Human-Robot Trust: Just a Button Press Away

Daniel Ullman and Bertram F. Malle
Cognitive, Linguistic, and Psychological Sciences
Brown University
Providence, RI, USA
{daniel_ullman, bertram_malle}@brown.edu

ABSTRACT

Many of the benefits promised by human-robot interaction require successful continued interaction between a human and a robot; trust is a key component of such interaction. We investigate whether having a person “in the loop” with a robot—i.e., the mere involvement of a person with a robot—affects human-robot trust. We posited that people who press a button on a robot to permit its plan execution would exhibit greater trust than people who merely observe a robot’s autonomous execution of the same plan. We assessed trust both toward the robot that participants interacted with and toward robots in potential future use contexts. We found (a) a marginally significant and medium-sized effect of the button press on people’s trust in the observed robot ($p = .12$, $d = .52$), and (b) a significant and large-sized effect on people’s trust in potential future robots, but only in social use contexts ($p = .04$, $d = .68$).

Keywords

trust; involvement; human-robot trust; social robotics; human-robot interaction

1. INTRODUCTION

Robots are increasingly used in contexts with complex social interaction, especially in socially assistive applications [1]. Many such applications require repeated interactions and evolving relationships between a human and a robot, where trust is an essential feature. Lee and See define trust as “the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability” [2]. Trust enables two agents to cooperate with each other, in turn yielding potential benefits of such cooperation [3].

A meta-analysis in 2011 identified factors that affect trust in human-robot interaction (HRI), noting at the time a marked lack of research on the human role in human-robot trust [4]. Subsequent work has begun to further investigate this human role. A meta-analysis in 2016 on trust in automation more generally identified user-centric factors that contribute to the development of trust—from human traits and states to cognitive and emotive dimensions [5]. The importance of appropriate trust in robots cannot be emphasized enough; recent work demonstrates the potential danger associated with an overtrust of robots in emergency evacuation scenarios [6].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

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

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

<http://dx.doi.org/10.1145/3029798.3038423>

Figure 1. Side view of Thymio and participant in involved condition.

In the present study we strip away many of the complexities of human-robot interaction to investigate a basic factor that may foster trust: the simple act of pressing a button to permit a robot to execute its task. The power of such simple actions has been illustrated in a game of dice where individuals’ physical act of throwing the dice themselves prompted them to overestimate their probability of success [7]. We posited that participants would exhibit greater human-robot trust after a comparable minimal involvement with a robot.

2. METHOD

A total of 42 adults recruited from Brown University and the Providence, RI community participated in the study. We excluded data from two participants who reported a failure to comprehend the experiment instructions. The mean age of the remaining 40 participants was 22.20 years ($SD = 4.53$). Participants self-reported gender as follows: 16 “Male”; 23 “Female”; 1 “Other.” Participants received \$5 as compensation for an expected 15 minutes of participation time.

We created as simple a test scenario as possible, with a simple robot (Thymio) completing a simple movement task. Participants were told that Thymio is able to generate and execute programs to achieve its goals, and that it has the task of navigating from its starting location to an end location; if Thymio detects an obstacle, it stops its program and then autonomously generates a program for a new path. Participants then experienced the between-subjects manipulation of *interaction type* (autonomous, involved). In the autonomous condition, Thymio executed the new path plan on its own, whereas in the involved condition Thymio waited for a button press from the participant (see Figure 1) before executing the new plan. It then successfully navigated around the object.

Participants then completed two measures of robot trust: *observed-robot trust* and *future-robot trust*. Using 8-point rating scales (from 0=“Not at all” to 7=“Very”), they evaluated the robot

they just saw on the items “trustworthy” and “dependable” (*observed-robot trust*, Cronbach’s $\alpha = .88$). Then participants completed a questionnaire featuring future robots in nonsocial and social use contexts, presented in random order. Participants were instructed to imagine that the future robots have very similar software as the robot they just saw (i.e., it generates a program and executes it on its own or waits for a button press to execute it) and are able to competently complete the particular tasks. Using 8-point rating scales (from 0=“Not at all” to 7=“Very”), participants answered two questions: “How much would you trust the robot in this scenario?” and “How willing would you be to use the robot in this scenario?” (*future-robot trust*).

The eight scenarios featuring future-use robots were selected from a pretested pool of scenarios in order to represent four nonsocial and four social *use contexts*. In the nonsocial contexts, the robot performs a task without direct interaction with a human; in the social contexts, the robot performs a task with direct interaction with a human. For example:

Nonsocial: A robot with very similar software as the robot you saw works in a nuclear reactor facility as an engineer. The robot needs to decide whether to shut down the reactor after it starts to overheat.

Social: A robot with very similar software as the robot you saw works in an airport as a transportation security officer. The robot needs to decide whether to select a suspicious person for a full-body pat-down.

3. RESULTS

There was a marginally significant and medium-sized effect of interaction type on trust in the robot that people interacted with, and a significant and large-sized effect on trust in imagined potential future robots in social use contexts.

On the *observed-robot trust* measure, people in the involved robot condition indicated more trust ($M = 5.53$, $SD = 1.38$) than people in the autonomous robot condition ($M = 4.75$, $SD = 1.59$), $F(1, 36) = 2.61$, $p = .12$, $d = .52$ (Figure 2). Because of the small sample size the study’s observed power was only .35; though the significance level was not below the typical .05 level, the medium-sized effect is notable.

On the *future-robot trust* measure, people in the involved condition also showed greater trust, albeit not in robots in general but specifically in future robots in social use contexts, $F(1, 38) = 4.65$, $p = .04$, $d = .68$ (Figure 3). Whereas participants in the involved condition had as much trust in future *nonsocial* robots ($M = 5.12$, $SD = 1.03$) as participants in the autonomous condition ($M = 4.89$, $SD = 1.25$), participants in the involved condition had relatively higher trust in *social* robots ($M = 4.57$, $SD = 1.12$) than participants in the autonomous condition ($M = 3.61$, $SD = 1.56$).

One possible explanation for these findings relates back to [7], such that simple involvement may prompt a greater sense of control in a given context, thereby allowing trust to develop.

4. CONCLUSION

Despite a small sample, the sizeable effect of simple involvement on trust is promising. Permitting a robot’s plan execution seems to lead to greater trust in the specific robot at hand and in future robots, at least in social use contexts. Follow-up studies are needed with other robots, various modes of involvement, and repeated interactions. These findings offer preliminary evidence that having a human “in the loop,” even via the mere press of a button, fosters human-robot trust.

Figure 2. Effect of Interaction Type on Observed-Robot Trust. Error bars = SE.

Figure 3. Interaction of Interaction Type and Use Context on Future-Robot Trust. Error bars = SE.

5. ACKNOWLEDGMENTS

The authors thank Oriel FeldmanHall, Stefanie Tellex, Stuti Thapa Magar, and Elizabeth Phillips for their contributions to this project. This work is supported by Office of Naval Research grant #N00014-14-1-0144, by the Brown University Humanity-Centered Robotics Initiative, and by HRI Pioneers. Daniel Ullman is supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program.

6. REFERENCES

- [1] Feil-Seifer, D., & Mataric, M. J. (2005). Defining socially assistive robotics. *Proceedings of ICORR*. Chicago, Illinois, June 28-July 1.
- [2] Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46, 50-80.
- [3] Gambetta, D. (1988). *Trust: Making and breaking cooperative relations*. Oxford, UK: Basil Blackwell.
- [4] Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53, 517-527.
- [5] Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors*, 58, 377-400.
- [6] Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). Overtrust of robots in emergency evacuation scenarios. *Proceedings of HRI*. Christchurch, New Zealand, March 7-10.
- [7] Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32, 311-328.