Investigating human perceptions of trust and social cues in robots for safe HRI

Alessandra Rossi*, Kerstin Dautenhahn†, Kheng Lee Koay‡ and Michael L. Walters§
Email: *‡§[a.rossi, k.l.koay, m.l.walters]@herts.ac.uk, †kerstin.dautenhahn@uwaterloo.ca

Abstract—Human-Robot Interaction (HRI) research has established that trust has a key role in human acceptance of robots in human-oriented environments. However, we can expect that robots will exhibit occasional mechanical, programming or functional errors, as with other electrical consumer devices. This research aims to identify the factors that undermine people's trust in robots, and to provide robots with coping mechanisms that will allow them to gain people's trust and to enhance a successful interaction. In this paper, we give a short overview of the state of our work.

I. INTRODUCTION

The autonomous robots will enable to collaborate with people in their daily living activities in the future. This means that robots will need to be able to share the same physical space and engage human users in social interactions. Therefore, it is important that people accept and trust robots to be able to look after their well-being and assist them in a safe way. Trust can be affected by several factors [1] in HRI, which can be: Human-related, such as self-confidence and prior experience with robots; Robot-related, such as robot's embodiment and failure rates; and Environmental, such as communication and team collaboration. In this study, we explore some of these aspects which can affect and undermine human trust in robots. We provide strategies that can help the recovery of trust after a trust breach. In particular, we investigate the effects of a robot that exhibits social cues and that affects robot failures on people's trust in them. This research [2] investigate the following research questions:

RQ1 How do robot errors with varying magnitudes of consequences affect human trust in a robot?

RQ2 Does the impact on trust change if the error happens at the beginning or end of an interaction?

RQ3 Can the trust of humans in a robot be regained more easily if it is a big error happening at the beginning or end of an interaction? Or is it easier to recover from a loss of trust caused by a small error happening at the beginning or the end of the interaction?

RQ4 Does awareness of the robots' real potential and limitations affect human perceptions of trust in the robot?

RQ5 Are the use of human social behaviours by robots sufficient for humans to trust a robot to look after their well-being?

RQ6 Can a human's trust in their robot change over time (if the robot starts to show erratic behaviours)?

II. STATE OF THE RESEARCH

In this Section, we provide an overview of our current research findings:

FQ1 People trust is affected more severely when the robot made errors having severe consequences ([3]–[5]).

FQ2 Participants tend to form their judgements at the beginning of an interaction and adjusted it later on, depending on the robot's performance ([3]–[5]).

FQ3 There is a greater tendency to not trust the robot when severe errors happen at the beginning of an interaction ([3]–[5]).

FQ4 For primary and secondary school students, we found that their awareness of being able to program different robots' behaviours affected their perception of the robots ([6], [7]).

FQ5 We conducted two different studies. One focused on a navigation task, and the other focused on tasks with different criticality levels. In the first study [8], we found that people preferred to interact with a social robot. They also tend to trust more a social robot to guide them than their initial expectations. In the second study, we did not observe differences between individual trust in a social or non-social robot.

FQ6 We are currently conducting a repeated study to assess how people's trust changes over several interactions. In particular, we aim to clarify research question RQ2.

REFERENCES

[1] 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.

[2] A. Rossi, K. Dautenhahn, K. L. Koay, and J. Saunders, "Investigating human perceptions of trust in robots for safe hri in home environments," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 375–376.

[3] A. Rossi, K. Dautenhahn, K. L. Koay, and M. L. Walters, "Human perceptions of the severity of domestic robot errors," in *International Conference on Social Robotics (ICSR)*, K. Abderrahmane, Y. Eiichi, S. G. Shuzhi, S. Kenji, C. John-John, E. Friederike, and H. He, Eds. Tsukuba, Japan: Springer International Publishing, 2017, p. 647–656.

[4] ——, "A study on how the timing and magnitude of robot errors may influence people trust of robots in an emergency scenario," in *International Conference on Social Robotics (ICSR)*, K. Abderrahmane, Y. Eiichi, S. G. Shuzhi, S. Kenji, C. John-John, E. Friederike, and H. He, Eds. Tsukuba, Japan: Springer International Publishing, 2017, pp. 42–52.

[5] A. Rossi, K. Dautenhahn, K. Koay, and M. L. Walters, "The impact of peoples' personal dispositions and personalities on their trust of robots in an emergency scenario," vol. 9(1), pp. 137–154, 2018.

[6] A. Rossi, P. Holthaus, K. Dautenhahn, K. L. Koay, and M. L. Walters, "Getting to know pepper: Effects of people's awareness of a robot's capabilities on their trust in the robot," in *Proceedings of the 6th International Conference on Human-Agent Interaction*, ser. HAI '18. New York, NY, USA: ACM, 2018, pp. 246–252.

[7] A. Rossi, S. Moros, K. Dautenhahn, K. L. Koay, and M. L. Walters, "Getting to know kaspar: Effects of people's awareness of a robot's capabilities on their trust in the robot," in *submitted at IEEE International Conference on Robot and Human Interactive Communication 2019*, ser. RO-MAN 19, 2019.

[8] A. Rossi, F. Garcia, A. Cruz Maya, K. Dautenhahn, K. L. Koay, M. L. Walters, and A. K. Pandey, "Investigating the effects of social interactive behaviours of a robot on people's trust during a navigation task," in *to be published in Proceedings of the*, ser. TAROS 2019, 2019.