Building the Mental Model with Trust in Human-Robot Collaboration

Daniel Majonica[1,2,*], Nardie Fanchamps[1], Deniz Iren[1] and Roland Klemke[1,2]

[1]Open Universiteit, Heerlen, Netherlands
[2]Cologne Game Lab, Technische Hochschule Köln, Cologne, Germany

Abstract

In this study, we explore human-robot interaction (HRI) with the focus on collaboration and trust. We use augmented reality (AR) technology such as the Hololens 2 to run an experiment with participants. Understanding the dynamics of human-robot collaboration becomes essential in a future with increasingly more robots in more diverse environments. We investigate an AR enhanced learning experience where participants were tasked with repairing robot components and navigating a game-like environment. Participants engaged with a robot in tasks that required joint decision-making. The study aimed to understand how trust is established and maintained in interactions with collaborative robots. Preliminary findings indicate significant individual variability in trust levels towards the robot, with perceptions ranging from complete trust to skepticism. The study also highlights a variance in participant behaviors, with some following the robot's guidance consistently, while others occasionally resisted it. Future work will further investigate these findings with larger participant groups and explore the effects of different robot designs on trust levels.

Keywords

human-robot interaction, immersive learning environment, mental model, robot readiness

1. Introduction

We transition step by step into a world where humans and robots collaborate and coexist in the same physical space. Both will need to learn how to communicate for successful collaboration and cooperation. To facilitate this interaction, it is essential for humans to get a realistic understanding of their robot partner's capabilities and limitations and to establish a certain degree of trust.

Robots are increasingly present in a variety of settings beyond the industrial sector. They are utilized for many purposes such as elderly care [1], as pills [2] in the medical domain, or delivery robots on the streets [3]. Our focus, however, is on robots in educational settings, where they can function as teachers [4] or learning companions [5]. Such educational robots can enhance the learning process by providing valuable input and feedback to users. Interaction with educational robots should be bidirectional, involving multiple layers of interaction and feedback between the learner and the robot.

Achieving effective learning from robots requires overcoming a threshold of acceptance, also called "robot readiness" [6, 7]. This includes understanding the mental model of the robot, the task at hand, and self-perception [8], necessitating dynamic interaction.

Immersive technologies such as augmented reality (AR) and virtual reality (VR) offer interesting concepts for robot learning. These technologies create virtual environments that ensure safety while modeling complex tasks. Immersive learning environments (ILEs) support rapid development and broad accessibility, making them suitable for enhancing human-robot interac-

2nd International Workshop of the EATEL SIG EduRobotX 2024: Exploring Educational Robots and Robotics for Learning and Development, ECTEL Conference, in Krems, Austria | 16-20 September 2024

[*]Corresponding author.

dm@colognegamelab.de (D. Majonica)
0000-0003-4792-0472 (D. Majonica); 0000-0001-7509-2251 (N. Fanchamps); 0000-0002-0727-3445 (D. Iren); 0000-0002-9268-3229 (R. Klemke)

tion (HRI) learning. This study investigates the impact of ILEs on HRI and their potential to improve the overall learning experience.

Trust vital for human-robot collaboration. Trust increases when trustees are successfully performing a task and when expectations are met [9]. Trust can also be measured at different times. In this study, we look at post-interaction trust [10]. This means, that after the interaction, participants assess their trust level towards robots. This, combined with their recorded decision during the interaction, can be analyzed to understand more on how big the spectrum of trust in human-robot collaboration is.

2. Methodology

Participants were asked to sit down in front of a LEGO Mindstorms EV3 robot and put on an AR device, in this case, the Hololens 2. On the table, between the participant and the robot was a set of special cards (Figure 1). All the instructions were then proceeded to be given through the AR device.

This study was so far conducted through two different sessions with the same setup but different participants. The participants in both sessions were from a diverse group aged 18 and above, from various academic fields related to technology-enhanced learning. The study was conducted with eight individuals. The participants joined voluntarily and were selected at random. They were also not paid or reimbursed through other means. The data was collected using an anonymized post-study survey with closed and open-ended questions. Additionally, each action with the AR device was recorded so the path taken could be reconstructed. Each session was structured to allow individual participants to interact privately with a robotic counterpart. The questionnaire included a system usability scale which was analyzed in a previous study [11] as well as three specific open-ended questions. The relevant three questions asked in the questionnaire were:

1. What perceived types of ILE components presented in this study (e.g. video pop-up) affected (both positively and negatively) the HRI?
2. To what extent do you trust the robot in terms of collaboration?
3. How confident do you feel about the decisions of the robot and about your own decisions?

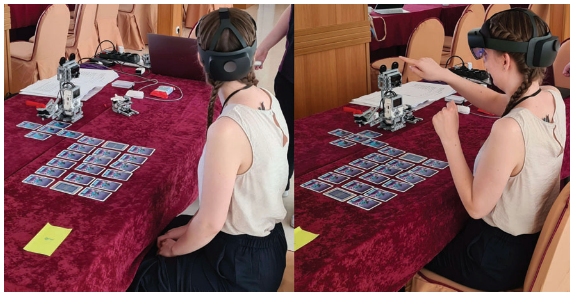

Figure 1: Setup of the game and interacting through Augmented Reality.

Prior to interacting with the robot, participants were briefed through the AR device with instructional videos, developed utilizing the Immersive Multimodal Psychomotor Environments for Competence Training (IMPECT) framework [12, 13, 14], showed the game rules as follows:

- Navigation through the spaceship is restricted to movements between rooms connected by doors.
- The primary task is to uncover a keycard within a concealed box and proceed to the engine room to conclude the game.

- The game introduces special 'robot rooms', where decision-making is jointly conducted by the participant and the robot.
- A maximum of 15 turns are allotted to locate the keycard and reach the engine room. If the number of turns exceeds 15, the time has run out and the spaceship is deemed to explode.

In the robot rooms (playing card in the middle of figure 2), the decision of the robot was scripted via an algorithm. The robot had two states dependent on the number of moves left until the end of the game. The switching happened when the estimated perfect path through the maze from the current position of the player would be within a critical range to not be able to complete the maze in the given 15 moves. Before this switch happened, the robot would always agree with the decision of the player even if that would mean they would go in the wrong direction. The second state was then to always point in the correct direction which is where disagreement can happen. Additionally, the last rooms were set up to have two valid paths to take while the robot was programmed to always disagree with the player's choice. This would result in a forced disagreement to be able to see if the participant chooses to trust the robot or go the other path. Afterward, the participants were not informed if the other path would work as well, so if they chose to disagree with the robot, they would only know that their chosen path was valid and might assume that the path the robot chose in the end was invalid. This assumption happened from observation during the study and questions of participants to the conducting researcher afterward.

In order to play the game participants first had to select the virtual card in the AR environment by clicking it. Then, participants had to flip over the selected physical equivalent card. This way, they could navigate the maze. In the special robot rooms, the movement changed slightly. First, participants selected the virtual card, however, the system then waited for the robot's algorithm to create a suggestion as described above. After that participants had full control over which room they would go to. They could select 1) their first choice, 2) the robots algorithms suggestions if it would be different, or 3) another room if they decided to go somewhere else. The turn ended with participants flipping over the selected physical equivalent card.

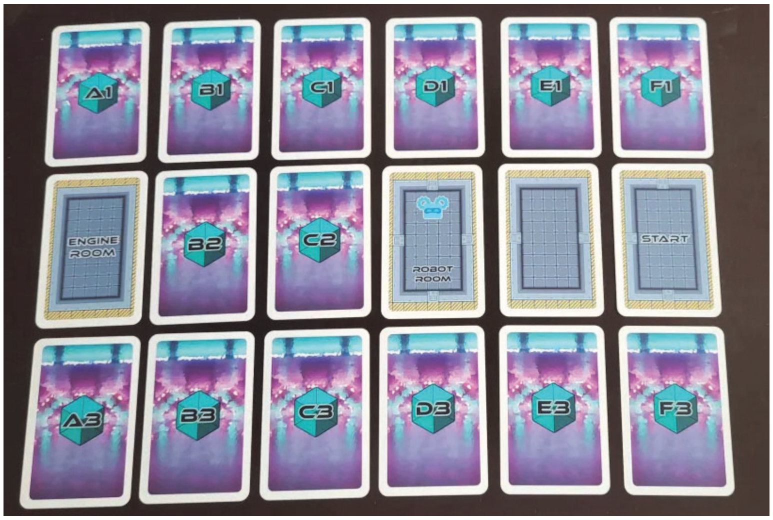

Figure 2: Game Cards on the table in the middle of a typical game.

2.1. Interaction with the robot

The interaction with the robot happened on different levels [15] with different interaction directions [16]. The first interaction was human-led. In this interaction, the robot first had to be assembled after it broke down. The participants first saw a video on how to effectively repair the robot on the AR device. Then, the participants could assemble the missing parts of the robot, leading the interaction. Possible mistakes or missing parts were then given as feedback through the AR device using pop-up feedback. The participants interpreted this feedback correctly and assembled the robot after some time. This interaction is partly there to give time to the participant to familiarize themselves with the robot and also look closer at it.

After the first interaction was done, the game, as described above began. Participants selected different rooms until they went into a robot room. In these robot rooms, the interaction was as follows: First, participants had to make a decision about which room to go next. Then they had to wait for the robot's algorithm to give a suggestion. These suggestions were displayed on the physical robot. Then, participants had to make their final decision based on their own first choice and the robots suggestion. This interaction was designed in this way to create a human-led approach. If the robot gave the suggestion before the participants made a decision, the participants would always follow the robot. This was shown in previous tests with this setup. In order to avoid this, we changed the interaction to have this two-step process. Because of this process, the interaction became more collaborative in nature. The planned disagreement of the algorithm at the last decision had the effect that people who reported higher trust in the robot decided against their own first decision.

Another crucial interaction with the robot happened when participants found the keycard. The keycard was stored in a special box, which, in the constrains of the game, can only be opened by the robot. However, to do this correctly, the participants first had to hand over the box to the robot in the correct manner. This interaction was first shown using the AR device as a video and if mistakes happened through feedback pop-ups. The interaction ended with the participants finalizing it with a press of a button on the robot. The robot then opened the special box for the participants to be able to obtain the keycard.

3. Main Findings

The following observations can be made:

Regarding trust, the data shows a wide spectrum of answers. On the question of "To what extent do you trust the robot in terms of collaboration?" the answers received range from "I trusted it [the robot] fully and did not doubt it" over to "[...] sometimes I trusted him [the robot] & sometimes not" up until the answer "I didn't trust the robot at all [...]". The participants varied also in behaviour while playing the game. While some participants followed the robot at every step, some participants refused to take the advice of the robot. However, within the study group, nobody always disagreed with the robot. This might be because in the beginning the robot's algorithm was programmed to always agree with the participant.

The data from the reconstructed path, combined with the participants' questionnaire responses, indicated that those who reported fully trusting the robot consistently followed the robot's decisions. Conversely, none of the participants who expressed distrust or difficulty in trusting the robot consistently went against the robot's decisions. This may suggest that they ignored the robot's decisions rather than attributing malicious intent to the robot.

According to the game rules, participants first needed to find the keycard to access the engine room. However, some participants, confused by the objective, headed straight for the exit, unaware of their mistake. When the robot suggested they should return, participants assumed it was an error and ignored it. The robot, limited in its communication abilities, couldn't explain the reason for its suggestion. This limitation was not anticipated or programmed into the robot's behavior during the study's design.

4. Discussion & Limitations

Due to the currently limited data, the findings are almost impossible to generalize. Further research and data collection has to be made to collect more findings in this area.

The study aimed to examine different types of HRI. Participants first engaged in a task where they repaired robot parts using an instructional video on the AR device. This activity helped them closely examine the robot and create a scenario where the robot needed assistance. This would then be flipped when the robot was consulted in the respective robot rooms where the human received assistance from the robot.

The results showed significant variation in trust levels among participants, even though they experienced the same environment and interaction knowledge. Some saw the robot as infallible, while others were skeptical. This suggests that trust in the robot's abilities is highly individualized and influenced by personal perceptions, especially when no prior information about the robot's intent is given.

According to the literature, the terms confidence and trust share similarities [17] and are close concepts. Even though that are often used interchangeably, these terms might not mean the exact same for every participant. While confidence refers to a specific reference, trust can be broad [18].

By chance, we found a technical limitation of the device which hindered one data collection. One participant was visually impaired which created a incompatibility with the built-in eye tracking of the Hololens 2 as well as the overall intractability through the device. The data of this participant was not used in the presented data.

5. Future Work

In future work, we will explore how trust is related to the mental model in HRI. We will run the same experiment but expand the questionnaire to use validated questionnaires [19] and validate if the current findings still persist with more participants.

Other ways how this study could be extended by researchers are the usage of different robots. The robot used was semi-humanoid which means it had human-like features while being very low on the human-likeness scale. Other robot designs might, especially humanoid robots could result in higher levels of trust as suggested by the literature [20].

6. Conclusion

This study is about HRI in an educational context, focusing on trust and collaborative decision-making facilitated by AR technology. Although the data is currently limited and not yet generalizable, several key insights emerged. Participants exhibited a broad spectrum of trust levels towards the robot, ranging from full trust to complete skepticism. These varying perceptions were influenced by individual interactions and experiences with robots, suggesting that trust in autonomous robots might be highly subjective and personal. Behavioral patterns also varied, with some participants consistently following the robot's guidance while others occasionally disregarded it. Initial agreement from the robot likely contributed to this variability, as it may have influenced early interpersonal dynamics.

Ultimately, this research underscores the complexity of developing trust in human-robot collaborations and points to the potential of personalized approaches to enhance HRI through means like Generative AI. Further investigations with a more extensive dataset will be crucial for validating these preliminary findings.

Acknowledgments

This research was funded by the BMBF, the German Federal Ministry of Education and Research, under the MILKI-PSY (ger. abb. for Multimodal immersive learning with artificial intelligence for psychomotor training; https://www.milki-psy.de) name and the grant code: `16DHB4013`.

We would also like to acknowledge the dedication and hard work of the development team of the software running on the AR device. Namely, Patrick Handwerk and Suhana Biswas.

References

[1] R. Bemelmans, G. J. Gelderblom, P. Jonker, L. De Witte, Socially assistive robots in elderly care: a systematic review into effects and effectiveness, Journal of the American Medical Directors Association 13 (2012) 114–120.

[2] R. Mundaca-Uribe, N. Askarinam, R. H. Fang, L. Zhang, J. Wang, Towards multifunctional robotic pills, Nature Biomedical Engineering (2023) 1–13.

[3] A. Gujarathi, A. Kulkarni, U. Patil, Y. Phalak, R. Deotalu, A. Jain, N. Panchi, A. Dhabale, S. Chiddarwar, Design and development of autonomous delivery robot, arXiv preprint arXiv:2103.09229 (2021).

[4] A. Edwards, C. Edwards, P. R. Spence, C. Harris, A. Gambino, 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 (2016) 627–634.

[5] C.-W. Wei, I. Hung, L. Lee, N.-S. Chen, et al., A joyful classroom learning system with robot learning companion for children to learn mathematics multiplication., Turkish Online Journal of Educational Technology-TOJET 10 (2011) 11–23.

[6] R. R. Galin, R. V. Meshcheryakov, Human-robot interaction efficiency and human-robot collaboration, in: Robotics: Industry 4.0 issues & new intelligent control paradigms, Springer, 2020, pp. 55–63.

[7] F. Barravecchia, M. Bartolomei, L. Mastrogiacomo, F. Franceschini, Redefining human–robot symbiosis: a bio-inspired approach to collaborative assembly, The International Journal of Advanced Manufacturing Technology 128 (2023) 2043–2058.

[8] T. Keller, D. Majonica, A. Richert, R. Klemke, Prerequisite knowledge of learning environments in human-robot collaboration for dyadic teams, Proceedings. ISSN 1613 (2020) 0073.

[9] C. Esterwood, L. P. Robert, The theory of mind and human–robot trust repair, Scientific Reports 13 (2023) 9877.

[10] K. Schaefer, The perception and measurement of human-robot trust (2013).

[11] K. A. M. Sanusi, D. Majonica, P. Handwerk, S. Biswas, R. Klemke, Evaluating an immersive learning toolkit for training psychomotor skills in the fields of human-robot interaction and dance., in: MILeS@ EC-TEL, 2023, pp. 70–78.

[12] K. A. M. Sanusi, D. Majonica, L. Künz, R. Klemke, Immersive training environments for psychomotor skills development: A student driven prototype development approach, in: Multimodal Immersive Learning Systems 2021, CEUR, 2021, pp. 53–58.

[13] K. A. M. Sanusi, M. Slupczynski, M. Geisen, D. Iren, R. Klamma, S. Klatt, R. Klemke, Impect-sports: Using an immersive learning system to facilitate the psychomotor skills acquisition process., in: MILeS@ EC-TEL, 2022, pp. 34–39.

[14] A. Samanta, H. Kotte, P. Handwerk, K. Asyraaf Mat Sanusi, M. Geisen, M. Kravcik, N. Duong-Trung, Impect-pose: A complete front-end and back-end architecture for pose tracking and feedback, in: Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, 2024, pp. 142–147.

[15] R. Schulz, P. Kratzer, M. Toussaint, Preferred interaction styles for human-robot collab-

oration vary over tasks with different action types, Frontiers in neurorobotics 12 (2018) 36.

[16] D. Majonica, N. Fanchamps, D. Iren, R. Klemke, Exploring immersive learning environments in human-robot interaction use cases, in: International Conference on Games and Learning Alliance, Springer, 2023, pp. 267–276.

[17] M. Lupoi, Trust and confidence, Sweet & Maxwell, 2009.

[18] B. D. Adams, Trust vs. confidence, Defence Research and Development Canada-Toronto, 2005.

[19] T. Nomura, T. Suzuki, T. Kanda, K. Kato, Measurement of negative attitudes toward robots, Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems 7 (2006) 437–454.

[20] J. Pinney, F. Carroll, P. Newbury, Human-robot interaction: the impact of robotic aesthetics on anticipated human trust, PeerJ Computer Science 8 (2022) e837.