Trust Development and Repair in Human-Robot Teams

de Visser, Ewart Jan; Pak, Richard; Neerincx, Mark A.

DOI
10.1145/3029798.3038409

Publication date
2017

Document Version
Final published version

Published in
HRI'17

Trust Development and Repair in Human-Robot Teams

Ewart J. de Visser
George Mason University
4400 University Drive, 3F5
Fairfax, VA 22030
+1.910.200.8596
edevisse@gmu.edu

Richard Pak
Clemson University
418 Brackett Hall
Clemson, SC 29634
+1.864.656.1584
richpak@clemson.edu

Mark A. Neerincx
Delft University of Technology
Mekelweg 4
2628 CD Delft
+31.15.278.7141
M.A.Neerincx@tudelft.nl

ABSTRACT

This paper is a call to the field of human-robot interaction to focus research efforts on the development and repair of trust within human-robot teams. To guide this effort, this paper describes the initial development of a framework for trust development in human-robot interaction research with a specific focus on trust repair. This framework identifies several unique trust-dyads within an example domain of Urban Search and Rescue Operations (USAR) that are suitable relationships for study. We conclude with several areas of research that should be addressed under a trust repair framework including trust measurement, model development and validation, mutual dynamic trust calibration, and long term trust development.

Keywords

Trust in autonomy; trust repair; long-term human-robot relationships; mixed-initiative teaming.

1. INTRODUCTION

Autonomous capabilities of robots are increasing daily, causing the eventual rise of true human-autonomy teams. Relationships between robots and people within these teams may mimic those between humans [1] and face similar challenges. For instance, autonomous robots may not perform perfectly, fail to meet expectations, or have goals that are miss-aligned with human team members. Surprises may occur on a regular basis causing potential threats to maintaining a productive trust relationship. Thus, there is a need to design and equip robots with abilities to repair trust when it is broken.

Currently, there are no frameworks that describe the development and repair of trust within human-robot teams. Some have suggested that to characterize and understand these mixed-initiative teams, research in human-human teams can be used as a template [2], [3]. Several human-human studies have examined which methods are most effective for repairing trust. This research has focused primarily on the role of apologies from the transgressor's perspective [4] and predictors that stimulate forgiveness from the victim's perspective such as likeability, blame attribution, probability of future violations, and generating empathy [5].

As a thought experiment, we use the urban search and rescue domain to think about how trust repair situations may develop for different trust-dyads within that team context. This will allow for scenario development and a prediction of future research issues.

HRI '17 Companion, March 06-09, 2017, Vienna, Austria
ACM 978-1-4503-4885-0/17/03.
http://dx.doi.org/10.1145/3029798.3038409

2. TRUST DEVEVELOPMENT IN USAR

2.1 Human-Robot Teams

The USAR domain is an active area of research and development. Robots have been critical in assisting rescue workers to retrieve victims. Operators control robots remotely, communicate with commanders on the ground, and sometimes even communicate with victims through robots [6].

2.2 Trust-Dyads

Several trust dyads can be identified that are unique to a USAR scenario. Figure 1 shows the people involved with robots within a typical USAR hierarchical organizational [7]. Each relationship represents a unique interaction in which trust can develop, break down, or repaired.

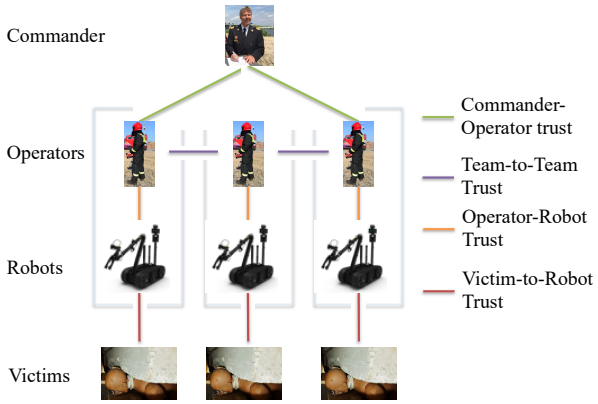

Figure 1. Initial framework of human-robot trust development.

2.2.1 Operator-Robot Trust

With increases in levels of autonomy and number of vehicles to manage, trust will play a more critical role and will mediate the relationship between robot and operator. Breakdowns in trust can possibly be restored by increasing the transparency of a system, providing trust cues, or situation-specific training to operators.

2.2.2 Commander-Operator Trust

Less is known about the trust relationship between the commander of a USAR team and the operator controlling the vehicles. Trust in both the operator's ability and trust in the various robotic platforms play a role in this setting. Some unique issues can occur that affect the trust relationship such as a delay in communications. Trust repair may be challenging when it is unclear if the robot or the operator is to blame for sub-optimal performance.

2.2.3 Team-to-Team Trust

A distinguishing feature of a good team is the ability to develop and sustain report between teammates [3]. It is not clear how autonomous entities affect this type of report building. It may be

that these teams have varying trust dynamics analogous to human-dog teams or management teams.

2.2.4 Victim-Robot Trust

This relationship involves a robot that is actively interacting with a victim for an extended period of time. It differs from the first trust dyad in that the human is not an expert and does not control the robot. A victim encountered by a robot may be in a highly vulnerable state both physically and psychologically. Trust development in this situation can reveal information to the robot that may expedite the rescue process and the survivability of the victim.

2.3 Challenges for Human-Robot Trust Development

In addition to the challenges identified in each of the trust dyads, three additional overarching challenges require specific mention.

2.3.1 Mutual Dynamic Trust Calibration

Robots of the future will have many abilities to adjust their behavior towards human partners. This would facilitate *mutually adaptive trust calibration,* or the re-calibration of trust of both the robot and the human partner. Previous work has indicated the need for describing the adaptation of the robot to the human as well as the human to the robot [8], [9].

2.3.2 Long-term Team Trust Development

Robots will be integrated in human-teams. Of great interest is how these relationships change over an extended period of time [10]. For example, the trust dynamics between victim and robots is much shorter and immediate than between the operator and robot. Trust dynamics between humans also develop differently because human team members function in a social organizational structure.

2.3.3 Interactions between Trust Dyads

Lastly, trust between each of the dyads will also affect each other. For instance, a breakdown in operator-robot trust may affect commander-operator trust. These types of interactions are unique and should be explored in more detail.

3. Recommendations for Future Research

We propose several methodologies and approaches that HRI researchers can use to start research in this area [10].

3.1 Field Exercises and SME Interviews

Field exercises with robots often reveal unique insights into human-robot collaboration. HRI researchers should use field exercises to increase a sense of risk and realism to their studies, essential for the emergence of trust relationships.

3.2 Ontology and Model Development

There is a great need to develop a method that allows for reasoning about humans and robots alike. We are developing such an ontology. This ontology defines the vocabulary and semantics of the multimodal human-robot communication and is the basis of the reasoning of the robot agent [11]. Furthermore, a computational model of trust development and repair that makes quantitative predictions would help to focus and integrate research within the HRI community.

3.3 Trust Measurement Development

Trust measurement remains a key issue for human robot interaction [12], [13]. It is important to know how trust is initiated, how it develops, how it breaks down, and how it recovers. This requires a convergence of behavioral, subjective, and physiological measures and correlates [13].

4. REFERENCES

[1] C. Nass, B. Fogg, and Y. Moon, "Can computers be teammates?," *Int. J. Hum. Comput. Stud.*, vol. 45, pp. 669–678, 1996.

[2] E. J. de Visser, R. Parasuraman, A. Freedy, E. Freedy, and G. Weltman, "A Comprehensive Methodology for Assessing Human-Robot Team Performance for Use in Training and Simulation," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 50, no. 25, pp. 2639–2643, Oct. 2006.

[3] E. Salas, D. E. Sims, and C. S. Burke, "Is there a 'Big Five' in Teamwork?," *Small Gr. Res.*, vol. 36, no. 5, pp. 555–599, Oct. 2005.

[4] K. T. Dirks, P. Kim, D. L. Ferrin, and C. D. Cooper, "Understanding the effects of substantive responses on trust following a transgression," *Organ. Behav. Hum. Decis. Process.*, vol. 114, no. 2, pp. 87–103, 2011.

[5] E. Tomlinson, B. R. Dineen, and R. J. Lewicki, "The road to reconciliation: Antecedents of victim willingness to reconcile following a broken promise," *J. Manage.*, vol. 30, no. 2, pp. 165–187, 2004.

[6] V. Srinivasan, C. L. Bethel, and R. R. Murphy., "Evaluation of head gaze loosely synchronized with real-time synthetic speech for social robots," *IEEE Trans. Human-Machine Syst.*, vol. 44, no. 6, pp. 767–778, 2014.

[7] J. De Greeff, K. Hindriks, M. A. Neerincx, and I. Kruijff-Korbayova, "Human-robot teamwork in USAR environments: the TRADR project," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, 2015, pp. 151–152.

[8] J. E. Mercado, M. A. Rupp, J. Y. C. Chen, M. J. Barnes, D. Barber, and K. Procci, "Intelligent Agent Transparency in Human-Agent Teaming for Multi-UxV Management.," *Hum. Factors*, vol. 58, no. 3, pp. 401–15, May 2016.

[9] J. Y. C. Chen and M. J. Barnes, "Human–Agent Teaming for Multirobot Control: A Review of Human Factors Issues," *IEEE Trans. Human-Machine Syst.*, vol. 44, no. 1, pp. 13–29, Feb. 2014.

[10] R. Looije, T. Mioch, and J. Van Erp, "Modeling relationships in human-robot teams and their performance effects during long-term operations : Theories , example and challenges *," 2016.

[11] T. Bagosi, K. V. Hindriks, and M. A. Neerincx, "Ontological reasoning for human-robot teaming in search and rescue missions," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 595–596.

[12] A. Freedy, E. J. de Visser, G. Weltman, and N. Coeyman, "Measurement of Trust in Human-Robot Collaboration," 2007.

[13] P. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. de Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Hum. Factors*, vol. 53, pp. 517–527, 2011.