Human-Robot Interaction: Developing Trust in Robots

Deborah R. Billings
Kristin E. Schaefer
Institute for Simulation & Training
University of Central Florida
4000 Central Florida Blvd.
Orlando, FL 32816
407.823.5560
dbillings@ist.ucf.edu;
kristinoleson@knights.ucf.edu

Jessie Y. C. Chen
Human Research & Engineering
Directorate Army Research Laboratory
12423 Research Parkway
Orlando, FL 32826
407.384.5435
jessie.chen@us.army.mil

Peter A. Hancock
Institute for Simulation & Training and
Department of Psychology
University of Central Florida
4000 Central Florida Blvd.
Orlando, FL 32816
407.823.2310
peter.hancock@ucf.edu

ABSTRACT
In all human-robot interaction, trust is an important element to consider because the presence or absence of trust certainly impacts the ultimate outcome of that interaction. Limited research exists that delineates the development and maintenance of this trust in various operational contexts. Our own prior research has investigated theoretical and empirically supported antecedents of human-robot trust. Here, we describe progress to date relating to the development of a comprehensive human-robot trust model based on our ongoing program of research.

Categories and Subject Descriptors
A.1. [**General Literature**]: Introductory and Survey

General Terms
Design, Human Factors, Performance, Reliability

Keywords
Human-Robot Interaction, Trust.

1. INTRODUCTION
The value of unmanned systems in the modern combat environment (across air, ground, and maritime domains) has been stressed by both the U.S. military and its allies [1]. In fact, the success associated with the deployment of these systems in current military operations has led to an expectation of their expanded roles in the future. One of the challenges involved in the implementation of such unmanned, robotic systems is facilitating appropriate levels of trust in the system to ensure effective collaboration. The human must be able to trust these systems to behave as expected, especially in battlefield situations.

2. HUMAN-ROBOT TRUST RESEARCH
Existing human-robot trust research focuses primarily on the outcomes of trust (e.g., reliance, success, failure, etc.) in a wide variety of application environments not constrained to military operations. Trust can impact the success of human-robot collaborations and may determine future robot usage [2]. Additionally, existing work often cites the importance of calibrating trust, or matching the human's perception of the robot to reality [3]. If trust is precisely calibrated, then the interaction will presumably be optimal. Therefore, research needs to identify the conditions necessary for the facilitation of these appropriate levels of trust in robots. The goal of our research is to identify antecedents of trust in human-robot interaction (HRI). We have identified numerous potential factors impacting trust and subsequently evaluated the current literature through a meta-analytic assessment (see [4] for a detailed description).

2.1 Identification of Antecedents of Trust
First, a literature review of existing theoretical and empirical human-robot trust research was conducted. Subject matter experts (SMEs) were also consulted. Based on the collected information, potential factors impacting trust were then identified and classified as robot-related, human-related, or environmental characteristics. This three-factor approach provided the framework for our human-robot trust model.

2.2 Meta-Analysis of Human-Robot Trust
We quantitatively reviewed our three-factor model framework (human, robot, and environmental factors) for HRI trust by conducting a formal meta-analysis of the empirical studies. Ten papers containing 60 correlational effect sizes and 11 papers containing 47 experimental effect sizes were included in the analysis. Results from this analysis revealed that overall trust is important to human-robot teams (d=+0.71; r=+0.25), and that robot characteristics (e.g., performance, appearance, proximity) are the most important factors, to date, in trust development (see Figure 1). Environmental factors provided a moderate effect, while a small effect was found for human-related factors. These results imply that manipulating design aspects of the robot can most help us to calibrate trust, or bring trust to the appropriate level to facilitate successful interactions. The trends in the literature indicate that higher trust is associated with higher reliability; the type, size, proximity, and behavior of the robot also affect trust. There is a need to develop specific design heuristics, trust mitigation strategies [5].

3. CURRENT RESEARCH
3.1 Trust in Other Domains
Our current work is exploring HRI trust further by assessing the similarities and differences between HRI trust and trust in other related domains (i.e., human-interpersonal trust, human-animal trust, and human-automation trust). Our goal is to determine how research in these areas can contribute to our human-robot trust

HRI'12, March 5-8, 2012, Boston, Massachusetts, USA.
ACM 978-1-4503-1063-5/12/03.

model, as the existing research in HRI remains limited. In particular, one question we are looking to resolve is: Is human-interpersonal trust a good analog for human-robot trust?

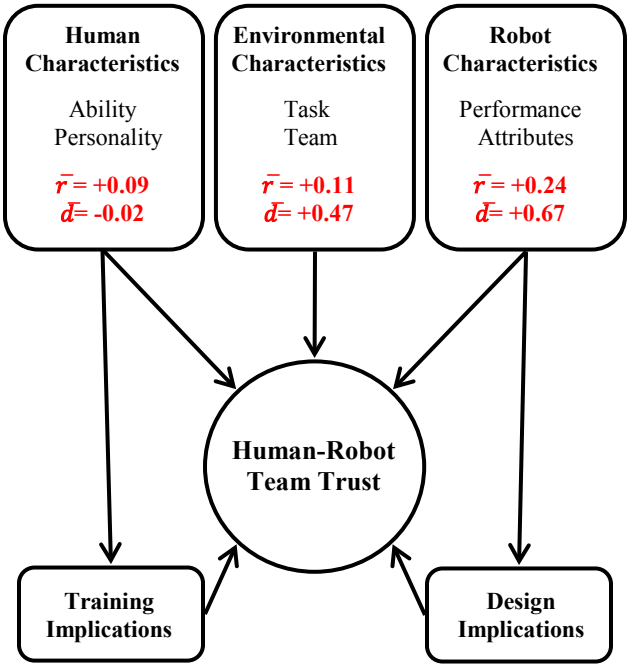

Figure 1. Three-factor model of trust

3.2 The Measurement of Trust

Another aspect of our current research is the measurement of trust. To date, several trust measures for human-automation interaction and human-interpersonal trust exist, but only one is available for HRI [6]. Our efforts will build on this existing research, as well as findings from our exploration of trust in the various domains, in order to develop and validate a subjective measurement scale allowing for trust calibration.

Before attempts can be made to quantify dynamic trust in HRI, several questions must be addressed. First, *what* exactly is being measured? Some researchers have distinguished differences between "trust," "trustworthiness," and "trust propensity" [7]. Others have discriminated between "dispositional trust" and "history-based trust" [8]. This list is certainly not exhaustive, as researchers continue to define even more types of trust (e.g., swift trust, initial trust, cognitive trust, affective trust, etc.). The most appropriate type of measure to use may depend on which type(s) of trust is being studied. Second, *how* can trust be measured (e.g., subjectively, objectively, physiologically)? Third, *when* should trust be measured (e.g., prior to the interaction, during the interaction, after the interaction, multiple times)? Addressing these questions represents an important step in quantifying trust in HRI.

4. CONCLUSION

Robots continue to transition from being perceived as tools to being perceived by humans as interdependent teammates [9]. Humans must accept and trust a robot before effective interaction can occur. Our prior meta-analytic findings highlight the importance of robot design factors in the development of trust. This provides a basis for future experimentation, yet there are other unknowns that must also be addressed. For example, robots do not operate at 100% all the time. When inevitable failures do occur, how does a human's trust level fluctuate? How can trust be repaired after it is violated? How can these dynamics be incorporated into our human-robot trust model framework? Our continued research efforts will help lay the foundation for answering these and other related research queries.

5. ACKNOWLEDGMENTS

The research reported in this document was performed in connection with Contract Number W911NF-10-2-0016 with the U.S. Army Research Laboratory, under UCF Task #3, P.A. Hancock, Principal Investigator. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Laboratory, or the U.S. Government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

6. REFERENCES

[1] U.S. Department of Defense. 2011. Unmanned Systems Integrated Roadmap FY2011-2036.

[2] Freedy, A., de Visser, E., Weltman, G., and Coeyman, N. 2007. Measurement of trust in human-robot collaboration. In *Proceedings of the 2007 International Conference on Collaborative Technologies and Systems* (Orlando, FL, May 21-25, 2007). CTS '07. IEEE, New York, NY, 106–114.

[3] Lee, J. D. and See, K. A. 2004. Trust in automation: Designing for appropriate reliance. *Human Factors* 46, 1 (Spring 2004), 50-80.

[4] Hancock, P.A., Billings, D.R., Schaefer, K. E., Chen, J.Y.C., de Visser, E.J., and Parasuraman, R. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* 53, 5 (Oct. 2011), 517-527.

[5] Billings, D. R., Oleson, K. E., Chen, J. Y. C., and Hancock, P. A. 2011. Mitigating inappropriate trust in human-robot interactions: A review of trust calibration strategies in the literature. Poster presented at the *2011 APA Annual Convention* (Washington, DC, August 04-07, 2011).

[6] Yagoda, R. E. 2011. *WHAT! You want me to trust a ROBOT? The development of a human robot interaction (HRI) trust scale* (Master's Thesis). NC State University.

[7] Colquitt, J. A., Scott, B. A., and LePine, J. A. 2007. Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *J. Applied Psychol.* 92, 4 (July 2007), 909-927.

[8] Merritt, S. M. and Ilgen, D. R. 2008. Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors* 50, 2, 194-210.

[9] Sanders, T. Oleson, K. E., Billings, D. R., Chen, J. Y. C., and Hancock, P. A. 2011. A model of human-robot trust: Theoretical model development. In *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting* (Las Vegas, NV, Sept. 19-23, 2011). HFES '11. HFES, Santa Monica, CA, 1432-1436.