second step, the logic of selection, a selection theory, e.g. SEU, is used to determine the appropriate action of different actors. Esser applies a selection rule form classical rational choice theory (RCT). However, SEU adds a subjective element to RCT which typically presumes objective rationality. Because of different preferences and different definitions of the situation actors may select different actions although they share the same situation.[1]

In the last step, the logic of aggregation, actions of many individual actors are usually merged by means of transformation rules, thus leading to the explanandum, the successor state of the social system. Especially this last step can be well accomplished via computer simulation.

The logic of selection is the central element of Esser's model of sociological explanation (MSE). It can be formalized as follows: Every actor has a set of alternative actions $a_i \in \{a_1, a_2, \ldots, a_n\}$, evaluated goals $u_j \in \{u_1, u_2, \ldots, u_m\}$ and expectations. These expectations can be modelled as probability values $p_{i,j} \in [0,1]$ which connect every action a_i with every goal u_j. $p_{i,j}$ denotes the subjectively estimated expectation that the selection of action a_i leads to the fulfilment of goal u_j. The actor selects the action a_i with the highest value of subjective expected utility. The SEU value for a specific action is calculated as

$$\mathrm{SEU}(a_i) = \sum_{j \in \{1, \ldots, m\}} p_{i,j} \cdot u_j$$

Esser's MSE refers to Coleman's (1990) micro-macro-model, which refers to actions of single actors. The interaction of several actors thus can be analysed either by sequential chaining of decision-making processes or by combining parallel processes of actors, which collaborate in a social system and that way produce common effects.

Referring to the second case, Esser constructs a multi-layer model with a meso level "between the overall macro structures of society and the micro actions of individual actors" (1993b: 112). This meso level is constituted by the collaboration of different decision-making processes on the micro level, namely as "aggregated effect of the situation-oriented action of actors" (ibid.).

3.2 Symmetrical construction of agency

We transferred the model of Esser to the collaboration of humans and technology, who both, according to Rammert and Schulz-Schaeffer, are elements of a distributed system. We assume that actions of human actors as well as of technical systems can be described in a symmetrical manner. Hence, we apply SEU theory similarly to human and nonhuman parts of the hybrid system, assuming that both have a set of actions, evaluated goals and probability values which combine actions and goals. Each component of the hybrid systems, with regard to its responsibility, selects the action with the highest SEU value.

Our starting point is a simple hybrid system consisting of a human actor A_H and a nonhuman actant A_{NH}. Both are in the situation S_t in the midst of a sequence of actions, which are running

[1] Of course, the logic of action could also be modeled by using more simple concepts such as KISS („keep it simple, stupid!"), cf. (Epstein/Axtell 1996). However, we assume a micro-sociological foundation of action, based in sociological theory, will provide a better starting point for modeling human-computer- or human-robot interaction – an issue that has rarely been investigated systematically.

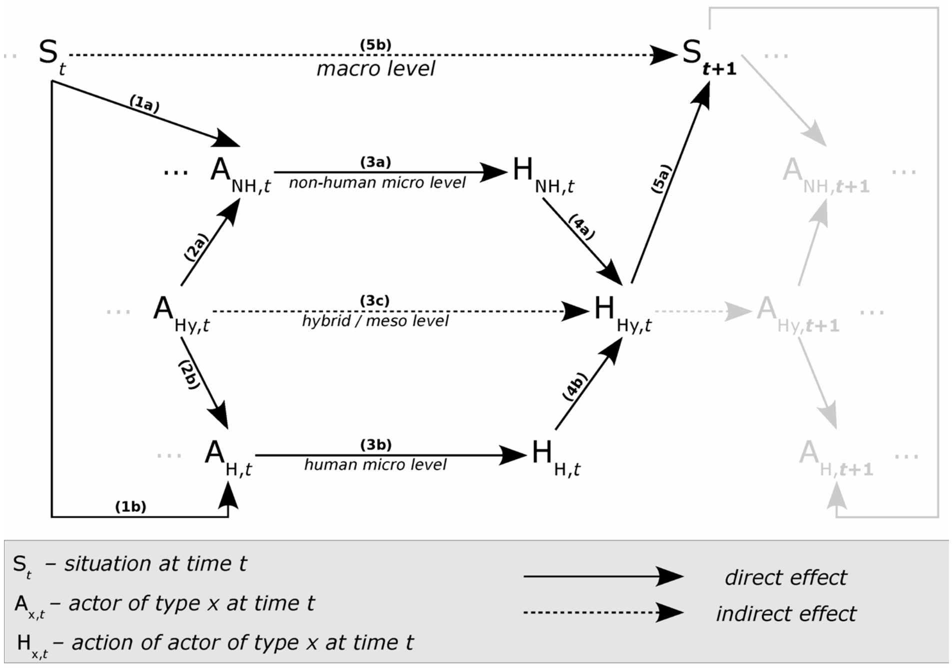

Figure 1: The model of sociological explanation of hybrid systems (HMSE)

in short periods of time. Both actors, or actants respectively, make an autonomous and thus subjective definition of the situation, indicated by the initial state $A_{H,t}$ and $A_{NH,t}$ (cf. Figure 1).

Now we borrow the idea of Esser and of Latour, that the cooperation of A_H and A_{NH} has constituted a meso level with a new hybrid actor A_{Hy} resulting from the interactions, which occurred before the moment t. The definition of the situation, performed by both partners (arrows 1a, 1b), thus is additionally shaped by the existence of this hybrid level (2a, 2b). Referring to the definition of the situation and the available options, both parts (A_H and A_{NH}) perform actions on the respective micro level (3a, 3b). The idea is that both, human actor and nonhuman actant, act on the micro level according to the SEU logic.

The actions of A_H and A_{NH} result (4a, 4b) in an aggregated effect on the meso level (3c). From an outside perspective one cannot determine the single contributions, but can only observe the composite overall action of the hybrid actor A_{Hy}. This coaction finally leads to aggregated effects on the macro level, which is beyond the hybrid system. Of course, other human, technical or hybrid actors contribute to these macro effects as well, which can be described as the transformation of the whole system from situation S_t to situation S_{t+1} (arrow 5b).

Please note that situation S_t does not affect the hybrid actor directly, because only human actors or technical actants are able to define situations. However, the coaction of A_H and A_{NH} leads to macro effects - hence the continuous arrow 5a. Additionally, the short sequence described, is part of a sequence of actions, which may continue for a while.

3.3 Intentionality of technology – a feasible assumption?

An integral part of the HMSE is the symmetrical application of a sociological theory of action to human actors and nonhuman actants. This opens up the question if the assumption of intentionality is feasible for inanimate technology. We are well aware of the fact that technological systems do not have intentions by themselves, but are coded by programmers who incorporate their intentions into the design of the system. In doing so they assume that the system will behave in the pre-programmed manner even if its constructor is absent. With other words: They assume that technological systems will perform actions that are compatible with the constructor's intentions.[2]

However, the question remains how to design the interaction of actors and agents properly, referring to a sociological theory of action. At the crucial moment, when a technology is released to its users, the interaction between the designer and the technological system ceases, and the main interaction takes part between the human actor, acting intentionally, and the technological system, accomplishing actions intentionally designed by the constructor.

In order to make things easier, we therefore decided to move along the way of multi-agent research. Computer sciences as well as research on multi-agent systems usually equip software agents with a BDI architecture, i.e. the ability to process believes, desires and intentions (Malsch 1998, Wooldridge 2001). By that way, software agents can behave in a way similiar to human interaction - or to phrase it more carefully: that can be interpreted by humans with the aid of patterns that are taken from experiences with human-human interaction (Geser 1989, Turkle 2005, Takayama / Nass 2008).

When implementing the HMSE as an interactive agent-based simulation we decided to equip the nonhuman actant A_{NH} with the ability to act intentionally according to the rules of the SEU theory. This allows us to monitor the interaction between humans and nonhumans and to compare these data with the self-assessments of the probands. Above all we can analyse whether the level of agency and the intentions, which humans attribute to nonhuman actants, is in accordance with the technically implemented level or not. Additionally, this experimental setup and its theoretical basis allows us to distinguish between goals and actions. Referring to Coleman (1990) and Esser (1993b) we define agency by the ability to plan *and* to act. By means of our software model we can empirically observe and measure whether people attribute either the performance of *actions*, the pursuit of *goals* or both to their nonhuman partners. To this regard the experiments produced the most surprising results.

3.4 Demonstration of the HMSE - an illustrative example

The concept of HMSE can be illustrated by a scenario, in which a human driver has to keep a certain distance towards another car running ahead, supported by a driver assistance system. According to the terms from the MSE we can distinguish three phases:

Cognition of Situation (Logic of Situation)

The human driver observes other cars running ahead and assesses whether separation is sufficient or s/he has to brake. The nonhuman assistance system, e.g. adaptive cruise control (ACC), does almost the same: observing traffic via its sensors and assessing if action is necessary. However, cognition of situation may be different, for example, if the driver recognises a car on the next lane as a potential conflict, because this car indicates lane change by its turn signal, whereas ACC doesn't react, because it only recognises cars on the same lane. Maybe it even accelerates, because from its point of view the lane is free.

Decision-Making (Logic of Selection)

Both parts of the hybrid system make their decisions based on their goals (e.g. avoiding an accident) and select the action with the highest SEU value: They take action which most likely

[2] We are grateful to Michaela Pfadenhauer and Knud Böhle, who helped us to phrase this argument more precisely.

leads to the desired result. By that way both act intentionally: humans literally, nonhumans rather mechanically, according to design goals and rules implemented in their software.

The overall behaviour of the hybrid actor is the result of the cooperation of A_H and A_{NH}, which sometimes may generate surprising effects if the driver decelerates and the assistance system accelerates, as in the case described above. By means of the hybrid (meso) level these actions are mutually recognized and consequently influence the behaviour of both partners in the next sequences. The outside observer, however, can only observe the behaviour of the hybrid actor A_{Hy}, which dynamically adapts speed to the speed of the car ahead.

Aggregation (Logic of Aggregation)

A mechanism is needed to transform a number of singular actions (of human drivers, hybrid cars etc.) into collective structures, such as the current state of traffic on a highway. The method of agent-based modelling and simulation (ABMS) is well suited for conducting and analysing the aggregation of a large number of actions. Using this method, we can observe emergent effects, structural dynamics, path dependencies, non-linear processes in complex systems etc., which can hardly be examined using other methods of social research (Resnick 1995, Sawyer 2005, Epstein 2007).

3.5 Strengths and weaknesses of our approach

Our approach, implementing a model of sociological explanation of hybrid systems and using it as a basis for an interactive computer simulation, does not allow answering fundamental ontological questions, for instance, if humans and nonhumans are equal. Furthermore, we cannot decide if smart technology deceives us and only simulates agency.

However, by means of our method we are able to capture not only the perspective of the human actors, e.g. by interviews, but also the perspective of nonhuman actants, e.g. by recording interaction data and having knowledge about their internal functioning – a task where other approaches, claiming nonhuman agency, have failed until now (Collins/Yearley 1992).[3] Thus, we are able to analyze the interaction of human actors and nonhuman actants empirically and compare attribution processes with real performance data. We can not only observe the feedback of human-automation interaction on humans, as Sherry Turkle (2005) did in her field experiments. In a laboratory experiment the setup of the nonhuman actant as well as the different parameters of the hybrid system can be changed in a controllable manner.

4 The HMSE as a basis for an interactive computer simulation

In this chapter we describe the SIMHYBS model as well as the experimental setting. The simulation model SIMHYBS was created in order i) to test the theoretical framework offered by the HMSE and ii) to observe the interplay of humans and nonhumans. We applied a simple, realistic scenario, which probands could use without much training. Additionally, it should allow the investigator to select different modes of distribution of agency between humans and nonhumans.

The scenario consists of a road and cars driving on it, whereas the traffic is only one-way (Figure 2). The drivers are software agents, most of them driving automatically with randomly selected speed and without regard of their environment. All in all, they are only obstacles for the car we are mainly interested in. This car is con-

[3] For instance, Callon/Law (1989) have been unable to grasp the perspective of he scallops, since they neither could be interviewed nor delivered any data. In our experiments, the agents couldn't be interviewed as well, but we could gather a large amount of data on their „behaviour".

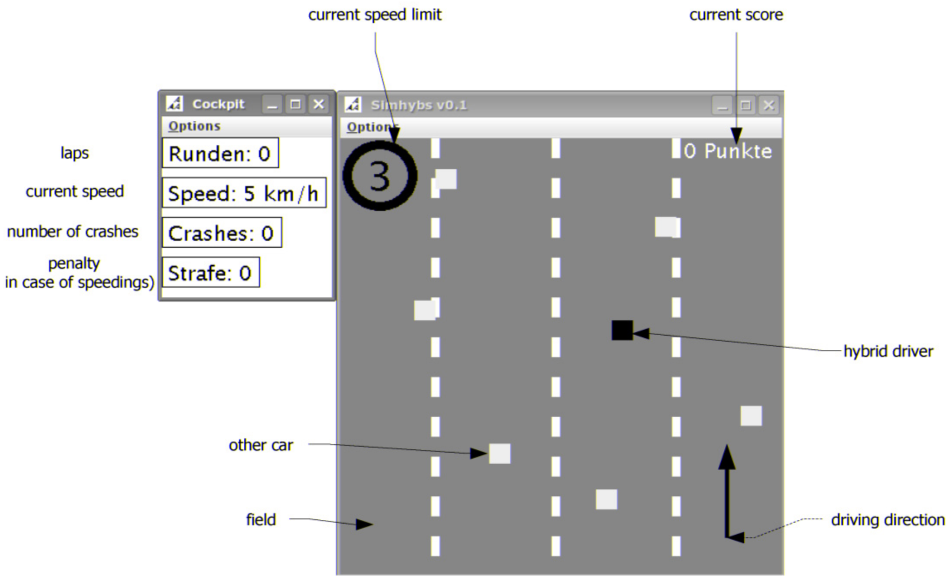

Figure 2: Screenshot of the interactive simulation SIMHYBS

ducted by a hybrid driver, consisting of a human actor A_H and a nonhuman actant A_{NH}. The latter is constructed as a driving assistance system, which can sense its environment, define a situation and finally make an appropriate decision, as demonstrated above in the case of ACC. According to the preselected driving mode (see below), the nonhuman component of the hybrid driver can perform different tasks, for example speed regulation or steering.[4] In the automatic mode it can also perform all tasks.

The hybrid car gets scores for each lap (defined by crossing the upper border of the screen); it loses points in case of a crash with another car or when it exceeds the speed limit. Cars can move into three directions: to the left (NW), to the right (NO) and straight on (N - towards the top of the screen).

According to the idea of the HMSE, decisions of the hybrid driver are assessed by means of the SEU theory, which refers to the subjective evaluation of alternatives, based on individual goals and subjective preferences. The basic decision rule is: actors try to maximise utility, i.e. they select actions with highest SEU value (see section 3.1).

This calculation can be done by humans as well as by nonhuman software agents. Both analyse the given situation from their individual perspective and select the action with the highest SEU value, e.g. accelerating/decelerating (G+,G-) or steering left/right (L,R,G).[5] However, actions are not performed immediately since the agency manager first has to check who is responsible for the respective action, before he accepts it.

4.1 Elements of the SEU model

The SEU model, as we have seen in the excursus above, consists of a set of feasible options/actions, evaluated goals, and expectations:

Options

steer to the left (L)
steer to the right (R)
no steering (G straight)
accelerate (G+)
decelerate (G-)

[4] Additionally, the hybrid driver has a software component, the agency manager, which moderates the actions of the human and nonhuman components.

[5] Since SIMHYBS has been implemented at a German research institute, some German relics remain in the software such as the abbreviation „G“ (geradeaus) or „FAS“ (Fahrerassistenzsystem).

Goals[6]

avoid crashes (c)
comply with the speed limit (g)
make laps (r)

Expectations

Expectations $p_{i,j}$ are important, because they comprise the ideas of the respective actor to what extent a certain action will help to achieve a given goal. For example, if a slow car is straight ahead, then the probability that accelerating will help to achieve the goal of crash avoidance is low (0.25), even if this action may help to gain high scores (1.0 – values in brackets are the probabilities we used in the SEU model). We cannot present the complete and therefore large $p_{i,j}$ matrix of expectations in detail here (cf. Fink/Weyer 2011: 103).

4.2 Experimental setup

SIMHYBS was implemented with the agent-based simulation software Repast (Repast 2010, Fink 2008). It can be operated in four modes, which differ regarding the distribution of roles/agency (see Table 1).

We made experiments with 31 probands; 30 of them could be used for analysis. Before starting the experiments, each proband got a short instruction, especially concerning the different modes and the distribution of responsibilities for different *actions*. In advance, we told all probands that the assistance system in any case supports them in reaching the overall *goal* (making a score as high as possible, in other words: account for all goals of the game). We will come back later to this distinction of actions and goals.

Every proband made seven simulation runs of about 3 minutes as depicted in Table 2.

Questionnaires were used in between the runs (FE) and at the end of the first six runs (FG) to gather additional information. The last questionnaire (FA) was used for the fully-automated mode. Probands were asked to evaluate the driver assistance system and to assess, to which degree both parts had contributed to the achievement of the goal. The final questionnaire no 7 furthermore asked for issues such as loss of control. An open interview completed the experiment.

Data Recording

During the runs we collected different types of data: questionnaires asked for self-assessment and for attributions on part of probands. Additionally, we recorded background data on total scores, laps, crashes, violations of speed limits, and keystrokes. This way we are able to compare the self-

Table 1: Modes of distributed agency

Mode	Type	Description
FAS-STEERING	semi-automated	The driver assistant is responsible for actions left, right, straight on. (L,R,G)
FAS-SPEED	semi-automated	The driver assistant is responsible for acceleration and deceleration system. (G+,G-)
MANUAL	manual	The driver assistant does not intervene, but only warns in case of violation of speed limit. ( )
FULL-AUTO	fully-automated	The driver assistant is responsible for all actions. The proband has the authority to intervene and to switch off the system for a short period of time. (L,R,G,G+,G-)

[6] The abbreviations refer to German words: "g" (Geschwindigkeit einhalten), "r" (Runden machen)

Table 2: Experimental sequence with appropriate number of records

Run	*Mode*				*Questionnaire*
1	FAS-STEERING				FE
2		FAS-SPEED			FE
3			MANUAL		
4	FAS-STEERING				FE
5		FAS-SPEED			FE
6			MANUAL		FG
7				FULL-AUTO	FA
Number of questionnaires	N=60 (2*FE)	N=60 (2*FE)		N=30 (FA) N=30 (FG)	

FE – questionnaire per experiment (only for FAS-STEERING and FAS-SPEED)
FG – questionnaire for overall experience
FA – questionnaire fully automated mode

assessment of probands with recorded data. Additionally, we can compare the attribution of agency to technology, done by our probands, with the real implementation of the nonhuman actant. In this respect, the results were surprising.

5 Results

The following sections mainly deal with the methodological benefits of the HMSE and present some empirical results on the issue of distributed agency.

5.1 Distribution of agency

After each simulation run, probands were asked to answer the question to which degree they had contributed to the overall goal of the game (cf. Table 3). We used an interval scale with five ranges of values (0-20%, 20-40%, 40-60%, 60-80%, 80-100%) that were presented to the probands.[7] For the calculation of an agency metric we mapped the groups to the interval [0,1]: "0-20%" → 0.1, "20-40%" → 0.3, …. Let N_{Mode} denote the number of questionnaires for a specific mode, then a mode-specific agency value evaluated by human actors can be calculated as follows:

$$Agency_H(Mode) = \frac{1}{N_{Mode}} \sum_{i=1}^{N_{Mode}} m_i$$

Table 3 presents the mean values for agency for the two semi-automated modes.

In the mode FAS-STEERING, in which the assistance system is responsible for the task steering (and probands for speed regulation), probands ascribe themselves an agency value of 0.433. In the mode FAS-SPEED, where the assistance system is responsible for the task speed regulation (and probands for steering), probands ascribe themselves an agency value of 0.580, indicating different perceptions of the distribution of agency. Several statistical measures like t-tests and confidence intervals confirm that this difference is significant.

[7] Although the questionnaire only provided five agency ranges the assumption of an interval scale is appropriate because the scale sections have the same size and are ordered. For future research we propose the use of a visual analogue scale (Reips/Funke 2008).

Table 3: Mode-specific agency values estimated by human actors

	Mean / $Agency_H$	Standard deviation	Median	0%/25%/50%/75% 100%-quantile
FAS-STEERING	0.433	0.159	0.5	0.1/0.3/0.5/0.5/0.7
FAS-SPEED	0.580	0.170	0.5	0.1/0.5/0.5/0.7/0.9

Both modes mentioned above are complementary to each other. If, for example, people ascribe themselves a share of 43.3 percent in reaching the overall goal of the game, they also – indirectly – define the share of the other part, the assistance system.

Consequently, we can calculate the agency of the nonhuman for a specific mode as follows:

$$Agency_{NH}(Mode) = 1 - Agency_H(Mode)$$

As Table 4 shows, agency of different tasks has been attributed almost symmetrically.

Concerning the task *speed regulation*, the agency value is 0.433 in the mode FAS-STEERING (directly calculated), in which the human is responsible for this task and the nonhuman for steering.[8] An almost identical value of 0.420 can be found in the mode FAS-SPEED (indirectly calculated), where the nonhuman is responsible for this task and the human for steering. Agency values obviously are similar, regardless of which part performs the task, the human or the nonhuman driver.

The same observation can be made for the task *steering*, where the agency value is 0.580 in the mode FAS-SPEED (directly calculated), in which the human is responsible for this task and the nonhuman for speed regulation.[9] Again an almost identical value of 0.567 shows up in the mode FAS-STEERING (indirectly calculated), where the nonhuman is responsible for this task and the human for speed regulation (Table 4).

These data seem to serve as an experimental proof of Latour's assertion of symmetry of humans and nonhumans – at least regarding a symmetrical attribution of agency (done by humans).

5.2 Delegation of actions or of goals?

After each test run in semi-automated modes we asked probands for the goals, which the assistance system had been pursuing. They could choose multiple entries from the following three goals: crash avoidance (c), laps (r) and keep speed limit (g) and combine them arbitrarily. As the

Table 4: Agency values for specific modes

Mode (actions performed by driver assistance system)	$Agency_H$ (calculated directly)	$Agency_{NH}$ (calculated indirectly)
FAS-STEERING (L,R,G)	0.433	0.567
FAS-SPEED (G+,G-)	0.580	420

8 Mathematically: $Agency_H(FAS-STEERING) \approx Agency_{NH}(FAS-SPEED)$

9 Mathematically: $Agency_H(FAS-SPEED) \approx Agency_{NH}(FAS-STEERING)$

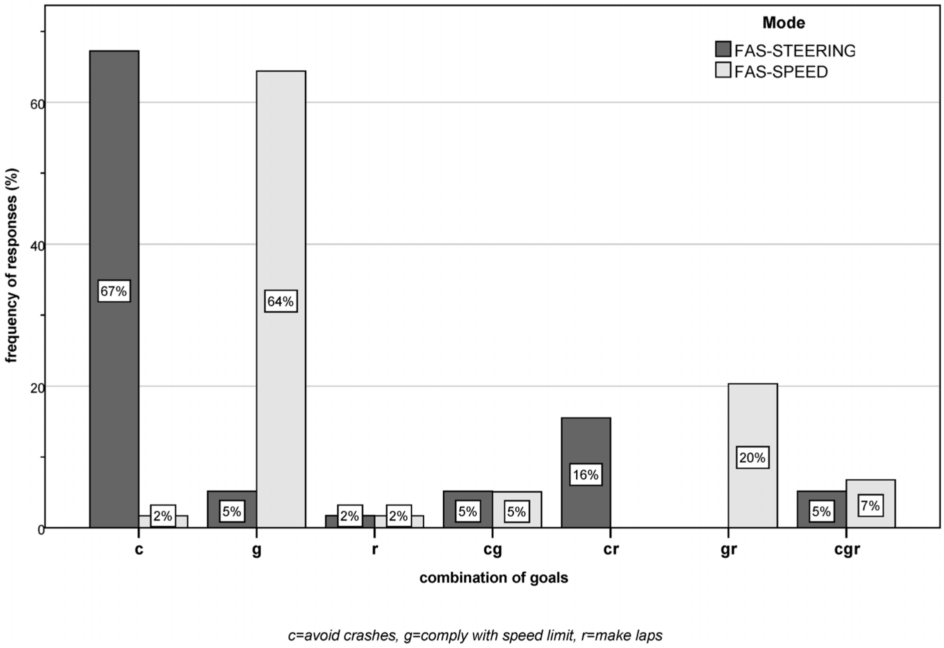

Figure 3: Which goals did the assistance system pursue?

chart in Figure 3 demonstrates, the assessments are extremely different, according to the respective mode.

This is surprising, since human and nonhuman had been instructed, respectively programmed to pursue the overall *goal* (consider all goals c, r and g) in all modes. Only the responsibility for *actions* (steering, speed regulation) had been distributed to a different degree. Nevertheless, probands obviously freed themselves of the task to pursue certain goals, when taking over a certain task:

For example, in the mode FAS-STEERING, where the assistance system steers the car (actions L, R and G), 67 percent of probands ascribed only the goal of crash avoidance to the assistance system. Presumably, they assumed that one cannot follow the two other goals with means of steering.

On the contrary, in the mode FAS-SPEED only 2 percent of probands guessed that the assistance system pursues this goal, even though the investigator had instructed them that the system supports probands in achieving the overall goal.

As an unexpected result of our inquiry, we can point to the fact that delegation of actions to nonhumans obviously goes hand-in-hand with the ascription of goals.

5.3 Interim conclusion

The preceding chapters have demonstrated the *methodological* value of HMSE. We do not claim that all of our findings will hold out against future testing. We rather assume that much more experiments will be needed to sustain or to refute these results. However, by programming the nonhuman actant as an intentionally acting player we have found a method to empirically observe the interaction of humans and nonhumans as well as processes of goal and action attribution. Additionally, we can differentiate between distribution of actions and of goals. Our methodology allows identifying sets of actions and ascribing an agency value to them. From the perspective of human probands it is obviously irrelevant whether certain tasks are performed by a human or a nonhuman. The agency value for respective sets of actions was almost identical. Furthermore, we could show that probands do not clearly distinguish between delegation of actions and of goals.

6 Conclusion

In this paper we presented a sociological model, which describes the co-action of technology in a way that is open for empirical investigations of distribution of agency. By this means we offered a proposal on how to fill the theory gap of current research, which mostly refers on the empirical observation of human-machine or human-robot interaction, but heavily lacks a theoretical foundations in terms of a sociological theory of action – as in the case of Turkle or Suchman (cf. chapter 2). On the other hand, models of the interaction of humans and nonhumans in sociology and related fields (e.g. Latour) are mostly based on single case stories and lack a possibility to investigate these issues by well-established methods from empirical social sciences. The HMSE is an attempt to develop a sociological model as well as a method to tackle theses questions experimentally.

Referring to our three hypotheses we now can conclude:

(H1) Latour's assertion of nonhuman agency can be empirically investigated by means of the HMSE model, which extends the common model of sociological explanation (MSE) to autonomous technology. The HMSE allows us to analyse the interaction of humans and nonhumans, to confirm the symmetry the-sis empirically and to produce novel results such as the mixture of delegation of actions and of goals.

(H2) Test runs have shown that human actors attribute agency to technical systems and perceive the relation of human and technology as a symmetrical relation.

(H3) Computer simulation is a practical method i) to investigate human-machine interaction, ii) to measure agency, and iii) to make attribution processes visible. The latter is done by comparing the perception of role distribution of our probands with the experimental setup and the recorded data.

Our data confirm the (very general) perception of nonhuman agency (Latour 1998). They also support attribution theory (Rammert/Schulz-Schaeffer 2002)and imply further considerations: Human actors not only ascribe agency to nonhuman actants. By taking this attribution, they also redefine their own role, e.g. when concentrating on a certain task and getting rid of the responsibility for pursuing other goals.

By interacting with autonomous technology human probands obviously tend to construct a role distribution, which remarkably differs from the distribution implemented in the software program. In some settings, humans obviously tend to attribute responsibility to the technical system and to overtrust technology – a fact already observed by human-factors research in psychology (Manzey 2008), which until now could not be explained by means of sociological theory of action.

Future research on HMI issues should analyse this point in more detail. If our findings can be confirmed and reproduced in further experiments in different scenarios, this might have an impact on the construction of user interfaces in advanced systems.

The HMSE can gain new insights into the interplay of humans and nonhumans and provide a deeper understanding of this kind of hybrid interaction, grounded on a sociological theory of action. Its findings, especially concerning implicit role distribution, thus may be a step to better understand human-machine interaction in real driving situations. However, prior to this more basic research is needed. The model and the method applied thus may also serve to better comprehend the issue of social co-operation in human-machine and human-robot interaction. Our approach may help to improve the design of sociable robots, whose autonomous actions are always part of a hybrid constellation, consisting of a human actor and a nonhuman agent, who perceive each other from their respective point of view. Both attribute properties to each other and act and interact on the basis of their specific preferences. Only if we learn to understand these processes of hybrid interaction theoretically and practically, we may be able to design sociable robots in a way that they become real (artificial) companions.

References

Bornmann, Lutz, 2010: Die analytische Soziologie: Soziale Mechanismen, DBO-Theorie und Agentenbasierte Modelle. In: *Österreichische Zeitschrift für Soziologie* 35 (4): 25-44, <http://link.springer.com/article/10.1007/s11614-010-0076-6>.

Braun-Thürmann, Holger, 2003: Künstliche Interaktion. In: Thomas Christaller/Josef Wehner (eds.), *Auto-nome Maschinen*. Wiesbaden: Westdeutscher Verlag, 221-243.

Breazeal, Cynthia L., 2004a: *Designing sociable robots*. MIT press.

Breazeal, Cynthia L., 2004b: Social interactions in HRI: the robot view. In: *Systems, Man, and Cybernetics, Part C: Applications and Reviews,* IEEE Transactions on 34 (2): 181-186.

Brooker, Peter, 2008: The Überlingen accident: Macro-level safety lessons. In: *Safety Science* 46: 1483-1508.

Brooks, Rodney, 2002: *Menschmaschinen. Wie uns die Zukunftstechnologien neu erschaffen*. Frankfurt/M.: Campus.

Callon, Michel/John Law, 1989: On the Construction of Sociotechnical Networks: Content and Context Revisited. In: *Knowledge and Society: Studies in the Sociology of Science Past and Present* 8: 57-83.

Coleman, James S., 1990: *Foundations of Social Theory*. Cambridge/Mass.: Harvard University Press.

Cummings, Mary L./Sylvain Bruni, 2009: Collaborative Human-Automation Decision Making. In: Shimon Y. Nof (eds.), *Handbook of Automation*. Heidelberg: Springer, 437-447.

Epstein, Joshua M., 2007: *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton, NJ: Princeton University Press.

Epstein, Joshua M./Robert Axtell, 1996: *Growing Artificial Societies*. Social Science from the Bottom Up. Washington, D.C.: Brookings Inst. Press.

Esser, Hartmut, 1991: *Alltagshandeln und Verstehen. Zum Verhältnis von erklärender und verstehender Soziologie am Beispiel von Alfred Schütz und 'Rational Choice*'. Tübingen: Mohr.

Esser, Hartmut, 1993a: The Rationality of Everyday Behavior: A Rational Choice Reconstruction of the Theory of Action by Alfred Schutz. In: *Rationality and Society* 5: 7-31.

Esser, Hartmut, 1993b: *Soziologie. Allgemeine Grundlagen*. Frankfurt/M.: Campus.

Esser, Hartmut, 2000: *Soziologie. Spezielle Grundlagen*, Bd. 3: Soziales Handeln. Frankfurt/M.: Campus.

Fink, Robin D., 2008: *Untersuchung hybrider Akteurskonstellationen mittels Computersimulation (Diplomarbeit)*. Dortmund.

Fink, Robin D./Johannes Weyer, 2011: Autonome Technik als Herausforderung der soziologischen Handlungstheorie. In: *Zeitschrift für Soziologie* 40 (2): 91-111, <http://www.zfs-online.org/index.php/zfs/article/view/3061>.

Gad, Christopher/Caspar Bruun Jensen, 2010: On the consequences of post-ANT. In: *Science, Technology & Human Values* 35: 55-80.

Geser, Hans, 1989: Der PC als Interaktionspartner. In: *Zeitschrift für Soziologie* 18: 230-243.

Graeser, Stefan/Johannes Weyer, 2010: Pilotenarbeit in der virtuellen Welt des zukünftigen Luftverkehrs. Erste Ergebnisse der Pilotenstudie 2008. In: Gerhard Faber (ed.), *Virtuelle Welten. Simulatoren in der Aus-, Fort- und Weiterbildung von Verkehrspiloten. Proceedings des 12. FHP-Symposium. Darmstadt*: FHP, 41-52.

Grote, Gudela, 2009: *Management of Uncertainty. Theory and Application in the Design of Systems and Organizations*. Berlin: Springer.

Hahne, Michael et al., 2006: Going Data in Interaktivitätsexperimenten: Neue Methoden zur Analyse der Interaktivität zwischen Mensch und Maschine. In: Werner Rammert/ Cornelius Schubert (eds.), *Technogra-fie: Zur Mikrosoziologie der Technik.* Frankfurt/M.: Campus, 275-309.

Huettenrauch, Helge et al., 2006: Investigating spatial relationships in human-robot interaction. *Intelligent Robots and Systems, 2006 IEEE/RSJ*

International Conference on Intelligent Robot Systems (October 9 - 15, 2006), Beijing, China: 5052-5059.

Krummheuer, Antonia L., 2010: *Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts.* Stuttgart: Lucius & Lucius.

Latour, Bruno, 1988: Mixing Humans and Nonhumans Together: The Sociology of a Door-Closer. In: *Social Problems* 35: 298-310.

Latour, Bruno, 1996: On actor-network theory. A few clarifications. In: *Soziale Welt* 47: 369-381.

Latour, Bruno, 1998: Über technische Vermittlung. Philosophie, Soziologie, Genealogie. In: Werner Rammert (ed.), *Technik und Sozialtheorie*. Frankfurt/M.: Campus, 29-81.

Lee, John D./Katharina A. See, 2004: Trust in automation: designing for appropriate reliance. In: *Human Factors* 46: 50-80.

Luff, Paul/Jon Hindmarsh/Christian Heath, 2000a: Introduction. In: Paul Luff/Jon Hindmarsh/Christian Heath (eds.), *Workplace studies: Recovering work practice and informing system design*. Cambridge/England: Cambridge University Press, 1-26.

Luff, Paul/Jon Hindmarsh/Christian Heath, (eds.), 2000b: *Workplace studies: Recovering work practice and informming system design*. Cambridge/ England: Cambridge University Press.

Macy, Michael W., 1998: Social Order in Artificial Worlds. In: *Journal of Artificial Societies and Social Simulation 1* (1), <http://www.soc.surrey.ac.uk/JASSS/1/1/4.html>.

Macy, Michael W./Robert Willer, 2002: From Factors to Actors: Computational Sociology and Agent-Based Modelling. In: *Annual Review of Sociology* 28: 143-166.

Malsch, Thomas (ed.), 1998: *Sozionik. Soziologische Ansichten über künstliche Sozialität*. Berlin: Edition sigma.

Manzey, Dietrich, 2008: Systemgestaltung und Automatisierung. In: Petra Badke-Schaub et al. (eds.), *Human Factors. Psychologie sicheren Handelns in Risikobranchen*. Heidelberg: Springer, 307-324.

Moray, Neville/Toshiyuki Inagaki/Makoto Itoh, 2000: Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. In: *Journal of Experimental Psychology*: Applied 6: 44-58.

Parasuraman, Raja/Thomas B. Sheridan/Christopher D. Wickens, 2008: Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. In: *Journal of Cognitive Engineering and Decision Making* 2: 141-161, <http:// archlab.gmu.edu/people/rparasur/Documents/ParasuramanJCEDM08.pdf>.

Parsons, Talcott, 1967: *The Structure of Social Action* (1937). New York: Free Press.

Rammert, Werner/Ingo Schulz-Schaeffer (eds.), 2002: *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*. Frankfurt/M.: Campus.

Reeves, B./C.I. Nass, 1996: *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge/Mass.: Cambridge University Press.

Reips, Ulf-Dietrich/Funke, Frederik, 2008: Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator. In: *Behavior Research Methods* 40, S. 699-704.

Repast, Developers Group, 2010: *Website*. Abrufbar unter <http://repast.sourceforge.net>.

Resnick, Michael, 1995: *Turtles, Termites, and Traffic Jams. Explorations in Massively Parallel Microworlds (Complex Adaptive Systems)*. Cambridge/Mass.: MIT Press.

Sawyer, Roberth Keith, 2005: Social *Emergence: Societies as Complex Systems*. Cambridge/Mass.: Cambridge University Press.

Sheridan, Thomas B., 1999: Human supervisory control. In: Andrew P. Sage/William B. Rouse (eds.), *Handbook of systems engineering and management*. Hoboken, NJ: John Wiley & Sons, 591-628.

Stanton, Neville A./Mark S. Young, 2005: Driver behaviour with Adaptive Cruise Control. In: *Ergonomics* 48: 1294 – 1313.

Sturma, Dieter, 2001: Robotik und menschliches Handeln. In: Thomas Christaller (ed.), *Robotik. Perspektiven für menschliches Handeln in der zukünftigen Gesellschaft*. Berlin: Springer, 111-134.

Suchman, Lucy A., 2007: *Human and Machine Reconfigurations: Plans and Situated Actions, 2nd Edition*. Cambridge/Mass.: Cambridge University Press.

Suchman, Lucy et al., 1999: Reconstructing Technologies as Social Practise. In: *American Behavioral Scientist* 43: 392-408.

Takayama, Leila/Clifford Nass, 2008: Driver safety and information from afar: An experimental driving simulator study of wireless vs. in-car information services. In: *International Journal of Human-Computer Studies* 66: 173-184,

<http://www.sciencedirect.com/science/article/pii/S1071581906000851>.

Turkle, Sherry, 2005: *The Second Self*: Computers and the Human Spirit. Cambridge/Mass.: MIT-Press.

Turkle, Sherry, 2006: A Nascent Robotics Culture: New Complicities for Companionship. AAAI Technical Report Series, <http://www.aaai.org/Papers/Workshops/2006/WS-06-09/WS06-09-010.pdf>.

Turkle, Sherry et al., 2006: Relational Artifacts with Children and Elders: The Complexities of Cybercompanionship. In: *Connection Science* 18: 347-361, <http://web.mit.edu/sturkle/www/pdfsforstwebpage/ST_Relational Artifacts.pdf>.

Weyer, Johannes, 2006: Modes of Governance of Hybrid Systems. The Mid-Air Collision at Ueberlingen and the Impact of Smart Technology. In: *Science, Technology & Innovation Studies* 2: 127-149, <http://www.sti-studies.de>.

Weyer, Johannes/Robin D. Fink/Fabian Lücke, 2012: Complexity and controllability of highly automated systems. How do drivers perceive and evaluate the co-operation of driver assistance systems? In: *Safety Science* (submitted).

Wooldridge, Michael, 2001: *Introduction to Multiagent Systems*. Hoboken, NJ: John Wiley & Sons.