Trust and Autonomous Systems: Papers from the 2013 AAAI Spring Symposium

Trust During Robot-Assisted Navigation

Erika Mason[1], Anusha Nagabandi[2], Aaron Steinfeld[3], Christian Bruggeman[3]
[1]Rochester Institute of Technology, Dept. of Electrical and Microelectronic Engin., 77 Lomb Memorial Drive, Rochester, NY 14623
eam4947@rit.edu
[2]University of Illinois at Urbana Champaign, Dept. Electrical and Computer Engineering, 1406 W. Green St., Urbana, IL 61801
anusha.nagabandi@gmail.com
[3]Carnegie Mellon University, School of Computer Science, 5000 Forbes Ave, Pittsburgh, PA 15213
steinfeld@cmu.edu, cgbruggeman@gmail.com

Abstract

Robotics is becoming more integrated into society and small user-friendly robots are becoming more common in office spaces and homes. This increases the importance of trust in human-robot interaction, which is essential to understand in order to design systems that foster appropriate levels of trust. Too much or not enough trust in a robotic system can lead to inefficiencies, risks, and other damages. The robot in this experiment was used as a navigational system to guide a participant through an arrow maze. This experiment examined human trust in robots, the decision between doing a task or relying on a robot, and inconsistencies between human awareness and robot guidance.

Introduction

Trust is becoming increasingly relevant in the field of human-robot Interaction. Knowing the factors that contribute to a person's trust in small robots will define how robot operating systems are designed as they are further integrated into more personal, home and office settings. In Desai et al (2012) on creating trustworthy robots, the main focus was to collect data on how people perceive the independence of automated systems, which was used to obtain an informed view of how robots elicit appropriate levels of trust from humans. As with many other papers, the rationale is based on the consequences of placing either too little or too much trust in an automated robotic system (Desai et al. 2009; Lee and See 2004). Too little trust in a robot may lead to poor judgment in situations such as maintaining a safe distance from other objects, properly perceiving signs, or executing tasks. This lack of trust can lead to humans opting for manual control when autonomy may be easier and more efficient. If a person places too much trust in a robot, they may not be able to recognize when the robot is performing poorly, which could place the user or bystanders into dangerous situations.

A recent study by team members and colleagues explored this issue in depth (Desai et al. 2012). Participants were asked to complete a slalom course as quickly as possible without hitting any obstacles, while passing gates on specific sides and completing other secondary tasks. Participants could switch between the autonomous mode, where the robot drives itself, and the shared mode, where the participant drives with some safety guarding. The robot was intentionally programmed to sometimes turn in the wrong direction which allowed examination of how reliability impacted mode choice and trust.

Lee and See (2004) examined the relationship between autonomy and trust. "It considers how the context, automation characteristics, and cognitive processes affect the appropriateness of trust." This serves as a reference to how trust can be measured, analyzed, and built to help integrate automated robots into society. This is important because, "the context in which the automation is used influences automation performance" (Lee and See 2004). This demonstrates that while the technology behind computers and robots is very important, it cannot completely be used as intended without proper user context and trust.

Using the Desai et al (2012) experiment as a basis, the study in this paper also examined how people respond to a robot that does not work perfectly. However, the user is co-located with the robot in this study, the robot is only driven manually, and the robot provides directional advice. The research questions were:

- Will a person trust the directions given by a robot over their own judgment and perception?
- Does the user's general trust in robots affect the way they approach the task?
- How do users respond to directions given by robots?
- How does a mistake by a robot impact the user's trust?
- How does trust change over time?

Methods

This experiment was designed to look at factors in HRI that deal with trust in smaller and non-threatening robots. The robot used was the iRobot Turtlebot, which is built on a Create platform and has a Kinect sensor (Figure 1). The task in this study was to navigate an arrow maze (Figure 2) as quickly as possible using the robot as a guide. The payment scheme emphasized time so that participants would focus on efficiency. The participant began and ended in the center of the maze and could only travel in the direction of each link's arrow (all links were one-way). The mazes were constructed with tape on large tarps. This allowed multiple, rapidly deployed mazes that could be rotated to give the participant another perspective of the same maze. This allowed stimuli variation within a single experiment.

Participants drove the robot with a Playstation 3 gamepad while walking alongside. The robot assisted by displaying recommended directions on a tablet screen (Figure 3, top of Figure 1). In reality, this was done via Wizard of Oz. There were two buttons on the tablet screen labeled "recalculate" and "re-do." "Recalculate" allowed the participant to move back to the prior node, while "re-do" allowed the participant to restart the maze from the center.

The robot assistance was similar to an in-vehicle navigation system suggesting a route to the driver. Likewise, the participant was not required to follow the directions given. Recommended directions were given manually from an experimenter laptop, but participants were told that the robot was autonomously giving directions. This was important because we wanted the participant to believe the robot was perceiving the maze and giving directions.

Participants were exposed to 3 reliability levels. The most reliable level (A), guided the participant through the shortest and most efficient path of the maze. Reliability B took the participant through a long route. Reliability C guided the participant the wrong way, to a dead end or a never-ending loop, before recalculating and advising a correct and short path. The robot initially appeared to suggest a good route, but the experimenter subsequently altered the route to achieve the designated reliability behavior.

Participant were given an initial trial run to see the robot, practice driving it through a maze, learn to interpret advice, and practice with interface. To further convince the participant that the robot was perceiving the maze, a non-functional, forward facing camera was mounted on the robot and participants watched the robot "scan" the maze before the run. Before the trial run, the participants were asked to complete a pre-experiment questionnaire with basic demographic questions about their experience with robots, puzzles, and navigational systems. This was used to examine how trust was impacted by prior experience. During the study, the participants completed a post-run questionnaire after each of four runs in order to capture opinion changes over time.

The questions asked after each run were:

- How many times did you not follow the Turtlebot?
- To what extent can you trust that the Turtlebot will give you directions that will help you get through the maze most efficiently?
- How has your trust changed since the last run?
- How much do you trust the Turtlebot in general?

The sequence of three different mazes was counterbalanced for the first three runs. The maze for the fourth run repeated the maze of the run that had a reliability of "B." This pattern for the maze numbers was repeated four times to make a sequence for each of the 12 participants.

The reliability ordering for the four runs was designed slightly differently. The first run and last run for each participant was reliability "A." This was intentional since we wanted to every participant to experience a good, bad, good sequence so we could examine a U-shaped experience. Half of the 6 participants experienced B then C for their middle runs, while the other half experienced the reversed order.

When the experiment was complete, participants were asked to complete an exit questionnaire. The post-experiment questions were:

1. Would you use the Turtlebot again?
2. Would you trust the Turtlebot to direct you in other areas that have already been mapped out?
3. Would you be comfortable having a robot like the

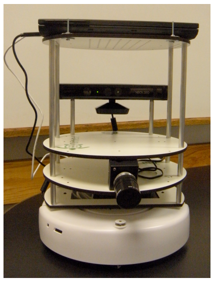

Figure 1. The Turtlebot with the Create base, the XBOX 360 Kinect, a small camera, and a laptop.

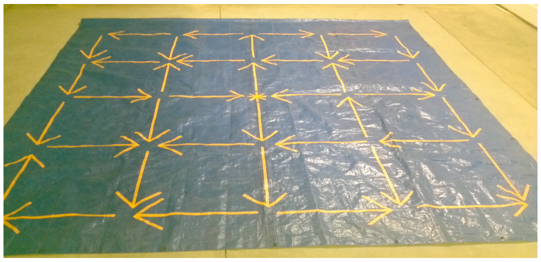

Figure 2. An arrow maze on a blue tarp. All mazes were 10' x 10' squares.

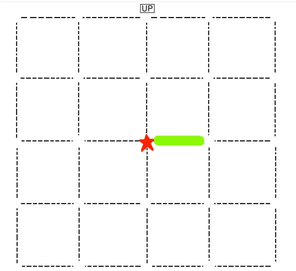

Figure 3. Interface shown on robot tablet screen. The green line indicated the advised path of travel.

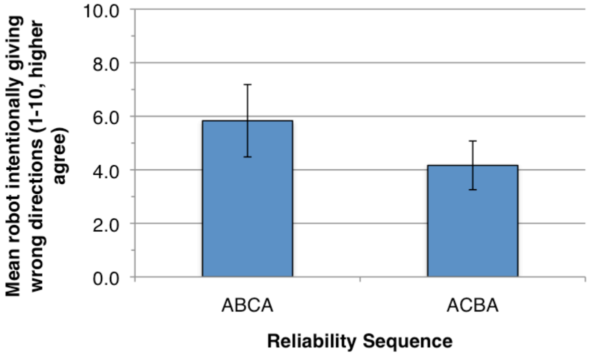

Figure 4. Intentionally wrong directions

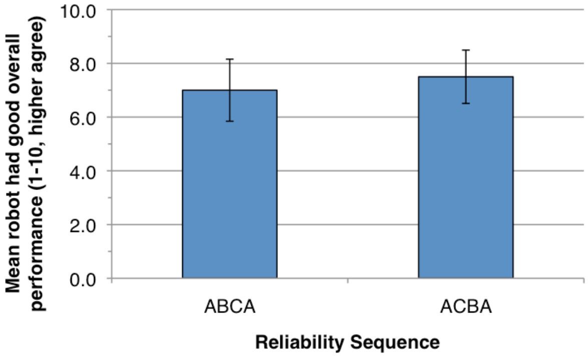

Figure 5. Perceived robot performance

Turtlebot navigating near you autonomously?

4. Would you be comfortable having the Turtlebot in your home or office?
5. How was the robot's overall performance?
6. How was the Turtlebot's overall perception?
7. Is this robot dependable?
8. Did you feel that this robot was intentionally giving you the wrong directions?
9. Do you trust this robot?
10. Do you trust robots in general?

The mazes were simple enough to be solved without the robot's assistance, thus allowing the participant to detect bad advice. However, it was assumed that some participants would not bother to look ahead and see if the robot was misleading them, thus measuring trust within a decision to do a task manually versus relying on a robot.

Results

Initial Trust

Each participant's initial trust in robots generally affected the way in which they approached the assigned tasks in the experiment. During the first run, all but two of the participants readily followed the robot's directions. This strong initial trust was reinforced by most of the first run's trust ratings, which were above 7 (on a scale of 1 to 10). For the two participants that started out by attempting to find their own way, one participant had experience with robots and directly remarked that he did not trust it because he assumed that we had programmed the robot to give false directions. Some of the participants made a mistake because they did not fully understand the orientation, but subsequently understood the interface after a quick explanation.

There was a significant negative correlation between a participant's age and their trust in robots in general (-0.61, p=0.034). There was also a significant positive correlation between the participant's trust in robots in general and trust in the robot by the end of the run (0.84, p<0.001).

Following Directions vs. Using Own Judgment

During the study, approximately 7% of the directions given were not followed by the participant (38 out of 514). This is low due to a nuance in the protocol. Once a participant strayed from the recommended path, the directions were modified to accommodate the new location of the robot and guide the participant from that point forward. However, this finding still shows that participants followed directions throughout most of the experiment.

For participants exposed to the reliability sequence ACBA, all but one participant faithfully followed the instructions for the reliability C run, where the robot led them to a dead end. This is likely due to the participants' high trust in the robot after the first run. These participants' subsequent run (B) showed more participants straying from the directions. However, people in the other order (ABCA) appeared to feel the robot was giving incorrect directions than for the ones with this order (Figure 4). While this was not significant, a least significant number analysis suggests this difference would reach significance with additional participants (47 samples).

This last finding is counter to our initial prediction. We expected that it would be harder for participants to detect bad advice for reliability B (long route) than for C (incorrect route). Looking back at our detailed notes we think the routing advice may have led to a bias. The C error led participants closer to the center of the maze before leading them into a dead end, so the participants may have inherently felt they were getting good advice. However, B initially led participants further from the center, so most participants decided to try to shorten their distance by cutting a corner of the path or looking for an alternate route. It is apparent that, compared to the ACBA sequence, more people thought that robot was intentionally giving incorrect directions during the ABCA sequence. Regardless of the se-

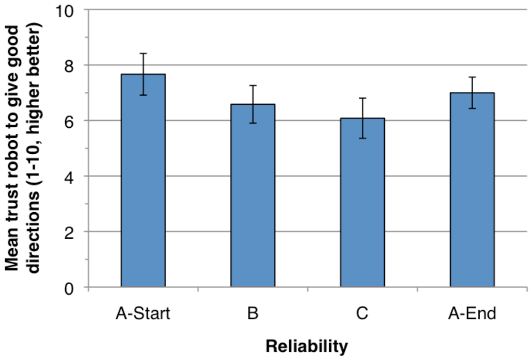

Figure 6. Trust ratings

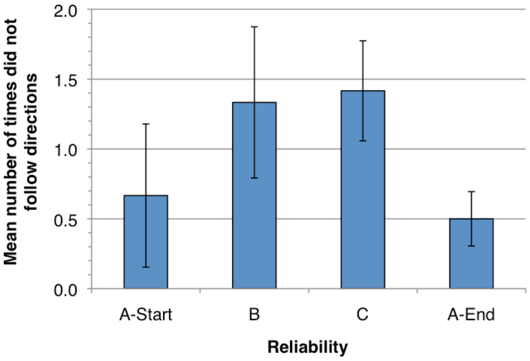

Figure 7. Number of times not following directions

quence, it is clear the participants did not view overall robot performance differently (Figure 5).

Recovering Trust

Whether the participant recognized a shorter path as the robot was taking them on a long path, or whether the robot led them to a dead end, the low reliability led to a decrease in the participants' trust. Trust fell during low reliability (B and C, Figure 6). Likewise, the number of times that participants did not follow the robot increased during low reliability (Figure 7). While neither result was significant, the combined evidence suggests a change in trust during low reliability. This is similar to drops in trust during low reliability found in related studies (e.g., Desai et al 2012).

However, trust did recover by the end of the final run for the second reliability A. Both trust and the number of followed directions moved towards the reliability A run at the beginning of the experiment. However, mean trust ratings did not dip below the medium range. This could be due to the fact that it was generally easier to let the robot lead rather than look ahead for quicker routes. While the advice during the final run was optimal, it is still notable that the robot's earlier mistakes resulted in no trust rating of 10. The multiple 10 ratings seen for the initial reliability A were not repeated.

By the end of the experiment, most people trusted the robot more than before (Figure 8). A mistake by the robot led to the participants disobeying the directions more often and an overall decrease in their trust, but their trust was reestablished once they saw the robot function correctly in the final reliability A run.

Also, the post-experiment questions on trust in the robot, robot dependability, good perception, good performance, and the inverse of whether the robot was intentionally appear to have potential as an index of positive attitude towards robots. A Chronbach's alpha analysis of these four questions resulted in 0.82, which is above the typically accepted 0.7 threshold for survey indices. However, this index did not show significant differences across the two different reliability sequences. In the future, it would be interesting to ask these questions after each run instead of at the end of the study.

Discussion

The main finding for this study is that participants generally trusted the directions given by the robot more than their own judgment. There appeared to be three core reasons for not following the advice and decreasing trust when they opted to trust their own judgment more than the robot. One reason was simply that the participant saw a slightly different path that the one intended, by either cutting a corner or finding an alternative path. The second was that the participants saw the robot miscalculate in the previous run and was more doubtful in a subsequent run. The third was that the participants intentionally looked for a path on their own because they thought that the experiment was designed to intentionally lead them on an incorrect path. One participant explicitly told us that he thought we programmed the robot to be incorrect. This participant was, more than likely, biased by previous experience with robots. Others with doubt in the robot were simply suspicious about the experiment because they likely sensed there was more of a purpose than just to observe maze driving. Overall, unless the robot had intentionally made a mistake or the participant had suspicions about the nature of our experiment, most of the people tended to rely on the robot for directions.

The overwhelming recovery of trust is likely tied to familiarity, which is known to foster trust. For example, an autonomous car driving in the lane next to you or parking in the spot next to your car would likely make you nervous or uncomfortable due to your lack of familiarity with autonomous vehicles (e.g., Desai et al 2009). Once humans are introduced to a new, well-performing technology and have time to adjust, they tend to acquire trust and confi-

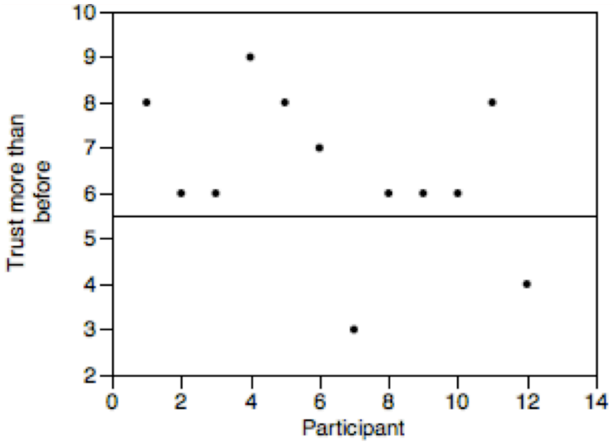

Figure 8. Perceived change in trust by participant number (1-10, lower to higher)

dence in the system over time. People know there is a possibility of error, but build an appreciation and sense of value in the system. The closest analogy to this work is in-vehicle navigation systems. These are not always the most reliable systems, but they are usually more efficient and accurate than planning and following a route manually. Additional features, like identifying nearby gas stations or dining options, can increase value and enhance reliance and trust.

A sense of control is also important. Airplanes are a good example of this effect. There are a number of issues that can arise, leading to crashes and death, which capture the public's attention and heighten concern over air travel. However, awareness of advances in technology, statistically safe performance, and expressions of confidence by peers and experts can offset such worries. The inclusion of highly trained pilots is especially important. As with in-vehicle navigation systems, humans feel more comfortable with autonomy when they, or their proxy, have the power to intervene in the event of a malfunction. This sense of control was also visible in our study, as users could ignore robot advice. We suspect participants would have reported much lower trust if they were not allowed to disregard robot advice. It would be interesting to test this hypothesis in a future study.

Familiarity is another reason that people can be hesitant with newer technology. Lack of familiarity with interfaces and mental models can foment distrust. For example, many older adults express distrust in smartphones due to confusion and lack of familiarity. Similar effects can occur with in-vehicle navigation systems, especially for new users. This behavior was also seen in this experiment; Participant 9 went the wrong way because the task of driving properly was emphasized over following directions. The participant proceeded to give low trust ratings saying that the robot was heading in the wrong direction, when in fact, it was advising the most efficient route.

Familiarity can also work against the robot. For example, we repeated a maze for each participant and one participant recognized the map. They remembered the long path from their B run during the final A run. The participant instantly took the longer path even though the robot would have recommended a shorter path.

In general, deciding between doing a task manually or relying on a robot is impacted by the nature of the task. If users believe that it is easy to do a task by themselves, they will usually not feel the need to rely on a robot. All of the participants of the experiment said that they felt comfortable around the robot, but when asked if they would use the robot again or want it in their own home or office, a few said no simply because they did not "need" the robot. There was awareness that the robot could have value in another setting. One participant stated that the robot would be useful for someone with a disability. If, for instance, the person could not see, the robot could be equipped to provide guidance through unfamiliar places. In fact, this functionality is a stated need within the blind community (e.g., Morton & Yousuf 2011).

The concept of users doing a task versus relying on a robot for assistance can be a blurry line. For example, the concept of joint control was recently explored in a study by Dragan and Srinivasa (2012). In this experiment, the input from the user's arm motion combined with internal path planning to help a robot reach towards and grasp an object. When the task was difficult, like an object being in a less accommodating position, the robot entered an autonomous mode and moved itself directly toward the object. Most people preferred this mode, commenting that, "they appreciated the precision of the autonomy." On the easy tasks, however, "opinions were split" because when they were not in the autonomous mode, "they felt more in control of the robot." The users also commented that after seeing the robot be wrong, they did not trust a more aggressive mode where robot is in full control because they could not intervene and ensure the robot did the correct thing. This suggests that although autonomy and technological advances can make tasks easier, matters of correctness, behavior, and trust must be addressed for effective usability.

As explained in Buxton and Sniderman (1980), issues such as user trust in any system are important, "In attempting to design a system to 'fit' the end user, behavioral issues must be considered and understood." They follow the premise that "if technologically-based tools are adopted, it is because the scope or magnitude of the user's problems have outgrown current techniques for dealing with them." Therefore, adopting the new tool must "eliminate problems, rather than create additional, or alternative ones." Also, "the new tool must adapt to the user, rather than force the user to adapt to it (which is too often the case today)." Since it is unlikely that the first implementation of any user interface is going to function as well as it could and suit the users, Buxton and Sniderman (1980) also suggest "suitable prototyping tools, techniques of observation, and methods of evaluation."

Even the most advanced, cutting-edge technology offers little to users if it not developed and designed in a way that elicits the correct level of comfort and trust to the user. This study explores participants' interaction with robots, their decisions between doing something themselves or relying on the robot, and how trust is impacted by reliability and familiarity. As expected, participants regained trust after system failures and finished the study appreciating the robot. Likewise, participants were generally willing to let the robot assume authority and rely on the robot to guide them through the maze. While this is promising for the future of robotics, it also forces robot developers to recognize the importance of high autonomy performance due to the potential for end users to place too much trust in robots.

Acknowledgments

This work has been supported in part by the National Science Foundation (IIS-0905228 and IIS-0905148). Funding for EM was provided through the Quality of Life Technology ERC REU program (EEEC-0540865). Thanks to Poornima Kaniarasua (CMU) and members of the UMass Lowell Robotics Laboratory for their assistance with robot and experiment logistics.

References

Buxton, W. & Sniderman, R. (1980). Iteration in the design of the human-computer interface. In *Proceedings of the 13th Annual Meeting, Human Factors Association of Canada*, 72-81.

Desai, M., Medvedev, M., Vázquez, M., McSheehy, S., Gadea-Omelchenko, S., Bruggeman, C., Steinfeld, A., and Yanco, H. 2012. Effects of changing reliability on trust of robot systems. In *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM.

Desai, M.; Stubbs, K.; Steinfeld, A.; and Yanco, H. (2009). Creating trustworthy robots: lessons and inspirations from automated systems. In *Proceedings of the AISB Convention: New Frontiers in Human-Robot Interaction*.

Dragan, A. D., and Srinivasa, S. S. (2012). Assistive teleoperation for manipulation tasks. In *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction,* 123-124. ACM.

Lee, J., and See, K. 2004. Trust in automation: designing for appropriate reliance, *Human Factors*, 46:50–80.

Morton, T., and Yousuf, M. (2011). *Technological Innovations in Transportation for People with Disabilities Workshop Summary Report, FHWA-HRT-11-041*. Office of Operations Research and Development, Federal Highway Administration, U.S. Department of Transportation.