Why Collaborative Robots Must Be Social (and even Emotional) Actors

Kerstin Fischer

Abstract: *In this paper, I address the question whether or not robots should be social actors and suggest that we do not have much choice but to construe collaborative robots as social actors. Social cues, including emotional displays, serve coordination functions in human interaction and therefore have to be used, even by robots, in order for long-term collaboration to succeed. While robots lack the experiential basis of emotional display, also in human interaction much emotional expression is part of conventional social practice; if robots are to participate in such social practices, they need to produce such signals as well. I conclude that if we aim to share our social spaces with robots, they better be social actors, which may even include the display of emotions. This finding is of empirical as well as philosophical relevance because it shifts the ethical discussion away from the question, how social collaborative robots should be, to the question, what kinds of human-robot collaborations we want.*

Key words: social robotics, collaboration, social cues, emotion, simulation, robots as social actors

1. Introduction

It is commonly believed that anthropomorphizing robots, i.e. treating them as if they had human-like capabilities, is due to a kind of misunderstanding, misconception or mindless error (initially for human-computer interaction by Nass and Moon 2000, then also for human-robot interaction, e.g. by Sparrow and Sparrow 2006). Correspondingly, philosophers have argued that social robots are inherently problematic because they pretend to possess. For instance, Robert Sparrow and Linda Sparrow argue that the use of robots in social roles is unethical:

> Insofar as robots can make people happier only when they are deceived about the robots' real nature, robots do not offer real improvements to people's well-being; in fact, the use of robots can be properly said to harm them. The desire to place robots in caring roles is therefore foolish; worse than that, it is actually unethical. (2006, 155)

The reason for the suspected harm thus lies in the suspected deception, which is taken to keep people from understanding the robots' real nature (cf. also Wallach and Allen 2009, 44-45 for a slightly weaker suggestion).

While this position is not uncontested in the philosophical literature (e.g. Sharkey and Sharkey 2012; Matthias 2015) and while it is not certain, either, that the underlying psychological process is really one of deception (see, for instance, the difference between children's beliefs about robots and their behavior towards robots documented in Melson et al. 2009), also in human-robot interaction research, scholars have asked that robot design should emphasize that robots are actually machines in order to prevent people from making unjustified attributions. A recent example of this position can be found in the call for papers to a workshop on Explainable Robotic Systems at the Human-Robot Interaction Conference in Chicago in March 2018:

> The implementation and use of explainable robotic systems may *prevent the potentially frightening confusion over why a robot is behaving the way it is*. Moreover, explainable robot systems may allow people to *better calibrate their expectations* of the robot's capabilities and be *less prone to treating robots as almost-humans*. (Call for: HRI 2018 Workshop on Explainable Robotic Systems,[1] emphasis mine)

At the same time, there is much work in human-robot interaction that aims to endow robots with increasing amounts of human-like capabilities; examples comprise, for instance, robot gaze behavior in conversation (e.g.

[1] https://explainableroboticsystems.wordpress.com/

Andrist et al. 2014) or during robot approach (e.g. Fischer et al. 2016), human-like proxemics (e.g. Walters et al. 2007), human-like timing (e.g. Lohan et al. 2011), among many others. Furthermore, robots are explicitly designed to take over human social roles, such as the role of a teacher (e.g. Edwards et al. 2016) or as a companion (e.g. Dautenhahn et al. 2005), and they are being developed for dementia care and autism therapy (cf. Scassellati, Admoni, and Matarić 2012).

In theory, the two positions do not exclude each other; it is theoretically conceivable to build robots that make use of social signaling mechanisms because these signals facilitate interaction while resisting the temptation to make robots human-like in other respects. Moreover, it seems theoretically possible to let the human interaction partner know at any given moment that the respective robot is an artificial agent and thus *per se* limited in terms of the biological processes that underlie the behaviors it may display. That is, we could create robots that do not use deceptive social cues. However, I will show that this is practically not possible. Thus, in this paper, I argue that if we want to be able to collaborate with robots, whether we like it or not, they have to be socially savvy, both in the perception of social cues and in their production. With collaboration I mean "to work jointly with others or together especially in an intellectual endeavor" (Merriam-Webster) or to “work in conjunction with another or others, to co-operate” (Oxford English Dictionary.) Thus, my considerations concern any kind of joint action (Clark 1996) between humans and robots.

In particular, I argue that collaboration generally involves coordination at many levels simultaneously and that therefore explicit coordination would be slow and tiresome. Having robots understand already established, conventional coordination systems, such as the social tools developed over thousands of years in human interaction, is thus not only the rational choice for the design of novel collaborative systems, such as robots, but other means of collaboration are also likely to be much slower, require more effort and training and will thus restrict human-robot collaborations to specific situations of use. The example with which I illustrate my point is linguistic feedback, which constitutes a highly optimized system for complex multilayer coordination, both explicitly and implicitly.

I will then suggest that both the production and the perception of social cues by all interaction partners are required for successful collaboration. That is, just having robots understand but not produce social signals is infeasible. This raises issues regarding the simulation of capabilities that are not actually there, which I address in section 5, using emotional expression as an example. I conclude that we have no choice but to make collaborative robots social, and even emotional, actors if we want collaboration to succeed. The philosophical discussion of what robots should and should not do needs to take into account what is feasible during collaboration.

2. Collaboration in Human Interaction

2.1. Collaboration, to the extent that it is joint action, requires coordination at many different levels simultaneously (Clark 1996). For instance, when listening to a conversation partner, we have to let our partner know that we hear her, that we hear what she is saying, that we understand what she is saying, that we understand it as a contribution to a particular topic or in a particular context, and finally how we think about it and about her (cf. Allwood, Nivre, and Ahlsén 1992). Correspondingly, languages possess large inventories of feedback signals that allow speakers to signal to each other delicately how they are doing with respect to these tasks (cf. Fischer 2000).

One example is the feedback signal *mm*; Gardner (2001) has analyzed the social practices in which *mm* is embedded in great detail. He finds two different uses of *mm* depending on its prosodic realization, that is, depending on the speech melody it is associated with. In particular, *mm* with a fall-rising intonation contour is regularly found after turns that are articulated unclearly or are conceptually difficult to understand, and it frequently occurs in a sequentially incomplete position (e.g. in the middle of a story). In contrast, *mm* with a falling intonation contour is regularly found after turns that exhibit no problems in articulation, are conceptually simple or straightforward, are not emotionally or judgmentally strongly expressive, and which are pragmatically, grammatically and intonationally complete. Gardner (2001) illustrates the function of the feedback signal *mm* with falling intonation in the following exchange:

> Ron: We had an appalling meeting tonight,
>
> (1.7)
>
> Sally: So did we:.
>
> Ron: Mm:.
>
> Sally: what was yours about-.

As we can see from Sally's response, Ron's use of *mm* with falling intonation indicates to his communication partner that he considers the current topic, her own appalling meeting, to be complete, conceptually simple and straightforward, and that he does not need to hear more about it. *Mm* is thus a very short word to fulfill highly complex functions that, if explicated, would correspond to a long list of statements. Furthermore, explicating the functions a feedback signal fulfills would not allow the speaker to fulfill these functions; for example, if you want to signal the communication partner that she should continue with what she is doing, then it is counterproductive to assert this in multiple statements. Thus, linguistic inventories provide shortcuts to solving complex tasks in highly economic ways. At the same time, it is not only *mm* by itself that fulfills all of these functions, but its conventional embedding into particular social practices (cf., for instance, Heritage 2005). For instance, for *mm* to have the functions outlined, it has to occur at specific places in interaction at transition relevance places, i.e. at places of potential pragmatic, semantic, grammatical and prosodic completion, at which a switch of speaker roles becomes possible (cf. Sacks, Schegloff, and Jefferson 1974). Correspondingly, during language acquisition, the child is socialized into mastering these social practices (see Filipi 2009) on how this may work, for instance, with respect to feedback signals). Learning a language thus means learning to navigate complex social collaboration processes, such as signaling what exactly is understood and how much more information one wishes to hear about a particular topic. Using language is therefore necessarily social action (cf. also Clark 1996).[2]

At the same time, language is the most economic tool for coordination such that coordination and joint action are most efficient using language. Language is a human tool developed and optimized over thousands of years for the complex negotiation of the *who, what, when, how, where* and possibly the *where-to* people may want to communicate. In addition, it fulfills many other functions, such as interpersonal relationship regulation, speech management or discourse structuring, just to name a few (e.g. Sacks, Schegloff, and Jefferson 1974; Clark 1996, 2002; Fischer 2000). Nevertheless, especially in cooperation, language is often smoothly interconnected with other kinds of social cues; for instance, Herbert Clark and Meredyth Krych (2004) describe how pointing and placing contribute to collaboration, and how they can take over linguistic action if the context is sufficiently clear. Similarly, Mutlu et al. (2012) summarize how gaze can support coordination in human-human and in human-robot interaction; Ambra Bisio et al. (2014), for instance, show how people exploit their partners' gaze behavior to anticipate their next

[2] This explains why language use by robots increases so-called 'mindless transfer' in human interactants (see Nass 2004), i.e. that if computers or robots use natural language, people are more inclined to understand them as social actors.

actions, and Anna-Lena Vollmer et al. (2010) demonstrate how children's proactive gaze is used by their parents as information as to the extent to which the child has understood the instruction.

To sum up, human social cues, including language, enable highly efficient and smooth collaboration between humans. Some of these signals are implicit, and by themselves, highly ambiguous; their functions are mostly due to their situational grounding and sequential placement in interaction, i.e. their interactional embedding in social practices. Other social signals are more straightforward in their interpretation and less context-sensitive, such as symbols like stop signs. Crucially, however, explicit and implicit modes of communication interact during successful and effortless cooperation.

3. Interacting with Robots

There are many possible ways of interacting with robots, and not all of them involve cooperation in the way outlined above; instead, interactions with robots can also take place using fewer, specialized interaction modalities, and they may also be restricted to interactions between robots and specialists, such as computer programmers. Therefore, such interactions do not rely on social cues and often require special training. Furthermore, there is also the option to separate humans and robots, so that coordination is not necessary; this is the approach currently taken in factories where robots and humans work in separate areas, but even in industry, people are striving for increasing human-robot collaboration (e.g. Philipsen, Matthias, and Thomas 2018).

If untrained personnel or the general population is involved, human-robot interactions generally rely on social cues since people have been found to understand social cues from computers and robots in similar ways as human social cues (see also Cerulo 2009); this has been ubiquitously demonstrated on many different cues. For instance, people have been shown to respond to blame attribution from robots in similar ways as they respond to blame attribution from other people (Groom et al. 2010), they attribute female characteristics to robots with female voices (Nass and Brave 2005), they feel closer to robots that are introduced them as in-group than as out-group members (Eyssel and Kuchenbrandt 2012), and they interpret a robot’s camera movements (i.e. ‘eye gaze’) as signs of attention (Lohan et al. 2011), just to mention a few results from the literature demonstrating this effect ubiquitously. While the purpose of these experiments often is to illustrate some sort of ‘mindless transfer’ from the human realm to the realm of computers and robots (Nass and Moon 2000), they do demonstrate that when robots use human social practices, they are also responded to in line with these social practices – at least overwhelmingly so. In fact, Kerstin Fischer (2011a, 2016) shows that people differ considerably concerning the degree with which they anthropomorphize robots and treat robots as social actors, though people who refuse to respond to robots on the social level at all are rare overall.

Now, robots may communicate via various different kinds of interfaces which allow different input modalities; for instance, touch screens are excellent to communicate the *what, where* and *where-to* from human to robot. Demonstration can be used (one way) for the *how* and the *what*, yet negotiation about meanings is non-trivial (cf. Chernova and Thomasz 2014); that is, if an instruction is not clear, the robot may have to resort to other modalities to clarify the instruction (Cakmak and Thomaz 2014). In contrast, language and other social cues can not only signal them all (partly implicitly and hence) highly economically, they also allow the interactive negotiation and collaboration about these issues. That is, if an instruction is not fully understood, the listener can signal this by means of a clarification question, a feedback signal (e.g. *huh*?), or even only by means of a quizzical look. The speaker can repair her utterance while speaking, expanding it to make it clearer, or reply to the clarification question (see Schegloff, Jefferson, and Sacks 1977). In contrast, in other modalities, often a switch to another modality or type of information is required; for instance, when using a touch screen that displays a map on which the user can select a certain location, the robot would have to devise a special method to disambiguate the instruction in case something is unclear (for example, the robot could display its visual field to the user); thus, other input channels are not always reciprocal, slowing negotiation down.

Furthermore, language and social cues do not presuppose any particular training of the human user, which is especially important in situations in which a training phase is not feasible, as in first time encounters with robots, for instance, when the robot serves as a guide in a shopping mall, or when users cannot be expected to learn new interaction methods, for instance, because they are cognitively challenged. Correspondingly, the endowment of robots with social cues has been shown to have a facilitative effect on usability, task efficiency, ease of use etc. (e.g. Admoni et al. 2014; Andrist et al. 2013, 2014; Fischer et al. 2016; Jensen et al. 2017; Nass 2010, among many others); much work in human-robot interaction is therefore currently dedicated to making robots understand and produce social signals. We can conclude that endowing robots with the ability to understand social cues (including language) enhances collaboration because of familiar implicit and explicit mechanisms to communicate information and to quickly and easily disambiguate, negotiate or repair it in case of problems arising. These considerations suggest that robots should understand human social signals if they are to collaborate with humans smoothly and efficiently. We can of course still decide for non-social robots only and deprioritize ease of use. However, as Andreas Matthias (2015) argues, the use of social cues in human-robot interaction serves to empower users to interact with technologies that they otherwise would need extensive training for and which would thus most likely remain inaccessible to them. From that perspective, a decision against the use of social signaling systems may not only make robots tiresome to use, but also prevent large groups of the population from interacting with them at all, which has ethical consequences too.

4. Robots Using Social Signals

In addition to understanding social signals, robots should also be able to use them; just processing social signals without using them themselves would be possible, and it would ensure that the robot interprets the human correctly while using other kinds of signals itself in order not to pretend to have capabilities it actually does not have. However, such a solution is problematic for several reasons: First, refraining from actively using capabilities that the robot masters passively would be misleading to users since they will not be able to build up an accurate mental model of the respective robot's capabilities. That is, if the robot understands social cues but does not produce them, people may constantly underestimate its capabilities since its affordances are not visible. People have been found to make use of all aspects of robot appearance and behavior to infer its capabilities and to design their behavior accordingly (Fischer 2016). For instance, if a robot uses a particular word, people expect it to also understand that word when they use it (e.g. Richards and Underwood 1984). The use of certain capabilities is therefore taken as indirect evidence of related capabilities, especially more basic ones (Fischer and Moratz 2001; Fischer 2011b, 2018), like an agreement providing indirect evidence that the utterance has been perceived and understood. In this way, robots are like all kinds of technology in that people look for cues that help them understand how the technology is intended to be used (cf. Norman 1988).

Second, if robots do not use social signals themselves, they will have to make information explicit that is usually signaled implicitly, such as the fact that an instruction was heard and understood. However, this can be very disruptive, like in the case of feedback signals illustrated above: While *mm* with rising intonation is used in situations in which more information from the communication partner is expected (Gardner 2001), that function cannot be fulfilled by stating explicitly that the partner should continue, since in that case, the robot has already taken the turn and prevented the partner from continuing.

In addition, explicit signals of successful understanding may have devastating effects on users' mental models of the robot. For instance, Fischer (2011b) had participants teach a robotic wheelchair the names of locations in an apartment for handicapped people, where the robot signaled either explicitly or implicitly what it had understood. The user would, for instance, steer the robot to the refrigerator and say "and this is the fridge". In the explicit condition, the robot would then say "I understood fridge. Is this where you want to be to open it?", whereas in the implicit condition, the robot would only pose the question: "Is this where you want to be to open it?". Since participants were free to steer the robot to any location they thought relevant and also to as many as they wanted, they heard either of the two different responses between one and three times, depending on whether they steered the robot to the three locations for which the different responses were provided; thus, the interactions in the two conditions, which lasted between 20 minutes and half an hour, differed only minimally. Nevertheless, participants in the explicit condition steered the robot to significantly fewer locations (6 compared to 10 on average), talked to it significantly less and finished significantly earlier. So the effects of hearing one, two or three explicit confirmations about what was understood, while everything else was identical across conditions, had considerable effects on the interactions. The reason is most likely that a robot that makes successful understanding explicit provides the signal that understanding is generally a problem. That is, by communicating explicitly that an utterance was understood, the robot implicitly communicated that understanding was not self-evident and may also have been unsuccessful.

This example illustrates that using explicit information about a robot's capabilities is far from trivial; robots cannot simply communicate what they can and cannot do since every information they provide gives rise to further inferences – which is extremely useful in interactions between humans (cf. Clark 1998) as a way to build up common ground (Clark 1996), yet which may hinder transparency in human-robot interaction. Furthermore, as the significantly shorter interactions with the 'explicit' robot show, a robot that communicates its understanding explicitly is perceived as much less pleasant to interact with. Thus, in order to warrant pleasant and successful interactions, robots should use the same socially implicit encoding of information as humans do in interaction, backgrounding what is not at issue (see also Fischer 2018), if interactions are to be efficient and pleasant.

Moreover, robots should not only understand but also use social practices in order to increase the readability of their behavior to users. For instance, regarding gaze, much previous work shows that social gaze behavior puts people at ease because they feel that they can predict the robot's behavior better (e.g. Admoni et al. 2014; Fischer et al. 2016). To conclude, it is not enough that robots understand human social signals to facilitate communication, they must also use them themselves if collaboration is to succeed.

5. Limits to Social Signaling: The Case of Emotional Expression

So far I have argued that collaboration between humans and robots can profit considerably from using social practices from human interaction. However, as indicated above, there are also reasons against endowing robots with social cues, and there may be limits to what serves the purposes of smooth collaboration. In particular, if robots use social signals, this raises issues of robot simulation and potential discrepancies between what the robot signals and how the robot 'really' works (cf. Seibt 2017). One problem is that robots' understanding of human social signals is currently very restricted; for instance, dialog interfaces for human-robot interaction do not allow the same range of possibilities as human language interaction does with respect to the available inventory, timing and processing speed and accuracy. Furthermore, dialog systems so far concentrate on the content side, which is still dealt with imperfectly; all language functions beyond the direct transfer of information are ignored, perhaps apart from basic politeness issues (cf. Gunkel 2016). Thus, addressing a user with a perfect "hello, how are you feeling today?" may invite the conclusion that the robot will fully understand the user's reply even though it may simply spot the one or other keyword in her answer.

Also with respect to other capabilities, if robots are currently endowed with the one or other social behavior, these remain isolated capabilities, implemented in research environments and evaluated in controlled user studies to provide proof of concept; so even if robots follow a speaker's eye gaze, these robots do not also understand speech, produce feedback signals or gesture, just to mention a few other potential areas of multimodal coordination. Furthermore, very few of these behaviors are robust enough to be taken out of the lab and to be implemented in commercial robots. When we speak of robots processing and using social signals, then we are discussing future technologies.

Nevertheless, irrespective of how perfect or imperfect robot technologies currently are, there may be discrepancies between what the robot signals and what the signal means to the human and the robot respectively. I am going to address the issue using emotional displays as an example. For instance, much recent research aims at implementing emotional signals into robots, where the challenge is taken to consist in producing emotional displays unambiguously, given that most robots lack expressive means and have very different morphologies; thus, this research concentrates on developing platform (i.e. robot morphology) independent expressive inventories, often for the core emotions anger, happiness, fear and sadness (see, for instance, Löffler, Schmidt, and Tscharn 2018) and uncritically assumes that robots should display emotional stance (e.g. Song and Yamada 2017; see Jung 2017; Fischer et al. 2019). However, obviously, these emotional signals do not correspond to emotional states in the robot, creating a mismatch between what is signaled and the basis for these signals. Such an approach ignores both the functional and the social basis of emotional display in human interaction, where emotions serve to manage delicate personal and interactional needs (e.g. Couper-Kuhlen 2009; Ekberg et al. 2016). Since robots are artifacts, they do not have those needs, unless one may want to endow them with a sense of self-preservation, in which case they may want to respond negatively to being out of power or having their memories erased - ‘convincing’ robot needs, which have been successfully exploited in HRI experiments such that experiment participants tended to respond favorably to the robots’ displayed needs (e.g. Seo et al. 2015; Kahn et al. 2015; Westlund et al. 2016). In such a situation, when attention to robot ‘needs’ is relevant, emotional signals can provide intuitively understandable indicators that action should be taken without disrupting another, main activity, such as a collaborative task. Otherwise, having a robot drive around in ‘happy’ or ‘sad’ states violates the nature of signaling, where the signal refers to a state that the robot is not in. Such a signal is thus misleading at best, if not manipulative. Nevertheless, even in this case, signals of positive emotions may serve to make people feel good, and as Andreas Matthias (2015) suggests, one may ask whether such an effect is not only desirable but also morally implicated, even if it entails deception.

In any case, what makes emotional displays necessary in human-robot interaction is the fact that emotional expression is largely socially defined. Hence, providing robots with the ability to decode emotional signals and to produce them where expected is the only rational choice if interaction quality is the goal. That is, emotional displays are socially required as integral parts of human activities (Jung 2017; Fischer et al.

2019). For instance, the delivery of bad news is conventionally associated with signs of empathy (e.g. Maynard 1997; Ekberg et al. 2016) in many cultures. Furthermore, as Erving Goffman (1978) has argued, conventional signals like interjections (‘response cries’, like *oops*), serve to indicate that a certain mishap, such as stumbling, is an exception, and that the speaker generally conforms with socially accepted norms of walking in a straight and predictable manner. Thus, social practices conventionally comprise emotional expression, which is independent of the respective speakers’ emotional states.

One such socially defined emotional display is the listener’s response during storytelling; here, as Margret Selting (1994) has demonstrated, listeners are expected to match the tellers’ signs of involvement, which in turn are expected to increase the closer the speaker is getting towards the climax of the story. If listeners fail to produce the expected signals of involvement, this has an impact on the speakers’ task performance, as demonstrated by Janet Bavelas, Linda Coates and Trudy Johnson (2000), who distracted people who were listening to a story to different degrees. They find that when listeners were so distracted that their feedback signals arrived not exactly at the expected time or not with the expected level of involvement, the stories speakers produced were worse, as rated *post hoc* by independent coders. Thus, emotional display may be interactionally required.

Similarly, during the delivery of bad news, signs of empathy are normatively required (see Maynard 1997); that is, if such signals are not provided, they are noticeably absent (in the ethnomethodological sense, see, for instance, Heritage 1988). For example, a robot that is providing information about store hours or about the availability of goods in a shopping mall, about train schedules in a railway station or about the availability of employees at a reception desk – tasks that are not unlikely to be fulfilled by robots in the near future – such a robot will have to provide appropriate signs of empathy when goods or contact persons are not available or trains are delayed in order to be rated as acceptable (Jung 2017); in a study in which a robot provided bad news either with or without such signs of empathy, the robot was rated as significantly less friendly, warm, polite and engaging when it did not use emotion expression (Fischer et al. 2019). Thus, while one can argue that a robot cannot be expected to produce signs of empathy because as a machine, it does not have that capability, the decision not to endow a robot with such signals has considerable interactional consequences, which may reflect back negatively onto the company or organization the robot represents.

To sum up, social and emotional signals are human societies' shortcuts to solving complex coordination tasks at different levels simultaneously. If robots are to cooperate with people, then having robots both understand and display social signals has the advantage that understanding may become intuitive (since it relies on conventional inventories of behaviors) and that coordination becomes seamless because negotiation is implicit and does not require extra effort or attention. Emotional display, if it serves coordination functions such that it indicates a robot's real needs (e.g. battery status), or if it is conventionally required as part of how a social practice works, will facilitate collaboration and thus be useful. In contrast, if robots fail to understand social signals, they will be understood as impolite, cold and uncooperative and thus as inacceptable, and they will be in the way all the time because they fail to coordinate implicitly and thus require extra effort – effort people may not be willing to spend as long as other humans are available.

6. Discussion

The point I have tried to make in this paper using empirical evidence from both human and from human-robot interaction is that collaboration between humans and robots has to make use of human collaboration systems, i.e. social, interactional systems, in order to succeed, to run smoothly and to be perceived as an asset and not as a burden. We may now ask whether collaboration necessarily involves such social signaling, or whether one can also conceive of human-robot collaborations in which robots refrain from using social signals. The latter may desirable from a perspective that takes the social signaling by robots to disguise their true mechanical nature. In the following, I therefore discuss the implications of the arguments presented above for a philosophical discussion.

First, let us consider whether collaboration necessarily involves social signaling. Self-evidently, there are also situations in which social cues are not necessary. For instance, in a chess game, participants can collaborate using one specific interaction modality, namely the movement of the chess pieces; additional coordination is not necessary, and consequently also no additional coordination by means of social cues. Whether or not social cues are required to ensure the quality of collaboration therefore depends on the interaction modality and on the complexity of the collaboration. Moreover, it depends on the human interaction partners; for example, while computer scientists may be able to interact with a robot by means of computer code, most people do not. Similarly, people differ to the extent to which they engage in social interaction with robots (Fischer 2011a); while some people easily enter in social exchanges with robots that address their human communication partners in a conversational manner, others respond in a task-oriented manner and refuse to reciprocate the social signaling by the robot. The degree of sociality employed may thus also be a matter of taste, capability or personality. While the richness of social signaling systems facilitate the coordination between human and robot in both cases, people may be willing to different degrees to enter social interactions with robots (see also Fischer 2016). At the same time, people may also be willing to different degrees to invest effort into the collaboration, for instance, by learning to adjust to the robot over time (cf. also Matthias 2015). From that perspective, my arguments above do not hold for all collaborations between humans and robots, but for those that take place in human space, concern human activities (i.e. activities otherwise carried out by humans) and that do not require any other training than the usual human socialization. Furthermore, some people may also be willing to put up with tiresome, effortful interactions with clumsy robots. Consequently, the decision we as a society are faced with (cf. Seibt, Damholdt and Vestergaard 2018) is not whether a robot should or should not use social signals, but rather whether or not we want to collaborate with robots, and if so, what kinds of collaborations we want and who we want to collaborate with robots.

The second implication of our discussion concerns the fact that social, and especially emotional, signals used by robots are possibly problematic due to the fact that they originate from very different mechanisms than human social signals and thus that they can lead to conclusions about capabilities that robots do not have. In the current paper, I want to restrict myself to the suggestion that emotional display should be opted for in order to facilitate collaboration only; that is, the robot design team (cf. Seibt, Damholdt and Vestergaard 2018) has to consider what kinds of activities the robot is supposed to be engaged in and what kinds of displays are necessary to facilitate the collaboration. Furthermore, as Andreas Matthias (2015) suggests, robots should always be both able and willing to disclose their true nature and the nature of the social, and especially emotional, signals if so requested by its users. However, as discussed above, signaling positive emotions may have positive effects on users (see Matthias 2015), which may suggest that the use of emotional displays beyond the cues necessary to coordinate may be useful. Here the question really is whether the signaling of emotions is necessarily deceptive or not; the equation between the use of social, and especially emotional, signals with deception lies at the heart of the criticism voiced by Sparrow and Sparrow (2006). However, whether or not people are deceived is an empirical issue, and in fact much research shows that social responses to robots are largely automatic but not uncontrollable (see Roubroeks 2014), that there is considerable interpersonal variation (Fischer 2011a), that people respond differently when they have time (Fussel et al. 2008), and that even children know quite well that they are interacting with a machine, in spite of their behavior (Melson et al. 2009). In Clark and Fischer (in preparation), we are therefore currently working on an alternative model that suits the data better than the current model that relies on the notion of mindless error. If people are however not deceived by social signals produced by machines, then also the ethical problems with respect to deception disappear. Under these circumstances, only the empirical problem remains that social robots notoriously fail to signal their real capabilities; (entirely justified) requests for transparency of robots' capabilities may need to consider that this is practically hard to achieve.

To conclude, even though the lion share of this paper is based on empirical work, and the argument, that collaboration on human activities with robots that do not actively employ social signals is not going to be successful, is of a practical nature, it nevertheless has a considerable theoretical impact. Robots that are not socially savvy will not be perceived as useful by most users and hence not be bought and consequently not built to an extent that they become a noticeable factor in our lives.[3] If we assume that we can just conceive of robots as smart tools, we disregard the fact that the spaces they occupy are really social spaces, governed by social rules. The questions to be addressed instead are who will want to collaborate with robots and on what activities – not whether or not robots should be social actors.

References

Admoni, Henny, Anca Dragan, Siddhartha S. Srinivasa and Brian Scassellati. 2014. "Deliberate Delays during Robot-to-human Handovers Improve Compliance with Gaze Communication." Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction – HRI 14, doi:10.1145/2559636.2559682

Allwood, Jens, Joakim Nivre and Elisabeth Ahlsén. 1992. "On the Semantics and Pragmatics of Linguistic Feedback." *Journal of Semantics* 9(1): 1-26. doi:10.1093/jos/9.1.1

Andrist, Sean, Erin Spannan and Bilge Mutlu. 2013. "Rhetorical Robots: Making Robots More Effective Speakers Using Linguistic Cues of Expertise." 8th ACM/IEEE International Conference on Human-Robot Interaction – HRI, doi:10.1109/hri.2013.6483608

Andrist, Sean, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu. "Conversational gaze aversion for humanlike robots."

[3] In fact, we can already see that social robots are running into problems because of their limited capabilities, see, for example, https://www.fastcompany.com/90315692/one-of-the-decades-most-hyped-robots-sends-its-farewell-message

In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 25-32. ACM, 2014.

Andrist, Sean, Micheline Ziadee Halim Boukaram, Bilge Mutlu and Majd Sakr. 2015. "Effects of Culture on the Credibility of Robot Speech." Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction – HRI 15, doi:10.1145/2696454.2696464.

Bavelas, Janet B., Linda Coates and Trudy Johnson. 2000. "Listeners as Co-narrators." *Journal of Personality and Social Psychology* 79(6): 941-52. doi:10.1037//0022-3514.79.6.941.

Bisio, Ambra, Alessandra Sciutti, Francesco Nori, Giorgio Metta, Luciano Fadiga, Giulio Sandini and Thierry Pozzo. 2014. "Motor Contagion during Human-Human and Human-Robot Interaction," *PLOS ONE* 9(8). doi:10.1371/journal.pone.0106172

Cakmak, Maya and Andrea L. Thomaz. 2014. "Eliciting Good Teaching from Humans for Machine Learners." *Artificial Intelligence* 217: 198-215, doi:10.1016/j.artint.2014.08.005

Cerulo, Karen A. 2009. "Nonhumans in Social Interaction." Annual Review of Sociology 35(1): 531-52, doi:10.1146/annurev-soc-070308-120008

Chernova, Sonia and Andrea L. Thomaz. 2014. "Robot Learning from Human Teachers." Synthesis Lectures on Artificial Intelligence and Machine Learning 8, no. 3: 1-121. doi:10.2200/s00568ed1v01y201402aim028

Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge Univ. Press.

Clark, Herbert H. 1998. Communal lexicons. In *Context in Language Learning and Language Understanding*, ed. Kirsten Malmkjaer and John Williams, 63–87. Cambridge: Cambridge University Press.

Clark, Herbert H. 2002. "Speaking in Time." *Speech Communication* 36(1-2): 5-13, doi:10.1016/s0167-6393(01)00022-x

Clark, Herbert H. and Kerstin Fischer. in preparation. “Robots as dynamic depictions.”

Clark, Herbert H. and Meredyth A. Krych. 2004. "Speaking While Monitoring Addressees for Understanding." *Journal of Memory and Language* 50(1): 62-81. doi:10.1016/j.jml.2003.08.004.

Couper-Kuhlen, Elizabeth. 2009. "A sequential approach to affect: The case of ’disappointment’." In *Talk in interaction - comparative dimensions*, ed. Markku Haakana, Minna Laakso, and Jan Lindström, 94-123. Helsinki: Suomalaisen Kirjallisuuden Seura.

Dautenhahn, Kerstin, Sarah Woods, Christina Kaouri, Michael L. Walters, Kheng Lee Koay and Iain Werry. 2005. "What Is a Robot Companion - Friend, Assistant or Butler?" IEEE/RSJ International Conference on Intelligent Robots and Systems, doi:10.1109/iros.2005.1545189

Edwards, Autumn, Chad Edwards, Patric R. Spence, Christina Harris and Andrew Gambino. 2016. "Robots in the Classroom: Differences in Students’ Perceptions of Credibility and Learning between “teacher as Robot” and “robot as Teacher”." *Computers in Human Behavior* 65: 627-34, doi:10.1016/j.chb.2016.06.005

Ekberg, Stuart, Alison R. G. Shaw, David S. Kessler, Alice Malpass and Rebecca K. Barnes. 2016. "Orienting to Emotion in Computer-Mediated Cognitive Behavioral Therapy." *Research on Language and Social Interaction* 49(4): 310-24, doi:10.1080/08351813.2016.1199085

Eyssel, Friederike and Dieta Kuchenbrandt. 2011. "Social Categorization of Social Robots: Anthropomorphism as a Function of Robot Group Membership." *British Journal of Social Psychology* 51(4): 724-31, doi:10.1111/j.2044-8309.2011.02082.x

Filipi, Anna. 2009. *Toddler and Parent Interaction*. Amsterdam: John Benjamins.. doi:10.1075/pbns.192.

Fischer, Kerstin. 2000. *From Cognitive Semantics to Lexical Pragmatics: The Functional Polysemy of Discourse Particles.* Berlin: M. De Gruyter.

Fischer, Kerstin. 2011a. “Interpersonal variation in understanding robots as social actors.” In *Proceedings of the HRI'11 Conference, Lausanne, Switzerland, March 6-9th, 2011,* 53-60.

Fischer, Kerstin. 2011b. "How People Talk with Robots: Designing Dialog to Reduce User Uncertainty." *AI Magazine* 32(4): 31, doi:10.1609/aimag.v32i4.2377

Fischer, Kerstin. 2016. *Designing Speech for a Recipient: The Roles of Partner Modeling, Alignment and Feedback in So-called Simplified Registers.* Amsterdam: John Benjamins Publishing Company.

Fischer, Kerstin. 2018. "When 'transparent' does not mean 'explainable'." *Workshop on Explainable Robotic Systems, HRI'2018 Conference*, Chicago.

Fischer, Kerstin, Lars C. Jensen, Stefan-Daniel Suvei and Leon Bodenhagen. 2016. "Between Legibility and Contact: The Role of Gaze in Robot Approach." *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN),* doi:10.1109/roman.2016.7745186

Fischer, Kerstin, Malte Jung, Lars C. Jensen and Maria aus der Wieschen. 2019. "Emotional Expression by Robots: When and Why. " *Proceedings of the International Conference on Human-Robot Interaction*, Daegu, Korea.

Fischer, Kerstin and Rainer Moratz. 2001. "From Communicative Strategies to Cognitive Modelling. " *Proceedings of the First International Workshop on 'Epigenetic Robotics,'* Lund, Sweden.

Fussel, Susan R, Kiessler, Sara, Setlock, Leslie D. and Yew, Victoria. 2008. "How People Anthropomorphize Robots." *Proceedings of HRI'08*, Amsterdam, p. 145-152.

Gardner, Rod. 2001. *When Listeners Talk: Response Tokens and Listener Stance.* Amsterdam: J. Benjamins Pub.

Goffman, Erving. 1978. "Response Cries." *Language* 54(4): 787, doi:10.2307/413235

Groom, Vicotira, Jimmy Chen, Theresa Johnson, F. Arda Kara and Clifford Nass. 2010. "Critic, compatriot, or chump? Responses to robot blame attribution." In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction* (HRI '10). IEEE Press, Piscataway, NJ, USA, 211-18.

Gunkel, David J. 2016. "Computational Interpersonal Communication: Communication Studies and Spoken Dialog Systems." *Communication +1* 5(1). https://scholarworks.umass.edu/cpo/vol5/iss1/7/

Heritage, John. 1988. "Explanations as accounts: A conversation analytic perspective." In Analysing Everyday Explanation: A Casebook of Methods, ed. Charles Antaki, pp. 127–144. London etc.: Sage.

Heritage, John. 2005. "Cognition in discourse. " In Conversation and Cognition, ed. H. Te Molder and J. Potter, 184–202. Cambridge: Cambridge University Press.

Jensen, Lars Christian, Fischer Kerstin, Suvei Stefan-Daniel and Bodenhagen Leon. 2017. "Timing of Multimodal Robot Behaviors during Human-Robot Collaboration. " *Proceedings of the International Symposium on Robot and Human Interactive Communication, Ro-Man*, IEEE.

Jung, Malte F. 2017. "Affective Grounding in Human-Robot Interaction. " In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. ACM, New York, NY, USA, 263–73.

Kahn, Peter H., Takayuki Kanda, Hiroshi Ishiguro, Brian T Gill, Solace Shen, Heather E Gary, and Jolina H Ruckert. 2015. "Will people keep the secret of a humanoid robot? Psychological intimacy in HRI. " *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 173-80.

Löffler, Diana, Nina Schmidt and Robert Tscharn. 2018. "Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound*." Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction – HRI 18*, doi:10.1145/3171221.3171261

Lohan, Katrin S., Katharina J. Rohlfing, Karola Pitsch, Joe Saunders, Hagen Lehmann, Chrystopher L. Nehaniv, Kerstin Fischer and Britta Wrede. 2011. "Tutor Spotter: Proposing a Feature Set and Evaluating It in a Robotic System." *International Journal of Social Robotics* 4(2): 131-46, doi:10.1007/s12369-011-0125-8

Matthias, Andreas. 2015. "Robot Lies in Health Care: When is deception morally permissible?" *Kennedy Institute of Ethics Journal* 25(2): 169-92.

Maynard, Douglas W. 1997. "The News Delivery Sequence: Bad News and Good News in Conversational Interaction." *Research on Language & Social Interaction* 30(2): 93-130, doi:10.1207/s15327973rlsi3002_1

Melson, Gail F., Peter H. Kahn, Jr., Alan Beck and Batya Friedman. 2009. “Robotic Pets in Human Lives: Implications for the Human–Animal Bond and for Human Relationships with Personified Technologies.” *Journal of Social Issues* 65(3): 545-67.

Mutlu, Bilge, Takayuki Kanda, Jodi Forlizzi, Jessica Hodgins and Hiroshi Ishiguro. 2012. "Conversational Gaze Mechanisms for Humanlike Robots." *ACM Transactions on Interactive Intelligent Systems* 1(2): 1-33, doi:10.1145/2070719.2070725

Nass, Clifford. 2004. "Etiquette Equality: Exhibitions and Expectations of Computer Politeness." *Communications of the ACM 47*(4): 35-37.

Nass, Clifford. 2010. The Man Who Lied to his Laptop: What Machines Teach us about Human Relationships. New York: Penguin.

Nass, Clifford and Steve Brave. 2005. *Wired for Speech. How Voice Activates and Advances the Human- Computer Relationship*. Cambridge, MA., London: MIT Press, doi:10.1108/02640470610660459

Nass, Clifford and Youngme Moon. 2000. "Machines and mindlessness: Social responses to computers." *Journal of Social Issues 56*(1): 81-103.

Norman, Don. 1988. *The Design of Everyday Things*. USA: Basic Books.

Philipsen, Mark P., Rehm Matthias and Moeslund Thomas T. 2018. "Industrial Human-Robot Collaboration." In Proceedings of the FAIM/ISCA Workshop on AI for Multimodal Human Robot Interactio, 35-38, Stockholm: Association for Computing Machinery. Doi:10.21437/AI-MHRI.2018-9

Richards, M.A. and K.M. Underwood. 1984. "How should people and computers speak to each other?" *Proceedings of Interact* 84:215-218.

Roubroeks, Maaike A. J. 2014. *Understanding social responses to artificial agents: Building blocks for persuasive technology*. PhD Thesis, Eindhoven: Technische Universiteit Eindhoven, doi: 10.6100/IR774470

Sacks, Harvey, Emanuel A. Schegloff and Gail Jefferson. 1974. "A Simplest Systematics for the Organization of Turn-taking for Conversation." *Language* 50, no. 4: 696-735. doi:10.1353/lan.1974.0010.

Scassellati, Brian, Henny Admoni and Maja Matarić. 2012. "Robots for Use in Autism Research." *Annual Review of Biomedical Engineering* 14(1): 275-94, doi:10.1146/annurev-bioeng-071811-150036

Schegloff, Emanuel A., Gail Jefferson and Harvey Sacks. 1977. "The preference for self-correction in the organization of repair in conversation." *Language* 53(2): 361–82.

Seibt, Johanna. 2017. "Towards an Ontology of Simulated Social Interaction: Varieties of the "As If" for Robots and Humans." In Sociality and Normativity for Robot: Philosophical Inquiries into Human-Roboto Interaction, ed. Raul Hakli and Johanna Seibt, 11-40. Switzerland: Springer.

Seibt, Johanna, Marlene Damholdt and Christina Vestergaard. 2018. "Five principles of integrative social robotics: Five principles of integrative social robotics." In Envisioning Robots in Society. Proceedings of Robophilosophy 2018, ed. Mark Coeckelbergh et al., 28-42. Netherlands: IOS Press.

Selting, Margret. 1994. "Emphatic speech style: with special focus on the prosodic signaling of heightened emotive involvement in conservation." *Journal of Pragmatics* 22(3/4): 375-408.

Seo, Stela H., Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E. Young. 2015. "Poor Thing! Would You Feel Sorry for a Simulated Robot?" *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction – HRI 15,* doi:10.1145/2696454.2696471

Sharkey, Amanda and Noel Sharkey. 2010. "Granny and the Robots: Ethical Issues in Robot Care for the Elderly." *Ethics and Information Technology* 14(1): 27-40, doi:10.1007/s10676-010-9234-6

Song, Sichao and Seiji Yamada. 2017. "Expressing Emotions Through Color, Sound, and Vibration with an Appearance-Constrained Social Robot." In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI'17)*. ACM, New York, NY, USA, 2–11.

Sparrow, Robert and Linda Sparrow. 2006. "In the hands of machines? The future of aged care." *Minds and Machines* 16(2):141-61.

Vollmer, Anna-Lena, Karola Pitsch, Katrin Lohan, Jannik Fritsch, J., Katrina Rohlfing and Britta Wrede Britta. 2010. "Developing feedback: How children of different age contribute to a tutoring interaction with adults." *Development and Learning (ICDL), 2010 IEEE 9th International Conference on Development and Learning,* 76-81.

Wallach, Wendell and Colin Allen. 2010. *Moral Machines: Teaching Robots Right from Wrong.* Oxford: Oxford University Press.

Walters, Michael L., Dag S. Syrdal, Kerstin Dautenhahn, René Te Boekhorst and Kheng Lee Koay. 2007. "Avoiding the Uncanny Valley: Robot Appearance, Personality and Consistency of Behavior in an Attention-seeking Home Scenario for a Robot Companion." *Autonomous Robots* 24(2): 159-78, doi:10.1007/s10514-007-9058-3

Westlund, Jacqueline M. Kory, Marayna Martinez, Maryam Archie, Madhurima Das and Cynthia Breazeal. 2016. "Effects of Framing a Robot as a Social Agent or as a Machine on Childrens Social Behavior." *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, doi:10.1109/roman