Robo-Tooltips: Understandable Robots for Trustworthy Interactions

Svenja Y. Schött
svenja.schoett@ifi.lmu.de
LMU Munich
Munich, Germany

Andreas Butz
butz@ifi.lmu.de
LMU Munich
Munich, Germany

ABSTRACT

Trust is a factor that can positively influence human-robot interactions. Transparent robot design and appropriately timed explanations have the potential to help users calibrate their trust in a robotic system. We introduce the idea of robo-tooltips, physical markers that signal to users that additional information about the robot is available, which they can access at their own discretion. We showcase the concept, which is similar to the tooltips used in desktop environments, by attaching QR codes to a robot. When the user interacts with the QR code, descriptions on the capabilities and behavior of the respective part of the robot are displayed. We present scenarios during which the user retrieves information using robo-tooltips. Further, we discuss some opportunities and challenges of offering optional explanations during human-robot interaction and outline directions for future work.

KEYWORDS

human-robot interaction, human-robot trust, transparency

1 INTRODUCTION

Personal and social robots are emerging in different contexts from healthcare [6] to the home [4]. Therefore, humans are increasingly confronted with robots in everyday situations. In essence, robots are computing systems with an additional embodiment [1]. While the area of human-computer interaction (HCI) aims at facilitating interactions with computers in general, human-robot interaction (HRI) also aims to understand and evaluate social interactions between humans and robotic systems [1]. The question of how to establish appropriate trust in robots is one of the central research areas of HRI [8], as users may hesitate to use a system they do not trust, and it thus becomes underutilized [14]. There is a basic trust problem between humans and robots, because robots are perceived as a different species [16]. Therefore, human-robot trust is influenced by factors that play a role in automation, as well as factors that impact human-human interactions. Consequently, human-robot trust is multi-dimensional, influenced by performance and moral aspects [12]. In their 2011 meta-analysis, Hancock et al. established that robot-, human- and environment-related factors influence human-robot trust [7]. They found that robot-related factors, especially robot performance, have the biggest impact on trust [7].

If possible, trust is rooted in an accurate understanding of a systems capabilities [12]. How much trust users put in a system is intertwined with the systems transparency and consequently the users' understanding of its performance [14]. Previous research has established that robot transparency affects trust in error scenarios [13]. Thus, enabling users to understand robots is especially important when the performance varies. For algorithmic interfaces, the relationship function between transparency and trust is bell-shaped [10]. This means that low and high understanding lead to decreased trust.

Making systems understandable and explaining the behavior of intelligent systems is the main focus of the research field of Explainable Artificial Intelligence (XAI) [11]. XAI aims to explain the inner reasoning behind black box machine learning [5]. Taking a step beyond explaining algorithms, there is potential to build on XAI methods to not only make decision-making, but also robot behavior and physical capabilities understandable.

In traditional graphical user interfaces (GUIs), additional information and explanations are offered using tooltips. Previous work suggests that in the real world, annotating objects using, for example, projection could make them de-facto self-explaining [2]. Transferring this idea to robots has the potential to increase user understanding in HRI. In this paper, we present a concept for providing such optional explanations integrated in the robot design: the robo-tooltips.

2 THE CONCEPT OF ROBO-TOOLTIPS

We define robo-tooltips as physical markers which inform human users that further explanations about the robot are available. They provide this information upon interaction. We focus on optional explanations, as this gives users themselves the power to access information about the robot and thus increase its transparency. Correctly sensing when exactly a user needs additional information is another challenge in itself, which we do not attempt to tackle. However, we argue that robo-tooltips, by allowing users to retrieve this information at their own discretion, represent a promising middle ground between not providing additional information and always providing information. Giving the human user access to knowledge about the robot creates interesting opportunities to get to know robots as conversation partners. Ultimately, by increasing robot transparency, the goal is to enable humans to calibrate their trust appropriately.

Figure 1 shows an example of tooltips added to a robot. To showcase the concept, we use QR codes attached to different parts of the robot. Upon scanning these codes with a smartphone, it displays a

MuC'22, 04.-07. September 2022, Darmstadt

https://doi.org/10.18420/muc2022-mci-ws13-452

Figure 1: Physical tooltips in the form of QR codes on a Kinova robot arm.

website with explanations. By utilizing phones as an assistive tool in making robots self-explaining, other modalities like audio output are possible. QR codes (Quick Response codes) are 2D matrix codes which store information, commonly used to encode URLs [15]. We selected QR codes for our showcase as they by nature compress larger amounts of information in a small space, which is ideal for adding them to other systems. Further, due to the prevalence of QR codes e.g. when accessing menus at restaurants or signing up for a corona test, even non-expert users have become aware of their functionality during the last few years. Users know that by scanning QR codes with their phone camera they receive additional information. Thus, QR codes seemed like an appropriate means to indicate to user that they can acquire further information. Due to the ubiquity of smartphones, anyone interacting with a robot can thus have access to this further information, making the robot in essence self-explaining. Of course, this is only a technical solution which allowed us to build a working prototype and explore our concept with little effort and today's technology. In the future, such explanations might be displayed in the field of view of AR glasses, retina implants, or whatever other technologies might emerge. The challenge will then be to appropriately and unobtrusively signify the fact that additional information is available.

3 USAGE SCENARIOS FOR ROBO-TOOLTIPS

To illustrate the use of robo-tooltips, we describe three simple interaction scenarios between a human and a robot.

3.1 System Failure

Robot transparency is especially important when it comes to system failure [13]. As robots are not 100% reliable, users need to be able to assess why the robot made a mistake and/or how to proceed. Imagine a kitchen-based robot that has the task of washing a glass, yet, after moving towards the glass, fails to grab it due to a gripper malfunction. A confused user could now retrieve information on the gripper specifically by scanning its robo-tooltip. The website linked to the QR code displays base information on the gripper functionality, as well as a situation-specific error message. Using the recommendations from the tooltip, the user is empowered to resolve the situation and help the robot to recover.

3.2 Usage Refusal

In an example described by Theodorou et al., a health-care robot used to assist the elderly remains unused, as the users do not trust the robot and consequently refuse interactions [14]. They point out the potential for transparency via robot design in such a scenario. We envision that the user who distrusts a system because they do not understand it, can scan the robo-tooltip and read through (or listen to) the information provided. In this case, the understanding that the robot has the capabilities to perform a helpful medical treatment is most relevant. Due to the bell-shaped function between understanding and trust, the trust would become higher after acquiring more information. If all goes well, the patient then agrees to the medical treatment by the hand of the robot, because of the increase in trust. Transparency, in this case, thus has the potential to increase effectiveness of such a health-care robot.

3.3 Social Games

Not all facets of human-robot trust focus on robot performance. During social interaction scenarios, humans may want to know about the moral standing of their interaction partner. Imagine a human and a social robot playing a competitive game [9]. Would the robot use every trick in the book to win at all cost? Is the robot always honest or can it lie to win the game? Can the robot make hollow promises to the human to get itself an advantage in the game? Our human wants to find out before making their next move. They scan the robo-tooltip and find the following information: "The robot has been programmed to imitate human strategies to win this game. As humans can lie, this robot also has the capability to lie to their opponent to win the game. However, the robot never lies unrelated to achieving the aim of winning." With this information and deepened understanding, the human player does not blindly trust the robot in the game context, but would still keep their trust in non-game situations, thus developing a much more differentiated kind of trust.

4 CONTENT DESIGN FOR ROBO-TOOLTIPS

Robo-tooltips are easy to add to new and existing robots. As portrayed in the scenarios, robo-tooltips can be used in different settings and interaction-scenarios to offer missing information when a user requests it. However, a big design challenge is the decision of *what* to explain. Theodorou et al., who argue that robots should be implemented in a transparent way and as such offer information to the user, suggest to communicate "...robot's capabilities, goals, and current progress in relation to its goals, its sensory inputs, and its reliability, as well as reports of any unexpected events."[14] In our scenarios, the robo-tooltips offer information on 1) unexpected events, 2) robot capabilities and 3) robot goals.

In analogy to linguistics, we can differentiate between nouns (subject or object), verbs (actions) and adjectives (robot properties) as components of the explanation. The example illustrated in Figure 2 offers explanations on a robotic limb and highlights the different explanation features.

Figure 2: An explanation of a robotic arm. The subject of the explanation is highlighted in red, the actions in blue and the properties in green.

Another challenge is that different people may require a different level of knowledge to successfully interact with a robot. While novice users may be satisfied with the subject explanation, developers may require additional knowledge necessary for debugging. Eiband et al. formulated a framework of different researcher perspectives on understanding in the context of intelligent systems [5]. This framework is not directed at robots specifically. However, we believe that similar system qualities are at play in understanding between humans and robots. Thus, this framework can act as a base for further exploration of explanation content in HRI. While we only took some first steps towards assessing what information humans require to understand a system, we argue that this challenge needs to be addressed further to ensure that the information provided is appropriate for a diverse set of users and contexts.

5 OUTLOOK

In this workshop, we present the vision and open challenges of offering optional explanations during human-robot interaction. Robo-tooltips are a simple, low-cost way to make robot capabilities and behavior more transparent. We believe that giving users the option to access information about the robot they are interacting with facilitates understanding. From a technical side, our showcase is constrained by the need for mobile phones, as these are used to access the information hidden behind the QR codes. However, the idea transcends technical constraints and could be implemented using more advanced technical means. Both steps, a) making the user aware of optional explanations and b) providing the explanation, will then have to be solved depending on the modality chosen.

Using technical tools available today, mobile AR applications seem promising to depict visual explanations in place. Similar to the projector-based annotations of real-world objects by Butz and Schmitz [2], the explanations on the robot would appear in physical proximity to the explained robot component, causing the explanation to seem less disjointed from the robot. In the future, we can envision going beyond the visual markers and using other modalities. For example, HRI designers could use mid-air haptics [3] in close-proximity interaction scenarios to act as tooltips. Here, a user would hover their hand above the robotic component they require more information about, and audio speakers integrated in the robot would provide a verbal explanation.

In the far future, we envision that intelligent robot systems are able to offer optional explanations dependent on user need without intermediate user action. With advanced sensing methods, a robot would recognize not only the basic affective state of users but assess user understanding. Step a) of making the user aware of the available additional information, which they then need to retrieve (e.g. by scanning a QR code) would be obsolete. For now, however, the control of whether and to what degree they require explanations remains in the hands of the user.

From a more general perspective, further research is needed on the impact of more accurate mental models for trust and user experience in the context of emotional interactions with robots.

We believe that robo-tooltips have the potential to improve HRI by increasing transparency. Improved human-robot understanding builds the basis for human-robot connectedness and appropriate trust. Eventually, with a heightened understanding of the robot capabilities, dangerous situations that are the result of dis- and overtrust can be mitigated.

6 ACKNOWLEDGEMENTS

Svenja Schött is part of the One Munich Strategy Forum at LMU Munich, funded under the Hightech Agenda Bavaria. This research was executed in collaboration with the project PERFORM (BU 1402/7-1).

REFERENCES

[1] Christoph Bartneck, Tony Belpaeme, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Šabanović. 2020. *Human-Robot Interaction: An Introduction.* Cambridge University Press. https://doi.org/10.1017/9781108676649

[2] Andreas Butz and Christian Schmitz. 2005. Annotating real world objects using a steerable projector-camera unit. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops.* IEEE, 109–109. https://doi.org/10.1109/CVPR.2005.412

[3] Tom Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian. 2013. UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th annual ACM symposium on User interface software and technology.* 505–514. https://doi.org/10.1145/2501988.2502018

[4] Boris de Ruyter, Privender Saini, Panos Markopoulos, and Albert van Breemen. 2005. Assessing the effects of building social intelligence in a robotic interface

for the home. *Interacting with Computers* 17, 5 (Sept. 2005), 522–541. https://doi.org/10.1016/j.intcom.2005.03.003

[5] Malin Eiband, Daniel Buschek, and Heinrich Hussmann. 2021. How to support users in understanding intelligent systems? Structuring the discussion. In *26th International Conference on Intelligent User Interfaces*. 120–132.

[6] Connor Esterwood and Lionel P. Robert. 2020. Personality in Healthcare Human Robot Interaction (H-HRI): A Literature Review and Brief Critique. In *Proceedings of the 8th International Conference on Human-Agent Interaction*. ACM, Virtual Event USA, 87–95. https://doi.org/10.1145/3406499.3415075

[7] 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: The Journal of the Human Factors and Ergonomics Society* 53, 5 (Oct. 2011), 517–527. https://doi.org/10.1177/0018720811417254

[8] Chan Hsu, Ching-Chih Tsao, Yu-Liang Weng, Cheng-Yi Tang, Yu-Wen Chang, Yihuang Kang, and Shih-Yi Chien. 2022. A Machine Learning Approach to Model HRI Research Trends in 2010~2021. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (Sapporo, Hokkaido, Japan) *(HRI '22)*. IEEE Press, 4. https://doi.org/10.5555/3523760.3523882

[9] Markus Häring, Dieta Kuchenbrandt, and Elisabeth André. 2014. Would you like to play with me?: how robots' group membership and task features influence human-robot interaction. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, Bielefeld Germany, 9–16. https://doi.org/10.1145/2559636.2559673

[10] René F. Kizilcec. 2016. How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2390–2395. https://doi.org/10.1145/2858036.2858402

[11] Q. Vera Liao, Moninder Singh, Yunfeng Zhang, and Rachel Bellamy. 2021. Introduction to Explainable AI. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 127, 3 pages. https://doi.org/10.1145/3411763.3445016

[12] Bertram F. Malle and Daniel Ullman. 2021. A multidimensional conception and measure of human-robot trust. In *Trust in Human-Robot Interaction*. Elsevier, 3–25. https://doi.org/10.1016/B978-0-12-819472-0.00001-0

[13] Birthe Nesset, David A. Robb, José Lopes, and Helen Hastie. 2021. Transparency in HRI: Trust and Decision Making in the Face of Robot Errors. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Boulder CO USA, 313–317. https://doi.org/10.1145/3434074.3447183

[14] Andreas Theodorou, Robert H. Wortham, and Joanna J. Bryson. 2017. Designing and implementing transparency for real time inspection of autonomous robots. *Connection Science* 29, 3 (July 2017), 230–241. https://doi.org/10.1080/09540091.2017.1310182

[15] Sumit Tiwari. 2016. An Introduction to QR Code Technology. 39–44. https://doi.org/10.1109/ICIT.2016.021

[16] Daniel Ullrich, Andreas Butz, and Sarah Diefenbach. 2020. The Eternal Robot: Anchoring Effects in Humans' Mental Models of Robots and Their Self. *Frontiers in Robotics and AI* 7 (Dec. 2020), 546724. https://doi.org/10.3389/frobt.2020.546724