Investigating Human Perceptions of Trust in Robots for Safe HRI in Home Environments

Alessandra Rossi
a.rossi@herts.ac.uk

Kerstin Dautenhahn
k.dautenhahn@herts.ac.uk

Kheng Lee Koay
k.l.koay@herts.ac.uk

Joe Saunders
j.l.saunders@herts.ac.uk

School of Computer Science, University of Hertfordshire, College Ln
Hatfield, AL10 9AB, UK

ABSTRACT

In an era in which robots take a part in our lives in daily living activities, humans have to trust robots in home environments. We aim to create guidelines that allow humans to trust robots to be able to look after their well-being by adopting human-like behaviours. We want to study a Human-Robot Interaction (HRI) to assess whether a certain degree of transparency in the robots actions, the use of social behaviours and natural communications can affect humans' sense of trust and companionship towards the robots. However, trust can change over time due to different factors, e.g. due to perceiving erroneous robot behaviors. We believe that the magnitude and the timing of the error during an interaction may have different impacts resulting in different scales of loss of trust and of restoring lost trust.

Keywords

Human-robot interaction; Social robotics; Robot companion; Trust in robots; Trust recovery.

1. INTRODUCTION

Robots are beginning to take part in our daily living activities, and using them as companions may require the development of a more natural HRI. To accept the presence of the robots in their lives humans need to believe and trust that a robot is able to look after humans' needs in a safe environment. Indeed, it is important that humans trust their robot companion e.g. to be able to not start a fire if it has been left to boil water for a tea or to not open the door to strangers when the humans are not at home or are asleep. We believe that a robot should be able to use humans' communication modes in order to provide support to people. According to Bethel et al. [1], naturalistic social interaction in robots can be designed through body movement, posture, orientation, color, and sound. Martelaro et al. [7] established that a robot showing vulnerability, through facial expressions, color and movements, increased trust and companionship. However, trust is a very complex feeling even between humans, and it could be induced and improved according to different factors [3] (i.e. social rules), and it can change after a violation. It could be very hard to estabilish again the trust after a breach. Trust recovery not only depends on personal characteristics [5], but also on the length of the relationship [12] and on other aspects, i.e. apology [2]. Several research groups both in Human-Computer Interaction and HRI are investigating how to promote trust and how to recover it after a violation. However, the current literature has produced different and sometimes contradictory results, e.g. Salem et al. [11] and Robinette et al. [10] highlighted that no matter the erratic behaviour of the robots, participants trusted the instructions of a faulty robot; while Muir and Moray [8] and Desai et al. [4] showed that errors produce an evident drop in the trust in the robot. These studies did not investigate the effects of both the magnitude and the timing of the errors. Instead, we believe that both the magnitude and the timing of an error have different impacts resulting in different scale of loss and gain of trust. Moreover, we believe that humans can recover trust in robots if the robots exhibit social cues, e.g. apologizing and promising to do better in the future [9].

HRI '17 Companion March 06-09, 2017, Vienna, Austria

ACM ISBN 978-1-4503-4885-0/17/03.

DOI: http://dx.doi.org/10.1145/3029798.3034822

2. RESEARCH QUESTIONS

The following research questions and hypotheses will be addressed:

R1 Which kind of erratic behaviors impact human's trust in a robot? We believe that there is a correlation between the magnitude of the error performed by the robot and the loss of trust of the human in the robot.

R2 Does the impact on trust change if the error happens at the beginning or end of interaction? We believe that there is a correlation between the timing in which the error is performed during the interaction and the loss of trust.

R3 Can the trust of humans in a robot be recovered more easily if it is a big error happening at the beginning or end of 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? We believe that there is a correlation between the timing of when the error occurred and the magnitude of the error.

R4 Are the use of human social behaviors enough for humans to trust a robot to look after their well-being? We believe that social cues make robots more human-like, and

better accepted by humans, then humans can be more inclined to rely on them.

R5 Can a human's trust in her robot change over time? We believe that trust could change if the initial conditions of trusting a robot change, e.g. the robot starts to show erratic behaviors.

3. IMPACT OF SMALL AND BIG ERRORS IN HUMANS TRUST INTO ROBOTS

The research questions **R1**, **R2** and **R3** are currently being investigated. Since people are very different from each other, in age, gender, cultural and social habits, we had to establish first what people consider to be small or big errors of their home companion robot.

3.1 Evaluating possible errors of robot companion

Procedure Participants were asked to imagine that they live with a robot companion in their home. However, the robot might exhibit some mistakes. The participant has to complete a questionnaire rating the magnitude of the errors illustrated in different scenarios, e.g. *"Your robot leaves your pet hamster outside the house in very cold weather"*. The questionnaire is composed of 20 questions, plus two optional in which the participant is free to add their own examples of errors not already included in the scenarios proposed.

Method This study has been organized as a within-subjects experiment. Each participant has been shown the same questions rated using a 7-point Likert scale [1= small error and 7=big error].

Analysis According the resulting answers of 50 participants - (32 men, 18 women), 19 to 63 years old [mean 41, std 11.59]. All the questions with values < 4 are considered small errors and those with values > 4 are considered big errors. We identify 3 big errors and 3 small errors picking the ones with the majority of ratings gained. We did not find any significant differences between gender or age of the participants and their rating of the errors.

3.2 Evaluating the impact of errors on human perceptions of robots

This study has not yet concluded. However, participants will be tested using an interactive storyboard accessible through a web application. They will be presented 10 different scenarios, in which the robot shows flawless and erroneous behaviours. Each experiment is executed using 5 different conditions: 10 different tasks executed correctly, 10 different tasks with 3 small errors at the beginning and 3 big errors at the end of the interactions, 10 different tasks with 3 big errors at the beginning and 3 small errors at end of the interactions, 10 different tasks with 3 big errors at the beginning and at the end of the interactions, and 10 different tasks with 3 small errors at the beginning and at the end of the interactions. At the end, the participants will be presented with an emergency situation, i.e. *"a fire in the kitchen"* to finally test the evolution of the trust in the robot.

4. FUTURE WORKS

In further studies we aim to investigate the change of trust over time and how to make robots more trustworthy and reliable for humans. A first study will address the research question **R4**, and a second one the research question **R5**. Both studies will be conducted using a real robot in a fully sensorised house belonging to the University, since humans interact differently with real and virtual robots [6].

5. REFERENCES

[1] C. Bethel and R. Murphy. Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(1):83–92, 2008.

[2] A. W. Brooks, H. Dai, and M. E. Schweitzer. IâĂŹm sorry about the rain! superfluous apologies demonstrate empathic concern and increase trust. *Social Psychological and Personality Science*, 5(4):467–474, 2014.

[3] J. A. Colquitt, B. A. Scott, and J. A. LePine. Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, 92(4):909–927, 2007.

[4] M. Desai, P. Kaniarasu, M. Medvedev, A. Steinfeld, and H. Yanco. Impact of robot failures and feedback on real-time trust. In *ACM/IEEE International Conference on Human-Robot Interaction*, pages 251–258, 2013.

[5] M. P. Haselhuhn, M. E. Schweitzer, and A. M. Wood. How implicit beliefs influence trust recovery. *Psychological Science*, 5:645–648, 2010.

[6] K.-B. Hatice, F. Ester, D. Kerstin, S. Dag Sverre, and N. Chrystopher L. Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot. *Advanced Robotics*, 23(14):1951–1996, 2009.

[7] N. Martelaro, V. C. Nneji, W. Ju, and P. Hinds. Tell me more: Designing hri to encourage more trust, disclosure, and companionship. In *Proceeding HRI '16 The Eleventh ACM/IEEE International Conference on Human Robot Interation*, pages 181–188. IEEE Press Piscataway, 2016.

[8] B. M. Muir and N. Moray. Trust in automation: Part ii. experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39:429–460, 1996.

[9] P. Robinette, A. M. Howard, and A. R. Wagner. Timing is key for robot trust repair. *Social Robotics. Lecture Notes in Computer Science*, 9388:574–583, 2015.

[10] P. Robinette, W. Li, R. Allen, A. M. Howard, and A. R. Wagner. Overtrust of robots in emergency evacuation scenarios. In *Proceeding HRI '16 The Eleventh ACM/IEEE International Conference on Human Robot Interation*, pages 101–108. IEEE Press Piscataway, 2016.

[11] 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. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 141–148, 2015.

[12] O. Schilke, M. Reimann, and K. S. Cook. Effect of relationship experience on trust recovery following a breach. *Proceedings of the National Academy of Sciences*, 110(38):15236–15241, 2013.