Trusting a Robot as a User Versus as a Teammate

Paul Robinette, Michael Novitzky, Michael R. Benjamin
Laboratory for Autonomous Marine Sensing Systems (LAMSS)
Department of Mechanical Engineering
Massachusetts Institute of Technology, Cambridge, MA, USA
paulrobi@mit.edu, novitzky@mit.edu, mikerb@mit.edu

***Abstract*—The human's role in human-robot interaction likely affects their trust in the robot. For example, prior work has shown that humans who are expected to use features of an autonomous robot tend to be more trusting than humans who operate a robot. In this paper, we discuss this prior work and its extensions into human-robot teaming. Our test environment is an on-water game of capture the flag where humans in motorized kayaks work with autonomous surface vehicles against a similar opposing team.**

I. Introduction

The factors that affect a human's decision to trust a robot change due to the environment and the human's role in the interaction, among others. In some cases, the prior performance of the robot may be a factor in this decision, in others it is ignored. In some roles, a human is more critical of the robot's abilities and less likely to trust it. The study of human-robot trust has found several examples of situations that tend to produce trusting or not trusting behavior in humans, but the field is still largely unexplored. The trust a human operator places in a robot, for example, seems to be highly dependent on the robot's prior performance [1]. In contrast, a participant in an experiment who is expected to use features of or information from a robot will almost always choose to trust the robot [5, 3]. Results from experiments that place the human in a similar role but use a virtual simulation of the environment; however, show that participants do, sometimes, base their decision to trust on the robot's prior behavior [4].

We consider two roles for a human while interacting with the robot in a situation that requires trust: user and teammate. In the user role, the human can choose to use some feature of the robot. The decision to use that feature indicates that they trust it to perform the action or provide the information correctly. The decision to not use the feature indicates that they do not trust the robot in this instance. In the teammate role, humans and robots are on a team with a shared goal. The humans will encounter many situations in the extended interaction that require them to trust or not trust a robot. In general, we define trust in these situations as the decision, by a human, to allow a robot to perform an action that has risk for the human. In our two examples below, the risks are in providing correct information to the human in an emergency and performing well on a team during a game.

II. User Trust in Emergencies

Prior work has found that people tend to trust directions from robots in real-world simulations of fire emergencies, regardless of the robot's prior behavior [3]. Participants were first asked to follow a robot guide through an unfamiliar building. In some conditions, the robot took an efficient route to the participant's destination. In others, it took a circuitous route, indicating that it did not know the way. After the participant reached their destination and was briefly distracted, artificial smoke and fire alarms were used to simulate a fire emergency. Participants left the room they were in, presumably to find safety. They encountered the robot pointing them to a new, unexplored hallway (Figure 1). Participants could choose to follow guidance from the robot or follow their own knowledge of the building's exit. All participants in either of the two conditions above chose to follow the robot.

Fig. 1. Emergency guidance robot pointing to previously unexplored exit.

Interestingly, a similar experiment in a virtual simulated environment of an emergency found different results [4]. In this experiment, participants tended to trust a robot that performed well in a prior interaction but not trust it if it performed poorly. It seemed that participants were able to make a rational choice, based on prior experience, in the calm environment of a computer simulation, but took the first option that presented itself in the real-world, physical environment.

In these two situations, participants were expected to use the robot to help them find safety. Efforts were made to ensure that participants saw the use of the robot as a choice: another, seemingly-good option was always offered. Thus, choosing to use the robot's information indicated that they trusted the robot. Participants in the virtual and physical experiments

Fig. 2. Human teammate in a motorized kayak and an autonomous robot teammate.

had the same role (potential robot user) but made different decisions in that role due to their environment.

III. Teammate Trust

In contrast to user trust in emergencies, teammate trust requires that humans and robots work together in a sustained interaction towards a common goal. In this situation, the humans and robots could be peers, each offering suggestions to the other to coordinate actions, but neither in control of the other. Conversely, one individual could be designated the team captain, and have a supervisory role over the others. To explore the different trust relationships between these roles, we developed a testbed (the Aquaticus project) where teams of two humans and two robots each play a game (capture the flag) against each other. The game is played on the water, in front of the MIT Sailing Pavilion. Humans pilot motorized kayaks (Mokai ES-Kapes) and work with autonomous surface vehicles (Clearpath M300 Herons) to defend their flag and attack the other team's flag (Figure 2). Humans communicate with their teammates through speech and robots respond using text-to-speech. Robots communicate over a wireless network.

The basic speech interaction framework leveraged in this game has already been evaluated in a previous effort to determine the effectiveness of human-robot teaming [2]. Participants were asked to pilot a motorized kayak while instructing their teammate to approach a buoy. Their teammate could either be a robot using voice commands or a human experimenter restricted to the same command and response phrases as the robot. Participants generally rated the robots favorably, but tended to prefer human teammates.

There are many ways to measure trust in this framework. We use numerous post-experiment surveys (notably the Shaefer Trust Scale [6]) to determine participant trust in the robot as a whole. Additionally, we record communications during the game. This allows us to determine if the human teammates are constantly asking their robot teammates for status, if the humans accept suggestions (or commands) given by robots, and the level of autonomy given to the robot by humans (e.g. does the human try to micromanage the robot, or allow it to decide its own objectives). We also monitor how often the human checks on the robot's status, both by recording explicit requests for status updates and by recording video that shows where the human is looking throughout the game. A teammate who is always looking over his or her shoulder at a robot most likely does not trust that robot.

In addition to trust, we also estimate human cognitive workload during our tests. We plan to correlate workload to trust and hope to affect both by modifying the robot teammate's behavior. To measure workload, we examine heart rate and head movements. A high heart rate and frantic scanning of the environment mean that the operator is likely very stressed, whereas a low heart rate and directed gaze indicate low stress.

IV. Ongoing Work

Currently, we are developing experiments with our human-robot teaming testbed for our Aquaticus project to determine the factors that produce trusting and not trusting behavior in humans. Our goal in this effort is to develop techniques for human-robot teaming that can be applied in other domains. To this end, we intend to explore the effect of roles, autonomy, and communication on the effectiveness of teams. One of the properties we use to identify an effective team is the level of trust that a human places in a robot (another is the cognitive workload that is required of the human teammates). We are performing experiments that place robots and humans in a number of different roles: as peers, with a human team captain, with a robot team captain, and assigning a robot wingman to each human (or vice-versa). The level of autonomy may also effect trust. Changing the robot's level of autonomy (e.g. from a basic, waypoint-following mode to an autonomous, strategizing mode) may have an effect on trust, but will likely have an effect on human cognitive workload. Finally, the amount of communication required and allowed to coordinate with robots will likely have an effect on the trust that a human places in the robot.

Acknowledgments

The authors would like to thank the MIT Sailing Pavilion staff for their professionalism and support in designing and maintaining our safety protocol. This work was supported by DARPA TTO, ONR 311, and Battelle.

References

[1] Munjal Desai, Mikhail Medvedev, Marynel Vázquez, Sean McSheehy, Sofia Gadea-Omelchenko, Christian Bruggeman, Aaron Steinfeld, and Holly Yanco. Effects of changing reliability on trust of robot systems. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, pages 73–80. IEEE, 2012.

[2] Michael Novitzky, Hugh RR Dougherty, and Michael R Benjamin. A Human-Robot Speech Interface for an Autonomous Marine Teammate. In *International Conference on Social Robotics*, pages 513–520. Springer, 2016.

[3] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M Howard, and Alan R Wagner. Overtrust of robots in emergency evacuation scenarios. In *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*, pages 101–108. IEEE, 2016.

[4] Paul Robinette, Ayanna M Howard, and Alan R Wagner. Effect of robot performance on human–robot trust in time-critical situations. *IEEE Transactions on Human-Machine Systems*, 2017.

[5] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 141–148. ACM, 2015.

[6] Kristin E. Schaefer. *The Perception and Measuerment of Human-Robot Trust*. PhD thesis, University of Central Florida, 8 2013.