

# Common (good) practices measuring trust in HRI

Patrick Holthaus

*Robotics Research Group*

University of Hertfordshire, Hatfield, United Kingdom.  
p.holthaus@herts.ac.uk

Alessandra Rossi

*PRISCA Research Lab*

University of Naples Federico II, Naples, Italy.  
alessandra.rossi@unina.it

## I. INTRODUCTION

Trust in robots is widely believed to be imperative for the adoption of robots into people’s daily lives. It is, therefore, understandable that the literature of the last few decades focuses on measuring how much people trust robots – and more generally, any agent - to foster such trust in these technologies. Researchers have been exploring how people trust robot in different ways, such as measuring trust on human-robot interactions (HRI) based on textual descriptions or images without any physical contact [1], [2], during and after interacting with the technology [3], [4].

Nevertheless, trust is a complex behaviour, and it is affected and depends on several factors, including those related to the interacting agents (e.g. humans, robots, pets), itself (e.g. capabilities, reliability), the context (e.g. task), and the environment (e.g. public spaces vs private spaces vs working spaces)[5], [6]. In general, most roboticists agree that insufficient levels of trust lead to a risk of disengagement [7] while over-trust in technology can cause over-reliance and inherit dangers [8], for example, in emergency situations [9]. It is, therefore, very important that the research community has access to reliable methods to measure people’s trust in robots and technology. In this position paper, we outline current methods and their strengths, identify (some) weakly covered aspects and discuss the potential for covering a more comprehensive amount of factors influencing trust in HRI.

## II. CURRENT MEASURES OF TRUST

Current trust measures often aim to quantify the amount of trust that is subconsciously exhibited towards a robot. Subjective questionnaire scales, such as [10], [11], [12], consist of several items, which when combined can capture different aspects of trust. Individual aspects can, for example, reflect people’s trust in performance or morality [10]. Such questionnaire scales are widely adopted, but sometimes concerns regarding validity or positive bias emerge [13]. To account for such biases, subjective measurements are often complemented by objective measurements of trust. Researchers, for example, successfully use compliance with a robot’s suggestions [14] or record whether people are sharing information that makes them potentially vulnerable – such as secrets [3] – with a robot as indicators of whether people trust a robot. These choices strongly relate trust to the task that the robot is undertaking, such as its criticality of the task, whether it implies a cognitive or a physical task, and the magnitude of its consequences. Trust is usually measured within a specific scenario where the

human-robot interaction takes place, sometimes also asking people to relate to other environments by imagining a different scenario or environment [15], [3], for example, by questioning whether their responses would change in another situation, with another agent, in a public vs private environment.

To calibrate subjective and objective measurements of subconsciously exhibited trust and to account for potential external biases and situational influences, trust measures are often combined with other questionnaires that measure people’s general tendency or propensity to trust [16], or their potentially negative attitude towards robots [17]. Such questionnaires are usually administered before any interaction to avoid side effects of the experiment itself. When calibrating against or comparing individual robots, trust measures are also often combined with scales that are able to measure the social attributes of a specific robot (e.g. [18]). Moreover, to account for potential influences of novelty, researchers employ familiarisation phases where participants can get accustomed to the robot [19] to minimise such an effect, or look into repeated and/or long-term experiments to directly observe any changes (e.g. [3], [20]).

## III. DRAWBACKS AND OPPORTUNITIES

The number of different approaches to measuring trust shows that there is currently no catch-all method providing a solution that would be viable across different experimental setups; individual solutions are required that are attuned to the participants, situation and research questions. Henceforth, a mix of methods is currently regarded as the gold standard. Appropriate questionnaire scales measuring more universal aspects of trust are usually paired with individual questions targeted at identifying further influences that emerge from the specific experimental situation. If employed, additional open-ended questions about people’s choices or behaviours can complete that mix of methods and allow for further qualitative insights to support reasoning on top of the quantitative analysis.

However, even such a mix of methods cannot comprehensively account for the magnitude of factors that potentially affect the trust people put into an interaction with a robot. Firstly, we would like to mention that most quantitative analyses only measure trust subconsciously, and researchers often avoid direct questions about how much people trust a robot. Since we believe that the trust people put forward might also be related to conscious concepts of a robot’s (social) credibility [21], we would like to encourage researchers to

incorporate direct questions about a robot’s trustworthiness and reasoning why people would trust or not trust a robot into their experimental scenario. Open-ended questions can also help to capture conscious aspects of trust when asking to explain if a person’s behaviours were influenced by their trust in the robot.

In line with [22], we propose to engage in a discussion to complement existing trust measures by looking at the wider situation and work together to find new *ad hoc* measures to address common side effects that might influence the amount of trust people put into a robot. Such measures would aim to assess: (I) the impact of repeated interactions and long-term human-robot interactions [3] with particular attention to the novelty effect and how it might interplay with trust metrics; (II) how imperfect interactions [23] might overshadow results; (III) how subconscious interpretations [24] might have been formed, and, as a consequence, influence the perception of behaviours; (IV) how the conformity of robot behaviours to social norms [25] in accordance with the situation might affect findings; and (V) how the surroundings or experimental environment prevent or facilitate feelings of trust. We, therefore, suggest specifically looking into five different aspects as a result of the current research that we believe are key for developing future measures of trust in HRI. These key aspects are the following:

- Human This aspect includes all factors related to the human, such as personality, predisposition to trust, previous experience and expectations.
- Robot We suggest measuring aspects related to the robot itself, such as people’s perception of its appearance, its functionalities, and its performance.
- Task This aspect includes factors depending on the task’s characteristics. For example, it would include concepts like outcome, severity and criticality.
- Environment The environment aspect aims to measure factors depending on the environment characteristics where the HRI is developing and how it may affect trust. These could include public, private and working environments.
- Other agents This aspect could include factors depending on the presence and actions of other agents, such as pets, children, and older people.

#### IV. CONCLUSION

Trust is a fundamental aspect that drives not only human interactions but any human-agent interaction in the humans’ day-to-day activities. Yet, this is one of the topics in HRI that still does not have well-defined and *ad hoc* measures to evaluate to which extent people trust robots, and, as a consequence, to calibrate such trust to have a successful and effective relationship. Accordingly, this paper suggests investigating five key factors that affect trust within a situation to be able to assess and reduce common side effects influencing how people put their trust in robots.

#### ACKNOWLEDGMENTS

This work has been supported by the Italian PON R&I 2014-2020 - REACT-EU (CUP E65F21002920003).

#### REFERENCES

- [1] W.-Y. Lee, M. Sakashita, E. Ricci, H. Claire, F. Guimbretière, and M. Jung, “Interactive vignettes: Enabling large-scale interactive hri research,” in *2021 30th IEEE international conference on robot & human interactive communication (RO-MAN)*. IEEE, 2021, pp. 1289–1296.
- [2] T. Law, M. Chita-Tegmark, and M. Scheutz, “The interplay between emotional intelligence, trust, and gender in human–robot interaction: A vignette-based study,” *International Journal of Social Robotics*, vol. 13, no. 2, pp. 297–309, 2021.
- [3] A. Rossi, K. Dautenhahn, K. L. Koay, M. L. Walters, and P. Holthaus, “Evaluating people’s perceptions of trust in a robot in a repeated interactions study,” in *International Conference on Social Robotics (ICSR 2020)*, ser. Lecture Notes in Computer Science, A. R. Wagner, D. Feil-Seifer, K. S. Haring, S. Rossi, T. Williams, H. He, and S. S. Ge, Eds. Golden, Colorado: Springer Cham, 2020, vol. 12483.
- [4] N. Lingg and Y. Demiris, “Building trust in assistive robotics: Insights from a real-world mobile navigation experiment,” in *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, 2023, pp. 1–7.
- [5] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, “A meta-analysis of factors affecting trust in human-robot interaction,” *Human Factors: The Journal of Human Factors and Ergonomics Society*, vol. 53, no. 5, pp. 517–527, 2011.
- [6] R. Alessandra, D. Kerstin, K. Kheng Lee, and M. L. Walters, “The impact of peoples’ personal dispositions and personalities on their trust of robots in an emergency scenario.” vol. 9, 2018.
- [7] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, “Towards safe and trustworthy social robots: ethical challenges and practical issues,” in *Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings 7*. Springer, 2015, pp. 584–593.
- [8] A. M. Aroyo, J. De Bruyne, O. Dheu, E. Fosch-Villaronga, A. Gudkov, H. Hoch, S. Jones, C. Lutz, H. Sætra, M. Solberg *et al.*, “Overtrusting robots: Setting a research agenda to mitigate overtrust in automation,” *Paladyn, Journal of Behavioral Robotics*, vol. 12, no. 1, pp. 423–436, 2021.
- [9] P. Robinette, W. Li, R. Allen, A. M. Howard, and A. R. Wagner, “Overtrust of robots in emergency evacuation scenarios,” in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2016, pp. 101–108.
- [10] D. Ullman and B. F. Malle, “Measuring gains and losses in human-robot trust: Evidence for differentiable components of trust,” in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 618–619.
- [11] K. E. Schaefer, “Measuring trust in human robot interactions: Development of the “trust perception scale-hri”,” in *Robust intelligence and trust in autonomous systems*. Springer, 2016, pp. 191–218.
- [12] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, “Foundations for an empirically determined scale of trust in automated systems,” *International journal of cognitive ergonomics*, vol. 4, no. 1, pp. 53–71, 2000.
- [13] R. S. Gutzwiller, E. K. Chiou, S. D. Craig, C. M. Lewis, G. J. Lematta, and C.-P. Hsiung, “Positive bias in the ‘trust in automated systems survey’? an examination of the jian et al.(2000) scale,” in *Proceedings of the Human Factors and Ergonomics Society annual meeting*, vol. 63, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2019, pp. 217–221.
- [14] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, “Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust,” in *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*, 2015, pp. 141–148.
- [15] A. Rossi, G. Perugia, and S. Rossi, “Investigating customers’ perceived sensitivity of information shared with a robot bartender,” in *Social Robotics*, H. Li, S. S. Ge, Y. Wu, A. Wykowska, H. He, X. Liu, D. Li, and J. Perez-Osorio, Eds. Cham: Springer International Publishing, 2021, pp. 119–129.

- [16] G. M. Alarcon, J. B. Lyons, J. C. Christensen, S. L. Klosterman, M. A. Bowers, T. J. Ryan, S. A. Jessup, and K. T. Wynne, "The effect of propensity to trust and perceptions of trustworthiness on trust behaviors in dyads," *Behavior research methods*, vol. 50, pp. 1906–1920, 2018.
- [17] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "Psychology in human-robot communication: An attempt through investigation of negative attitudes and anxiety toward robots," in *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE catalog No. 04TH8759)*. IEEE, 2004, pp. 35–40.
- [18] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas) development and validation," in *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*, 2017, pp. 254–262.
- [19] M. Edmonds, F. Gao, H. Liu, X. Xie, S. Qi, B. Rothrock, Y. Zhu, Y. N. Wu, H. Lu, and S.-C. Zhu, "A tale of two explanations: Enhancing human trust by explaining robot behavior," *Science Robotics*, vol. 4, no. 37, p. eaay4663, 2019.
- [20] A. Ayub, Z. Francesco, P. Holthaus, C. Nehaniv, and K. Dautenhahn, "Continual Learning through Human-Robot Interaction – Human Perceptions of a Continual Learning Robot in Repeated Interactions," *International Journal of Social Robotics*, under review.
- [21] P. Holthaus, "How does a robot's social credibility relate to its perceived trustworthiness?" in *RO-MAN 2021 Workshop on Trust, Acceptance and Social Cues in Human-Robot Interaction - SCRITA*, 2021. [Online]. Available: <https://arxiv.org/abs/2107.08805>
- [22] M. Chita-Tegmark, T. Law, N. Rabb, and M. Scheutz, "Can you trust your trust measure?" in *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction*, 2021, pp. 92–100.
- [23] T. Schulz, R. Soma, and P. Holthaus, "Movement Acts in Breakdown Situations - How a Robot's Recovery Procedure Affects Participants' Opinions," *Paladyn, Journal of Behavioral Robotics: Special Issue Trust, Acceptance and Social Cues in Robot Interaction*, vol. 12, no. 1, pp. 336–355, 2021.
- [24] P. Holthaus, T. Schulz, G. Lakatos, and R. Soma, "Communicative Robot Signals: Presenting a New Typology for Human-Robot Interaction," in *International Conference on Human-Robot Interaction (HRI 2023)*. Stockholm, Sweden: ACM/IEEE, 2023, pp. 132–141.
- [25] P. Holthaus, C. Menon, and F. Amirabdollahian, "How a Robot's Social Credibility Affects Safety Performance," in *International Conference on Social Robotics (ICSR)*, ser. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, vol. 11876, pp. 740–749.