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

ISSN: 1861-3675

STI Studies
www.sti-studies.de

When is a Robot really Social?

An Outline of the Robot Sociologicus

Martin Meister (University Duisburg-Essen, martin.meister@uni-due.de)

Abstract

The article explores the idea of understanding the "social" in the emerging field of Social Robotics from an explicitly sociological perspective, and more specifically from the viewpoint of sociological theory of action.[1] It suggests to found the basic architecture of the "social robot" and the interaction with it on generalized expectations, to solve the main problem of Social Robotics – the problem of finding an adequate way of reducing the complexity of social situations. I argue in this paper on empirical grounds that Social Robotics, unlike the heterogeneous field of Service Robotics, has developed into a distinguished field of research. And I present some evidence that the problem of the complexity of social situations is a central issue in the field itself, not least regarding the methodological problem of the comparability of performance of specific technical solutions and human reactions to these. By drawing on this evidence and applying a sociological model of the reasoning process of social actors, an architectural blueprint is developed that tries to catch central aspects of a "really social" robot from a sociological perspective while working with central issues from the discourse of Social Robotics itself. This basic idea of a transfer of principle from sociological theory of action is positioned against social constructivist approaches and the tradition of AI-critique. Finally, some possible uses of the robot sociologicus are sketched out, both from a sociological perspective and as a possible contribution to the interdisciplinary field of Social Robotics and human-computer interaction research.

[1] I would like to thank Knud Böhle and an anonymous referee for very useful comments on earlier versions of this article. I would also like to thank the members of our Berlin working group on the sociology of technology (called the "tekkies-group"), especially Jochen Gläser, Cornelius Schubert, Valentin Janda and Julian Stubbe, for an intensive discussion of an early draft of this article. In a discussion following the presentation of the paper at the colloquium of the Institute of Sociology at the University Duisburg-Essen, Gregor Bongaerts, Marcel Erlinghagen und Ingo Schulz-Schaeffer pointed me to the necessity of various sociological clarifications of the approach. Astrid Weiss, at the stage of an early presentation, warned me about confusing some pretty phrasings from Social Robotics with the field's practice.

1 Introduction

Social Robotics is an emerging field of interdisciplinary research, which recently parallels the established field of Service Robotics (or possibly only the term). Usually, so-called robot companions – robots that can serve individual human users like pets – are seen as an important, highly socially relevant part of the field of Social Robotics, with special focus on long-term, emotional and trusted human-machine-relations (Breazeal et al. 2008, Krämer & Rosenthal-von der Pütten, in this thematic issue). So it is quite astonishing that the discipline specialized on dealing with social relations and social structures does not, except for rare exceptions[2], contribute to this field at all: sociology. As a consequence, sociology is absent from the list of scientific disciplines towards which the formerly purely engineering stance in robotics has opened up recently, as can be seen in the following list from one of the forewords to the actual edition of the "Handbook of Robotics":

> "In advancing robotics further, scientific interest was directed at understanding humans. Comparative studies of humans and robots led to new approaches in scientific modeling of human functions. Cognitive robotics, lifelike behavior, biologically inspired robots, and a psychophysiological approach to robotic machines culminated in expanding the horizons of robotic potential" (Inoue 2008, p. X).

The title of my article is inspired by the title of the only article dealing explicitly with Social Robotics from an explicitly sociological point of view: "When a robot is social: Spatial arrangements and multimodal semiotic engagement in the practice of social robotics" (Alac et al. 2011). The authors take the radical social constructivist stance that only what is enacted in social practice or perceived by the actors as social is, in fact, "social". I strongly doubt that this is a reasonable starting point for any investigation of or contribution to the field of Social Robotics (compare below). It seems more adequate to raise the open question "When is a robot social?", and then to relate it to discussions in the field of Social Robotics. That is exactly what I am going to try.

Of course, any attempt to answer this question has to take into account that the term "social" has different meanings in different scientific disciplines. To mention but two, extremely contradicting examples from the fields of advanced computing and robotics: The well-known "media-equation" theory (Nass/Reeves 1996), drawing on the observation that humans tend to react to cues sent by machines as if these were other human actors, is summarized in the so-called CASA-paradigm: "Computers as social actors". This paradigm has a strong influence on robot and companion design, especially on the design of interfaces and 'human-like appearance' of technical apparatus. At least as concerns the application of the term "social", quite the opposite is true for the well-known critique of "human factors research" in design, usability and requirements engineering, which calls for a shift "from social factors to human actors" (Bannon 1991) to be able to grasp the complexity of users' intentions and situations.

And even in different strands of sociological theory and research there are very different meanings of "the social". Below I will try to apply an understanding from the actual sociological theory of action. The proposed conception models decisions of socialized actors for specific types of actions based on perceptions of the situation at hand, and based on a calculation of the likely consequences of this choice. Despite a lively discussion about the details of the modeling of this reasoning process (including the very mean-

[2] See the works of Sal Restivo for one of these exceptions (Restivo 2001), which is mainly oriented towards a theory of social cognition as opposed to the mainly individualistic stance of cognitive science.

ing of "calculation" in human reasoning), most proponents of the sociological theory of action agree that the huge majority of human actions are routine actions[3]. Then, an action that turned out as sufficiently adequate in past situations is performed in a present situation perceived as sufficiently similar without further reasoning.

The proposed model for solving the potentially infinite situational complexity can be summarized in the following steps[4]:

- In their choices of actions, social actors are oriented by the perception of the relevant aspects of the situation, including expectations about the intentions and the influence of other actors involved in the situation.
- In the further course of events these initial perceptions and expectations are confirmed (or denied), which leads, over many interactions, to a consolidation of these perceptions and expectations – they are generalized.
- Social order (or social structure) is made up from nothing other than these generalized expectations. Typically, three levels of expectations are differentiated: On the micro-level, these are expected patterns of interaction including cues that indicate in which type of interactional order the situation is embedded. On the meso-level, expectations concern e.g. formal or informal roles that regulate the division of labor in organizations; and on the macro-level, these are institutions: beliefs, attitudes and norms that are shared across society.
- Expectations from all three levels, checked and consolidated via many interactions, enter the initial perception of the situation and the following choice of action. Step one is never a calculation of every possible relevant aspect of the situation because this would render any social action impossible.

Because social actors, according to this model, apply generalized expectations, situational complexity is not a major problem for their reasoning in almost all situations. In the vast majority of cases social actors follow routines because they base their choice of appropriate interpretation of the situation and of appropriate action on proved and tested generalized expectations.

I do not see any principal reason against an attempt to realize this model on machines. To put the core of the model in words that are more suitable for the transfer to a technical design problem: Social actors are *optimized* for successfully dealing with the problem of reducing the vast complexity of social situations.

My basic aim in this article is to explore the potential of this concept for an understanding of the term "social" in Social Robotics. Can this thesis, despite the substantial differences between human socialization and technical optimization, be used as an abstract principle – or a blueprint – for the design of robots, or for an explicit modeling of human-robot interaction based on this blueprint? In this line of thought the question "when is a robot really social?" is specified as: "when is a robot social in a sociologically meaningful way"? To construct the reasoning process of robots or the modeling

[3] Rational Choice Theory is one of these exceptions. Here the homo oeconomicus is presented as an actor who permanently calculates every aspect of the situation – and who has access to all relevant information about the situation. See for an early and prominent discussion of these shortcomings Simon (1997: 291-295), and for a summary of the narrowness of the theoretical figure of the homo oeconomicus Schimank (2010): 102-127.

[4] This of course is a crude summary that hopefully expresses the basic point in a way accessible outside of sociology. I accepted neglecting important differentiations in the theory of social action that of course are important within the discipline.

of man-robot interaction by following this general model could be an attempt to solve the problem of environmental complexity for robots – especially for those robots built to interact with socialized humans. This is the question about the *robot sociologicus*.

I proceed as follows. First, I briefly summarize Social Robotics as a distinguished field of research and the understanding of "the social" in this field. Next, I present some evidence that the problem of dealing with the complexity of social situations is a central issue in the field itself, especially methodologically. Relevant approaches and findings from different strands of research in the field and in the social sciences are presented that could contribute to a discussion of generalized expectations on the micro-, meso- and macro-level in Social Robotics. Based on this illustration and a brief summary of a specific sociological model of action (Esser's model), the different thoughts and pieces of evidence are, in an inevitably sketching way, drawn together to form the rough blueprint of a possible architecture of the robot sociologicus. Then, the approach proposed is depicted in contrast to the two dominating paradigms in the humanities dealing with robotics: AI-critique and social constructivism. The final section sketches three different possible uses of the architectural blueprint of the robot sociologicus.

2 Social Robotics as a distinguished field of research

To make a sociological contribution to the interdisciplinary field of Social Robotics on the principal idea of social reduction of complexity can only work out if the term "social" has a serious meaning in the field, and contributions from non-technical disciplines are not only seen as nice-to-have, but as part of the inner core of this field (which also presupposes that that field has a core at all). Social Robotics, then, could form a new research program and a possible agenda for a new and integrated research practice to which a sociological contribution would evidently make sense.

As one prominent application area of the 'New Robotics', the idea of developing service robots, machines suited for serving ordinary people in their everyday domestic or public environments, has a history reaching back at least twenty years. At least since then it has been common to divide the overall field into three strands of research, with Service Robotics as opposed to Industrial and Field Robotics the latest (but historically oldest) and most challenging part of the robotics endeavour (see cf. Kawamura et al. 1996). This classification of three strands of robotics research might be exaggerated or 'unfair', but stems from the field itself. All three areas have their own conference series, journals, market leaders for equipment, and so on[5]. Unlike Industrial Robots, which repeatedly do the same things in an accurately defined surrounding, and unlike Field Robots, which operate far away from humans, Service Robots are thought to operate in the habitat and in the presence of the most disturbing and unpredictable elements imaginable: ordinary human beings. Everyday human activities present tremendous challenges for a robot, concerning self-localization and navigation, steering model- and decision-making, sensors and interface design, to name but a few of the technical difficulties that have to be solved. Moreover, all of these single tasks have to be integrated in one architecture and on one

[5] This classification is at least 'unfair' with respect to newer developments in industrial robotics, where man-machine-interaction has become an important issue. And besides this basic division, there are at least three other strands of robotics research and application: robotics in entertainment, in arts, and intelligent extensions of the human body: intelligent exoskeletons for soldiers (or disabled people), and intelligent prostheses (mainly for disabled people).

hardware platform, and have to be processed altogether in real-time.

The agenda and the research practice of Service Robotics treats this challenge as a bundle of purely technical problems. From a technical point of view, settings crowded with ordinary humans are the most complex environments and thus the biggest technical challenge for an advanced robot. Empirical investigations of real man-robot-interaction and studies on usability and acceptance are only conducted in rare cases and not systematically integrated in research practice, but engineers imagine the attractiveness of applications from their own point of view – they simply imagine themselves as users, the so-called "I-Methodology" (Akrich 1995). The same holds true for the conceptualizations of the "sociability" of the robots. This is often built on everyday assumptions about "the human" or "the user", and mainly treated as a question of interface design, as summarized in the following quote:

> "It is still not generally accepted that a robot's social skills are more than a necessary 'add-on' to human–robot interfaces in order to make the robot more 'attractive' to people interacting with it, but form an important part of a robot's cognitive skills" (Dautenhahn 2007: 682).

Thus the service robot, despite being conceptualized and constructed for application in everyday situations and interaction with humans, remains a *robot technologicus*.

Moreover, the field of Service Robotics is massively heterogeneous. Everyday environments only served as the most complex and demanding domain for a wide spectrum of disciplinary traditions like mechanical engineering or electrical engineering, different and often competing schools of computer sciences or AI, materials science, biology, and so forth. Scholars from these traditions often do not understand or accept each other's theoretical tradition or even their understanding of "theory", and the families of mathematical calculation they use. And they do not agree at all on application visions[6], test beds or criteria for evaluation or comparability. Thus a core research and development field has never been established[7].

Both with respect to the purely technocentric approach and to heterogeneity, this situation seems to have changed with the emergence of Social Robotics as a distinguished research program. Originating from an association of robotics scholars with an interest in human domains, and scholars from the man-computer-interaction community (in which psychological and social sciences approaches have always played an important role) and, in recent years, its subfield Human-Robot-Interaction Research (HRI), Social Robotics seems to integrate the conceptualization and empirical investigation of man-robot-interaction into the core of its research agenda. So statements like the following seem to be typical for characterizing this field:

> "Social Robotics is a new research program and a possible agenda for research practice, which for the first time regards social and societal issues as an integral part of the agenda of robotics research and development" (Steinfeld et al. 2006: 34),

or:

> "Social robotics researchers agree that the design of social robots poses both social and technical problems" (Sabanovic 2010: 444).

---

[6] For some researchers, especially from computer science and AI, grand visions (e.g. computers will beat the human chess champion in five years, or: a team of robots will beat the human football champion in fifty years) were and are an important driver of development, while most of the more engineering-oriented researchers believe this fixation on grand visions harms the development of useful machines as well as debates within society.

[7] See for an application of the sociological concept of boundary objects to the empirical case of the massively heterogeneous field of Service Robotics Meister (2011a), and for an attempt to apply this reconstruction to technology assessment and robo-ethics Meister (2011b, 2012).

One can and should be careful not to take this programmatic stance as a description of the collaborative practice in this field, nor assume that interdisciplinary cooperation across the two cultures has suddenly become smooth. Moreover, the initial definition in the first issue of the "International Journal of Social Robotics" (IJSR) is very wide:

> "Social Robotics is the study of robots that interact and communicate among themselves, with humans, and with the environment, within the social and cultural structure attached to their roles" (Ge/Mataric 2009: 1).

Furthermore, the list of issues addressed in the journal is nearly as wide as in Service Robotics. Its range covers (ibid):

- The human-robot-interaction issue itself (e.g. "models of human and animal social behavior as applied to robots", "affective and cognitive sciences for socially interactive robots" and "applications in education, entertainment, games, and healthcare");
- typical societal issues (e.g. "robot-ethics in human society" or "social acceptance and impact in the society");
- issues from general AI (e.g. "knowledge representation, information acquisition, and decision making" or "learning, adaptation and evolution of intelligence");
- issues from biologically inspired machines (e.g. "biomechatronics, neuro-robotics, and biomedical robotics"), and
- purely technical issues (e.g. "multimodal sensor fusion and communication" or "software architecture and development tools").

Nonetheless, two features of the field indicate that in Social Robotics there is common skepticism about a purely technology-driven development (the robot technologicus). The conceptualization and evaluation of "interaction with robots" and of a realization of appropriate "skills" of the robot – or at least an appropriate realization of an appropriate "adaptability" of the robots to humans and social situations – seem to form a widely accepted common ground (a "going concern" in terms of interactionism, Strübing 1998), not to say a kind of core understanding, in the field. If this assumption holds true, the field of Social Robotics would differ substantially from Service Robotics, both regarding consideration of non-technical (in some sense: "social") issues and degree of heterogeneity.

The first indicator for this is the pure distribution and frequency of the central issues, rated by the central themes in the field's leading journal (the "International Journal of Social Robotics"; IJSR)[8]. By my own rule of thumb, this distribution looks as shown in figure 1.

As can be seen, the typical descriptions of robot components and architectures are presented just like in any other robotics journal, and with only marginal reference to any possibility of comparing these approaches and individual realizations.

Especially noticeable is that no attempt has been made to develop a kind of reference architecture for a social – or sociable – robot. An architecture is the backbone of any robotics approach because it defines how the components of the robotics system are interconnected (as variants of the chain "sense-think-act"; see Murphy 2000)[9].

---

[8] This is in fact only a rule of thumb for illustrating the main issues. Many of the relevant articles of course are published in other journals, and only a further examination of the social dynamics of the field of Social Robotics could foster the preliminary observations presented here. But I think the evidence for the difference between Service Robotics and Social Robotics is strong enough to be more than just an ad-hoc impression - and the successful introduction of a specialized journal is part of this evidence.

[9] Murphy (2000) describes the history of robotics approaches at a high level of abstraction as the succession of three differ-

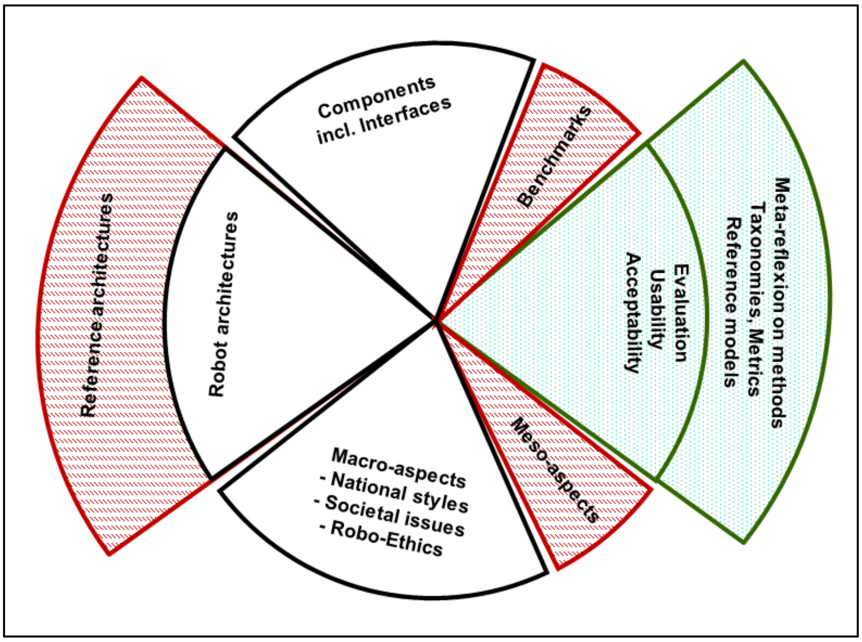

Figure 1: Thematic blocks in the IJSR (my own illustration). Reference architectures, benchmarks and meso-aspects are issues that astonishingly seem to be missing.

There is a striking amount of articles which tackle the importance of issues on the level of society at large– especially questions of societal impact of advanced robots (robo-ethics), and the question whether this is shared or different in national settings of development and use (and acceptance) of robots. What is evidentially missing are articles dealing with the meso-level, that is the consequences of an integration of robots in organizational settings. The introduction of a care-giving robot (e.g. Paro) will evidently not only create new human-robot- interactions, but will also change the organizational setting in nursing homes with respect to workload, work description and hierarchies.

But what really differs from Service Robotics is the high amount of articles that deal with the conceptualization and empirical investigation of the robots' acceptability and usability, and of patterns of man-robot interaction. The importance of this thematic block for many participants in the field is also evident from a meta-reflection on methods, which aims at taxonomies and metrics to ground a comparison of robotic approaches and empirical results. I will turn to this point in the next section.

The second indicator for a substantial difference from Service Robotics is the general treatment of the relation between technical and nontechnical aspects in Social Robotics. There, not only the sheer amount of research into nontechnical aspects is much higher, but a conceptual space is opened up to relate approaches explicitly to one another. There are many more or less elaborated versions of this conceptual space, and respectively different versions of what is defined as "the social". One of the most often cited elaborations is Dautenhahn (2007). She distinguishes three principal perspectives on human-robot-interaction (ibid: 683pp): a robot-centred, a human-centered, and a robot cognition-centered perspective (that focusses on cognitive models and social skills of

ent design philosophies: The hierarchical paradigm (playing chess), the reactive paradigm (starting from building insect-like behaviors), and the hybrid paradigm (with is kind of a compromise between the two), which in recent years seems to be the mostly accepted design philosophy in robotics.

the robot). Within this space, she distinguishes five conceptual approaches to HRI, as shown in figure 2.

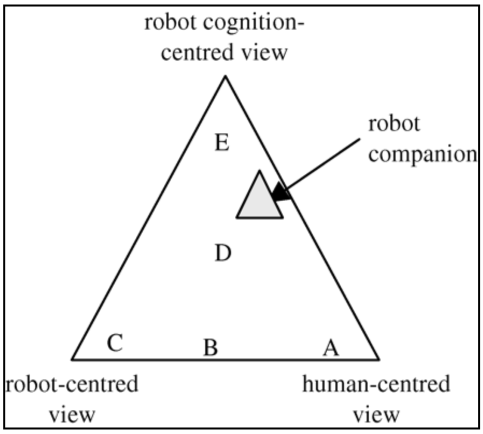

Figure 2: The conceptual space of HRI approaches, with positioning of the robot companion (Dautenhahn 2007: 686)

In the robot-centered corner (C in the figure) is the *"sociable robot"* that is equipped with a built-in drive to engage with human users – this is the "robot-as-creature view". The only requirement here (B in the figure) is that it can act in and react to a societal environment. This is the conceptually weakest approach and close to the usual approach in Service Robotics (see above).

On the opposite side of the spectrum, in the human-centered corner, is the *"socially evocative robot"* (A in the figure) that should evoke positive feelings by the users and a perception as being useful. In this approach the reasoning process of the robot and its concrete behavior do not matter in principal as long as the evocation occurs. But in the field it is widely assumed that a human-like shape, size and behavior of the robot will make the occurrence of evocation more likely – that is one reason for the popularity of anthropomorphism in robotics.

In the robot cognition-centered corner (E in the figure) is the *"socially interactive robot"*, that "possesses a variety of skills to interact and communicate, guided by an appropriate robot control and/or cognitive architecture" (ibid: 684). It requires a "deep modeling" (ibid) of human cognition. This definition forms kind of a docking point for the robot sociologicus.

But Dautenhahn introduces another definition, the *"socially intelligent robot"* (D) and gives it a more specific meaning, which explicitly stems from the traditional AI view of intelligent machines that "behave similarly to a human" (ibid). This is quite obviously an approach that does not fit into any of the other approaches. So staying in the logic of the figure, it would make more sense to extend the figure to a square with the classical AI approach as another corner in the figure - as far as AI includes social behavior as an important part of understanding (or building) intelligence[10]. This perspective is robot-centered, but it differs from the more technical view of the "sociable" robot as it uses the robot as a tool for understanding the grand themes of AI like intelligence, evolution and the mind. The figure, then, would have two axes and look as shown in figure 3.

To sum up with respect to the question of the outline of the field: With the inclusion of concepts and empirical investigations of human-robot interaction in the core of the field (instead of "I-Methodology"), and a conceptual space which allows to relate different approaches to one another, the field of Social Robotics surely looks different from the robot technologicus and the massive heterogeneity of Service Robotics. But looking at the figure also reveals that concepts (or metaphors) of "the social" involved are very different.

[10] Most of the approaches to the "Novelle AI" – the "artificial life route to artificial intelligence" (Steels/Brooks 1994) – are no longer inspired by models or metaphors from the philosophy of mind or psychology, but from biology, from the theory of evolution or from anthropology dealing mainly with animal intelligence or with early stages of human societies (like the widely discussed "social brain" hypothesis), but not with actual societies.

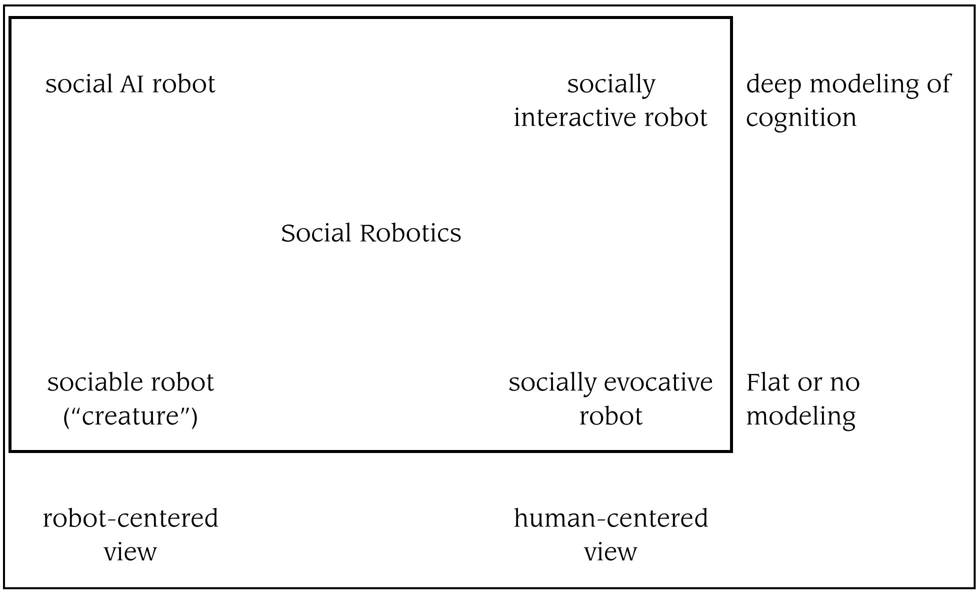

Figure 3: Extended version of the conceptual space of Social Robotics

With the possible exception of the sociable robot, which is only interesting from a technology-assessment perspective, for all approaches a sociological contribution could make sense. For instance, the investigation of the socially evocative robot could take its starting point at the concept of attribution of agency to the robot, a question that can be empirically investigated. But I think that there is good reason to link the socially interactive robot with an explicit modeling inspired by the key point of sociological theory of action: the reduction of the complexity of social situations, which in Social Robotics appears in the first place not as a problem of theory or concept, but as a methodological problem.

3 The complexity of social situations and the problem of comparability of HRI-investigations

As depicted in the graphical overview of the field above, there are many conceptualizations and empirical investigations of the robots' acceptability and usability, and of patterns of man-robot-interaction in Social Robotics. And there are many explicit critiques of purely machine-centered approaches. Sabanovic (2010), for example, envisions an integrated practice of what she terms "designing from the outside in" (ibid: 447):

> "Iterating between real world observation, technology design, and interactive evaluation allows for emergent meanings and interactions to drive the development of robotic technologies. In the process of outside-in design, the constraints are defined by empirical social research and the social context of use, rather than technical capabilities, and the final evaluation is based on the subjective experiences and opinions of users, rather than internal measures of technical capability and efficiency" (ibid).

This is kind of a radical version of the human-centered approach outlined above, that in some sense could also be understood as an application of constructive technology assessment with iterative steps between developers and users, and respective "promise-requirements-cycles" (van Lente 1993, Rip/Shot 2002: 160pp). But such an iterative approach is only suitable for single projects in transdisciplinary cooperation where a societal (and not a scientific) goal is the main focus – and where this goal is undisputed, which is seldom the case in a purely scientific context.

Unlike in transdisciplinary cooperation, for an interdisciplinary field to emerge in a distinctive sense it is firstly important to balance the disciplinary perspectives involved, and secondly to determine criteria for a comparison of different robotic solutions and the findings of different investigations of user experiences and different settings of human-robot interaction. Only in this way, a state of research or a state of technology can be reached.

This necessity to determine such criteria for comparison is widely acknowledged in the field of Social Robotics. There is a call for metrics and taxonomies in many articles, and a broad meta-discussion on related methodological issues. But to determine criteria for comparison is not easy at all for a technical apparatus that is not built to be useful in standardized settings (which can be judged by clear-cut criteria for good system performance like goal achievement). So the problem for evaluation and comparison is not only the "incredibly diverse range of human-robot applications" (Steinfeld et al. 2006: 33). Even from a purely "robot-centred" view, there is a variety of physical characteristics of the settings of investigation. And not least there are human actors in these settings – whose prior experiences, actions and roles are hard to standardize, and who interact not only with the robots, but also with each other. These are typical dimensions of the complexity of social situations. And this not only holds true for the obviously challenging list of characteristics of a socially interactive robot as shown above, but also for rather simple devices that no one would characterize as intelligent in a human way. A good example for this is the empirical study of Sung et al. (2010) that shows convincingly how complex the interplay of a robot, the physical environment and human actors is even in an – at first glance - easy situation: the introduction of standard vacuum cleaning robots in domestic homes.

Acknowledging the complexity, statements about social situations are quite common in the field:

> "Evaluating the interaction [with a robot] is complicated by the fact that there is a whole plethora of ways in which the interaction can be considered, from task-orientated to social and evaluated quantitatively or qualitatively. Therefore, it can prove difficult to find standardized dimensions to analyze different HRI experiments" (Salter et al. 2010: 405).

There are different ways to tackle the problem of comparability of HRI-studies. One way only seldom mentioned in Social Robotics (and never really exemplified in depth) would be to develop a benchmark for optimizing human-robot interaction. This would be a standardized setting, or a test bed, combined with a measurable goal for different robotic solutions, just as it was established in for Search and Rescue Robotics and of course in RoboCup (soccer playing robots evaluated by the simple benchmark to score a goal). These play-like settings with their rules are a way of reducing complexity for the sake of comparability. It is obviously very demanding to find a standardized setting *and* a common goal that is directly measurable as an indicator for success in complex situations. Nonetheless, from my view it is astonishing (and maybe only explicable by the cultural gap between classic AI, from which RoboCup emerged, and HRI and Social Robotics) that RoboCup@Home[11], a tournament setting in which the robots have to solve the same tasks in a domestic setting, is not considered at all in the discussions in Social Robotics.

In Social Robotics, there are two main approaches to the problem of achieving comparability: a stricter modeling based on quantitative data and a more interpretative sorting of data mainly from qualitative observations. Many of these approaches are imported into Social Robotics from HCI, but there is a broad agreement that the domain of interaction with robots is more complex than interacting with computer systems via an interface. Hence, many

[11] See http://www.robocupathome.org/.

authors suggest that the models of HCI have to be extended appropriately.

A prominent voice from robotics calls for a combination of both approaches to foster the strengths of different methods to counterbalance their possible weaknesses (an approach known in sociology as triangulation). Bethel/Murphy 2010 summarize the existing approaches to HRI in five methodological types (which they term "primary methods"): self-assessments, behavioral observations, psychophysiological measures, interviews, and task performance metrics. Drawing on that, they recommend to apply "three or more methods of evaluation" (ibid: 358) in each empirical investigation (for the same robot examined in the same situation)[12]. This recommendation is, as described above, directly connected to the advance of the field of Social Robotics as a whole:

"The use of ... three or more methods of evaluation can provide validity and credibility to the human studies that are performed associated with HRI. This will improve the overall field, but also will result in stronger public acceptance of robots. ... Additionally, the engineering community will be able to use the information obtained from well conducted user studies to design and build better robots" (ibid: 358).

Taking aside the notorious methodological problem of combining quantitative and qualitative studies (a gulf in many sciences, and certainly in sociology), both sides face specific problems with the complexity of social situations. I will proceed by giving one example for each side to illustrate what seems to be typical.

[12] In addition, Bethel/Murphy (2010) suggest to increase the sample sizes (number of probands) of the empirical cases. This is good advice in principal, but often hard to achieve in project-driven (and financed) research. And of course important insights or hypotheses that direct further research emerge quite often from individual projects or observation that do not fit methodological requirements like adequate sample size: "Media equation" or" uncanny valley" are but two examples for such influential hypotheses for the field of Social Robotics.

The Quantitative Side

To start with the quantitative side, a typical example is the extension of the TAM-model ("Theory of Acceptance Model") for robotics applications proposed by Heerink et al. (2010). They aim at the proof of a model that consists of the variables that are crucial for the acceptance and the actual use of a robot, in their case an assistant robot for care of the elderly. In a first step, they present a universal model for the influences on acceptance of computer technology called the UTAUT model ("Theory of Acceptance and Use of Technology") as depicted in figure 4.

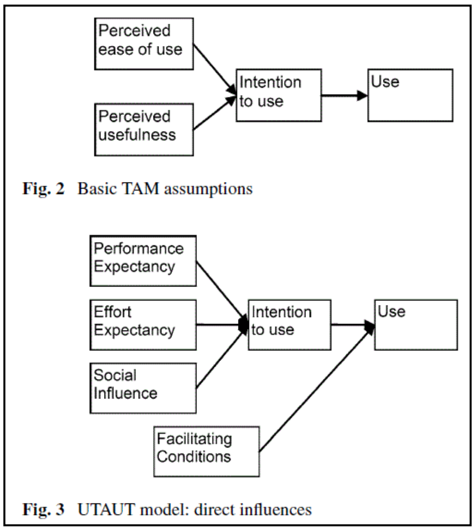

Figure 4: TAM and UTAUT models (Heerink et al. 2010: 363)

In the next step the authors claim that this model has to be adapted to the specific characteristics of the domain assistive robotics. Drawing on the results of many other studies, additional variables are added to the model, especially "perceived enjoyment", "social presence" and "perceived sociability" of the robot and "trust" in the robot (ibid: 363pp). All these variables are then operationalized as items in questionnaires for probands who interact with different robots. The resulting empirical data (answers of probands respectively "measures") are computed using multivariable statistics. The overall model resulting from a series of empirical investigations, including the significance of statistical correlations, looks as shown in figure 5.

According to the authors this resulting model "can be used to predict and explain acceptance of assistive social robots" (ibid: 373). Because the variables are expectations (intentions and perceptions of the situation), the resulting model can be understood as a cognitive model that can be empirically tested and extended by inclusion of the results of other research projects. So it seems it can do a good deal with respect to the problem of comparability. But this potential strength comes at a prize: First, an average (or ideal) user is constructed by statistical aggregation, while of course real users might dramatically differ. Also, a model that does not take differences in kind of intention, expectation or perception into account may be dramatically oversimplifying. Even more importantly for the meaning of the "social" robot, the model must be kept sufficiently simple regarding the number of variables to allow multivariable statistics to work – which is a conceptual reduction of the complexity of the social situation. And this reduction here is somewhat arbitrary – as in many of the examples mentioned above, there might be a huge amount of other influences (possibly important variables) that shape any situation at hand. It seems that, in order to keep the model calculable, complexity is faded out by the determination of the items in the questionnaire. For instance, the variable "perceived sociability", described as "the perceived ability of the system to perform sociable behavior"; ibid: 364), is operationalized only through the following items of the questionnaires:

- "I consider the robot a pleasant conversational partner.
- I find the robot pleasant to interact with.
- I feel the robot understands me.
- I think the robot is nice" (ibid).

This obviously is not sufficient for what is meant by any of the approaches to the "social" in Social Robotics.

So without playing down the general strengths of quantitative approaches, there is no criterion for keeping the model simple enough to avoid an explosion of variables and items. An architectural backbone that could link the cognitive model with the problem of comparability seems to be missing.

The Qualitative Side

In Social Robotics, there are some attempts to fix a state of the art also for more qualitative HRI-studies, which typically present a huge list of necessary aspects of or determinants for "good human-robot interaction", divided into main dimensions. Inter-

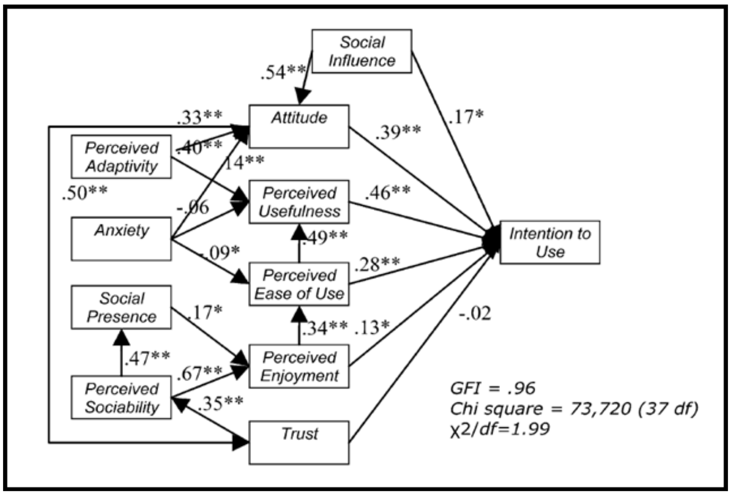

Figure 5: Resulting model of robot acceptance (Heerink et al. 2010: 372)

Table 1: The Methodological Mix

	Methods	Expert Eval	User Studies	Questionnaires	Physio. Measures	Focus Groups	Interviews
Research Objectives							
Usability							
	Effectiveness	X	X				
	Efficiency	X	X				
	Learnability	X	X				
	Flexibility	X	X				
	Robustness	X	X				
	Utility			X			X
Social Acceptance							
	Performance Expectancy			X		X	
	Effort Expectancy			X		X	
	Attitude toward Using Technology			X			
	Self Efficacy			X		X	
	Forms of Grouping			X		X	
	Attachment			X		X	
	Reciprocity			X			
User Experience							
	Embodiment			X		X	
	Emotion			X	X	X	
	Human-Oriented Perception			X			
	Feeling of Security			X	X	X	
	Co-Experience			X		X	
Societal Impact							
	Quality of Life			X		X	X
	Working Conditions			X		X	X
	Education			X		X	X
	Cultural Context			X		X	X

Figure 6: Overview of the USUS framework (Weiss et al. 2009: 6)

estingly, the two main views on Social Robotics sketched above (following Dautenhahn's account) find their twin here. In the "robot-centred view", dimensions of technical performance are the core dimensions, as in Steinfeld et al. (2006). Dimensions there are (1) navigation, (2) perception, (3) management, (4) manipulation, and, added at the end of the row, (5) social. On the opposite side, in the "human-centred view", e.g. Bartneck et al. (2009) present the following dimensions: (1) anthropomorphism, (2) animacy, (3) likeability, (4) perceived intelligence, and (5) perceived safety, leaving all technical aspects out of the picture at least on this highest level of categorization. Again, I will only shortly present one example for the latter type of sorting of relevant aspects.

Weiss et al. (2009) present an overview of approaches to the evaluation of human-robot interaction. Their focus is on the question "if people experience robots as a support for cooperative work and accept them as part of society" (ibid: 2) and thus claim to give a holistic view on the evaluation of humanoid robots. Their framework has the acronym USUS meaning "usability, social acceptance, user experience, and societal impact" (ibid), and combines these major dimensions with appropriate methods of empirical investigation (see figure 6).

In contrast to the stricter modeling and the quantitative measures depicted above, this framework is explicitly meant to support "formative evaluation" (ibid: 5). It sorts possibly relevant factors for achieving better robotic solutions, where "better" is judged by the human users. So this approach does not aim at kind of metric. But there is no principle for sorting the potentially important aspects, and thus the range of possibly relevant aspects cannot be restricted. So the individual findings of diverse investigations of human-robot interaction cannot be compared.

To sum up briefly: There is awareness of the problem of the complexity of social situations both on the quantitative and the qualitative side of HRI-investigations, but there seems to be no principal solution in sight for the 'complexity gap'. In her encompassing overview of studies of robots in elder-care robotics Flandorfer (2012)[13] sur-

[13] The special interest of Flandorfer (2012) is to show the manifold and interrelated influence of sociodemographic factors on the acceptance of robots for care of the elderly. But it turns out that not only the classical sociodemographic factors like

renders faced by the exploding number of factors:

"We may assume that the more research will be done, the more methods will be developed" (ibid: 9).

4 Examples for generalized perceptions and expectations from the field of Social Robotics

As briefly summarized above, human actors, at least from a sociological perspective, do not face the problem of exploding complexity when confronted with all the potentially relevant aspects of social situations – in most cases they simply follow generalized expectations, and even their perceptions of the situations are very selective and just as generalized. Evidence for this can be found in many of the listings of relevant aspects in the Social Robotics and HRI-literature. In Weiss et al. (2009) for example, "forms of grouping" and "cultural context", especially the national style of practical perception and handling of technology (exemplified by the case of Japan; ibid: 3) are mentioned, but only conceptionalized as some influential aspects of many. But belonging to a group or culture means to narrow the space of perception of and reactions to a new technology based on prior experiences of the collective – again a possible (and in social reality practiced) means of reducing complexity. In what follows I will only give three examples for this general idea.

The Wildness of Situations and Trust

One important dimension of the methodological problem of complexity left aside so far is often mentioned in the HRI-literature: The problem of the "wildness" of the situation of investigation. On the one side, there are laboratory experiments, where the whole situation is thoughtfully arranged to be as methodologically clear as possible. On the other side are empirical investigations in realistic settings that are hardly methodologically controllable. In Social Robotics, this issue is described as a trade-off between methodological reliability (e.g. clearly distinguishing the dependent variable from all the possibly infinite independent variables) and realism:

"Experimenting in real-world environments can provide both many benefits and also its share of difficulties. Certain experimental settings may create difficulties, such as the environment may be too challenging for the capabilities of a robotic device. ... Changing or engineering the environment may be necessary to address specific research questions and experimental methodologies. However, this may have varying effects on users or participants. For instance, controlled conditions help to conduct rigorous, quantitative, statistically significant analysis, but may also create an effect on the outcome. … All the difficulties involved in real-world experimentation may explain why it is difficult to replicate experimental HRI scenarios" (Salter et al. 2010: 406).

As a possible solution, a taxonomy (as precursor for a metric) is presented in terms of control. It looks as shown in figure 7.

Again, it is obvious that the six dimensions of control, especially in their combination, include so many possibilities that it is unclear how this could guide architectural or methodological decisions.

But with the discussion of a positive side of "wildness" the whole idea of total controllability of the robot, the human and the situation becomes questionable. How do human actors solve the problem of uncontrollability of situations? One solution widely

age, gender, family status and income are important, but also technological experience or cultural background. Moreover, the study is well aware of methodological problems like changing of results depending on whether the probands had prior experiences with robots or not, or the shaping of the setting of investigation by the ageing-and-innovation discourse, especially by stereotypes common in the engineering discourse (see Peine/Neven 2011 for this point), and generally of the problem of comparability of these studies. This was a strong inspiration for this article.

Int J Soc Robot (2010) 2: 405–415 407

Table 1 Levels of control in relation to Participant and Robot influences

	Level of control				
	None	Low	Medium	Moderate	High
	1 2	3 4	5 6	7 8	9
PA	Free	Natural	Comfortable	Directed	Controlled
PG	Large	Medium	Small	Paired	Singular
PE	Free	Natural	Familiar	Adapted	Sterile
RA	Autonomous	Fixed	Combination	Wizard of Oz	Remote-Controlled
RG	Plethora	Multi-Agent	Robot+Anim.	Robot+Inanim.	Singular
RE	Open	Secured	Challenging	Engineered	Controlled

Figure 7: Taxonomy of the wildness of situations of human-robot interaction, Salter et al. (2010): 407: P = human participant, R = robot, A = autonomy, G = group, E = environment

acknowledged in social theory is to trust interaction partners, and this idea is also discussed in HRI, for example by Yagoda/Gillan (2012). The authors cite the common sociological definition of trust as

> "the willingness of a party to be vulnerable to the outcomes of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (Mayer et al. 1995: 712).

They then ask for the conditions under which humans would trust robots. Before exploring this abstract consideration any further, the authors turn, again, to the development of a measurable scale which consists of rather conventional aspects of control from workflow-management like "dependability", "competence" and "reliability" (Yagoda/Gillan 2012: 242pp). Thus the potential of trust to reduce complexity is not considered at all on the side of the humans. Furthermore, looking at human-robot interaction, it seems viable to apply the abstract principle of trust to the modeling of the robot. A robot that by purpose is helpless in some respect and asks trusted humans for appropriate help would be – against the dream of the robot technologicus – a realization of this principle. The empirical investigation of "robots asking for directions" (Weiss et al. 2010) could be interpreted in this way, because here the functionality of the robot is dependent on people's willingness to help the robot achieve its task. This principle seems to guide many artistic approaches to human-robot interaction[14] (cf. Kac 1997).

Social Roles

Taking on roles is in sociology known as one major principle to reduce the complexity of social situations. Perceiving interaction partners via typical roles and sending cues that one is acting according to a recognizable role makes it unnecessary to take all the possibly relevant aspects of individual actors into account, and makes it possible to choose actions that fit the normal expectations that are attached to that role. In HRI, the principle of role-taking is mainly applied when modeling typical patterns of human-robot interaction. Like in HRI in general, these approaches come in a more explicitly modeled and quantitative version, and in a more empirically derived qualitative one (see above). Again, I will sketch out only one example for each version, both of them widely cited.

The first version, initiated by Scholtz (2003), is derived from a general

[14] These approaches of course are from the world of arts, aiming at performances (often poorly documented) and by no means are appropriate for a methodologically controlled investigation of human-robot interaction. Nonetheless, my personal favorite is Nam June Paik's Robot K-456; see http://cyberneticzoo.com/?p=3437.

framework of human action and distinguishes five principle roles as a basis for an empirical evaluation of human-robot interaction: The roles of supervisor, operator, mechanic, bystander and teammate, where only the last three ones can also be found in human-human interaction while the first and the second one are specific for human control of robots (think about the discussion of grades of "wildness" as opposed to total control of the machine outlined above). These five roles also determine principle types of action that are defined by these roles and aim at guiding empirical research, knowing that "a research challenge will be what generalizes between different domains" (ibid: 9). So this is by purpose a top-down approach.

The second version of a role-based approach, as initiated by Kahn et al. (2008), suggests to identify "design patterns" in a bottom-up way. These are "fundamentally patterns of human interaction with the physical and social world" (ibid: 98) which can be understood as episodes of perception of and interaction with technology that appear often (if not always, meaning these patterns are universal, a claim that is debated) in the same way. Patterns like "initial introduction", "in motion together", "recovering from mistakes" or "reciprocal turn-taking in game context" (some of the patterns observed by the authors in robotic experiments with children) define interaction roles for the humans and the robots involved in the episodes.

This approach originated in architecture and has been broadly imported to HCI, usability research and HRI. The approach does not draw on one abstract model, but derives types (patterns) from various sources, which comprise empirical investigations and engagement in an iterative design process, but also a "philosophical base of what counts as fundamental constructs in human-human interaction" (ibid: 99). The ultimate goal is to build up and extend a model kit of such patterns of human-robot interaction.

Again, the range of aspects that are possibly relevant for these patterns is large. But the authors are well aware of this for the aim of reusing patterns that have been tested (with other robots and in other situations) and therefore strongly stress the issue of levels of abstraction of the patterns: Patterns should be "specified abstractly enough such that many different instantiations of the pattern can be realized in the solution to a problem" (ibid: 98).

A "really social" robot in this sense should not only 'know' about interaction roles; it should also be able to 'read' signals to infer what roles or interaction patterns are relevant for its situation. Such a 'reading' of signals is not at all trivial for a machine even with a rather simple set of tasks (and requires more or less lifelong learning of humans). Kuo et al. (2011) tackle this problem with an extension of the interaction pattern approach. They introduce "cue-oriented design patterns" which start from "interaction cues (or social cues) that a robot can perceive and act upon or express in an interaction. These cues can be verbal, non-verbal or a combination of both" (ibid: 446). Just as in human social life, 'reading' such cues correctly would 'tell' the robot whether and when it is expected to take the roles of initiator or responder in a given situation. So while addressing a rather technical problem (task analysis), the authors work on a cognitive model of the interaction and thus the robot itself. Like Kahn et al. (2008) before, Kuo et al. (2011) emphasize the issue of level of generalization:

> "Setting the right abstraction level for design patterns is the key to ensure reuse of the pattern and construction of more complex design patterns" (ibid: 446).

Working on this issue could not only result in an ordering principle that could convert a sheer model kit of patterns into a sorted repository, and with respect to the problem of reuse could lead the way to the answer of the question of comparability. The issue of generalization of empirically derived interaction patterns can also be interpreted from a sociological point of view as an interesting operationalization for the analysis of interaction at the micro-level of sociality – and we do not have many concepts or methodological tools for determining what social scripts are in concrete.

From sociology we know that social roles can not only be conceptualized on the micro-level, but also on the meso-level, the level of organizations. Starting with Barley (1986) numerous studies from the sociology of technology and organization studies have shown that the introduction of new technology leads to major changes in the arrangement of professional roles and hierarchies in organizations, e.g. the distribution of professional expertise and power relations between patients, nursing staff, doctors and technical people in a hospital or nursing home. And for a robot to act "really social" one would expect that it is at least able to recognize patients, nursing staff or doctors – or just passers-by. So it is astonishing that any analysis of roles on this level is widely missing from Social Robotics[15].

Society at Large: The Macro-level

As on the micro- and meso-level of sociality, a "really social" robot should also be able to perceive and act upon generalized expectations on the highest level of scale, the level of expectations taken for granted by human actors in society at large. In Social Robotics, there are two broad strands of discussion on this level of scale, and I will only briefly mention them to complete the picture, because all of these (and presumably other) strands of discussion are of course equally worthwhile (and disputed).

The first strand of discussion is Robo-Ethics (Veruggio/Operto 2008 and Decker/Gutmann 2012). While there is a flourishing debate about the possible juridical and moral accountability of highly developed robots, the actual problem in robot development is more down to earth: to implement rules of socially acceptable robot behavior that go beyond the big red "Stop!"-button and obstacle avoidance sensors robots use today. The problem here is of course that in modern societies there are but few fundamental institutions that are undisputed. Moreover, different macro-level expectations might be in conflict with one another. To mention but one example: We would expect a robot not to cheat its users. But interesting experiments (Short et al. 2010) reveal that some cheating behavior makes the robot more "human-like" and thus adds more social possibilities to its overall behavior. So it might be good advice to address this issue only for the specifics of different domains (the solution of Veruggio/Operto 2008 and the existing Robo-Ethics roadmaps).

The second strand of discussion deals with the issue of different national robotics cultures both regarding the development and the use of robots. Almost everyone agrees that especially the East Asian robotics culture differs strongly from the western one (see cf. Matsuzaki 2010). There is quite a lot of quantitative, questionnaire-based research on question of different national styles, but the results are arbitrary or even contradictory. While e.g. Han et al. (2009) summarize:

> "Culturally Europeans recognize robots as machines for labor, while Japanese and Koreans consider them as friends" (ibid: 101),

[15] There are only some studies from a more managerial viewpoint that ask for changes in the work-flow due to the introduction of robots in work organization. One exception is Mutlu/Forlizzi (2008) who report that the job definition (including hierarchies) and workload of the professionals plus the interruptability of routines of collective work are main factors for acceptance of robots in hospital environments.

MacDorman et al. (2009), using basically the same methods, found no strong evidence that "Japan really has robot mania". From the qualitative side Wagner (2009) questions the three most prominent cultural arguments for a specific Japanese way of robotics: "Historical antecedents of robots in Japan", "religious preconditions of the Japanese interaction with robots", and "Astro Boy as a role model for a friendly robot companion". Though this interesting research question seems not to be settled yet, one would expect that a "really social robot" should be able to recognize the national culture in which it has to perform, to react adequately.

5 Drawing thoughts together: An outline of the robot sociologicus

Taking all the generalized expectations on different levels of scale collected above together (and further elaboration of course would add more of them), it seems to be possible to translate the question about the robot sociologicus into a blueprint of the architecture of this robot – or at least into a fundamental structure of its reasoning process. To do this, first of all a decision about the architectural principle (the "design philosophy") has to be made. Normally in robotics (and in AI) these are principles from cognitive science, biology or psychology. But understanding the term of the "really social robot" from a sociological point of view, this of course means to try to apply a sociological principle to the main architectural decisions.

As already mentioned in the introduction above, when describing the sociological concept of generalized expectations, Esser's general theory of social action can serve this purpose (see Fink/Weyer 2011 and this thematic issue for a similar approach, but with a different goal). Esser not only stresses the importance of routine action, but combines in his modeling the SEU-approach of rational choice (the individual calculation of "Subjective Expected Utility", SEU) with a richer concept of social situations from the tradition of symbolic interactionism as well as with Goffman's concept of frames and Schütz's concept of social action that is planned 'in the head' of the actor in "modo futuri exacti", which means that the action at hand is chosen by searching for past actions that are "typically similar" to the actual one[16].

Esser's model of action can be described in a condensed way as a seven-stage model:

(1) If a situation is perceived as a call for action, all relevant aspects of this situation are condensed to a "mental model of the situation", a so-called "frame".

(2) It is justified whether this actual frame "matches" sufficiently an already familiar frame in the memory of the actor. The result of this comparison is decisive for the attitude towards the situation, called the "mode". If there is a match, the "automatic-spontaneous mode" is selected and the known frame from the memory is applied without any further reasoning. If there is no match the "reflecting-calculating" mode is selected and a new frame is developed.

(3) Based on this framing of the situation, a mental model of action – a "script" – is selected, with consists of a model of an isolated episode of action combined with a respective expectation of successfully accomplishing that episode.

(4) As with the chosen frame, the script is also justified whether there is a suf-

[16] The famous original formulation is: "I base my projecting of my forthcoming act in the Future Perfect Tense upon my knowledge of previously performed acts which are typically similar to the prescribed one, upon my knowledge of typically relevant features of the situation in which this projected action will occur" (Schütz 1982: 69).

ficient "match" with an already known script in the memory of the actor. In case of match this script is applied in the "automatic-spontaneous mode" without any further mental activity – this is the case for routine action. In the case of no match in the "reflective-rational mode" a new script is developed.

(5) Only after this mental anticipation is completed, the visible action itself is conducted, which then is only an execution of the result of the inner reasoning.

(6) The success of this executed action is judged by the actor.

(7) The whole episode of reasoning and the judgment of interaction success, including all expectations about aspects of the situation and action episodes that make up a "match", are finally stored in the memory, which extends the repository of 'tested' frames and scripts (see Esser 1999: 165ff, 355ff, and Esser 2001: 239ff, 295ff).

In the sociology of action many aspects of Esser's model, as usual in sociology, are strongly contested, not least the SEU-approach in Esser's version of "expected model utility", but this is not relevant for the very general consideration about the architecture of a social reasoning process here[17]. Also the strict dichotomy of the two modes ("automatic-spontaneous" versus "reflecting-calculating") is criticized, with the suggestion to to either further develop the core model (see cf. Kroneberg 2005) or to put the basic model on different grounds (cf. Schulz-Schaeffer 2008 who suggests to replace the function of the two modes for frame selection with three different kinds of definition of the situation). For the question of the transferability of the basic model from sociology of action to the architecture of a "really social" robot it is only important that there is a principal modulation of the attitude towards the situation (Esser's "modes" or some architectural equivalent for these) while frames model the handling of the concrete situation at hand – both architectural considerations in combination describe a way to drastically reduce the complexity of the situation.

If we now just fill in the different forms of relevant generalized expectations outlined above (from the discussion in the field of Social Robotics) into this form of a reasoning process, the architecture of the robot sociologicus could look like shown in figure 8.

Despite being a rather crude picture this architectural blueprint tries to catch central aspects of a 'really social' robot from a sociological perspective while working with central issues from the discourse of Social Robotics itself.[18] It seems to be in line with Dautenhahn's "socially interactive robot" depicted above by explicit "deep modeling" of the cognitive preconditions of social interaction.

And by highlighting reduction of complexity as the central modeling principle the blueprint is opposed to the "socially evocative" as well as the "sociable robot" in Dautenhahn's terms – or, to put it in the more metaphorical terms I use throughout this aticle, it stands in sharp contrast

[17] Of course, following a SEU-approach would become relevant if not only the general architectural principle was applied to a robot's architecture, but if also the SEU formalism was used for the concrete mathematics of the reasoning process.

[18] But it remains a question for sociological theory of action whether the integration of more specific instantiations of generalized expectations into the overall Esser model theoretically really works out. It seems plausible to me that the application of roles, trust etc. on a higher level of abstraction can be modeled as the results of framing, script selection and judgment of the results of visible action only in a generalized way. But this is not part of the original concept and has to be verified – and will eventually have an influence on the modeling itself.

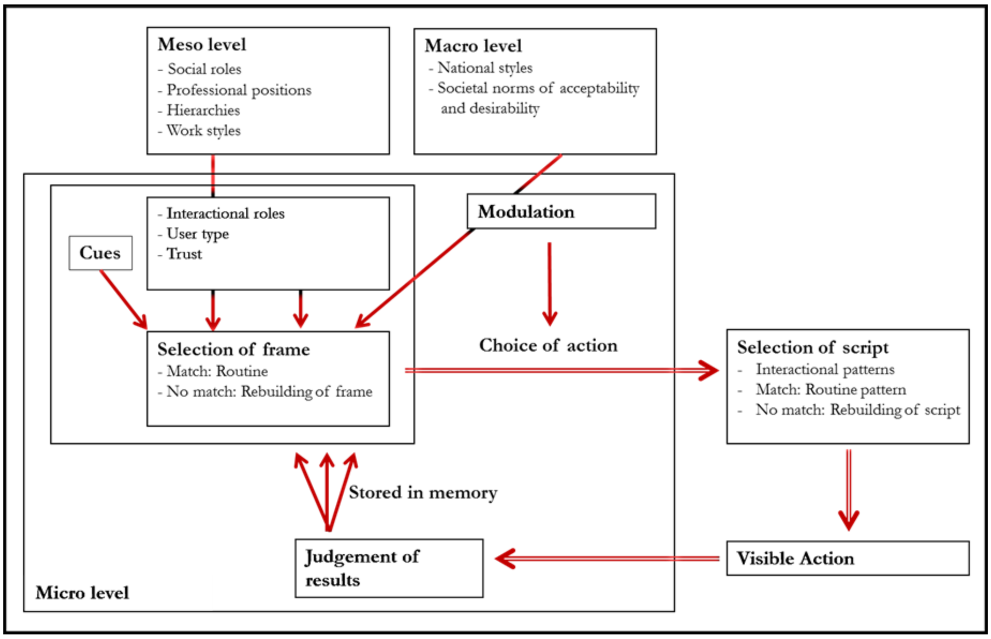

Figure 8: A possible blueprint of the architecture of the reasoning process of the robot sociologicus

to the robot technologicus[19] with it's problem of an explosion of potentially relevant "aspects of an unstructured environment". This general problem is especially obvious if the environment is "wild" and thus conceptually and methodologically challenging.

The blueprint specifies reduction of complexity as the general solution in two ways:

First, it highlights different forms of generalized expectations (on different levels of scale) as a central part of the framing of the situation. While generalization evidently is reduction of complexity, this effect is supported by the social solution for the problem of perception of adequate expectations: cues are sent, interpreted and institutionalized to point to an adequate perception of social roles, to hierarchies, to initiator or responder roles in interaction etc.

And second, the architectural blueprint applies the cognitive model of a "match" between the perceived actual situation and situations experienced (or tested) earlier and stored in the memory, leading to routine action which is possibly the most drastic form of reduction of complexity known in sociology.

6 Placing the proposed approach in the larger context of discussion

The suggested general approach can, from a social sciences' or humanities' point of view, be crudely positioned in equidistance from the two dominant poles in the discussion about the possibility of realizing 'intelligent' (or 'social') machines – AI-critique on the one hand, and social constructivism on the other hand.

In the well-known tradition of AI-critique, any claim of a full-fledged realization of human-like thinking or action on machines is criticized with the argument that substantial features of human thinking or acting can never be

[19] I should clarify that the metaphor of the robot sociologicus only works if opposed to the robot technologicus. It does not work as well in sociology itself, firstly because the homo sociologicus is a pure rule- and role-follower, which is not the same as following routines in most cases, and secondly because in the cases of reflective rebuilding of frames and scripts a robot – simply because it is a machine programmed for specific purposes, and calculates its utility – unavoidably shares features of the homo oeconomicus.

grasped, or even mimicked, in a meaningful way by any type of machine intelligence. To mention but two of the most important examples of this argument: It is claimed that machines (like robots) are not able to play chess successfully because they are only able to compute numbers, but not to understand the rules of the game. And it is claimed that machines are in principle not able to understand the semantics (and hence the sense) of the symbols they can process – the "Chinese Room Experiment" is the best-known formulation of this fundamentalist argument (Searle 1980; Searle 1986).

But if exaggerated visions are put aside, many of the features that AI-critique claimed to be impossible for machines in fact turned out to be technically achievable, not in the way envisioned as "hopeful monstrosities" (Schot/Rip 1996: 255), a point of departure for many innovations, but as a working solution that evolved over many steps, many negotiations and of course many failures. Moreover, with respect to the development of the R&D-fields of AI and especially robotics, major arguments from AI-critique often have been translated into straightforward technical challenges. For example, New Robotics with its focus on embodiment and situatedness of intelligence (and hence the strong orientation towards biological models) echoed many critiques of the Old AI (or GofAI: Good old fashioned Artificial Intelligence) simply because "elephants don't play chess" (Brooks 1990). And even the linguistic basis of Searle's critique of AI is taken as a constructive starting point to enable a robot to understand the intended meaning of a human user via a "symbol grounding" approach (see cf. Lemaignan et al. 2012). In this approach it is explicitly not claimed that a 'really semantic' understanding can be reached, but a technical solution that functions in principal in a comparable way: a model for a "correspondence between symbols and sensor data that refer to the same physical object" (ibid: 183).

In sociology itself, there are only some versions of AI critique. Probably the best-known claim is the distinction of "mimeomorphic" versus "polymorphic" action proposed by Collins/Kusch (1998). The first type of action is introduced as rule-based only and context-free (like swinging a golf club) and thus can be accomplished by humans and machines alike. The second type of action depends on the application of tacit knowledge of the cultural characteristics of the situation at hand – a capability no machine can ever achieve. This sociological critique of AI is not in the first place meant to be a critical contribution to technical developments, but warns against a wrong picture of human action to prevent treating humans like machines, especially a reduction of human skills and competence to "mimeomorphic action" in work settings, resulting in deskilling and alienation in practice.

Many of the contributions from philosophy and the social sciences to the flourishing debate about Robo-Ethics (see the overviews Veruggio/Operto 2008 and Decker/Gutmann 2012) point in the same direction. Conceptualized mainly for an advisory role for raising consciousness in the robotics discourse, it provides a long and without a doubt worthwhile list of possible negative implications of robots for societies and groups of humans. However, almost all of these issues are not specific to robotics, but can be formulated for any IT-technology. The only issue specific for robots and especially for potential companions, that is: for a situation where "we are going to be cohabiting with robots endowed with self-knowledge and autonomy" (Veruggio/Operto 2008: 1511), is formulated as the danger of "psychological problems" arising from a fundamental challenge or even breakdown of established categories: a "confusion between the real and the artificial" (ibid: 1512), resulting in "deviations in human emotions, problems of attachment … fears, panic … feeling of subordination towards robots" (ibid.).

In strict opposition to (or at least: ignorance of) the positions depicted, the sociological theory of action is totally agnostic with respect to these critiques about and warnings against losing the core meaning of 'the human'. Whereas in "mimeomorphic action" and most variants of AI-critique the point is to warn against any reduction of the richness and complexity of human reasoning and acting, the basic point in sociological theory of action is to model not the substance, but the abstract principle how actors are able to act at all faced with situations of potentially infinite situational complexity. And because it is only an abstract model that is transferred to the technical realm, this means that there is no equation of humans and machines in substance, especially not between human socialization and technical optimizing. So the whole idea of the robot sociologicus is not about artificial sociality in a substantial sense. The idea only relies on a transfer of an abstract principle to the architecture of a robot or the modeling of man-robot interaction. The implementation of any basic concept from sociology will always result in a more or less clever technical apparatus, with hardware, software architecture and algorithms, and with sensors (perception) and actuators (action/ behavior) embedded in its environment, which of course is quite different from human actors. Thus my overall argument may have its pitfalls (and of course has to be developed further), but is completely in line with the following statement:

"Relationships with computational creatures may be deeply compelling, perhaps educational, but they do not put us in touch with the complexity, contradiction, and limitations of the human life cycle. They do not teach us what we need to know about empathy, ambivalence, and life lived in shades of grey. To say all of this about our love of our robots does not diminish their interest or importance. It only puts them in their place" (Turkle 2006: 61).

Located on the other pole of the spectrum of discussion is social constructivism, which denies any substance in 'the human', nature and technique likewise, but treats literally everything that exists as the outcome of social processes of negotiation. Because this position is well-known, I concentrate here on one article from this camp that deals explicitly with Social Robotics. This article has already been mentioned above. Its title reads as follows: "When a robot is social: Spatial arrangements and multimodal semiotic engagement in the practice of social robotics" (Alac et al. 2011). Based on the ethnographic observation of experiments with human probands (preschool-children – toddlers – and their teachers) and robots in a classroom setting the authors depict in great detail how much the possibility and kind of interactions between humans and robots can change if there are even slight variations of the concrete observational setting.

But why do the authors characterize the robots they observed as "social" in the title of the article? The authors base their approach including the interpretation of the empirical findings in a strictly situational concept. The key point of this concept is the following: All parties engaged in the situation manage to reach a "multiparty interactional coordination [that] allows a technological object to take on social attributes typically reserved for humans" (ibid: 894). This stance consequently denies any substance of the nature of the robot (and towards all other elements involved including human agency):

"We claim ... that the robot is in fact social, but its social character does not exclusively reside inside the boundaries of its physical body or in its programming ... As the roboticists, toddlers, and their teachers engage in the design practice, the robot becomes a social creature in and through the interactional routines performed in the 'extended' laboratory" (ibid: 917).

Without any doubt it is an important insight that major changes of the situation, e.g. replacing other technological aids or people or even dogs with robots (as in the often mentioned scenario of robot pets for seniors) will alter the situation at hand (a household, a nursing home etc.) in a relevant way. And the authors convincingly point to the important role of the engineers respectively the roboticists themselves in the observational setting, an aspect mostly neglected in human-robot interaction research. But from the conceptual stance of rooting everything only in the situational dynamics stem, with respect to any investigation of or contribution to the field of Social Robotics, three conceptual shortcomings.

First, with a concept of complex and dynamic, ever-changing situations, it can only be shown for single cases *that* observational settings differ, but key factors leading to these differences cannot be identified, simply because there are too many candidates for such factors whose characteristics change permanently. In consequence, it seems near impossible to find a way to compare different empirical observations in different settings. This quite obviously creates a problem for almost every attempt to build methodological considerations on this general conception – the problem of comparability depicted for Social Robotics above.

Second, because the definition of literally every term is rooted in the details of the situation at hand, it is unclear from a sociological point of view how actors or the robots can orient themselves in situations – e.g. follow the "routines" (which are certainly generalized expectations) cited by the authors. More generally, from a sociological point of view it is hard to imagine actors that handle social situations without drastically reducing the situations' complexity by applying generalized expectations.

And third, from the viewpoint that neglects any substantial differences between humans, machines and other objects follows that literally everything can become "social" in nature, if only it is "enacted" in the situation at hand. Consequently, there is no principal difference between the "interactional achievements" that can be reached with a robot, a dog or a candy bar (drawing on Harraway, ibid: 915ff). This generalization might be criticized or not from a social science point of view. When applied to robotics as an interdisciplinary endeavor, it is surely worthwhile to remind engineers that they are not only creating artifacts, but in the same instance are creating society: Investigating the "robot's social character means one has to look beyond the robot's computational architecture and its human-like appearance and behavior" (ibid: 895). But engineers are trained to be engineers – for them, human environments are, in the case of robotics, the most complex and thus challenging context for an advanced technology. At this point, the authors' stance against any substantial attributes of robots or humans leads to advice that must sound strange in every engineer's ear, but also in the ears of everyone who has been ever involved in interdisciplinary cooperation with engineers: "Rather than controlling the machine, the robot's designers are called to participate in human-machine interactional and situational couplings" (ibid: 896).

These three shortcomings, consequences of the leveling of all substantial differences *and* any modeling decisions about principles guiding human (or robot) actions, make it dubitable that the phrase "when a robot is social" can be a reasonable starting point for any investigation of or contribution to the field of robotics. But these three shortcomings of a purely social constructivist stance also point to the benefits of the robot sociologicus for methodological considerations in the field, for sociology or interdisciplinary cooperation likewise.

7 Some possible uses of the robot sociologicus

If the architectural blueprint presented works to at least some extent, what are the possible uses of the robot sociologicus? There are at least three different answers depending on disciplinary perspective.

From the perspective of the development of the interdisciplinary field of Social Robotics, the conceptualization of generalized expectations could be a point of orientation for the problem acknowledged field-wide of comparability of empirical investigations in the light of the complexity of social situations. Instead of collecting an ever increasing list of possibly relevant situational aspects of human-robot interaction, grounding research on a containable amount of expectations that reduce situational complexity (like e.g. trust instead of controllability, roles on different levels of scale etc.) could orient empirical investigation towards a principle approved in a different domain – human societies. From the discussion of the examples above it seems to me that this could also be directly applied to down-to-earth methodological questions, e.g. the choice of appropriate issues for questionnaires in quantitative research or the focus of observation in qualitative research. The discussion has also shown that there are some points of contact between existing empirical studies and the principle of generalized expectations. But I am well aware that it is notoriously difficult to compare existing empirical studies by applying a new consideration. Nonetheless, given the acknowledgement of the overall problem in the field of Social Robotics, I think such an endeavor would be worthwhile.

From a purely sociological perspective, there are several interesting questions about the robot sociologicus. From a reconstructive perspective – the sociological reconstruction of the whole field as an interesting case for the sociology of technology and innovation – it would be very interesting to empirically investigate in greater depths how, how far and why a formerly massively heterogeneous field (Service Robotics) turned to a distinct field with at least partly shared goals over a vast array of disciplinary orientations. One crucial point here is not only the possible unification of concepts, but the further development of methods to deal with the issue of reduction of complexity, on the quantitative as well as on the qualitative side of research, and of a possible institutionalization of metrics and benchmarks for 'good' human-robot interaction. These issues are obviously of great interest both from a classical constructivist as from a socio-technical constellations (cf. Rammert 2012) point of view.

But from a sociological perspective the robot sociologicus could also serve as an experimental platform for an investigation of conceptual issues that are either particularly suited for formal modeling or are hard to investigate with common conceptual means (in sociology theories and concepts are usually formulated in natural language with its inherent vagueness). Two of these possible issues where mentioned above: First, the determination of a threshold for what counts as an "appropriate" match of frames in Esser's conception or as "typically similar" in the conception of Schütz, and second a concrete conceptual description of what a "script" is (an episode of action consisting of a typical interaction pattern). Both these issues could be, under the precondition of an adequate implementation of the basic reasoning architecture, quite straightforwardly examined, either in computer simulations or better, but more challengingly, with real robots.

Finally, from the interdisciplinary perspective, the most obvious use of the robot sociologicus is to simply build it and then to explore it in empirical HRI-studies. Any modeling of generalized expectations of course is only about the “deliberative” layer of a robot architecture, leaving the sensor and the actuator layer (in the “sense-think-act” chain; Murphy 2000) aside, but this could presumably be solved conventionally[20]. Interestingly, and also debatable according to many sociological approaches, the “reactive” layer for the robot sociologicus would only be a – without any doubt necessary – security measure (e.g. a proximity sensor to prevent the robot from hitting humans). However, what is part of the “reactive” layer in many robotics approaches – sheer bodily reactions modeled on biological conceptions – here would be part of the higher reasoning process, because routine action would become part of the “deliberative” layer. Looking at the picture at large, given the undisputed complexity of all domains (a nursing home, a household etc.) in which an exemplar of Social Robotics is to function in a way meaningful for humans, it would be very attractive to conceptually equip the robot with a technical equivalent of the principle by which human actors solve the problem of complexity of situations – and to empirically investigate the interplay of generalized expectations generated and applied by humans and by robots[21].

[20] But it is by no means trivial for a machine (nor for human actors) to interpret signals (or cues) adequately and to signal interpretations or intentions in a comprehensible way.

[21] The methodological problem of acquisition of appropriate data is then more prominent on the human side. While the reasoning process of the robot, an appropriate architecture and a sufficient storing of data given can be tracked and reconstructed from computer protocols (see Hahne et al. 2006 for a suggestion for integrating computer data into the “technographic” approach to technology usage), it is much more difficult to develop methods to track human behavior in a comparable way.

References

Akrich, M., 1995: User Representations: Practices, Methods and Sociology. In: A. Rip/T.J. Misa/J. Schot (eds.), *Managing Technology in Society: The Approach of Constructive Technology Assessment*, London/ New York: Pinter Publishers: 167-184.

Alac, Morana/Javier Movellan/Fumihide Tanaka, 2011: When a Robot is Social: Spatial Arrangements and Multimodal Semiotic Engagement in the Practice of Social Robotics. In: *Social Studies of Science* 41 (6): 893-926.

Bannon, Liam J., 1991: From Human Factors to Human Actors: The Role of Psychology and Human-Computer Interaction Studies in Systems Design. In: J. Greenbaum/M. Kyng (eds.), *Design at work: Cooperative Design of Computer Systems*, Hillsdale, NJ: Lawrence Erlbaum Associates: 25-44.

Barley, Stephen R., 1986: Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments. In: *Administrative Science Quarterly* 31: 78-108.

Bartneck, Christoph/Dana Kulic/Elizabeth Croft/Susana Zoghbi, 2009: Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. In: *International Journal of Social Robotics* 1 (1): 83-93.

Bethel, Cindy L./Robin R. Murphy, 2010: Review of Human Studies Methods in HRI and Recommendations. In: *International Journal of Social Robotics* 2: 347-359.

Breazeal, Cynthia/Atsuo Takanishi/Tetsunori Kobayashi, 2008: Social Robots that Interact with People. In: Bruno Siciliano/Oussama Khatib (eds.), *Handbook of Robotics*, Berlin: Springer: 1349-1370.

Brooks, Rodney, 1990: Elephants don`t Play Chess. In: *Robotics and Autonomous Systems* 6: 3-15.

Collins, Harry M./Martin Kusch, 1998: *The Shape of Actions. What Humans and Machines Can Do*. Cambridge, Mass.: MIT Press.

Dautenhahn, Kerstin, 2007: Socially Intelligent Robots: Dimensions of Human-Robot Interaction. In: *Philosophical Transactions of the Royal Society: Biological Science* 362 (1480): 679-704 <http://rstb.royalsocietypublishing.org/content/362/1480/679.full.pdf+html> (22.11.2012).

Decker, Michael/Mathias Gutmann (eds.), 2012: *Robo- and Informationethics. Some Fundamentals*. Münster: LIT Verlag.

Esser, Hartmut, 1999: *Soziologie - Spezielle Grundlagen. Band 1: Situationslogik und Handeln*. Frankfurt a.M.: Campus.

Esser, Hartmut, 2001: *Soziologie - Spezielle Grundlagen. Band 6: Sinn und Kultur*. Frankfurt a.M.: Campus.

Fink, Robin D./Johannes Weyer, 2011: Autonome Technik als Herausforderung der soziologischen Handlungstheorie. In: *Zeitschrift für Soziologie* 40 (2): 91-111.

Flandorfer, Priska, 2012: Population Ageing and Socially Assistive Robots for Elderly Persons: The Importance of Sociodemographic Factors for User Acceptance. In: *International Journal of Population Research* 2012: <http://downloads.hindawi.com/journals/ijpr/2012/829835.pdf> (21.11.2012).

Ge, Shuzhi Sam/Maja J. Mataric, 2009: Preface. In: *International Journal of Social Robotics* 1: 1-2.

Hahne, Michael/Eric Lettkemann/Renate Lieb/Martin Meister, 2006: Going Data mit Interaktivitätsexperimenten: Eine neue Methode zur Beobachtung und Analyse der Interaktivität von Menschen und Maschinen. In: Werner Rammert/Cornelius Schubert (eds.), *Technografie. Zur Mikrosoziologie der Technik*, Frankfurt a.M.: Campus: 275-309.

Han, Jeonhye/Eunja Hyun/Miryang Kim/Hyekyung Cho/Takayuki Kanda/Tatsuya Nomura, 2009: The Cross-cultural Acceptance of Tutoring Robots with Augmented Reality Services. In: *International Journal of Digital Content Technology and its Applications* 3 (2): 95-102.

Heerink, Marcel/Ben Kröse/Vanessa Evers/Bob Wielinga, 2010: Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. In: *International Journal of Social Robotics* 2: 361-375.

Inoue, Hirochika, 2008: Foreword. In: Bruno Siciliano/Oussama Khatib (eds.), *Handbook of Robotics*, Berlin: Springer: XII-XIII.

Kac, Eduardo, 1997: Foundation and Development of Robotic Art. In: *Art Journal* 56 (3): 60-67.

Kahn, Peter H./Nathan G Freier/Takayuki Kanda/Hiroshi Ishiguro/Jolina H. Ruckert/Rachel L. Severson/Shaun K. Kane, 2008: Design Patterns for Sociality in Human-Robot Interaction. In: *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction (HRI '08)*. ACM, New York, NY: 97-104 <http://delivery.acm.org/10.1145/1350000/1349836/p97-kahn.pdf> (27.03.2013).

Kawamura, K/T. Pack/M. Bishay/M. Iskarous, 1996: Design Philosophy for Service Robots. In: *Robotics and Autonomous Systems* 18: 109-116.

Kroneberg, Clemens, 2005: Die Definition der Situation und die variable Rationalität der Akteure. Ein allgemeines Modell des Handelns. In: *Zeitschrift für Soziologie* 34 (5): 344-363.

Kuo, I-Han/Chandimal Jayawardena/Elizabeth Broadbent/Bruce A. MacDonald, 2011: Multidisciplinary Design Approach for Implementation of Interactive Services. Communication Initiation and User Identification for Healthcare Service Robots. In: *International Journal of Social Robotics* 3: 443-456.

Lemaignan, Séverin /Raquel Ros/E. Akin Sisbot/Rachid Alami/Michael Beetz, 2012: Grounding the Interaction: Anchoring Situated Discourse in Everyday Human-Robot Interaction. In: *Internatinal Journal of Social Robotics* 4: 181-199.

MacDorman, Karl F./Sandosh K. Vasudevan/Chin-Chang Ho, 2009: Does Japan Really have Robot Mania? Comparing Attitudes by Implicit and Explicit Measures. In: *AI & Society* 23 (4): 485-510.

Matsuzaki, Hironori, 2010: Gehorsamer Diener oder gleichberechtigter Partner? Überlegungen zum gesellschaftlichen Status von humanoiden Robotern in Japan. In: *Technikgeschichte* 77 (4): 373-390.

Mayer, R.C./J.H. Davis/F.D. Schoorman, 1995: An Integrative Model of Organizational Trust. In: *Academy of Management Review* 20 (3): 709-734.

Meister, Martin, 2011a: Mensch-Technik-Interaktivität mit Servicerobotern. Ansatzpunkte für eine techniksoziologisch informierte TA der Robotik. In: *Technikfolgenabschätzung - Theorie und Praxis* 20 (1): 46-52.

Meister, Martin, 2011b: *Soziale Koordination durch Boundary Objects am Beispiel des heterogenen Feldes der Servicerobotik*. Dissertation, Technische Universität Berlin.

Meister, Martin, 2012: Investigating the Robot in the Loop. Technology Assessment in the Interdisciplinary Research Field Service Robotics. In: Michael Decker/Mathias Gutmann (eds.), *Robo- and Informationethics. Some Fundamentals*, Münster: LIT Verlag: 31-52.

Murphy, Robin R., 2000: *An Introduction to AI Robotics*. Cambridge, MA: MIT Press.

Mutlu, Bilge/Jodi Forlizzi, 2008: Robots in Organizations: the Role of Workflow, Social, and Environmental Factors in Human-Robot Interaction. *3rd ACM/IEEE International Conference on. Human-Robot Interaction (HRI 2008)* IEEE, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6249448> (27.03.2013).

Nass, Clifford/Byron Reeves, 1996: *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*. Cambridge, Mass.: Cambridge University Press.

Peine, Alexander/Louis Neven, 2011: Social-structural Lag Revisited. In: *Gerontechnology* 10 (3): 129-139.

Rammert, Werner, 2012: Distributed Agency and Advanced Technology. Or: How to Analyze Constellations of Collective Inter-agency. In: Jan-Hendrik Passoth/Birgit Peuker/Michael Schillmeier (eds.), *Agency without Actors. New Approaches to Collective Action*, New York: Routledge: 89-112.

Restivo, Sal, 2001: Bringing up and Booting up: Social Theory and the Emergence of Socially Intelligent Robots. *2001 IEEE International Conference on Systems, Man, and Cybernetics, Vol. 4*: 2110-2117 <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=972867> (08.12.2012).

Sabanovic, Selma, 2010: Robots in Society, Society in Robots. Mutual Shaping of Society and Technology as a Framework for Social Robot Design. In: *Internatinal Journal of Social Robotics* 2: 439-450.

Salter, Tamie/François Michaud/Hélène Larouche, 2010: How Wild is Wild? A Taxonomy to Characterize the 'Wildness' of Child-Robot Interaction. In: *International Journal of Social Robotics* 2: 405-415.

Schimank, Uwe, 2010: *Handeln in Strukturen. Einführung in die akteurtheoretische Soziologie.* Vierte völlig überarbeitete Auflage. Weinheim: Juventa.

Scholtz, Jean, 2003: Theory and Evaluation of Human Robot Interactions. In: *Proceedings of the 36th Annual Hawaii International Conference on System Sciences,* 2003, IEEE.

Schot, Johan/Arie Rip, 1996: The Past and Future of Constructive Technology Assessment. In: *Technological Forecasting and Social Change* 54: 251-268.

Schulz-Schaeffer, Ingo, 2008: Die drei Logiken der Selektion. Handlungstheorie als Theorie der Situationsdefinition. In: *Zeitschrift für Soziologie* 37 (5): 362-379.

Schütz, Alfred, 1982: *Collected Papers: The Problem of Social Reality. Vol. 1*. Dordrecht: Kluver.

Searle, John R., 1980: Minds, Brains, and Programs. In: *Behavioral and Brain Sciences* 3: 417-457.

Searle, John R., 1986: *Geist, Hirn und Wissenschaft*. Frankfurt/M.: Suhrkamp.

Short, Elaine/Justin Hart/Michelle Vu/Brian Scassellati, 2010: No fair!!: An Interaction with a Cheating Robot. In: *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2010)*: 219-226 <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5453193> (28.03.2013).

Simon, Herbert A., 1997: *Models of Bounded Rationality: Empirically Grounded Economic Reason*. Boston, Mass.: MIT Press.

Steels, Luc/Rodney Brooks (eds.), 1994: *The Artificial Life Route to Artificial Intelligence. Building Situated Embodied Agents*. New Haven: Lawrence Erlbaum.

Steinfeld, Aaron/Terrence Fong/David Kaber/Michael Lewis/Jean Scholtz/Alan Schultz/Michael Goodrich, 2006: Common Metrics for Human-Robot Interaction. In: *Proceedings of the First ACM International Conference on Human Robot Interaction*, Salt Lake City, UT: 33-40 <http://www.ri.cmu.edu/pub_files/pub4/steinfeld_aaron_m_2006_1/steinfeld_aaron_m_2006_1.pdf> (12.07.2010).

Strübing, Jörg, 1998: Bridging the Gap: On the Collaboration between Symbolic Interactionism and Distributed Artificial Intelligence in the Field of Multi-Agent Systems Research. In: *Symbolic Interaction* 21 (4): 441-464.

Sung, JaYoung/Rebecca E. Grinter/Henrik I. Christensen, 2010: Domestic Robot Ecology. An Initial Framework to Unpack Long-Term Acceptance of Robots at Home. In: *International Journal of Social Robotics* 2 (4): 417-429.

Turkle, Sherry, 2006: A Nascent Robotics Culture: New Complicities for Companionship. In: *AAAI Technical Report Series WS-06-09*: 51-61 <http://web.mit.edu/sturkle/www/nascentrobotics-culture.pdf> (28.02.2013)

Veruggio, Gianmarco/Fiorella Operto, 2008: Roboethics: Social and Ethical Implications of Robotics. In: Bruno Siciliano/Oussama Khatib (eds.), *Handbook of Robotics*, Berlin: Springer: 1499-1524.

Wagner, Cosima, 2009: 'The Japanese Way of Robotics': Interacting 'Naturally' with Robots as a National Character? In: *RO-MAN 2009. The 18th IEEE International Symposium on Robot and Human Interactive Communication*, Tayama, Japan, Sept. 27-Oct. 2, 2009: 510-515 <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5326221> (20.11.2011).

Weiss, Astrid/Regina Bernhaupt/Michael Lankes/Manfred Tscheligi, 2009: The USUS Evaluation Framework for Human-Robot Interaction. In: *AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction*. SSAISB: 158-165 <http://www.aisb.org.uk/convention/aisb09/Proceedings/NEWFRONTIERS/FILES/WeissABernhauptR.pdf> (12.07.2010).

Weiss, Astrid/ Judith Igelsböck/ Manfred Tscheligi/Andrea Bauer/Kolja Kühnlenz/ Dirk Wollherr/Martin Buss, 2010: Robots Asking for Directions - The Willingness of Passers-by to Support Robots. In: *5th ACM/IEEE International Conference on Human-Robot Interaction*, Osaka: 23-30 <https://www.lsr.ei.tum.de/fileadmin/publications/HRI2010final.pdf> (12.07.2010).

Yagoda, Rosemarie E./Douglas J. Gillan, 2012: You Want Me to Trust a ROBOT? The Development of a Human-Robot Interaction Trust Scale. In: *International Journal of Social Robotics* 4 (3): 235-248.