Truly Social Robots
Understanding Human-Robot Interaction from the Perspective of Social Psychology

Daniel Ullrich[1] and Sarah Diefenbach[2]

[1]*Institute of Informatics, LMU Munich, Amalienstraße 17, 80333 Munich, Germany*

[2]*Department of Psychology, LMU Munich, Leopoldstraße 13, 80802 Munich, Germany*

Keywords: Social Robots, Human-Robot Interaction, HCI Theory, Social Psychology, Perception, Trust, Design Factors.

Abstract: Human-robot interaction (HRI) and especially social robots play an increasing role within the field of human-computer interaction (HCI). Social robots are robots specifically designed to interact with humans, and already entered different domains such as healthcare, transportation, or care of the elderly. However, research and design still lack a profound theoretical basis, considering their role as social beings, and the psychological rules that apply to the interaction between humans and robots. The present paper underlines this claim by a list of central research questions and areas of relevance, and a summary of first results of own and others' research. Finally, we suggest a research agenda and dimensions for a framework for social robot interaction, which truly accounts for their social nature and relevant theory from social psychology.

1 INTRODUCTION

Social robots play an increasing role within the field of human-computer interaction (HCI). In contrast to industrial robots in the context of industry 4.0, social robots are specifically designed to interact with humans. Nowadays, the most popular areas of application are healthcare (for an overview, see Beasley, 2012), transportation, retail, care of the elderly (e.g., Paro Robots, 2016), housekeeping, or robots taking the role of a social companion or pet-substitute (e.g., Robyn Robotics, 2016). With technological advancements, further domains will surely follow, so that the domain of social robots is about to become one of the most important in human-robot interaction (HRI). In this context, psychological questions such as how a robot is perceived, whether we trust or distrust it, accept or reject it, are of central relevance (Taipale et al., 2015).

Already in the 90s, Nass and colleagues (1994) coined the "Computers-Are-Social-Actors" (CASA) paradigm, suggesting that people apply social rules during their interaction with computers, which naturally gains even more relevance in the particular domain of social robots. Nevertheless, current research and developments focus too much on technological borders and possibilities, but disregard social and psychological factors. Though HRI researchers generally acknowledged social robots as an important application domain, including studies on anthropomorphism in social contexts (e.g., Fussel et al., 2008), or specific relations between robot behaviour and human perceptions (e.g., Hoffman et al., 2014; Mok, 2016), an integrated view of these findings is still missing. Relatively little attention is paid to the essential nature of social robots as "social beings among us", and mechanisms of social perception and related phenomena of social psychology in sum. Altogether, it seems that the domain misses a theoretical grounding and framework that fully accounts for the social nature of social robots.

The present paper wants to make a first step towards a better understanding of underlying mechanisms related to the perception and interaction with social robots, and a stronger integration of psychological knowledge into research and design. We follow an interdisciplinary approach, utilizing a combination of theory and methods from HCI and psychology, in order to provide a basis for successful and human-centred robot design. More specifically, we want to stress a dedicated perspective that understands social robots "as a species" and highlights the psychological rules that apply to the interaction between humans and robots. We depict central research questions and

Ullrich D. and Diefenbach S.
Truly Social Robots - Understanding Human-Robot Interaction from the Perspective of Social Psychology.
DOI: 10.5220/0006155900390045
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 39-45
ISBN: 978-989-758-229-5

areas of relevance, summarize first results of own and others' research, and present a research agenda for the domain of social robots, that accounts for their social nature and relevant theory from social psychology.

2 THE INTERMEDIATE POSITION OF SOCIAL ROBOTS

One specific characteristic of social robots is their intermediate position between "usual" human-computer interaction (HCI) and human, interindividual interaction. When interacting with our environment, our behaviour often relies on scripted pattern. However, in the case of social robots, neither general models of human-computer interaction nor models about human interaction seem fully transferrable. While being a piece of technology on the one hand, their anthropomorphic shape, their ability to speak and interaction capabilities suggest robots to be “more” than technology and algorithms, as also our own recent studies showed in impressing ways (e.g., Männlein, 2016; Ullrich, 2017; Weber, 2016).

Thus, a central topic within the understanding of social robots is that of projection and classification. Humans classify objects with which they interact both through bottom-up and top-down processes. The former uses cues of the interaction artefact (it is small, round, has digits 1-12 and watch hands) and the latter prior learned knowledge (I know how a watch looks like, I have seen it before). The particular feature of a social robot is that it hold cues that qualifies for the class “intelligent living being” and sometimes even “human” (to an extent). Being perceived as a member of such class results in specific user expectations and behaviour (such as over-trust) which differs from those of other classes of technology

A lack of understanding of the mechanisms affecting our perception of intelligent technology can result in flawed designs with yet unknown consequences. For example, a less-than-ideal designed autopilot in the automotive context lead to over-optimistic expectations regarding its actual capabilities. The driver developed over-trust in the system and used it in situations that the system could not handle, which ultimately resulted in a fatal accident. In sum, with the current state of research, we seem to be unprepared for this challenge and are not yet exploiting the possibilities of social robots to full potential.

In order to design and act responsibly, we need to seriously acknowledge the social nature of social robots and relations to the general mechanisms of social psychology (e.g., responsibility attributions, judgments and decision making in social contexts). Only a thorough understanding of the social perception and reactions towards robots can enable adequate design decisions and exploit social robots to full potential. Otherwise, unintended, sometimes dramatic consequences, may occur, e.g., accidents due to over-trust in "smart" technology as outlined above. It needs a deeper exploration of the principles that determine what people perceive and how they behave when being confronted with social robots. Naturally, such questions gain even more complexity when we think of settings where more than one robot is involved. To foresee peoples' reactions and perceptions when being confronted with social robots, a thorough understanding of social robots as a "species" and the unfurling human-robot relation is required. To translate such insights into adequate design solutions, it needs insights about the consequences of specific robot properties, and the kind of reactions they afford.

With an interdisciplinary background in psychology and computer science, our vision is to bridge knowledge from the two areas towards a human-centred design of social robots, with an emphasis on the effects of social context. Both fields provide manifold theories and insights of interaction with technology or humans respectively that may be fruitfully combined. For example, in the HCI context, this may be general models of user experience and evaluation of interactive technology (Law and van Schaik, 2010) or approaches to model artificial intelligence (e.g., Cohen and Feigenbaum, 2014). In social psychology, theories about social roles, social identity, group dynamics and attribution mechanisms (Smith et al., 2014) could support the shaping of social robot behaviour and task suitability.

3 CENTRAL RESEARCH QUESTIONS AND FIRST INSIGHTS INTO SOCIAL ROBOT INTERACTION

The present section depicts exemplary research questions and first insights from own and others' studies, underlining the need for a stronger integration of (social) psychological research within the domain of social robots. After that, we extract three general dimensions of interest in social robot interaction, forming a basis for future research and a systematic link between design factors and psychological consequences.

3.1 Personality: What Character Do Humans Appreciate in One Situation or the Other?

One central question within any social situation is the perception of the others' character and consequences for liking and reactions. This of course, also applies to the interaction with social robots. HRI research already showed that robot personality is relevant, that subtle changes in a robot's appearance can lead differences in perceived robot personality, and further effects on social aspects like trust, acceptance or compliance (e.g., Goetz et al., 2003; Kim et al., 2008; Salem et al., 2015; Walters et al., 2008). However, a systematic view on these findings is still missing, leaving unclear whether there is a general kind of robot personality that promotes or diminishes liking and acceptance. There actually are two common, contradicting theories, with little empirical evidence in HRI research for both of them. The first is similarity attraction, i.e., a person chooses and prefers to interact with other people/robots similar to them (e.g., Byrne, 1971; Lee et al., 2006; Tapus et al., 2008). The second is the complementary principle, stating that a person is more attracted to people with personality traits that are contrary to their own (e.g., Leary, 2004; Lee et al., 2006; Sullivan, 2013).

While previous research explored robot personality and effects on liking as an isolated factor, also the task context could be of relevance, and personality and task context may interact with each other. Just like we expect different behaviours/shades of personality (e.g., encouraging, critical) from a friend between different situations, we may also judge different robot personalities as more or less appropriate from one situation to the other. Thus, design recommendations for robot personality may vary depending on the specific area of application. An own study found first evidence for this assumption (Männlein, 2016). We explored effects of three different robot personalities in four different usage scenarios. While in some scenarios, a neutral, conservative personality was preferred, in others, participants wanted a robot with strong character, which could be a positive (nice, friendly) or even a negative (stubborn, grumbling) personality. As a general tendency, differences in robot personality were more relevant in exploration-oriented scenarios (e.g., a social robot as house mate) and less relevant in goal-oriented scenarios (e.g., a social robot selling a train ticket).

3.2 How Much Do We Rely on Robots' Judgments - Compared to Human Judgments?

As already outlined above, trust in robots' judgments and capabilities is a central factor to foresee the reactions towards robots and to design responsibly. For a first exploration of the basis level of trust towards robots (compared to humans) we ran a replication of the famous Asch (1951) paradigm in the context of social robots (Ullrich et al., 2017). Asch explored peoples' reactions to majority opinions on their own perceptions and judgments. The experimental setting poses a simple task: Identifying a line out of three lines that matches a reference line. In the control condition, nearly all participants are able to perform the task correctly and pick the right line. Variations of social context then demonstrate the influence of group opinions on individual judgements. It showed that people begin to mistrust their own perceptions when their social environments comes to other "perceptions" then their own. If surrounded by confederates, instructed to pick a wrong line, people tend to adjust their judgments as well and pick a wrong line then, even if their perception probably tells them otherwise. However, in another experimental condition, already one among the many confederates who picks the right line could induce positive encouragement and a trend towards more correct judgments.

In our replication study, one of the confederates was a social robot, who participated in the experiment as well (see Figure 1). Participants entered their judgments through a computer interface, and were also displayed the (seeming) judgments of all other participants, including the robots'. The general trend of results was that the social impact of the robot on individual judgments was even higher than that of the other participants. Especially the effect of positive encouragement was more pronounced than if a human participant was the only one giving the correct answer. This shows, the generally high level of trust towards social robots, and, as in the present case, how this effect may be used for positive encouragement (e.g., in the field of therapy/rehabilitation. If the robot believes in my skills, I will do the same, and the robot's optimistic judgments may be even more powerful than what the doctor says.) On the other hand, it also hints at the high sensibility and responsibility related to the design of social robots. If there is such a high potential for trust in social robots, it is essential that such trust is used in an adequate way, and to avoid over-trust.

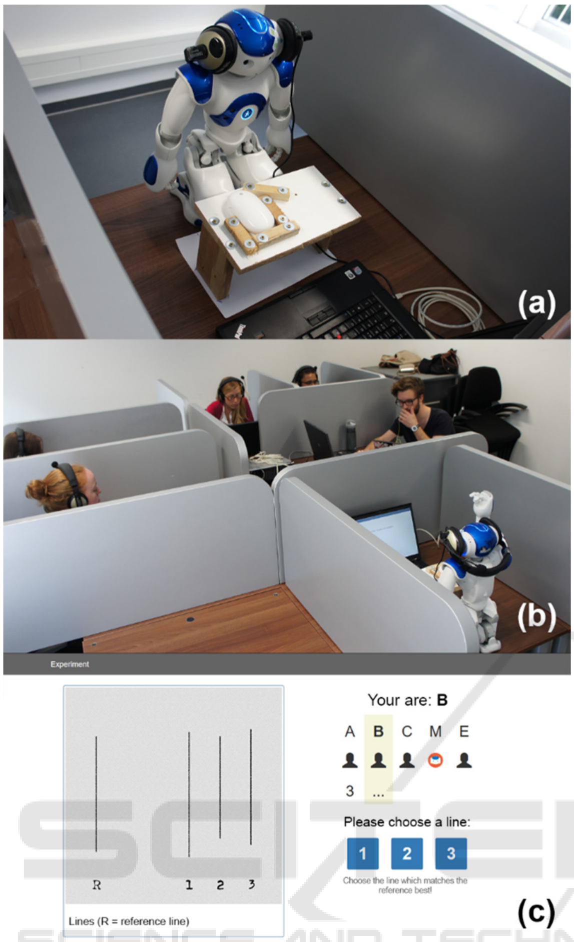

Figure 1: Replication study of the Asch paradigm on conformity and perception with a social robot as participant. (A) shows the robots' cubicle and custom user interface, (b) the experimental situation and (c) the on screen user interface for participants.

3.3 Responsibility: How Much Accountability Do We Assign to Robots – or Ourselves?

Closely related to the issue of trust and distrust is the topic of responsibility, and how much accountability humans assign to robots, compared to other humans. Again, mechanisms from social psychology appear as a helpful start to understand in which situations what level of accountability is assigned. Though trusting robots in general, human's attributions also reflect the concern for self-protection, and making others' accountable for mistakes. This effect has already been demonstrated in the HCI domain in various fields, but gains increasing importance in the domain of social robots, where accountability attributions have severe consequences for the following reactions towards robots as social agents. For example, a study by Moon (2003) in the field of consumer psychology explored responsibility attributions in the context of computer aided purchase decisions. In general, the results reflect a self-serving bias, where consumers tend to blame computers for negative outcomes but take personal credit for positive ones. However, this effect is also moderated by the personal history of self-disclosure ´between human and computer. In a more intimate relationship, consumers are more willing to credit the computer for positive outcomes, and more willing to accept responsibility for negative outcomes. Such effects, of course, are also highly relevant in the domain of social robots that even provides more room for relationship building than just "usual" human-computer interaction.

3.4 In- or Outgroup: What Makes Robots Being One of Us? What Are the General Dimensions of Social Robot Perception?

Finally, central to all the matters about trust, responsibility, and characterization, and the question to what degree mechanisms of social interaction may apply to the domain of social robot interaction, appears the question about what makes robots being one of us, and the general dimensions of social robot perception. As outlined in the introduction section, the interaction with social robots can be positioned somewhere between normal human-human and human-computer interaction. Subtle differences in their design may decide about mechanisms of projection and classification in one or the other direction, and in consequence, the activated psychological processes when entering the interaction. To consider this in design, an important prerequisite is to know the general dimensions along which we classify a robot as social being or not, and which design factors are relevant for the overall perceived human-likeness.

In an experimental study (Weber, 2016) we explored the relative impact on perceived human-likeness for two central factors in social robot design, namely, motion and speech. In our study, the role of the social robot was applied in the sports context, more specifically, being a karate teacher, giving instructions for specific karate moves (see Figure 2). Each factor (motion, speech) was realized in three degrees of differing fidelity by help of a Nao robot, and through systematic combination, the relative impact of these factors was tested. Overall, speech was found to be more relevant than motion for perceived human-likeness, global impression, and general preference. Of course, this finding cannot be generalized yet and further research with a wide range of settings and other design factors and robot-types is necessary. However, it already reveals the importance of dedicated knowledge on the specific effects of single design factors and their relative importance. Such insights allow concentrating design efforts on the most relevant parts and consequences from a psychological perspective.

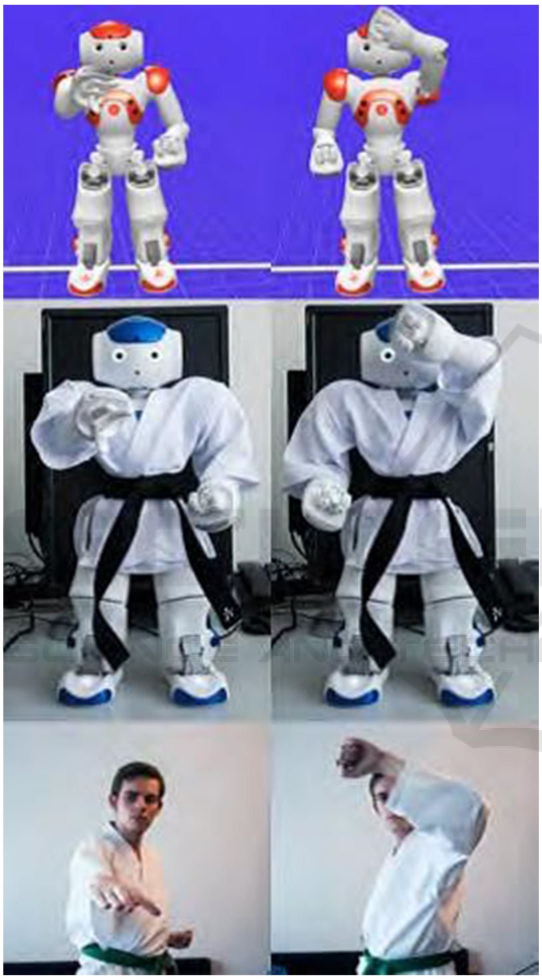

Figure 2: A social robot teaching karate moves.

4 OUTLOOK AND RESEARCH AGENDA

As exemplified by the questions raised in the preceding paragraphs, the overall aim of future research around social robot interaction must be a better understanding of the underlying psychological mechanisms and exploration of its impact on robot properties, design fundamentals, and dynamics in social contexts. More specifically, our research agenda suggests three fundamental directions. First, a thorough exploration of psychological mechanisms and dynamics of social interaction, through a series of experiments, with varying independent (e.g., personality, anthropomorphism) and dependent variables (e.g., trust, human-likeness, perceived will, behaviour correlates of over-/under-trust). In our experiments, we used a NAO robot as representative for the class of social robots. Although our own research as well as others' show that a high fidelity humanoid robot like Sophia (Hanson Robotics, 2016) is not necessarily needed to evoke social effects (e.g., social presence, Hoffman et al, 2015), a broader variation of fidelity within the same experimental settings is preferable to explore the range of effects.

Second, a systematic exploration of the design space and the relevance of single design factors for perceptions, perceived character, trust, and acceptance, with the goal to derive design pattern for an intended robot experience in different scenarios, areas of application, and contextual requirements (e.g., security-related issues).

Third, an exploration of group dynamics in settings with multiple social robots. As already noted above, designing for social robot interaction gains even more complexity in settings where more than one robot is involved. This for example, is already the case in the Japanese Henn na Hotel, where the human staff was almost fully replaced by social robots, who are now running the reception, doing cleaning services etc. (see Figure 3). In order to foresee the emerging dynamics in such settings, knowledge about the special characteristics in multi-robot interaction is crucial. This includes, for example to develop paradigms for multi-robot-collaboration studies, and to explore how findings from studies on single robot-human interaction might change when robots constitute the majority.

Figure 3: Social robots running the reception at the Japanese Henn na hotel (www.h-n-h.jp).

Finally, knowledge from all three research directions must be synthesized in an integrative model on social robots "as a species", providing an overview of relevant mechanisms and variables of social robot interaction, and their interrelations. Such knowledge will then allow design recommendations for specific domains and use cases.

5 CONCLUSION

As exemplified above, entering the domain of social robots, means entering a domain that asks for other, possible even more sensible and complex considerations, than HCI design per se. While social robots form a great potential to enrich our society, profound knowledge about the peculiarities of their species is needed, to bring them into our world with best effect, and support a fruitful collaboration between research and practice. We hope the present considerations may help to outline the importance of this endeavour, and that our studies will provide a basis to create better, trusted, and accepted social robots, in a way that positively contributes to human (robot) society.

ACKNOWLEDGEMENTS

We thank Simon Männlein, Thomas Weber, and Valentin Zieglmeier for their effort of planning and conducting experiments, as their help was crucial for our research.

REFERENCES

Asch, S. E., 1951. Effects of group pressure upon the modification and distortion of judgments. Groups, leadership, and men, 222-236.

Bartneck, C., Kulić, D., Croft, E., Zoghbi, S., 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics,* 1(1), 71-81.

Beasley, R. A., 2012. Medical Robots: Current Systems and Research Directions. Journal of Robotics 2012, 14.

Byrne, D. E., 1971. The attraction paradigm, volume 11. Academic Press.

Cohen, P. R., Feigenbaum, E. A. (Eds.), 2014. *The handbook of artificial intelligence* (Vol. 3). Butterworth-Heinemann.

Fussell, S. R., Kiesler, S., Setlock, L. D., Yew, V., 2008. How people anthropomorphize robots. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction,* 145-152. ACM.

Goetz, J., Kiesler, S., Powers, A., 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003*, 55–60. IEEE.

Hanson Robotics, 2016. Retrieved November 30, 2016 from http://www.hansonrobotics.com/robot/sophia/

Hoffman, G., Birnbaum, G. E., Vanunu, K., Sass, O., Reis, H. T., 2014. Robot responsiveness to human disclosure affects social impression and appeal. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction,* 1-8. ACM.

Hoffman, G., Forlizzi, J., Ayal, S., Steinfeld, A., Antanitis, J., Hochman, G., ... & Finkenaur, J., 2015. Robot presence and human honesty: Experimental evidence. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction,* 181-188. ACM.

Kim, H., Kwak, S. S., Kim, M., 2008. Personality design of sociable robots by control of gesture design factors. In *ROMAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 494–499. IEEE.

Law, E. L. C., van Schaik, P., 2010. Modelling user experience–An agenda for research and practice. *Interacting with computers, 22*(5), 313-322.

Leary, T., 2004. *Interpersonal diagnosis of personality: A functional theory and methodology for personality evaluation.* Wipf and Stock Publishers.

Lee, K.M., Peng, W., Jin, S.-A., Yan, C., 2006. Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human-Robot Interaction. *Journal of Communication, 56*(4), 754–772.

Männlein, S., 2016. *Exploring robot-personalities Design and measurement of robot-personalities for different areas of application.* Master-Thesis, LMU Munich.

Mok, B., 2016. Effects of proactivity and expressivity on collaboration with interactive robotic drawers. In 2016 11th ACM/IEEE International *Conference on Human-Robot Interaction (HRI),* 633-634. IEEE.

Moon, Y., 2003. Don't blame the computer: When self-disclosure moderates the self-serving bias. *Journal of Consumer Psychology, 13*(1), 125-137.

Nass C., Steuer, J., Tauber, E.R., 1994. Computers are social actors. *Computer Human Interaction (CHI) Conference: Celebrating Interdependence 1994*, 72–78.

Paro Robots USA. 2014. PARO Therapeutic Robot. Retrieved October 20, 2016 from http://www.parorobots.com/

Robyn Robotics Ab. 2015. JustoCat. Retrieved October 20, 2016 from http://www.justocat.com/

Salem, M., Lakatos, G., Amirabdollahian, F., Dautenhahn, K., 2015. Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. *Proceedings of the Tenth*

Annual ACM/IEEE International Conference on Human-Robot Interaction, 141–148. ACM.

Smith, E. R., Mackie, D. M., Claypool, H. M., 2014. *Social psychology*. Psychology Press.

Sullivan, H. S., 2013. *The interpersonal theory of psychiatry*. Routledge.

Taipale, S., Luca, F. D., Sarrica, M., Fortunati, L., 2015. *Social Robots from a Human Perspective*. In Springer International Publishing.

Tapus, A., Tapus, C., Mataric, M. J., 2008. User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics, 1*(2), 169–183.

Ullrich, D., 2017, accepted. Robot personality insights. Designing suitable robot personalities for different domains. *I-com Journal of Interactive Media*.

Ullrich, D., Butz, A., Diefenbach, S., 2017, under review. More than just human. The psychology of trust in social robots.

Weber, T., 2016. *Show me your moves, Robot-sensei! The influence of motion and speech on perceived human-likeness of robotic teachers*. Bachelor-Thesis, LMU Munich.

Walters, W. L., Syrdal, D. S., Dautenhahn, K., Boekhorst, R., Koay, K. L., 2008. 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), 159–178.