Why Criteria of Decision Fairness Should be Considered in Robot Design

Sonja K. Ötting
Bielefeld University
Bielefeld, 33602, Germany
soetting@uni-bielefeld.de

Jochen J. Steil
Technische Universität
Braunschweig
Braunschweig, 38106, Germany
jsteil@rob.cs.tu-bs.de

Sugeeth Gopinathan
Bielefeld University
Bielefeld, 33602, Germany
sgopinathan@techfak.uni-bielefeld.de

Günter W. Maier
Bielefeld University
Bielefeld, 33602, Germany
ao-psychologie@uni-bielefeld.de

Abstract

Contemporary robots increasingly become part of work groups and teams and assume new roles: formerly being tools, they now start to interact with their human team members and become social interaction partners. But factors considering social interaction are underrepresented in robot behavioral design, but research in this area is steadily increasing. One of the most important aspects of successful social interaction is perceived fairness, as it improves trust, social functioning and performance of a team. We give an overview on current design factors for social Human-Robot Interaction (HRI) and argue why insights from research on organizational justice contribute to better HRI and team effectiveness. We introduce the concept and importance of justice and apply this to social HRI. Finally, we propose requirements for the design of fair robot behavior and discuss further research questions.

Author Keywords

Fairness; organizational justice; robot design; sociality, Human-Robot Interaction (HRI)

ACM Classification Keywords

J.4 SOCIAL AND BEHAVIORAL SCIENCES: Psychology

Figure 1: TEGA, http://robotic.media.mit.edu/portfolio/robot-vocal-expressivity/

Figure 2: Probo, http://probo.vub.ac.be/Probo/RAT.htm

Figure 3: Flobi, https://aiweb.techfak.uni-bielefeld.de/flobi-head

Introduction

To ensure effectiveness in human-robot teams, robot design has to consider the robot's individual behavior and appearance as well as its social interactions with human team members. Robotics research often considers social science insights in robot design to meet this demand. One important factor for successful social interaction in teams, recognized in both robotics and social sciences, is trust [15, 17]. However, empirical evidence on robot factors that build trust is rare. To narrow this research gap, we introduce fairness as a fundamental basis for trust and team effectiveness [4]. Fairness is perceived when decisions comply with the principles of organizational justice. These principles are well known in work and organizational psychology, but as of yet, they are not or only fragmentarily included in robot behavioral design.

Therefore, we analyze and discuss current guidelines for social Human-Robot Interaction (HRI), including anthropometric and psychological design factors, and argue why organizational justice research contributes to better social HRI and team effectiveness. Subsequently, organizational justice is introduced and applied to social HRI. Finally, we propose requirements for fair robot design and discuss further research.

Current Design Factors for Social HRI

This section focusses on recent developments that enable robots to socially interact with humans and become part of a work group or team.

Fong and colleagues [11] define social robots as "embodied agents that are part of a heterogeneous group: a society of robots or humans. They are able to recognize each other and engage in social interactions". Therefore, robots that are part of a team (a society of humans and robots) should be able to engage in social interactions.

Recent developments in social robotics acknowledge anthropomorphic aspects in robot behavioral design. These aspects may refer to robots' shapes, behaviors and interaction with humans [7]. Hence, anthropomorphism is not limited to the physical appearance of the robot. In order to create socially acceptable robots, it is necessary to develop familiarity by anthropomorphic design and social characteristics pertaining to humans [10]. Goetz and colleagues [13] establish this relation between the social cues in robots and general acceptance: The users in this study tend to prefer robots with anthropomorphic characteristics that match the sociability required in the given scenarios.

One of the recent additions to the world of social robots is Tega from MIT Media Lab. This robot is designed to interact with children and establish long-term interactions with them [19]. Probo [24] is another robot that incorporates an appearance that is appealing to children. It is a huggable animal-like robot, designed to act as a social interaction partner. The robot head "Flobi" is a novel anthropomorphic robot head that combines state-of-the-art sensing functionality with an exterior that elicits a sympathetic emotional response. It can display emotions in a human-like way, to enable intuitive human-robot-interaction and is capable of displaying primary emotions as well as secondary emotions like shame [16]. EDGAR, Expression Display & Gesturing Avatar Robot by Nanyang Technological University, is another example of human like adaptation in social robotics [1]. The head of the robot has a rear projection system that is capable to display facial

Figure 4: EDGAR, http://www.mae.ntu.edu.sg/NewsnEvents/Pages/Detailed-Page.aspx?news=c417c561-4c48-4c77-8140-367a683ef93b

Figure 5: RoboThespian, https://www.engineeredarts.co.uk/robothespian/

Figure 6: PEPPER, https://www.ald.softbankrobotics.com/en/cool-robots/pepper

features and expressions on the robot face. This makes it suitable for social interactions as demonstrated with SociBot Head and RoboThespian from Engineered Arts [12]. Another advanced social robot is Pepper [9] from Aldebaran Robotics. It is equipped with an emotion engine that can perceive the user's emotions, by analyzing their voice and facial expressions, resulting in the robot trying to adapt its attitude to suit the user.

As we can see, roboticists make an effort to design human-like and socially acceptable robots by choosing the robot design to suit the task or targeted user group. Other studies suggest additional psychological factors for social HRI through human-likeness, including autonomy, imitation, intrinsic moral value, moral accountability, privacy and reciprocity [18].

Most of these factors tap into the human's natural understanding of human-like appearance and behavior, with the goal to build trust, facilitate social interaction and enhance effectiveness of the human-robot team. The importance of trust is clearly established in HRI research [14]: Meta-analytic results show that robot factors (in comparison to human or environmental factors) are most important for the enhancement of trust [15]. Nevertheless, empirical evidence regarding robot factors that build trust and enhance team performance is sparse.

To narrow research gap, we use insights from work and organizational psychology: A highly important basis for social interactions and effectiveness in teams through trust is fairness [4]. There has been research on fairness in HRI, but this is mostly concerning distributive fairness (e.g. how tasks are distributed) [8]. However, justice consists, besides the outcome of distribution decisions, of the procedure of decision-making and their communication. Particularly the latter are highly related to trust and team effectiveness [3].

In the following, we describe the concept of organizational justice, its importance in work groups and teams and its application to human-robot teams.

Fairness in Social Interactions

Research in work and organizational psychology shows that the perception of fairness depends on the principles of organizational justice (for a current overview: [6]). These principles consist of four dimensions: distributive, procedural, interpersonal and informational justice. Distributive justice focuses on the outcome of decisions (e.g. the distribution of money or tasks). An outcome is perceived as fair if the ratio of one's contributions to one's outcomes is equivalent to that of a comparison other (e.g. a team member). Procedural justice refers to the decision-making procedures. A procedure is perceived as fair if the recipients have voice during the procedure and if procedural justice criteria (consistency, lack of bias, accuracy, correctability, and ethicality) are given in the process. Interpersonal and informational justice concern the communication of decisions: The first describes the justice of the decision authority's behavior (i.e. the robot's behavior) towards the recipient. This behavior is perceived as fair, if it is respectful, polite and dignified. The latter refers to the adequacy of the information used to explain the decision. It is perceived as fair if information is truthful, well-reasoned, specific and timely.

Organizational justice is highly important for effective organizational behavior at the individual and the team

level [4]. It is related to a number of attitudes (e.g. supervisor satisfaction, outcome satisfaction, job satisfaction, organizational commitment, trust) and behaviors (e.g. counterproductive work behavior such as withdrawal, work performance, organizational citizenship behaviors) [2, 5] as well as outcomes at the team level (e.g. team efficacy, group cohesion, attachment to the team) [3, 4].

Robots as decision authorities are seldom considered in organizational justice research. But several findings justify the assumption that the principles of organizational justice are just as applicable to human-robot interaction as to human-human interaction: The "Computers As Social Actors" theory [22] validates the assumption that computers and robots are perceived as social interaction partners. Studies on computers and robots as decision authorities show that humans do perceive fairness or unfairness in reaction to computers [26] and that justice is an important factor for team members' attitudes and behaviors no matter who is making the decision (human, robot or computer) [23].

Requirements for Fair HRI

Several researchers included the principles of distributive justice in their robots' design. Sometimes fairness is included explicitly in the design of algorithms for fair allocation of collaborative tasks (e.g. [8]), sometimes rules are included that adhere to justice criteria but are named differently, e.g. equality [25].

Töniges and colleagues [28] have proposed additional requirements for fair cyber-physical systems concerning the other justice dimensions. These requirements can be applied to robotics as well.

Procedural justice requires that the robot is able to present a means for the user to raise objections and make suggestions for improvements or corrections. Additionally, procedural justice requires that the user is well informed and convinced that procedures are applied consistently, are free of bias, are based on accurately collected information, and conform to prevailing standards of ethicality. These requirements have not yet been considered or discussed in robot design. One way to go would be as discussed in [29], where advice giving by robots is considered and it is proposed to resort to natural language i.e. by speech.

Interpersonal justice requires a polite and respectful phrasing of the robot's communication. This justice dimension is sometimes included as politeness in robot behavior (e.g. [27]), but has, to our knowledge, not been considered outside the service context.

Informational justice requires the robot to communicate the decision, and the reasons behind it, in a timely, specific and well-reasoned manner. This means that it will not suffice to provide the sheer action or decision; the path the decision-making process takes has to be transparent. This requirement has not yet been fulfilled in robot design. For most robots, it is new to communicate that a decision took place [25] and even less common to communicate reasons behind it. Research in Knowledge Representation and Reasoning [30] makes first steps in that direction. However, this demands a vast knowledge base about tasks, containing symbolic representation [21] and includes the possibility to add and learn new scenarios.

Further Research on Fair HRI

The presented analysis shows that more and interdisciplinary research is needed on how to implement the principles of organizational justice in an optimal way for the given collaborative task or scenario. In addition to that, robotics and social science researchers need to further investigate the specific consequences of implementing fairness principles into robot design. The relationship between justice, trust and team effectiveness should also be investigated because it is not straightforward. Research on trust and reliance suggests that too much trust might lead to overreliance on automated systems, which ultimately might cause failures or injuries [20].

Conclusion

As robots are increasingly part of work groups and teams they should be able to successfully engage in social interactions. The literature overview shows that trust is an important factor for successful human-robot teams and that different anthropomorphic as well as psychological factors have started to appear in robot design. It also shows that additional research is needed to establish the assumed relationships. Hence, we infer that fairness should be included in robot design, as it is an important basis to build trust and foster team effectiveness, contributing to human-likeliness.

References

1. Pang W. Ching, Wong C. Yue, and Gerald S. G. Lee. 2016. Design and development of EDGAR. A telepresence humanoid for robot-mediated communication and social applications. In *Proceedings of 2016 IEEE International Conference on Control and Robotics Engineering (ICCRE)*, 90–93. DOI=10.1109/ICCRE.2016.7476147.
2. Yochi Cohen-Charash and Paul E. Spector. 2001. The role of justice in organizations. A meta-analysis. *Organizational Behavior and Human Decision Processes* 86, 278–321.
3. Jason A. Colquitt, Raymond A. Noe, and Christine L. Jackson. 2002. Justice in teams. Antecedents and consequences of procedural justice climate. *Personnel Psychology* 55, 83–109.
4. Jason A. Colquitt, Cindy P. Zapata-Phelan, and Quinetta M. Roberson. 2005. Justice in teams. A review of fairness effects in collective contexts. In *Research in personnel and human resources management*, Joseph J. Martocchio, Ed. Emerald Group Publishing Limited, Bingley, 53–94. DOI=10.1016/S0742-7301(05)24002-1.
5. Jason A. Colquitt, Brent A. Scott, Jessica B. Rodell, David M. Long, Cindy P. Zapata, Donald E. Conlon, and Michael J. Wesson. 2013. Justice at the millennium, a decade later. A meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology* 98, 199–236.
6. Jason A. Colquitt and Kate P. Zipay. 2015. Justice, fairness, and employee reactions. *Annual Review of Organizational Psychology and Organizational Behavior* 2, 75–99.
7. Carl F. DiSalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. 2002. All robots are not created equal. The design and perception of humanoid robot heads. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 321–326. DOI=10.1145/778712.778756.
8. Myriam El Mesbahi, Nabil Elmarzouqi, and Jean-Christophe Lapayre. 2014. Fairness properties for collaborative work using Human-Computer Interactions and Human-Robot Interactions based environment: “Let us be fair”. In *Distributed, ambient, and pervasive interactions*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Alfred Kobsa, Friedemann Mattern, …,

Eds. Springer International Publishing, Cham, 319–328. DOI=10.1007/978-3-319-07788-8_30.

9. *Find out more about Pepper*. https://www.ald.softbankrobotics.com/en/cool-robots/pepper/find-out-more-about-pepper.
10. Julia Fink. 2012. Anthropomorphism and human likeness in the design of robots and Human-Robot Interaction. In *Social robotics*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, …, Eds. Springer, Berlin, Heidelberg, 199–208. DOI=10.1007/978-3-642-34103-8_20.
11. Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42, 143–166.
12. Gheorghe Gîlcă, Nicu G. Bîzdoacă, and I. Diaconu. 2016. Using the rear projection of the Socibot Desktop robot for creation of applications with facial expressions. In *7th International Conference on Advanced Concepts in Mechanical Engineering. DOI=10.1088/1757-899X/147/1/012067*.
13. Jennifer Goetz, Sara Kiesler, and Aaron Powers. 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Proceedings. ROMAN 2003. The 12th IEEE International Workshop on Robot and Human Interactive Communication*, 55–60. DOI=10.1109/ROMAN.2003.1251796.
14. Victoria Groom and Clifford Nass. 2007. Can robots be teammates? Benchmarks in human–robot teams. *Interaction Studies* 8, 483–500.
15. Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* 53, 517–527.
16. Frank Hegel, Friederike Eyssel, & Britta Wrede. 2010. The social robot 'flobi': Key concepts of industrial design. In *19th International Symposium in Robot and Human Interactive Communication*, 107-112.
17. Bart A. de Jong, Kurt T. Dirks, and Nicole Gillespie. 2016. Trust and team performance. A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology* 101, 1134–1150.
18. Peter H. Kahn, Jr., Hiroshi Ishiguro, Batya Friedman, Takayuki Kanda, Nathan G. Freier, Rachel L. Severson, and Jessica Miller. 2007. What is a human? Toward psychological benchmarks in the field of human-robot interaction. *Interaction Studies* 8, 363–390.
19. Jacqueline Kory Westlund, Jin J. Lee, Luke Plummer, Fardad Faridi, Jesse Gray, Matt Berlin, Harald Quintus-Bosz, Robert Hartmann, Mike Hess, Stacy Dyer, Kristopher dos Santos, Sigurður Örn Aðalgeirsson, Goren Gordon, Samuel Spaulding, Marayna Martinez, Madhurima Das, Maryam Archie, Sooyeon Jeong, and Cynthia Breazeal. 2016. Tega. A social robot. In *HRI '16. Proceedings of the 11th ACM/IEEE International Conference on Human Robot Interaction*, 561.
20. John D. Lee and Katrina A. See. 2004. Trust in Automation. Designing for Appropriate Reliance. *Human Factors* 46, 50–80.
21. Hector J. Levesque. 1986. Knowledge representation and reasoning. *Annual Review of Computer Science* 1, 255–287.
22. Clifford Nass, Jonathan Steuer, Ellen Tauber, and Heidi Reeder. 1993. Anthropomorphism, agency, and ethopoeia. Computers as social actors. In *INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems*, 111–112. DOI=10.1145/259964.260137.
23. Ötting, S. K. and Maier, G. W. in prep. *The importance of procedural justice in Human-Machine-Interaction. Cyber-Physical Systems as new decision authorities in organizations*.
24. Jelle Saldien, Kristof Goris, Bram Vanderborght, Johan Vanderfaeillie, & Dirk Lefeber. 2010. Expressing emotions with the social robot Probo. *International Journal of Social Robotics 2*, 377-389. DOI=10.1007/s12369-010-0067-6.

25. Julie Shah, James Wiken, Brian Williams, and Cynthia Breazeal. 2011. Improved human-robot team performance using Chaski, a human-inspired plan execution system. In *HRI '11. Proceedings of the 6th ACM/IEEE International Conference on Human-robot Interaction*, 29–36. DOI=10.1145/1957656.1957668.
26. Daniel B. Shank. 2012. Perceived justice and reactions to coercive computers. *Sociological Forum* 27, 372–391.
27. Vasant Srinivasan and Leila Takayama. 2016. Help me please. Robot politeness strategies for soliciting help from people. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4945–4955. DOI=10.1145/2858036.2858217.
28. Torben Töniges, Sonja K. Ötting, Britta Wrede, Günter W. Maier, and Gerhard Sagerer. 2017. An emerging decision authority. Adaptive cyber-physical system design for fair human-machine interaction and decision processes. In *Cyber-physical systems. Foundations, principles and applications*, Houbing Song, Danda B. Rawat, Sabina Jeschke and Christian Brecher, Eds. Elsevier Academic Press, Amsterdam, 419–430.
29. Cristen Torrey, Susan Fussell, and Sara Kiesler. 2013. How a robot should give advice. In *HRI '13. Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, 275–282.
30. Shiqi Zhang, Mohan Sridharan, Michael Gelfond, and Jeremy Wyatt. 2014. KR³. An architecture for knowledge representation and reasoning in robotics. In *15th International Workshop on Non-Monotonic Reasoning. NMR 2014*. INFSYS Research Report, 233–241.