Constructing the Robot's Position in Time and Space

The Spatio-Temporal Preconditions of Artificial Social Agency

Gesa Lindemann (University of Oldenburg, gesa.lindemann@uni-oldenburg.de)

Hironori Matsuzaki (University of Oldenburg, hironori.matsuzaki@uni-oldenburg.de)

Abstract

Social robotics is a challenging enterprise. The aim is to build a robot that is able to function as an interaction partner in particular social environments, for example to guide customers in a shopping mall. Analysing the construction of social robots entails going back to the basic preconditions of social interaction, which are usually overlooked in sociological analysis. Surprisingly enough, they are overlooked even by approaches that theorize the agency of technological artifacts, such as Actor-Network Theory or the theory of distributed agency. Social robotics reveals the importance of a basic feature of social interaction: not only is matter/embodiment crucial for understanding the social, but we must also describe how embodied beings position and orient themselves spatially/temporally. This aspect is taken into account neither by ANT nor by the theory of distributed agency. Our analysis shows that two modes of positioning can be distinguished: reflexive self-positioning, and the recursive calculation of position in digital space/time.

1 Introduction[1]

Social robotics is a challenging enterprise. The aim is to build a robot that is able to function as an interaction partner in particular social environments, for example to guide customers in a shopping mall. Unlike industrial robots, which work within a controlled environment, social robots (S-R) must have a certain level of autonomy in order to operate in much less structured environments and work for or with ordinary people. S-Rs should care for the sick, watch the elderly, vacuum the carpet, collect the rubbish, guard homes and offices, give directions on the street, or function as communication mediators between humans (see, for example, Feil-Seifer/Skinner/Matarić 2007; Salvini et al. 2011; Sharkey/Sharkey 2011; Yamazaki et al. 2012).[2]

Analysing the construction of S-Rs means going back to the basic preconditions of social interaction, which are usually overlooked in sociological analysis. Surprisingly enough, they are overlooked even by approaches that theorize the agency of technological artifacts, such as Actor-Network Theory (ANT) (Latour 2005; Callon 1986) or the theory of distributed agency (TDA) (Rammert/Schulz-Schaeffer 2002; Rammert 2012). Social robotics reveals the importance of a basic feature of social interaction: not only is matter/embodiment crucial for understanding the social, but we must also describe how embodied beings position and orient themselves spatially/temporally. This aspect is taken into account neither by ANT nor by TDA. Unfortunately, those approaches which do include the problem of spatio-temporal positioning have the disadvantage of assuming only living human beings as social actors, and having a preference for time over space. This holds true for pragmatism (Mead 1932, 1934/1967; Joas 1989), the classic phenomenological approaches (Schütz 1932/1981; Berger/Luckmann 1966/1991) and ethnomethodology (Garfinkel 1967, 2002). Other authors include space, but they also refer only to human beings as social actors; examples are Bourdieu (1972/1977), Goffman (1974) or Giddens (1984). A promising candidate which meets all three criteria – taking account of time, space, and more than human actors – is Helmuth Plessner's theory of ex-centric positionality and shared world (*Mitwelt*). Being strictly formal, this theory does not exclude any entity in advance from being a member of a concrete shared world, i.e. social world. Furthermore, the theory of ex-centric positionality begins by asking how entities are positioned, or position themselves, spatio-temporally. This draws both time and space into the focus of the analysis.

Our argument here proceeds in three steps. We first sketch the theory of positionality and the shared world, then outline our project's methodological problems and present our data and its interpretation. On this basis,

[2] The nascent presence of those technologies outside the lab and their impacts on social lives are still underresearched in the social sciences. To name a few exceptions, Turkle (2011) interprets social robots as "relational artifacts" that can become an easy substitute for the difficulties of dealing with other people. Drawing on ethnographic observations, Šabanović (2010) proposes the framework of "mutual shaping" to explore the dynamic interaction between robotics and other social domains in robot development. Alač et al. (2011) offer an in-depth semiotic analysis of the coordinative interaction process between robots and humans in laboratory experiments. However, the aspects discussed in our paper are not recognized as problems in these previous studies on social robotics.

we argue that positioning of robots depends on what we call "recursive calculation", which must be distinguished from the "self-reflexive positioning" found in social actors.

2   Ex-centric positionality and the theory of the shared world

The theory of ex-centric positionality goes back to the German philosopher and sociologist Helmuth Plessner. He developed it to describe the difference between inanimate and living things, a problem that seems also to be crucial for S-R engineers: How is a thing, whether animate or inanimate, positioned spatio-temporally? According to Plessner, animate beings not only are positioned, but position themselves. The latter requires a particular structure of self-reference, which distinguishes animate from inanimate beings.

We began our project with a triadic concept of the social, developed from Plessner's theory of ex-centrically positioned selves. A self is here defined as a being that experiences its own states (pain, hunger, thirst), perceives its environment, and acts on the environment according to its perceptions. A bodily self thus performs a threefold mediation between its sense of its own condition, its perceptions, and its activities. A self is the practical accomplishment of this threefold mediation. If a self is related to itself, it creates a distance from the self, i.e. to the accomplishment of the current threefold mediation. This necessarily means that it is not completely absorbed in the execution or performance of experiencing its states, perceiving its environment, and acting, but maintains a certain distance. It is this distance, this being somehow outside, that Plessner (1928/1975: 292) refers to as ex-centric.

Ex-centric positionality is the precondition for taking the position of the other and expecting the expectation that another self places on one. An

ex-centric self not only experiences itself and its environment, but also experiences itself vis-à-vis other ex-centric selves, by which it is experienced as a self. Entities that live in such complex relationships are referred to as persons who live in a shared world. A shared world is a sphere of reciprocal reference where ex-centric selves can reciprocally adopt each other's positions; that is, an ex-centric self behaves towards itself and others from others' perspective. As a result, both self-reference and reference to others is mediated by the fact that an ex-centric being experiences itself as a member of a shared world (Plessner 1928/1975: 304; Lindemann 2010). By definition, this concept of the social is solely formal. Each entity – human or non-human – involved in these complex relationships is a social person. Nevertheless, a distinction must be made between social persons and other beings. It makes a practical difference whether the relationship with other beings is structured by expected expectations or not. If a self expects the expectations of another self, the expectations of the other entity have to be taken into account. If there are no expectations to expect, the relationship to the other entity is less complex.

The formal theory of the shared world suggests that a triadic structure is required to delimit the borders of the shared world. An ex-centric self (Ego) behaves towards itself and others (Alter) from others', i.e. third actors', perspective. Within this triadic structure, the interpretative relationship between Ego and Alter is simultaneously an observed relationship. Since it is an observed relationship, it is possible to distinguish between its current performance and a generalizable pattern that structures the relationship. A rule can thus be institutionalized that guides the distinction between those entities whose expectations have to be expected and other beings. This assumption has been

corroborated empirically (Lindemann 2005). The formal structure can be described as follows. Ego relates to other entities. If Ego expects expectations from Alter, it is up to Ego to interpret Alter's appearance as a communicative statement that indicates Alter's expectations placed on Ego. This interpretative relation is not only performed, but also experienced from a third actor's perspective. Since it is an observed performance, it reveals patterns that guide the interpretation of Alter's communicative statement. The triadic constellation can thus be interpreted as the condition for delimiting the borders of the social world (Lindemann 2005) and the emergence of social order (Habermas 1981/1995 Vol. II: 59–61; Luhmann 1972: 64–80; Lindemann 2012).

Our initial idea was to analyse how the status of the S-R is defined in triadically structured processes of communication. However, looking at our data, it turned out that field actors also had other problems, ones apparently more basic than that of how to define the S-R's status and, especially, whether the S-R was recognized as a social person either occasionally or generally. The data forced us to turn our attention to something we had previously more or less taken for granted: how entities orient and position themselves in space and time.

2.1 Spatio-temporal positioning

Sociology has an obsession with the social dimension of experiencing the world. Although ANT and TDA usefully include other entities as well as humans in the social, they are faithful to sociology in remaining clearly focused on this social dimension. Latour, for example, argues that the collective must be assembled and that institutionalized procedures must decide which entity is a proper member of the collective (Latour 2004, 2005).[3]

[3] Without mentioning or even knowing it, he is applying Luhmann's (1969/1983) notion of "legitimation by procedure" to a new field, the delimitation of the social.

But how can entities assemble if they do not have a position in time and space? The social requires a spatio-temporal structure that cannot itself be reduced even to a more encompassing social construction. We suggest that time and space are not merely a social construction of time and space, but that social actors exist as spatio-temporal beings. A socially functioning S-R therefore has to solve the problems of spatial and temporal positioning before it can function as a social actor.

To analyse problems of spatio-temporal positioning, it is useful to look at general theories. Most approaches in a phenomenological or pragmatist tradition distinguish between the localization of things in a measurable space-time and the position of a living body (in German a Leib). For example, the location of a thing is determined through its relationship to other locations. A table is in front of a window; its legs have a definite angle in relation to the tabletop, which is above the floor, etc. Things are objectified bodies (Körper), and as such they are incorporated into a system of relative spatial relations and relative distances. All locations in this system are determined solely on the basis of mutual references. This also implies that objectified bodies can never coexist at the same time in the same place. If they did, they would be absolutely identical with one another, that is, indistinguishable. GPS and Google Earth are global devices to define the relative spatial and temporal position of any single objectified body. In this respect, they make no distinction between tables, rats or humans – all are objectified bodies, and all can thus be positioned within a system of measurable relative locations. If objectified bodies are moving, the system needs to include time, so as to determine that at a particular point in time only one body occupies a particular space.

There are different views on how the living body should be conceptualized. We refer mainly to Plessner's, enhanced by the subtle phenomenological descriptions offered by Hermann Schmitz (1964–1980). As mentioned above, our major argument is that Plessner's model includes not only time (like Luhmann or Mead) but also space, and leaves open the question of who is to be recognized as social actor.

Plessner develops his concept of the living body with reference to his theory of living beings in general, which characterizes them as bodies that position themselves. To understand this, we must ask how the particular form of self-referentiality of inanimate and animate beings can be described. Inanimate things appear as independent from a perceiving consciousness only because they are constituted by an internal referential context of individuation. This referential context, according to Plessner, must be distinguished from the concrete "gestalt" (form) in which a physical thing appears. In the perception of the gestalt, the individual elements spontaneously come together to create a whole, a unified form (Gestalteinheit). But if the unity of the thing were equated with its unified form, it would be impossible to combine different forms into one whole. Only by distinguishing the two can we understand the form's transformation (Gestaltwandel) and change.

Plessner discusses change through the example of smoking a cigar. First the smoker holds the cigar in his hand, then he smokes it, and finally there is nothing left but a little pile of ash. If there were only the unified form, and not the overarching unity of the thing that creates a whole out of the two phenomena "cigar and ash", it would be impossible to say that the ash is the ash of the cigar (Plessner 1928/1975: 84–85). The unity of the thing is guaranteed as long as the point of unity, which turns the differ-

ent appearances into an appearance of something, remains distinct from the gestalt. The difference between thing and gestalt is also crucial for the assumption that there is a space that can be distinguished as such from a concrete gestalt occupying a particular space. Only if we differentiate the thing from its gestalt can we identify the space in which the cigar (as gestalt) formerly existed, but which is at present inexistent. The space once occupied by the cigar is empty. There is only a pile of ash left, which has a different spatial extension.

"Thing" in this context means a structuring principle of physically ascertainable appearances which constitute the gestalt, the concrete physical appearance. This must be distinguished from the structuring principle itself, which enables a differentiation between gestalt and thing. A thing cannot be completely perceived, but directs the perceiving observation around itself, to its sides that carry its properties – which in turn refer to it, to the thing. When one looks at an inanimate object, the sides with properties send the observer to the core, to the nonappearing inside, which in turn points to the sides with properties, the exterior of the thing. The exterior side of the inanimate thing forms its boundary contours.

Plessner (1928/1975: 127–132) formulates his hypothesis of the specific independence of living things based on the "passive" self-referentiality of the thing. In contrast, the living thing is distinguished by the fact that it executes this self-referential structure itself. For Plessner, this is the leap that distinguishes the phenomenon of the living from the phenomenon of the inanimate. The boundary contours of the living thing are not only its visible exterior sides, but also the evidence that the living thing, in a specific sense, has its own boundary.

In the case of the living body, the boundary has a dual function. The liv-

ing body uses its boundary to close itself off from its surroundings, to make itself into its own self-organizing domain. At the same time, the living body relates to its surroundings by means of its boundary. This boundary allows it to independently enter into contact with its surroundings. In terms of space, this means that the living being does not exist only at a defined spatial position, but relates itself to the space it occupies and its surrounding space. Plessner calls this boundary phenomenon (Grenzsachverhalt) "positionality". A living thing that sets its own spatial boundaries is its own self-regulating domain in relation to its surroundings. In this way, a living thing produces its own exterior surface, which is observable by an external observer. Living beings are therefore characterized by expressivity.

The living thing distinguishes itself from its surroundings by creating boundaries, and enters into contact with its surroundings by means of those boundaries. This is heightened by the fact that the living thing relates to the fact that it relates to its surroundings by means of its boundary. In other words, the living being not only realizes its own boundary, but experiences itself as realizing its boundary. It is thus that the living being experiences itself and its environment. Plessner calls this "centric positionality" (Plessner 1928/1975: 237–244).

The experienced/experiencing living being is characterized by a particular form of self-reference. It actively occupies a space by itself at present and it experiences its space as its present spatially extended states. Hunger, pain or pleasure are present experienced states and localized sensations experienced by a self. This self-reference means that a living body presently positions itself at a particular point in space and is simultaneously related to that and to the way it spatio-temporally positions itself. It

is in a present condition, which it experiences. This particular form of self-reference seems to be the precondition for what Plessner and Schmitz call "absolute location". To know where/when a living body is located, it is not necessary to place it within the system of spatial relations and relative distances. Without knowing the relative location of the objectified body, which I have, I know that my living body, which I am, is "here" and "now". If I feel pain, I do not need first to locate the site of the pain as above, below, approximately within the outline of my objectified body – indicating that it is probably my pain. The location of the living body is accessible without such relative spatial specifications. It spontaneously stands out, as from a background, and is spatially defined ad hoc (Schmitz 1964: 20–23). In other words, absolute location denotes how the living body differentiates itself from its environment.

The space in which objectified bodies exist does not inherently denote a centre; objectified bodies are reciprocally defined in their spatial determinedness and, as such, they make regular, mutual reference to one another. The living body, on the other hand, provides evidence that experiential space has a centre by structuring that space according to the practical demands of its relationship to the environment. For the relative spatial determinedness of "chair" and "wall", for instance, it is irrelevant which side of the wall the chair is on. But for the practical demands of an experiencing living body's global references, it is significant whether the body must first go into the next room to sit on the chair or if it can sit down immediately. This form of self-reference is the basis of ex-centric positionality.

Usually mere lip-service is paid to the relevance of the spatio-temporal aspects of selves. The analysis of building social robots reveals that there is much more at stake than simply say-

ing "we start from the assumption that actors operate from the here/now". This becomes obvious by reference to our field observation.

3   Methodology and data

Between 2011 and 2012, one co-author, Hironori Matsuzaki (HM), stayed for extended periods at several robotic research institutes in Europe and Japan, amounting to 14 months of participant observation in different labs. He also conducted around 30 expert interviews with robotic engineers, law experts and robot industry players and around 10 interviews with lay users of S-R. Additionally, HM gathered documents produced in the field. The interviews were paraphrased or transcribed, and those conducted in Japanese were (at least partially) translated into English or German. Documents were also translated as necessary. Field notes, documents and interviews were coded using procedures that could be described as a heretical deviation from grounded theory: according to Glaser/Strauss (1967), the code should be developed primarily with reference to data alone, but we also used an abstract theory (positionality theory, theory of space) as a reference point for coding. That is, both our field observation and the coding of the data were structured by concepts – such as the space of things, objectified bodies, and the space of living beings' self-positioning.

The major problem with this theory-guided approach is the generation of conceptual artifacts – i.e. data – which, due to the theoretical framework adopted, are only produced in the field notes. This danger cannot be avoided, but it can be controlled by making one's theoretical assumptions as explicit as possible. We call this a critical-reflexive method (Lindemann 2002), which has been fruitfully adopted in several empirical projects (Lindemann 2005, 2009). It is critical

in assuming that observation and interpretation are structured by theoretical concepts. By making these explicit, the observer self-critically delimits how s/he will construct his/her observations and interpretations. This first aspect may be somewhat unusual for sociologists, but the second one is more commonplace: sociologists expect that there are actors who interpret the world themselves; the observed social world is an already-interpreted world. Sociologists therefore see themselves as facing the task of reflexively making interpretations of interpretations.[4]

The analysis we present here draws especially on an ethnographic study of field experiments with S-R that HM conducted in Japan between November and December 2012. The experiments aimed to introduce more smoothly functioning assistive robot technologies into everyday life. Data were collected mostly at a robotics research institute in a Japanese college town, a shopping centre located close to the institute, and some robotics-related events. We pseudonymize the proper names of human actors, technical artifacts (robots), institutions and related entities to protect the privacy of individuals directly observed during the research.

The interviews and statements cited in this paper are not literally translated into English, because a word-for-word translation would hardly be understood due to the openness of Japanese grammar. For instance, in a Japanese everyday conversation, both subject and object are frequently omitted when the meaning can be de-

---

[4] We will not go into more detail here, since this aspect of sociological methodology is to some extent common sense. Georg Simmel first discussed it in 1908 in *Soziologie*. Later, Alfred Schütz (1932/1981) emphasized that sociologists always interpret the interpretations of the social actors they observe. Anthony Giddens (1984) presented the same insight, and Latour (2005) applied it to the problem of who can count as a social actor.

duced from the predicate or context. In this sense, the Japanese language requires much interpretation by the recipient. A strictly literal translation of an interview excerpt may illustrate this point[5]. Italicized passages indicate the interviewee's emphasis. Bracketed descriptions explain non-verbal cues:

Fujita: (with an amused smile) Do not know much about robots, well, have come here today with very little knowledge. Well, as regards the robot's own speech, nothing went beyond expectations. But then, was pleasantly surprised when I spoke and understood what said.

Interviewer: Understood what said?

Fujita: Yes, also today, when was asked, "Where would like to go?" said, "Utopia". Then received a prompt reply, "Okay, Utopia right?" (laughing) And figured that *listened to!* Of course, a robot, not a human being, so wondered about that point, for example, whether really would understand my words. And then, when spoke to, responded *so quickly*! Was delighted. That was a great surprise.

Interviewer: Thought understood. (both laughing)

Fujita: (in a joyful tone of voice) Yes, did.

To avoid further confusion, each sentence is not structured according to the Japanese word order (subject–object–verb), which is entirely distinct from that of English. The "it" that stands for the robot was not uttered during the actual interview. This is also true for "I" and "me", the words to express the interviewee's first-person perspective. Sometimes both speakers omit many sentence constituents and use only the verb, which may hinder a reader's understanding of the content. A literal English translation of spoken Japanese sentences thus does not always convey the ac-

curate sense, and may be misleading. For these reasons, we decided to adopt the paraphrase translations by HM, a Japanese native speaker. We are well aware of the risk of "double interpretation" that may result from this method.

3.1 Experimental participants

The experiments were conducted in the framework of an ongoing research project to implement S-R applications supporting the social participation of elderly and disabled people. According to the Japanese engineers, daily shopping was to be made an easier and more entertaining experience for senior citizens, though the S-R platform for this application is still in the pilot phase. The field experiments took place in a two-storey shopping centre.[6]

During HM's stay, three different types of mobile robot platforms were deployed.[7] The first platform (type A) consists of a black rectangular box on wheels with two arms and a head carrying two large cameras and a round speaker (these components are mostly perceived as the robot's eyes and nose). A shotgun microphone is mounted on a long pole protruding from behind its right shoulder. While it does not look humanoid or animal-like in a narrow sense, overall the robot evokes the image of a biological being.[8] The exterior of the second robot (type B) looks more sophisticated

---

[5] Personal interview, 17 December 2012.

[6] In the past few years, the research institute has developed a cooperative relationship with this commercial facility, albeit not on an equal footing. In negotiations, it is the researchers who have to struggle to maintain the relationship. The experimenters are taught to follow a myriad of rules on-site and not to be rude to customers.

[7] They were built during previous research projects of the institute. At the time of HM's field observations, the aim of the project was to implement a feasible support program for shoppers into these existing platforms.

[8] According to the researchers, the robot's exterior design is not popular with the general public. Some recipients label it as

due to the plastic shield that covers the aluminium frame of the robot body. It is about 110 cm high, and can, like the first robot, cruise on a wheeled base at a speed of 2.5 km/h (the experimenters consider this speed to best suit the target group). Finally, there is a smaller robot (type C). It is about 30 cm high and was originally developed as a communication device to be utilized in combination with cell phones; it therefore has no means of moving. In the field experiments, it was made mobile using an electric platform truck. Placed on the cart, it could move around the test site and approach test persons. All these robots are intended to guide elderly customers through the shopping centre and provide them with information on stores and products. A robotic wheelchair, developed as a support device for disabled people, was also tested on-site.

In one of the experiments, one robot (type A or type C) was supposed to identify and approach a target person, hold a short conversation, and then guide the person around the shopping mall. The focus was on the interaction process between robot and test person, with the aim of producing a convincing expressive surface for the S-R that could be presented as a successful project outcome at the final review, to which the media were also invited. The experiments aimed to ensure that the interactions followed the planned scenario. Each sequence of experimental human–robot interaction lasted a maximum of 20 minutes, though its preparation often took several hours.

The human personnel of the experiments consisted of robotics researchers and lay persons who were to interact with the S-R. The robotics researchers worked as a team with a roughly even mixture of Japanese and foreign members. They were postdocs, PhD students, MA students as

ugly, comparing the facial part with insects like the mantis.

assistants, and a female member of the institute's support staff. The team leader (Kuwata) is Japanese. Some researchers worked all day long (if necessary from early morning until the shopping centre closed); others did not appear regularly in the field because they had duties in other research projects.

The test subjects were lay people. Two elderly ladies were sent from a temporary employment agency specialized in senior citizens. They were on duty for three or four hours on average and earned 1,000 yen (about 8 euros) per hour. Conversations with them revealed that they were not participating only to make money, but also for pleasure. They thought of this as a way of being part of their local community, and also enjoyed interacting with the S-R.

To facilitate the experimental procedure, the engineers used external assistance. For one experimental session, two or three young people (mostly college students in their early twenties) were hired as part-timers for such tasks as installing technical devices, transporting the robots between the control station and the entrance area, monitoring the test site including protection of the robots, or responding to questions from passers-by. The support staff or one of the engineers took care of new part-timers, providing them with a brief introduction on the project and the setup of technical devices for the experiments. To avoid unnecessary effort, part-timers with previous experience were favoured and employed several times. Sometimes they were also hired as test persons or for other interaction experiments carried out in the mall or the lab.

In certain situations, shoppers or store staff also had an important impact on interaction among experiment participants. For instance, passers-by with small children often stood near the test site and watched the scene for a while. Some curious onlookers

talked to the party involved in the experiment or even tried to touch the robot body, in which case the student assistants had to stop them by asking them politely not to interrupt the engineers' work. Even the less interested shoppers required attention: they had to be kept out of the area, particularly when the robot was moving. For these purposes, the experimenters set up a sign reading "We are conducting experiments with service robots. Thank you for your cooperation."

The experiments had a kind of "back stage" (Goffman 1956), the control station, which was called "backyard" by the engineers and placed at the furthest end of the building. It consisted of two small rooms filled with desktop and laptop computers, monitors, desks, chairs, hand trucks, cables and devices, battery chargers, repair tools, spare parts for the robots, tripods, video cameras, removable external sensors, and so on – all the equipment needed for the experiments. From this back stage, the robot's "front stage", its expressive surface or behaviour, was produced and controlled. It was more than a minute's walk from the control station to the entrance area, so that the engineers often had to use cell phones or wireless transceivers to communicate with student assistants or each other.

3.2 Preliminary procedures of the field experiment

Two preliminary processes ran parallel to the technological preparation. The first was negotiation with the head of the commercial facility to make sure that the experiments could be performed. Kuwata, in charge of directing the experimental procedures, was also responsible for this. In the case of important events such as an on-site public presentation of the project, he was to give the store manager a blueprint in advance. To obtain consent, Kuwata had to demonstrate that the event would not interfere with sales activities or endanger the

safety of humans (experiment participants, customers, etc.). The power balance between the parties was lopsided; for example, during the briefing Kuwata "keeps bowing to the store manager" (field notes) – a behaviour clearly indicating the higher status of the other. A second preparatory process was making the human subjects familiar with the experimental setting, and vice versa: information on the facial shape of each lay participant was captured using an external camera and stored in the facial detection system. The robot used in the experiment was not presented to the two elderly women (Sakai and Takagi) at this stage. Kuwata talked to the women between experimental sequences. Sitting face-to-face at the entrance area of the shopping mall, he tried to give them easy, step-by-step instructions on what to do in each phase of interaction with the robot.

During the final demonstration, each woman was to act as a customer entering the shopping mall: At the entrance, the robot waits for her as the target person. When she appears, the robot detects her by reference to individual facial recognition information. The target person takes the designated route towards the robot and walks slowly enough for her face to be recognized. Soon after the robot has identified her as a target person, it comes up to welcome her. The team of robot and human then starts a short dialogue, in which the robot must take the initiative. The robot asks the test person what she has come to buy; she gives an appropriate answer and is guided to her favoured destinations by the robot:

Kuwata explains that the ladies will be led either to the bookstore Utopia or to the clothing store Denim Factory. The bookstore is at one end of the building and cannot be viewed from their present location. Kuwata describes the course the robot will take:

"On your right, there is a narrow corridor. Starting from that spot near the

mirrored column, the robot will head toward the corridor. Then you should just walk behind it at a little distance. At the end of the corridor, it turns right to reach the goal." The ladies are asked to comply with this instruction in concrete interaction situations. Kuwata is not entirely focused on practical issues, but sometimes makes small talk with them about topics irrelevant to the experiment (e.g. the forthcoming national election)... While taking facial images of one lady (Sakai), Antonis, an engineer from Cyprus, stands next to her and points to the exact spot where she should stand. Sakai is asked to look diagonally into the camera placed on her left. Then Antonis goes behind the camera to see live footage displayed on the laptop screen. Antonis and Sakai are now standing toe-to-toe. Checking the images, Antonis discusses the angle of her face with Kuwata. They look, over the camera, at her real face and then back to its representation on the monitor. Antonis asks Sakai to move her face a little to the right. The procedure is repeated several times. After saving selected pictures, they have the lady walk past the camera to test whether facial recognition works. (Field notes)

The complex technical system that enables this interaction scenario is segmented into small functional elements such as locomotion and localization of the robot, facial recognition and tracking of the target person, path planning through crowded spaces, speech recognition in a noisy environment, etc. The execution is distributed among software and hardware components of the robots, different sensors and external cameras embedded in the environment (at the entrance area), and a dozen computers running in parallel. The mediation of perception and actuation for the robot is based on the performances of these functional sub-units. The sub-units are integrated with each other by engineers. Afterwards it

should function automatically, but if problems occur they have to be solved by engineers working in the control room or at the test site.

These types of robots, "network robots", are designed to work in connection with different external components. Perceptive tasks are distributed to technical components installed in the environment (often grouped under the term "ambient intelligence"), whereas actuating tasks are entrusted to the robot body, which can move and behave within these environments. The splitting of sensory and motoric components is usually explained by the variety of functions the robot must accomplish. With the increasing complexity of tasks, it becomes difficult to integrate and coordinate all functions within the robot body.[9] Dividing the unity of the robot's activities is believed to be a better way of overcoming these technical problems and making the robot capable of interacting with lay users, who usually possess very limited knowledge of advanced technologies.

At the beginning of each experiment, a robot is spatially calibrated. Its initial point is determined as point zero, from which any movement or behavioural activity is calculated. This action is decisive for the robot's navigation, because it is the point from which movement direction and travel distance is derived. Only from a determined starting point can a robot of this kind begin to cruise. Within a three-dimensional physical environment, the robot moves with reference to a

[9] The experimenters need to operate multiple computers (sometimes more than ten) simultaneously in order to make the robot complete the interaction process. A team member described the dilemma: "Of course, nothing can beat having one computer that can accomplish everything. But we have enough trouble dealing with the enormous quantity of real-world data. The processing capacity of the robot's computer is still too low to run different resource-hungry applications, like facial recognition, at one time" (Field notes).

static topological model of the indoor space, formed only by the two co-ordinate axes (x, y). This model is represented as a two-dimensional floor plan with the geometric properties of the environment. The actual location of the robot is expressed in x- and y-positions, while the motion direction of the robot is defined through the variable "θ", from which the differentiation of directions – from the robot's viewpoint: to and fro, up and down – is derived mathematically (by calculating the emerging angle with reference to values in the x-axis and the y-axis). On the monitors in the control room or the laptop screens, the engineers can see the top-down view of the test site with abstract images of the "trajectories"[10] of the real entities (robots, humans, and other objects) moving in the space. This optical representation of binary data is designed for ease of operation by human actors (engineers, test persons). "Properly speaking," one researcher emphasized, "what can be seen on the GUI [graphic user interface] does not correspond to the visual space perception of the robot."[11]

According to the engineers, the robot can work autonomously in principle. This means that once the robot has started an operation, it can move alone and execute its tasks without continuous external control. Dealing with lay people in a real-life environment is, however, seen as one of the major challenges for S-R applications, because these environments are often unpredictable and the robotic system has to react to uncertain factors. To ensure a high level of safety and reliability, it is considered necessary for a remote operator to oversee and assist the robot's operation. This approach (semi-autonomous control of the robot) was taken by the researchers observed. The robot was to approach the target person and initiate conversa-

tion by itself. Once the robot had done this, the human operator took over control. The operator would drive the robot, assist its speech recognition, and trigger its utterances. The user interface prompted the operator to take action. It was up to him or her whether the robot should execute a certain action or not. Moreover, the mobile robot called on the engineers for help when something unpredictable occurred or it needed to handle correspondence problems between the predefined sequences of events and the data gained in real time from the environment. For instance, the robot sent signals to the operator's computer when its infrared sensors detected obstacles on its route that could not be synchronized with those on the preinstalled map of the environment.

In the field trials observed, two main types of virtual maps proved decisive for the robot's localization and navigation.[12] The first type is a preinstalled map. The second type is created during the S-R's operations: after being placed on point zero the robot (or

---

[10] "Trajectory" is a term used in the field to denote the path of an entity's movement.

[11] Personal interview, 25 December 2012.

[12] GPS, a space-based satellite navigation system often used to provide location and time information for the navigation of driverless cars, is not implemented in the mobile robots of our field, mainly due to the noise in indoor environments. The engineers also do not apply more challenging approaches to robot localization such as SLAM (Simultaneous Localization and Mapping), mainly because of their focus on dealing with practical problems in a real-world application. Alongside other methods for navigation and localization (Light Detection and Ranging, GPS, Digital Cartography), the automated "Google car" uses SLAM technology, which creates and updates a map of a vehicle's surroundings while keeping the vehicle located within the virtual map. To build up a SLAM map, however, the car needs first to be driven manually along a route while its sensors collect relevant data about the outdoor environment. The car then drives autonomously on the route, comparing the data acquired in real time to the previously recorded data so that it can capture changes within a known environment and update the map. See, for instance, Guizzo 2011; KPMG 2012.

rather, the computer on board the robot body) starts to measure the current distance between the robot body and the objects in the environment using two infrared sensors in its foot part, and creates a two-dimensional map of objects scanned in the area where it is to move. This second map should approximately match the pre-installed map, so that the robot can detect its present location and navigate along a predefined route without remote operation. Without such approximate matching, the robot loses its way and gets stuck at one spot. It sends a signal for help, and the engineers correct direction and route by inputting precise information on its present location. In combination with terrain mapping, the odometry method is employed to localize the wheeled robot. Here, the robot calculates its position in space relative to a starting point (point zero); using shaft encoders on its two wheels, it measures velocity and the rotations of the wheels in real time and computes how far it has travelled. Its current location is then estimated (not determined) from travel distance to the default position.

As these methods are sensitive to error arising from different noises in a real-world environment, the robot must continue to fine-tune its approximate location by a probability calculus referred to as "particle filter". For example, if the robot occupies a particular space, this position is defined by several parameters (90 degree angle to the wall, distance of 1.2 m, velocity of 2.3 km/h, etc.). A particular set of parameters that defines a particular position is called a variable or a particle. A set of possible particles is calculated for a specific point in time: it is calculated that at a particular point in time the S-R could possibly be at n-positions (particles or variables). Between 100 and several hundred such positions are calculated. The entire set of calculated variables displays a pattern from which the

probable position of the S-R at a specific point in time can be derived. Each position of the robot is thus deduced from a pattern of variables/particles. Its position is not determined precisely, but estimated as a probable position, on the basis of a set of possible positions. Diverse patterns of variables are simulated by the robot's computer in advance (random sampling). While moving in a real environment, the robot keeps updating the patterns of variables by comparing current data received by sensory input with previous data (resampling of probability), and calculates a region where the robot is probably currently located. The mean value of the resampled variables is then defined as the estimate of the robot's position at a particular point in time.

A visual representation of the robot's orientation in space (displayed on the computer monitor via the GUI) may help to make sense of this process (see Figure 1). On the map with a black background, oblongs depict the store areas. Bold lines around these areas express the walls and/or columns. The boundary between the corridors and the adjoining stores is represented by thin lines. Small dots scattered around in the store areas represent static objects scanned by the robot's sensors. Circles express moving entities (e.g. walking humans) tracked by the sensors installed in the robot's surroundings. At one corner of the corridor, there is a square object outlined in bold. This figure stands for the robot that is moving toward the identified target person (two footprints). From its front, two dotted lines radiate in the direction of forward movement. A number of dots enclosed by a polygonal shape overlaps with the rear of the robot figure. When the robot starts moving, the polygon filled with dots follows the S-R with a short time lag. This polygon and its dots represent a pattern of estimated variables (particles), i.e. the
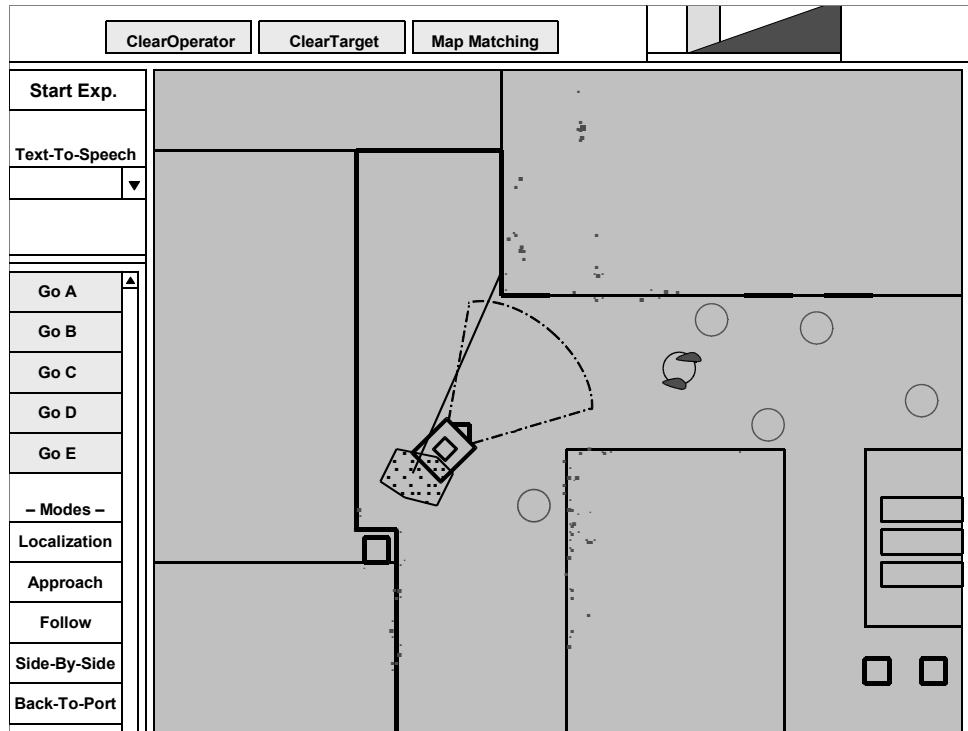
Figure 1: Visual representation of the robot localization via GUI

current region where the robot can probably be found.

Interacting with a target person is even more difficult for the robot than localizing itself. The robot must first identify one of the participants whose individual data (physiognomic attributes, family name) have been stored in the system in advance. This requires careful, time-consuming preparatory work – booting up the computers including the robot's on-board computers, setting up different devices on-site, calibrating the laser range finders and external cameras, registering facial images of the target person, integrating all the functional sub-units, test running the robot, etc. For instance, the different data sets of the two functional sub-units running outside the robot body, "facial recognition" and "human tracking", have to be combined so that the robot has relevant information regarding whom to address. A network of laser range finders, set at the four corners of the entrance, anonymously tracks the trajectories of the target person. Simultaneously, in the middle of the en-

trance area, a digital camera connected with face detection software matches the person's frontal facial images against his/her individual data within the subsystem. By associating this information with the trajectories observed, the location of the registered person is determined. This multi-sensor fusion is realized using data processing by the computers in the control room.

As a next step, the diverse sensory inputs of external components have to be related to the robot's behaviours. The coordination of sensory and motoric inputs at the preparatory stage mostly remains invisible for lay participants. During this process and a test run, the experimenters encountered different types of technical difficulties resulting from the complexity of the whole system and the large quantity of data on the robot's environment. In some cases, the experiment had to pause for an extended period to find out what was wrong with the system. Such situations were stressful and time-consuming.

I go back to the "backyard". There, Kuwata and two other researchers continue with a dry run of the guide robot (type A). Watanabe, a Master student sitting next to Kuwata, helps him to operate the robot using the interface for remote control. In front of both engineers, four computers are running. I can see a bunch of open windows cluttering up the screens. Tom, a Canadian colleague,[13] monitors at other computers whether the fusion of facial recognition and human tracking is working properly. They communicate in English, sometimes switching to Japanese for Oda, who is not good at English. ... It is more than six hours since they commenced their work. They look very tired. Watanabe, who is waiting for instructions from the team leader (Kuwata), takes off his glasses to wipe his face. Out of the blue, Kuwata gives a shout of surprise. He notices that the visual representation of human trajectories tracked around the robot has disappeared from the displays. Searching for possible explanations for this, Kuwata repeats in English, "Why?" After a thorough investigation of relevant system parameters and source codes in the compilers, he exclaims with a lost look on his face, "It's working, but it's not working." In answer to my question, Watanabe explains that the data obtained by the environmental sensors is not being sent to the computer on board the robot. "That is strange because that information is received by the other robot (the platform truck for type C without the robot body)[14] that works with the same program." ... After an approximately 30-minute struggle with the uncertain origin of the problem comes a Eureka moment. Kuwata calls out suddenly and starts to describe where the blame should be laid. It turns out that the odd phenomenon emerges from different time settings. The clock of the computer that integrates the information from the laser range finders is set several seconds earlier than that of the

robot's computer. Therefore the robot keeps throwing away all the data of human tracking, evaluating them as previous, thus irrelevant data. (Field notes)

Sometimes it took several hours to solve such problems. In extreme cases, planned interaction experiments had to be postponed despite the large amount of effort and time invested. For both researchers and paid lay participants, this was a waste of time and resources.

The robot's different tasks, including verbal communication with the interaction partner, are predefined and executed based on the action flowchart, a software program with diagrams that represent the sequences of behaviours the robot should perform. This program enables the developers to give the robot instructions without translating the whole process into programming code. On the chart, which consists of event blocks and lines connecting them, there are some decision points where the robot (or the operator) must choose a path to follow among the listed alternatives. The decision is made in the form of answers to "if/then" or "true/false" statements. Decisions are based on relevant information from the environment. For instance, if the value read by the sensors indicates that someone registered as a target person is standing in front of the robot, it welcomes him/her by name and/or says, "Nice to see you again. Do you remember me?" In the case of a non-target person or if the target person's name is not yet stored, the robot greets with a simple "Hello" before starting to introduce itself. Behaviours associated with the interaction with humans are mostly realized in this way. Situations covered by the prepared flowchart can be handled automatically by the robot itself – it executes designated behaviours according to the algorithmic patterns prepared by the engineers.[15] When unpre-

---

[13] Among the engineers, foreign colleagues from the USA, Europe and other distant countries are usually addressed by their first name, while Japanese, Korean and Chinese members call each other by their surnames. A person of higher position (e.g. Kuwata) is spoken to respectfully, by attaching the Japanese honorific "san" to his/her name (Kuwata-san). "San", commonly used as a title of respect, is comparable with the English honorifics Ms., Miss, Mrs. or Mr.

[14] During HM's field observations, the small robot (type C) often broke down. In such cases, the electric cart intended as a means of mobility for the robot was itself deployed as a robot platform.

[15] This embodies the notion of the "Chinese room" proposed by John Searle

pared situations occur, the human operator in the control station takes over. S/he conducts speech recognition and makes the robot provide appropriate answers. Responses to questions posed by the human participants are chosen from sample answers paired with particular questions.

In the trial, the robots sometimes confused the lay participants by guiding them in an unexpected direction. Even then, deviations from the predefined interaction protocol were generally not welcomed. The site supervisor, Kuwata, directed the test persons to follow the shopping suggestion offered by the robot, however incorrect. When the robot led Sakai to the wrong store, she was given the explanation that the robot is unable to distinguish clearly between the sound of "clothing store" (fukuya) and that of "bookstore" (honya).

4 Interpretation

When the experiments started, they were framed communicatively in two ways. The researchers had to negotiate with the shopping mall manager for permission to perform the experiments, and the human participants in the experiments had to be informed in advance about the experimental procedures – what they could expect the robot to do, and so on. The negotiations with the management were nearly finished when HM arrived, and only one meeting could be observed directly. HM also participated several times when the group leader, Kuwata, briefed the two test persons Sakai and Takagi. Both interactions were structured by a more or less explicit refer-

in *Minds, Brains, and Programs* (1980). In this thought experiment, a man in a closed room who speaks only English tries to converse with a recipient outside in written Chinese. Simply by following the program's instructions, the English speaker can give accurate answers without making sense of them, convincing the recipient that he is able to understand a Chinese conversation.

ence to absent third actors: the negotiations with the manager referred to the stores and their commercial interests, to customers and their safety; the meetings with Sakai and Takagi were determined by reference to the expectations of the future audience, because the experiments were not only experiments but also trial runs for the final presentation of the project. With this in mind, Kuwata did not want Sakai and Takagi to act spontaneously towards the robot. Instead, they were requested to follow a predefined choreography consisting of five steps:

1. The robot waits for the target person (customer) at the entrance;
2. S/he enters the shopping mall;
3. The robot detects him/her with reference to individual information provided by the networked sensors;
4. The robot approaches the target person and offers him/her shopping ideas;
5. The target person is accompanied to his/her favoured destinations by the robot.

Regardless of whether or not the robot's behaviours fit this scheme, the women were asked to proceed to the next step as if the robot had functioned properly. Even if the robot misidentified the store, they should follow it; although the robot's speech recognition sometimes mistook "bookstore" for "clothing store", the test person was to follow the robot to the suggested store. We interpret this instruction more as a theatre director's guidance to an actor than as information provided to a test subject. The director wants a perfect performance on stage in front of the public and the official reviewers of his project.

We will now look in more detail at the problem described at step 3 and 4. The S-R is set on point zero and has to compute incoming data and actuate its movements. This phase is not about acting, it is not about producing an effect in the sense of ANT or TDA.

Rather, it is about the robot's position in the situation.

4.1 Spatial positioning

The researchers seem to have assumed an empty space within which the position of each thing can be calculated. The S-R thing occupies a calculable position at a particular point in time. If it moves, the objectified body of the S-R thing will occupy a different space at a different point in time according to a planned trajectory. This space should be empty before the robot body moves into the particular position. "Empty space" should not be misunderstood as a philosophical term: it is simply a space that can be occupied by a particular gestalt at a particular point in time. As such, "empty space" is a practical precondition of planning a trajectory.

Within the empty space, each position can be defined by reference to the x/y-axis and to a measurement using discrete units, which can be infinitely divided into discrete sub-units (metre, centimetre, millimetre, nanometre, etc.). This allows each position to be calculated more and more precisely according to any current practical purpose. We call this digitally measurable space "digital space". Conceptualizing space in this way allows space and spatial extensions of objectified bodies within it to be measured at a particular point in time, for example by infrared sensors. The measured space can then be transformed into a map, which can be compared to a preinstalled map. If the maps match up, the S-R has a calculated position within digital space. The characteristics of the preinstalled map do not differ in principle from the features of the measured space around the S-R. On the contrary, infrared measurements result in an up-to-date digitalized map. There are two digitalized maps of space, which should match up. In fact, differences between the maps are likely to occur, and indicate,

for example, that there is a position defined as empty space on the preinstalled map, whereas on the updated map produced via inputs from the infrared sensors this position is defined as a space occupied by an objectified body.

Within digital space, the S-R must be set on point zero to calculate its trajectories and behavioural activities. Point zero is a space occupied by the S-R body at that time when it starts. It is an identifiable point on the two maps – the preinstalled map of the shopping mall and the map created in real time by measuring devices. Point zero must always be identified before the robot starts to work. It does not change; it is fixed and therefore every change of position can be calculated with reference to it. Different methods are used for this, such as odometry or particle filtering. In odometry, the revolutions of the wheels are counted and the angle of turns measured if the direction changes. The moving robot is always related back to point zero by a chain of calculations. This allows the robot's position to be approximately estimated on the preinstalled map at any point in time. This method of orientation is counterchecked by renewed infrared measurements and probability calculus through particle filtering, enabling data to be provided for an ongoing match between the two maps. For the robot's position to be estimated uninterruptedly, the matching between maps has to be continuous. If it fails, the S-R's positioning breaks down and it becomes lost in an empty space.

Particle filtering displays most clearly what we identify as the crucial principle of positioning the robot. It produces a set of parameters by different measurements (distance to wall, angle to wall, velocity, etc.), uses them to calculate possible positions, and refers to these sets of calculated positions to estimate a most likely position at a particular point in time. Here calculation takes a recursive loop,

culminating in a probable position. The recursiveness of calculation becomes even more complicated if the calculation is carried out for different points in time, ordered along the distinction previously/later. Using n-recursive loops of calculations of calculations of calculations, a trajectory of the S-R is calculated.

However, this form of positioning is not the only one possible. If we look at how Kuwata describes the experiments to Sakai and Takagi, positioning seems to function quite differently:

"On your right, there is a narrow corridor. Starting from that spot near the mirrored column, the robot will head towards the corridor. Then you should just walk behind it at a little distance. At the end of the corridor, it turns right to reach the goal." … While taking facial images of one lady (Sakai), Antonis, an engineer from Cyprus, stands next to her and points to the exact spot where she should stand. Sakai is asked to look diagonally into the camera placed on her left. (Field notes)

If we compare this form of positioning to recursive calculation, it seems to be very simple. What are the preconditions of this simplicity? Kuwata addresses Sakai with "on your right". Using the difference between right and left, Kuwata refers to a body that defines its own position. From a "here" directed to the front, a body can distinguish between right and left. This form of self-positioning must be presupposed for the words "on your right" to make sense. Kuwata recognizes that left and right have a different meaning if the distinction is actuated from a different "here". Kuwata must take Sakai's "here position" in order to say "on your right". The position of each body is thus determined by itself. And it demands some effort to take the position of the other or to treat each position as interchangeable.

Obviously, Kuwata and Sakai assume that they all, including Antonis, share a common space around them. This is corroborated by the way Antonis refers to Sakai. He simply points to the position "where she should stand". The space around them is a social space – a space common to all participants. How should we make sense of this social space? Here a difficult decision must be taken. We might assume that calculable mathematical space is common to all beings, but if this were true, social space would not be structured by being centred around different "here"s. Instead, centredness would be erased from social space. Our data give no indication that this conclusion is possible. The situation we have described seems to be determined by the fact that there is a common space within which different centres, different "here"s, exist.

To make sense of this, we refer to the analysis of space offered by Hermann Schmitz, in particular his analysis of the spatial structure of the pain experience (1964: 183-216). In an intense pain experience, the perception of the environment breaks down. There is only a living body experiencing its pain here and now, which stands out from an undifferentiated space around it. This spatio-temporal point is not defined by relation to other points, which is why Schmitz describes it as an absolute spatio-temporal positioning. This accords with other phenomenological characterizations of the here/now. The here/now indicates a reflexive self-positioning. It is not self-consciousness that is at stake, but simply the phenomenon of self-positioning. What is particular about Schmitz's analysis is that he relates the phenomenon of self-positioning to the phenomenon of an unstructured space from which the self as a living body stands out. "Here" stands out from an unstructured space, which can be experienced as a space common to each living body. The common space is unstructured and has to be set up from each centre (living body) by es-

tablishing directions like front, backwards, right, left, above, or below.

4.2 Spatio-temporal positioning

The difference between a position defined by recursive calculation and a reflexive self-positioning from which directions are set up becomes even more obvious if we take time into account. To become calculable, time too must be brought into a measurable form. The basic features of this process have been described by Norbert Elias (1992: 46–47). He understands time as a functionally tripolar relation between humans who link two series of discrete events with each other. One of these series is supposed to be the standard series, and functions as a framework for defining the other series of events. At present, the atomic clock, which refers to nuclear events, is considered to be the standard series of events. It enables discrete points to be defined one after the other, measurable as nanoseconds or even smaller units. Due to the form of measurement, we call it "digital time".

The series of discrete units is given an index of previously/later. Events determined according to this measure of time are thus defined by their positions relative to each other. Relative positions are more or less stable. To give an example: On 12 February 1913 at 14:15, Agathe Meyer had a heart attack. On 13 February 1913 at 09:21, Agathe Meyer died. The order of previously/later does not change. On 15 February 2013, the events are still in the same order of earlier and later. What is now earlier in relation to a later event will still be earlier tomorrow. This distinguishes measured time from the difference we experience between past, present and future: there, what is a future event now will have become an event in the past tomorrow.[16] Time here indicates a

modal difference with reference to an actual present. There seems to be no way out of one's actual present. The experience of pain exemplifies this well.

The S-R system bug described in the field notes is an indication what happens if the difference between present, past and future is simulated within the framework of digital time. Within the realm of recursive calculation there is no present. Presently incoming sensory inputs are not included in the calculation of the situation if there is no match between two measured series of previously and later. The series implemented in the system of the robot confronts the series implemented into the sensory system gathering data from the environment. The sensory system delivers data which are some seconds earlier than the measured time of the robot system. Data from 13:45:44 are irrelevant for calculating the robot's action at 13:45:46.

The robot works on the basis of digital space/time and recursive calculation. Its position is defined in time and space by matches of 1) digitalized spaces and maps, and 2) different digitalized time series. If there is no match, the robot is lost in empty space and time without positioning or orientation.

5 Conclusion and discussion

S-Rs are both similar to and different from social actors. They are similar in that robots and social actors are objectified bodies, which can be identified and referred to in spatio-temporal experience and in digital space/time. But a S-R differs from a social actor regarding its ways of existence in space and time. Being a social actor requires, for example, taking the position of another, the precondition of which is that an entity is able to accomplish self-positioning. As is well

---

[16] For the distinction between these two aspects of time, see McTaggart (1908). Schmitz offers an insightful discussion of

McTaggart's idea that time is unreal (Schmitz 1980: 476–479).

known in pragmatist and phenomenological traditions, taking one's own position means acting from a centre, which is understood as "now" (Mead) or "here/now" (Plessner, Schmitz). Mead's concept of "specious present" was coined to show that each living being organizes its own temporal order of past and future from its actual present (Mead 1932). This is how a self positions itself temporally. It is the precondition for taking the position of the other. Similarly, in a phenomenological tradition time and space play a crucial role. The theory of ex-centric positionality refers to a form of reflexive self-positioning, whereby a living body actively occupies presently a particular spatial position and as such stands out from an undifferentiated spatial background. This spatio-temporal self-positioning is the point of reference from which living bodies seem to set up their directions into a space shared by other living bodies as well. Ex-centric positionality is described as a reflexive loop, enabling this absolute self-localization to be relativized and thus the position of the other and of third parties to be taken up.

Whether we refer to Mead or to Schmitz and Plessner, each of these models assumes that there must be some form of reflexive self-positioning as a precondition for taking the position of the other. That this form of self-reflexive positioning exists is corroborated by our data. Robots apparently exist in a differently constructed time/space – a time without present and a space without centres, without spontaneous directions, and without the possibility of taking the position of the other. Within this digital space/time, it is an extremely complicated mathematical enterprise to position any kind of body concretely. Each body is only an objectified body, the position of which has to be calculated for particular points in time. Such bodies do not occupy a particular space by themselves. In-stead, their position has to be calculated externally.

If these bodies appear in the space common to living bodies, they may spontaneously be treated as social actors by living bodies. Although we did not present them here, there are interaction sequences involving lay people in our data that support this. Nevertheless, the engineers, at least among themselves, never refer to S-Rs as social actors. They seem quite aware of the fact that their creatures lack some crucial characteristics of what it is that makes a social actor. Thus the observed practices of social robotics are characterized by a two-fold reality: lay people may occasionally ascribe some features of social actors to S-Rs, whereas for the engineering experts S-Rs are nothing but a technical system, the agency of which is an engineered construction. This second reality is the main subject of our article.

To improve the simulation of social interaction, the problem of spatio-temporal positioning has to be solved. We assume there are two technical solutions. The first would be generating learning automata that can position themselves reflexively and interact spontaneously with a real-world environment including a centred space. The development of a radically new engineering approach to manage the paradoxes of self-positioning and self-reflexivity would be crucial to this alternative. Biologically-inspired robotics may have potential for such a breakthrough. The second possibility would be for robotics to drop the idea of constructing artificial social agency, and try instead to make maximal use of recursive calculation and/or ambient intelligence. Learning automata whose operations are based on recursive calculations already exist. Good examples are autonomous vacuum cleaners that can construct a map of a limited space and localize themselves within it. The reach of such robots could be

extended by taking full advantage of ambient intelligence. This would imply a constant monitoring of a larger space. In places where S-Rs would work, each moving or movable body (humans, rats or tables) must be continuously observed and their relative positions calculated. The more precisely all the bodies involved are traced, the easier it will become for S-Rs to simulate spontaneous actions of bodies that position themselves reflexively.

The first solution relies on further technological, especially mathematical, innovations, which could lead to less controllable machines. The second solution requires more effective high-performance computing, able to handle the enormous amounts of data emerging from seamless surveillance of bodies of all kinds. This second solution is probably easier to achieve and it is more compatible with streamlining social agency within a calculable digital space-time. However, it is a scenario likely to increase the risk of a surveillance society. How would lay users feel about an autonomous black box whose functioning is predicated on continuous surveillance? If such a technology is deployed in public and/or private spaces, it may be used for spying on personal information. Introducing S-Rs into everyday life will therefore require new kinds of legal regulations, in order to prevent an invasion of privacy by the misuse of robotic technology.

References

Alač, 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, 893-926.

Berger, Peter L./Thomas Luckmann, 1966/1991: *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Harmondsworth: Penguin.

Bourdieu, Pierre, 1972/1977: *Outline of a Theory of Practice*. Trans. Richard Nice.

Cambridge: University of Cambridge Press.

Callon, Michel, 1986: Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. In: John Law (ed.), *Power, Action and Belief: A New Sociology of Knowledge*. London: Routledge, 196-233.

Elias, Norbert, 1992: *Time: An Essay*. Trans. in part from the German Edmund Jephcott. Oxford: Blackwell.

Feil-Seifer, David/Kristine M. Skinner/Maja J. Matarić, 2007: Benchmarks for evaluating socially assistive robotics. In: *Interaction Studies* 8: 423-429.

Garfinkel, Harold, 1967: *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.

Garfinkel, Harold, 2002: *Ethnomethodology's Program Working Out Durkheim's Aphorism*, ed. Anne Warfield Rawls. Lanham, MD: Rowman & Littlefield.

Giddens, Anthony, 1984: *The Constitution of Society*. Cambridge: Polity Press.

Glaser, Barney G./Anselm L. Strauss, 1967: *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.

Goffman, Erving, 1956: *The Presentation of Self in Everyday Life*. New York: Doubleday.

Goffman, Erving, 1974: *Frame Analysis: An Essay on the Organization of Experience*. Cambridge, MA: Harvard University Press.

Guizzo, Erico, 2011: How Google's Self-Driving Car Works, <http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/how-google-self-driving-car-works> (accessed 11 February 2013).

Habermas, Jürgen, 1981/1995: *Theorie des kommunikativen Handelns*. 2 vols. Frankfurt a.M.: Suhrkamp.

Joas, Hans, 1989: *Praktische Intersubjektivität*. Frankfurt a.M.: Suhrkamp.

KPMG, 2012: Self-Driving Cars – The Next Revolution, <http://www.kpmg.com/US/en/IssuesAndInsights/ArticlesPublications/Documents/self-driving-cars-next-revolution.pdf> (accessed 11 February 2013).

Latour, Bruno, 2004: *Politics of Nature: How to Bring the Sciences into Democracy*. Cambridge, MA: Harvard University Press.

Latour, Bruno, 2005: *Reassembling the Social. An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Luhmann, Niklas, 1969/1983: *Legitimation durch Verfahren*. Frankfurt a.M.: Suhrkamp.

Luhmann, Niklas, 1972: *Rechtssoziologie*, vol. 1. Reinbek bei Hamburg: Rowohlt.

McTaggart, J.M. Ellis, 1908: The Unreality of Time. In: *Mind* 17: 457-474.

Mead, George H., 1932: *The Philosophy of the Present*, ed. Arthur E. Murphy. La Salle, IL: Open Court.

Mead, George H., 1934/1967: *Mind, Self, and Society*. Chicago: University of Chicago Press.

Plessner, Helmuth, 1928/1975: *Die Stufen des Organischen und der Mensch. Einleitung in die philosophische Anthropologie*. 3rd ed. Berlin: de Gruyter.

Rammert, Werner, 2012: Distributed Agency and Advanced Technology, Or: How to Analyze Constellations of Collective Inter-Agency. In: Jan-Hendrik Passoth, et al. (eds.), *Agency without Actors? New Approaches to Collective Action*. London: Routledge, 89-112.

Rammert, Werner/Ingo Schulz-Schaeffer, 2002: Technik und Handeln. Wenn soziales Handeln sich auf menschliches Verhalten und technische Abläufe verteilt. In: Werner Rammert/Ingo Schulz-Schaeffer (eds.), *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*. Frankfurt a.M.: Campus, 11-64.

Salvini, Pericle, et al., 2011: The Robot DustCart. In: *IEEE Robotics & Automation Magazine* 18: 59-67.

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

Schmitz, Hermann (1964–1980): *System der Philosophie*. Bonn: Bouvier.

Schmitz, Hermann, 1964: Die Gegenwart. In: *System der Philosophie*, vol. I. Bonn: Bouvier.

Schmitz, Hermann, 1980: Die Person. In: *System der Philosophie*, vol. IV. Bonn: Bouvier.

Schütz, Alfred, 1932/1981: *Der sinnhafte Aufbau der sozialen Welt. Eine Einleitung in die verstehende Soziologie*. Frankfurt a.M.: Suhrkamp.

Searle, John, 1980: Minds, Brains, and Programs. In: *Behavioral and Brain Sciences* 3: 417-424.

Sharkey, Amanda/Noel Sharkey, 2011: Children, the Elderly, and Interactive Robots. In: *IEEE Robotics & Automation Magazine* 18: 32-38.

Simmel, Georg, 1908: *Soziologie. Untersuchungen über die Formen der Vergesellschaftung*. Berlin: Duncker & Humblot.

Turkle, Sherry, 2011: *Alone Together. Why We Expect More from Technology and Less from Each Other*. New York: Basic Books.

Yamazaki, Ryuji, et al., 2012: Social Acceptance of a Teleoperated Android: Field Study on Elderly's Engagement with an Embodied Communication Medium in Denmark. In: *Lecture Notes in Computer Science* 7621: 428-437.