Measuring Relational Trust in Human-Robot Interactions

Theresa Law
Tufts University
Medford, United States
theresa.law@tufts.edu

CCS CONCEPTS

• **Computer systems organization** → **Client-server architectures**; • **Human-centered computing** → *Interaction design theory, concepts and paradigms*; *Empirical studies in interaction design.*

KEYWORDS

Trust, relational trust, human-robot interaction, human-robot trust

ACM Reference Format:
Theresa Law. 2020. Measuring Relational Trust in Human-Robot Interactions. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20 Companion), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3371382.3377435

1 INTRODUCTION AND BACKGROUND

Trust is an integral part of almost any human-robot interaction (HRI). The most technologically advanced robot can sit unused if a human interactant does not trust it. Conversely, a robot that is overly trusted may be assumed to be more advanced than it truly is, resulting in over-reliance on an imperfect system [8]. One of the most widely used definitions for trust in HRI is that trust is "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability." [6]. As robots initially emerged into society in factory roles, their goals were clear and their performance was concretely measurable with metrics such as time to completion, number of errors in a given behavior, and behavior consistency over time. The humans who worked with them, therefore, could base their trust in the robots on how effectively they achieved these clearly defined goals.

With the advent of social robotics, however, robots' goals, and the metrics used to determine how well they met those goals, are now more vague and less obviously measurable. When a robot's goal is to be a social companion, how can we tell when it is doing a good job versus a bad job? Consequently, what cues can we utilize to inform us of whether or not we should trust it? With social robots, we have introduced a trust beyond that of trusting a robot's capabilities in a measurable task; we need to trust the robot's ability to behave as a social agent, a task which necessitates a more relational trust than tasks performed by factory robots.

HRI '20 Companion, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-7057-8/20/03.
https://doi.org/10.1145/3371382.3377435

Trust can be defined as a multidimensional concept that can be affected by a large range of factors [10] [3]. In a survey I wrote of empirical HRI trust studies that were published in 2018 and 2019, I categorized trust experiments into measuring either performance-based trust, relation-based trust, or a mix of the two. Performance-based trust measures how much people trust a robot's technical capabilities to perform in a certain way or to a certain standard. This is the type of trust that is primarily relevant to factory robots. Relation-based trust, meanwhile, refers to trusting that a robot can behave in such a way that it conforms to some societal norms. Compared to performance-based trust, relation-based trust is understudied in HRI, but is becoming increasingly relevant as robots enter our social spheres. Additionally, the work on relation-based trust that has been done almost exclusively uses subjective questionnaires as a means to measure trust, rather than the objective measures that are more often seen in studies that measure performance-based trust [5]. Table 1 provides examples of robot factors that have had an effect on either performance-based or relation-based trust. These examples come from studies included in my survey about trust in HRI.

Table 1: Examples of robot factors in previous studies that have affected performance-based and relation-based trust.

Performance-Based	Relation-based
Task Completion	Robot Gender
Factual Mistake	Group Emotion Expression
Error Timing	Facial Expression
Error Severity	Group Conformity

My research seeks to elucidate the role of relation-based trust in HRI. Toward this goal, I will explore robotic social factors that affect human trust, and paradigms that can be used to measure trust objectively. In my preliminary work, I have utilized subjective questionnaires as my measurements of trust. Moving forward, I will also be incorporating objective trust measures as well. Objective measures of trust allow us to see how people actually behave during a trust-dependent task with a robot, rather than rely on the results of subjective questionnaires which may ask for speculation about possible behaviors and scenarios. Currently, there are very few studies that utilize an objective measure that gets at relational trust [5]. Ideally, a study will incorporate the results of a robust, validated subjective questionnaire with objective behavioral data. This combination will paint a fuller picture of how the participant viewed the situation as well as how they actually acted during it. These thus-far untapped HRI questions and paradigms will provide valuable insight into how we interact with robots as social agents rather than simply advanced tools.

2 PRELIMINARY RESULTS

To study relation-based trust, I have thus far investigated how emotional intelligence, a social factor, facilitates trust. Emotional intelligence (EI) can be broadly defined as one's ability to detect, understand, and react to one's own emotions and the emotions of others [9]. High levels of EI have been linked to positive behaviors such as more effective teamwork [2]. I ran a series of vignette-style studies on Amazon Mechanical Turk to see if a robot's level of displayed EI affected participants' trust in the robot. In the first study, participants watched a vignette in which a human was reprimanded by their boss, and their robot teammate responded to the human who was reprimanded in a highly empathetic or in non-empathetic manner to indicate high or low EI. Participants' trust in the robot was then measured with a questionnaire based on [7], and their interpretation of the robot's level of EI with a questionnaire based on [1]. The robot's gender and the manner in which the vignette was presented (in text or audio form) were manipulated along with the EI level.

The robot in the the high EI conditions, which exhibited high levels of empathy, was trusted more than robot in the low EI conditions. Because EI is not an indicator of how capable a robot is at performing a task, this increase in trust seems to be driven by a relation-based trust.

In the second study, the robot's reliability and capability were manipulated rather than its EI. After the human-robot team got reprimanded by their boss, rather than respond in an empathetic or non-empathetic manner, the robot explained that it was either always available to help the human and had a reliably accurate algorithm (high trustworthiness), or was only sometimes available and had a moderately accurate algorithm (low trustworthiness). Participants then answered the same trust and EI questionnaires. Again, the robot's gender and the vignette presentation style were manipulated, as well as the order in which participants answered the two questionnaires.

When the robot's level of trustworthiness was specifically manipulated, there was a questionnaire order effect. When participants were primed to think about EI by answering the EI questionnaire before the trust questionnaire, they rated their trust in the robot as higher than when the trust questions were asked first. This highlights the fragility of relying solely on subjective, self-reporting questionnaires as the only measure of trust [4].

3 FUTURE WORK

EI is an attribute that is not necessary to perform most tasks, but carries societal and social weight that can affect how people interact with and trust robots that display that characteristic. My future work will continue to explore behaviors such as this to further understand how humans trust robots in a manner beyond trusting the robot's performance and task capabilities. Additionally, the studies presented here were observational studies and will be replicated as in-lab studies; while these online studies provided us with insight as to how humans trust a robot that they observe, it is possible that people will react differently when they are the actual interactants. I will also be focusing on running studies in which I can utilize an objective measure of trust to supplement the subjective questionnaires. As shown in my EI study, subjective measures can be influenced by seemingly trivial factors such as the order in which questionnaires are presented. To this end, I will turn to fields such as behavioral economics, which offer a number of games such as the prisoner's dilemma and the ultimatum game, both of which can be used to measure relation-based trust because a player's strategy is dependent upon their perception of how fair and cooperative the other player will be. By having participants play these and other similar games with robots, I will have an objective measure of how much they trust that a robot will conform to certain social norms. These trust measures will be objective because they will be inherent in the actual results of the interactions, and will better inform us on how our manipulations affect relation-based trust. Additionally, my previous work on EI highlighted the fragility of current subjective questionnaires. The goals of my future work are therefore to 1) further investigate how social factors influence a person's relational trust in a robot; 2) develop paradigms to measure relational trust objectively; and 3) develop a robust and validated subjective questionnaire specifically looking at a person's relational trust in a robot.

4 CONCLUSION

Understanding relation-based trust in HRI - i.e., trust that a robot will behave as a social agent who complies with societal norms - is growing increasingly important as robots with social goals and purposes are becoming more common. HRI research needs to therefore be directed towards studying factors that affect our relation-based trust in robots, not just how much we trust that a robot is capable of performing a task well. To develop this line of research, we need both systematic, empirical studies of these factors, and the development and deployment of robust, objective paradigms and validated subjective questionnaires to measure trust. While my work has shown one of many possible examples of instances in which factors about how the robot behaves influence people's relation-based trust in the robot, there are many more rich social interactions that have yet to be explored.

ACKNOWLEDGMENTS

This project was funded in part by the National Science Foundation grant IIS-1316809 and the AFOSR grant FA9550-18-1-0465.

REFERENCES

[1] David R Caruso and Peter Salovey. 2004. *The emotionally intelligent manager: How to develop and use the four key emotional skills of leadership.* John Wiley & Sons.

[2] Nicholas Clarke. 2010. Emotional intelligence and its relationship to transformational leadership and key project manager competences. *Project Management Journal* 41, 2 (2010), 5–20.

[3] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors* 53, 5 (2011), 517–527.

[4] Theresa Law, Meia Chita-Tegmark, and Matthias Scheutz. Forthcoming. The interplay between emotional intelligence, trust, and gender in human-robot interaction. *International Journal of Social Robotics* (Forthcoming).

[5] Theresa Law and Matthias Scheutz. Forthcoming. Trust: Recent concepts and evaluations in human-robot interaction. In *Trust in Human-Robot Interaction: Research and applications*, Chang Nam and Joe Lyons (Eds.). Elsevier, San Diego, CA, Chapter 2.

[6] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[7] Roger C Mayer and James H Davis. 1999. The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of applied psychology* 84, 1 (1999), 123.

[8] Raja Parasuraman and Victor Riley. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human factors* 39, 2 (1997), 230–253.

[9] Peter Salovey and John D Mayer. 1990. Emotional intelligence. *Imagination, cognition and personality* 9, 3 (1990), 185–211.

[10] Daniel Ullman and Bertram F Malle. 2018. What does it mean to trust a robot?: Steps toward a multidimensional measure of trust. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 263–264.