Building Appropriate Trust in Human-Robot Teams

Scott Ososky, David Schuster, Elizabeth Phillips, and Florian Jentsch
University of Central Florida, Orlando
sososky@ist.ucf.edu; dschuster@ist.ucf.edu; ephillips@knights.ucf.edu, Florian.Jentsch@ucf.edu

Abstract

Future robotic systems are expected to transition from *tools to teammates*, characterized by increasingly autonomous, intelligent robots interacting with humans in a more naturalistic manner, approaching a relationship more akin to human–human teamwork. Given the impact of trust observed in other systems, trust in the robot team member will likely be critical to effective and safe performance. Our thesis for this paper is that trust in a robot team member must be appropriately calibrated rather than simply maximized. We describe how the human team member's understanding of the system contributes to trust in human-robot teaming, by evoking mental model theory. We discuss how mental models are related to physical and behavioral characteristics of the robot, on the one hand, and affective and behavioral outcomes, such as trust and system use/disuse/misuse, on the other. We expand upon our discussion by providing recommendations for best practices in human-robot team research and design and other systems using artificial intelligence.

Background

Robotic platforms are used in many civilian and military applications, for example, to conduct search and rescue operations (Murphy 2004), deal with IEDs in southwest Asia (Sharkey 2008), and perform tasks in other environments hostile to human beings. Despite the wide use of robotic platforms, however, the aforementioned tasks are currently accomplished with direct human oversight in which the robot is primarily a tool that is tele-operated, rather than a collaborating teammate. There is now a renewed emphasis in the military domain for autonomous systems that "extend and complement human capability in a number of ways" (Defense Science Board 2012).

Future robotic systems are expected to be able to process the complexity of our world and be active participants in human-system collaboration. In turn, soldiers will take on more of a supervisory role, as autonomous capabilities reduce the need for constant human supervision from a control station. In turn, this will allow soldiers to participate as active members in decision making and task execution. Therefore, the transition from *tools to teammates* is characterized by increasingly autonomous, intelligent robots interacting with humans in a more naturalistic manner, approaching a relationship more akin to human–human teamwork.

With the transition in robotics from *tools* to intelligent *teammates*, questions about the factors that influence the quality of a human–robot team must be answered. One of the most pressing questions is how human performance will be affected by the shift from operator to collaborator. As technology enables robots with greater intelligence and autonomy, human teammates will need to better understand robots and anticipate robotic behaviors (Philips, Ososky, Grove, and Jentsch 2012). Further, human teammates will have to bring forward both *trust* and a healthy dose of *skepticism* when interacting with these systems.

The tendency of humans to attribute sophisticated human-like qualities to non-human entities, including attitudes and motivations underlying judgments such as trust, comes easily (Nass and Moon 2000), and this tendency is easily influenced by seemingly superficial qualities of robots (Lee, Kiesler, Lau, and Chiu 2005). In fact, humans are increasingly willing to think of a robot as being "alive", due in part to the ubiquity of lifelike machines and agents found in movies, video games, and toys (Garreau 2007). However, this phenomenon may have unintended, negative consequences in dangerous situations, such as military applications — a Soldier once ran across a battlefield in Iraq under machine gun fire to "rescue" his robot (Singer 2009, p. 339).

Although there may be similarities in the way users of robotic systems regard intelligent robots and human teammates, it would be inappropriate to apply human-teaming constructs, such as trust, to human-robot teams without recognizing the differences between these two types of teams. Similarly, we posit that there will be critical differences between future human-robot trust and what is already known about trust in automation, in general. Therefore, while robot teammates will offer a unique form of human-robot interaction, this interaction will be similar to both exclusively human teams and current automated systems.

It is for these reasons that an effective human-robot team will be better characterized by *appropriate trust* than by maximizing trust. Appropriate trust minimizes negative performance outcomes at either end of the spectrum. Trust develops through the cognition of the human in the context of the robot and environment. Put simply, the better understanding by the human team member leads to more appropriate trust of the robot team member.

Mental Models: Building Blocks of Trust

For a human to appropriately trust a robot, he/she must understand the robot's capabilities and limitations in the context of the current mission's goals. Human understanding of robots as team members can be examined through the lens of mental models. Mental models are internal representations used by humans to understand world around them (Craik 1943). *Shared* mental models (SMM), then, apply specifically to teamwork as "knowledge structures held by members of a team" (Cannon-Bowers, Salas, and Converse 1993, p. 223), including understanding of relevant equipment, the task, of team members, and the interaction among team members.

Mental models are used to mentally examine various system state outcomes based on potential action selection, which Rasmussen (1979) described as "experiments on an internal representation or model." (p. 8) A purpose of an accurate mental model, then, is to allow users to make correct predictions about future system states (Wickens & Hollands 1999). Some common themes emerge from existing definitions: mental models are internalized representations of systems; they can also be run internally to generate system-state expectations. Further, humans maintain and select from multiple mental models held for different systems and situations.

Cannon-Bowers and colleagues (1993) identified four knowledge content models within SMM that need to be held by teams in order to be effective: The *equipment* model contains knowledge about the operation, functions, and limitations of tools and technologies. The *task* model consists of procedures, strategies, contingencies for completing the work. The *team interaction* model identifies team members' roles, communication patterns, and dependencies. Finally, the *team* model contains knowledge about other teammates' knowledge, skills, and abilities. SMMs help team members to predict and adapt to changing demands made on the team, as well as coordinate their actions to cope with said demands, especially in circumstances in which the team cannot overtly communicate with one another (Mathieu, Heffner, Goodwin, Salas, and Cannon-Bowers 2000).

The content models and overall framework were originally defined in the context of expert human teams. This framework can be adapted to human–robot teams, given a few important distinctions. Specifically, an intelligent, autonomous robot is unique in that it is both "equipment" and "team member". Therefore, a human's mental model of a robot teammate (as part of a SMM) includes knowledge about a robot's capabilities, limitations, and "personality". It follows that mental models are internally leveraged to guide human–robot interaction (Lohse 2011), including trust in robotic systems.

Similarly, prior research has suggested that other human—non-human teams involve interactions in which humans leverage mental models of both equipment and team member to accomplish work (e.g., human-animal teams). These teams can provide insight into fostering appropriate trust in future robotic teammates. For instance, because humans have more experience interacting with animals than with robots, they are generally able to recognize and accept an animal's capabilities. More specifically, animal partners may be skilled at performing some tasks and limited in others (Phillips, Ososky, Swigert and Jentsch 2012). This type of understanding often comes through experience and interaction with non-human teammates like animals. Additional research by Ososky, Phillips, Swigert and Jentsch (2012) provided support for this notion; they found that people tended to report more trust in a robotic teammate after observing it in a series of videos than a robot teammate they imagined. This lends further support for the role of experience and familiarization with robotic teammates in fostering human-robot trust. As with animals, opportunities for acclimation and exposure may be critical components for building trust within humans working as part of a human-robot teams.

Influence of Robot Characteristics on Mental Models

Mental models of robots are quickly formed around superficial characteristics, including physical form. Kiesler and Goetz (2002) found that participant ratings of a robot's reliability changed based on physical appearance. Participants were also willing to assign significantly different personality trait ratings (e.g., cheerful, responsible) to a humanoid robot versus a vehicle-shaped robot. Sims and colleagues (2005) found similar results in ratings of robotic forms viewed on a computer screen: Robots with spider-legs were rated as more aggressive than robots with two legs, wheels or treads. Additionally, vehicles with arms were rated as more aggressive and intelligent than those without. The researchers hypothesized that certain features served as affordances to bio- and anthropo-morphism in human perceptions of robots. Humans anthropomorphize in order to rationalize the actions or behaviors of non-human objects, including computers, animals, and robots (Duffy 2003).

Superficial influences on mental models are not limited to physical traits either. Lee and colleagues (2005) found a predictable pattern of robot knowledge estimation based on the spoken-language and country of manufacture of the robot. If the robot spoke Chinese, for example, participants estimated higher robot-knowledge of Chinese landmarks and locations than an English-speaking robot. Even within human–human relationships, mental models of individuals are formed based on relevant and pseudo-relevant information; while confidence in those models is determined by richness, independent of information accuracy (Gill, Swann, and Silvera 1998). Because a humanoid robot was used in the study, participants estimated the robot's knowledge using mental models based on human interaction metaphors, not machine knowledge.

Further, because of the propensity to anthropomorphize, humans attribute animacy and intention even in the simplest interactions. Already in 1944, Heider and Simmel asked participants to write about what they saw in an animated film in which two triangles and a circle moved about a rectangular box on a black and white display. Nearly all of the participants wrote a narrative of connected events about animated beings. These shape-entities were perceived as having intentions, and reacting to one another. Similarly, Ju and Takayama (2009) conducted a study in which people observed the motion of an automatic door. The findings suggested that the variations in the door's motion behavior were interpreted as gesture with intent: reluctant, welcoming, or urging. Finally, Saerbeck and Bartneck (2010) asked participants to observe the motion of an iRobot Roomba at different accelerations and path curvatures. The researchers reported that all but one participant used emotional adjectives to describe the robot's behavior, with nearly all of the participants attributing personality to the Roomba. In summary, "what matters for the human–robot relation is how the robot appears to human consciousness" (Coeckelbergh 2011, p. 199).

As technology enables robots with greater intelligence and autonomy, human teammates must possess a clear and accurate understanding of robots (Phillips, Ososky, Grove, and Jentsch 2011). The quality of a futuristic robot teammate will be of little benefit if humans misunderstand its purpose or misinterpret what it is doing. The problem is that the misinterpretation of robotic characteristics or behavior results in the incorrect assessment of robotic capabilities and functions, misaligned trust and reliance, and inappropriate social responses toward robots in complex and / or dangerous situations.

Proposition 1. Robot characteristics may impact on human perception, affecting both human-robot interaction and performance.

> **Proposition 1.1**. Humans are easily influenced by superficial characteristics of robots, both physical and social.
>
> **Proposition 1.2**. Humans tend to anthropomorphize technology, relating robots to perceived organic analogs.

Influence of Robot Characteristics on Trust

A recent meta-analysis (Hancock et al. 2011) suggested that *robot characteristics* have the strongest influence on trust within human-robot teams, followed by *environmental characteristics*. Robot related characteristics include performance-based factors such as behaviors, reliability, and transparency, in addition to attribute-based factors such as personality, type, and anthropomorphic qualities. Environmental characteristics include teaming factors and task demands. Likewise, the analysis also supported the notion that trust, in general, is a relevant aspect of human–robot teaming (Oleson, Hancock, Billings, and Schesser 2011). Trust in robots, like other aspects of mental models, is continuously refined through interaction.

Appropriate Trust in Robotic Systems

Trust alone in a robot's reliability does not guarantee better teamwork. An inaccurate or incomplete mental model may lead to an overestimation of a robot's abilities, creating a pitfall for automation *misuse.* Automation misuse describes failures resulting from a mis-calibrated, overreliance on automation capability (Parasuraman and Riley 1997). On the other hand, a human may choose not to work with an autonomous robot teammate at all or in a minimal capacity. The "underutilization of automation" (Parasuraman and Riley 1997, p. 233) is known as automation *disuse*. Given a task of sufficient personal risk, a human may decide not to leverage the abilities of a robot teammate at all if trust in that asset is low or diminished. However, given a task of sufficient complexity (e.g., IED defeat or urban patrol operations), some use of the robot may be imperative to reaching higher performance outcomes. Therefore, in human-robot teaming, appropriate trust is maintained when the human uses the robot for tasks or subtasks the robot performs better or safer while reserving those aspects of the task the robot performs poorly to the human operator. Humans must have a sufficiently developed mental model of the robot in order to appropriately trust the robot.

Proposition 2. Mental models provide a foundation for appropriate trust in robotic systems.

> **Proposition 2.1**. Incomplete or inaccurate mental models crate pitfalls for robotic system misuse or disuse.
>
> **Proposition 2.2**. Sufficiently developed and accurate mental models allow team members to appropriately use and not use robotic systems.

Trust in Robot Teammates versus Human Teammates

What, then, distinguishes robot reliability from human reliability? Sheridan succinctly highlighted the differences in human and machine (automation) reliability, respectively. He stated that, "humans differ enormously from machines, in that they are inherently variable and unreliable in their detailed behavior, while simultaneously being hyper-adaptable and metastable in their overall behavior because they perceive and correct their own errors" (Sheridan 2002, p. 166). Robots embodied in, and reacting to, a dynamic world, are more than simple machines. To this point, Sheridan noted that complex systems interacting with the world introduce an almost infinite number of failure points for which planning is practically impossible.

There are numerous ways in which current, task-oriented robots fail outside of the control of laboratory settings. Carlson and Murphy (2005) included a wide range of occurrences in their review of UGV reliability that can be classified as failures. Failure was defined as "the inability of the robot or the equipment used with the robot to function normally" (p. 424). Failures included: robot is stuck, sensor malfunction, communication breakdown, loss of power, controller issues, etc. The researchers also identified the outcomes of failures: either field-repairable or not, terminal or not. Not surprisingly, the combination of failure types, task environments, and actual UGV robots surveyed (Talon, Packbot, etc.) created a range of failure modes whose consequences varied wildly. The conclusion of the investigation was that current UGV robot reliability equated to between 6 and 20 hours between failures (i.e. low reliability), which did not include the time required to diagnose and correct the failure (hours, even days).

Do the reliability issues of current robots render them useless to humans? No, robots do not need to be perfectly reliable. Instead, they need to behave in understandable and/or predictable ways. Lee and associates (2010) investigated ratings of robot service performance given physical forms, robot reliability and error mitigation strategies. They found that errors on task negatively impact ratings of a robot's service performance. This negative impact was observed when interacting with both human-like and machine-like robots; however the negative impact was mitigated with forewarning of the robot's limitations. Wiegmann, Rich and Zhang (2001) investigated the effects of human interaction with a diagnostic aid that shifted reliability during trials. When the reliability of the aid shifted from 100% to 80%, human subjective ratings of the diagnostic aid's reliability were lower than that of participants interacting with an 80% reliable aid all along. While the researchers concluded that trust is easily lost and hard to regain, the argument can also be made that the participants had no knowledge of the nature or cause of the change in reliability.

Consider that the calibration of appropriate trust is more a function of a human's mental model of the robot's ability and limitations, than the ground-truth reliability of the robot itself. With the appropriate mental model, human teammates should be able to adapt to the limitations of the robot and adjust the utilization of the robot accordingly. Without this knowledge, however, it would prove difficult to coordinate within a task-environment given an unpredictable robot teammate.

Proposition 3. Human subjective assessment of trust in robots ultimately determines the use of robotic systems.

> **Proposition 3.1**. Robots need not be perfectly reliable, rather, they must perform in predictable and understandable ways.
>
> **Proposition 3.2**. Limitations of robot ability or reliability may be mitigated by forewarning human teammates of such limitations to calibrate trust appropriately.

Trust in Robot Teammates versus Trust in Automation

Examining trust in a robot teammate requires a slightly different approach from traditional *trust in automation* approaches. Previous trust in automation studies found that humans tend to use automation when their trust in the system exceeds their own self-confidence in performing a task (Lee and Moray 1994). However, the nature of the current context, the risk associated with *task* itself might be the determining factor in the decision to deploy a robot-teammate. Consider, for example, the real-world context of Soldier–robot teams tasked with either IED-defeat or patrol operations. A Soldier might always elect to allow a robot to attempt bomb disposal (see Greenemeier 2010), regardless of the Soldier's self-confidence in his or her ability to disarm a bomb. Alternatively, patrolling civilian environments, characterized by competing goals and uncertain conditions, may require more calculated approach (i.e., the Soldier might take the lead on the patrol task). Within military applications, team interdependency required by the task combined with the consequence of failure can dictate trust and reliance on robot-teammates. A critical difference, then, between traditional system automation and intelligent robots is that the robot is intended to be a collaborating teammate (see Hoeft, Kochan and Jentsch 2006), not merely an automated tool that assumes the place or duties of the human. This requires that the human must have a sufficient understanding of how the robot can contribute to mission goals and integrate this information with other mission factors such as task interdependency, time pressure, and safety.

Proposition 4. Under conditions of task complexity and uncertainty, human team members require an awareness of their own performance or risk to rely upon robots appropriately.

Proposition 4.1. In human-robot team tasks, there may be sub-tasks for which robot performance is poor, but human performance is worse.

Proposition 4.2. Assessment of human and robot performance given situation factors, is necessary to allocate tasks to the best team member.

Two Types of Trust in Robots

It is also important to recognize and clarify two different types of trust as they relate to human–robot teaming. Lee and See defined trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (2004, p. 51). The impact of trust in the human-robot team depends on how trust affects behavior. Hancock, Billings and Schaefer (2011) suggest two types of trust. There exists a notion of trust in *intention*, indicating that the human trusts that the robot is not deceptive in its directives or actions and will operate in a manner intended to benefit the goals of the team (Hancock, Billings and Schaefer 2011). Alternatively, there exists the quality of trust in *competency* or *ability* (Lee and See 2004). This type of trust aligns with the belief that the robot has the hardware or software necessary to complete a task and will reliably work toward completing team goals. Lee and Moray (1992) similarly defined this dimension as trust in *performance*.

In general, it is possible to have one type of trust (intention or ability) without the other, as well as both or neither. We typically think of trust in automated systems as a trust in ability or reliability. Conversely, knowledge of another person's tendencies or intentions is a relevant factor in exclusively human teams (Cannon-Bowers, Salas, and Converse 1993). However, both types of trust may be important to the human–robot relationship, given the human propensity to anthropomorphize technology. Robots may act with intentions to complete a task in a certain manner as designed by programmers or engineers. However, it is important not to overlook the human perception of robot intent, whether real or imagined, that also factors into human subjective assessment of trust in robot teammates.

Proposition 5. When evaluating trust in robotic systems, distinguish between trust in competency and trust in intention.

Proposition 5.1. Trust in competency is more important in automated systems. Trust intention is more important in human teammates.

Propostiion 5.2. Future robot teammates may embody elements of both types of trust.

Conclusions

As a collaborator, the human will have a greater responsibility for understanding the mission context and the robot's role within the mission. As we have described, this understanding can be understood as a series of mental models. Given the impact of trust observed in other systems, trust in the robot team member will likely be critical to prevent misuse or disuse, each of which can have critical consequences.

The factors that contribute most to trust in the system are characteristics of the robot itself. Even seemly surface qualities of the robot can have a large impact on the resulting mental model of the human. At the same time, models are dynamic and can be molded through continued interaction. The synthesis of trust and automation literature that we have provided here suggests that designers and researchers should not make maximizing trust their goal. Rather, trust must be appropriately calibrated to the capabilities of the robot and the task context. This can best occur when the human has a sufficiently complete and accurate mental model of the robot. We suggest that how robot factors can impact this mental model, and ultimately trust, should be the topic of future investigation. In the interim, we have provided research-based considerations for future robot systems based on the current state of the art and our initial results.

References

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan (Ed.), *Current issues in individual and group decision making* (pp. 221-246). Hillsdale, NJ: Erlbaum.

Carlson, J., & Murphy, R. R. (2005). How UGVs physically fail in the field. *IEEE Transactions on Robotics, 21*(3), 423-437.

Coeckelbergh, M. (2011). Humans, animals, and robots: A phenomenological approach to human-robot relations. *International Journal of Social Robotics, 3*(2), 197-204. doi: 10.1007/s12369-010-0075-6

Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.

Defense Science Board. (July 2012). *The role of autonomy in DoD systems*. Washington DC: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics.

Duffy, B. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems, 42*(3-4), 177-190. doi: 10.1016/s0921-8890(02)00374-3

Garreau, J. (2007). Bots on the ground. *The Washington Post*. Retrieved from The Washington Post website: http://www.washingtonpost.com/wp-dyn/content/article/2007/05/05/AR2007050501009.html

Gill, M. J., Swann, W. B., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology, 75*(5), 1101-1114. doi: 10.1037/0022-3514.75.5.1101

Greenemeier, L. (2010). Are military bots the best way to clear improvised explosive devices? *Scientific American*. Retrieved from http://www.scientificamerican.com/article.cfm?id=robot-ied-clearance

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 53*(5), 517-527. doi: 10.1177/0018720811417254

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243-259.

Hoeft, R. M., Kochan, J. A., & Jentsch, F. (2006). Automated systems in the cockpit: Is the autopilot "George," a team member? In C. Bowers, E. Salas & F. Jentsch (Eds.), *Creating high-tech teams: Practical guidance on work performance and technology* (pp. 313). Washington, DC: American Psychological Association.

Ju, W., & Takayama, L. (2009). Approachability: How people interpret automatic door movement as gesture. *International Journal of Design [online]*. Retrieved from http://icd110.dt.ntust.edu.tw/ojs/index.php/IJDesign/article/view/574/262

Kiesler, S., & Goetz, J. (2002). Mental models of robotic assistants. *CHI '02 extended abstracts on Human Factors in Computing Systems*, 576-577. doi:10.1145/506443.506491

Lee, M. K., Kielser, S., Forlizzi, J., Srinivasa, S., & Rybski, P. (2010). Gracefully mitigating breakdowns in robotic services. *Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction*, 203-210. doi:10.1145/1734454.1734544

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40*(1), 153-184. doi: 10.1006/ijhc.1994.1007

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*(1), 50-80. doi: 10.1518/hfes.46.1.50.30392

Lee, S., Kiesler, S., Lau, I., & Chiu, C. (2005). Human mental models of humanoid robots. *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, 2767-2772. doi:10.1109/ROBOT.2005.1570532

Lohse, M. (2011). Bridging the gap between users' expectations and system evaluations. *2011 IEEE RO-MAN*, 485-490. doi:10.1109/roman.2011.6005252

Murphy, R. R. (2004). Human-robot interaction in rescue robotics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 34*(2), 138-153. doi: 10.1109/TSMCC.2004.826267

Nass, C., & Moon, Y. (2000). Machines and mindlessness - social responses to computers. *Journal of Social Issues, 60*(1), 81-103. doi: 10.1111/0022-4537.00153

Oleson, K. E., Hancock, P. A., Billings, D. R., & Schesser, C. D. (2011). *Trust in unmanned aerial systems: A synthetic distributed trust model*. Paper presented at the 16th International Symposium on Aviation Psychology, Dayton, OH.

Ososky, S., Phillips, E., Swigert, B., & Jentsch, F. (2012). The influence of training and mental model formation on interpretation of robot behavior and intention in a simulated environment. Orlando: University of Central Florida.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 39*(2), 230-253. doi: 10.1518/001872097778543886

Phillips, E., Ososky, S., Grove, J., & Jentsch, F. (2011). From tools to teammates: Toward the development of appropriate mental models for intelligent robots. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55*(1), 1491-1495. doi: 10.1177/1071181311551310

Phillips, E., Ososky, S., Swigert, B., & Jentsch, F. (2012). Human-animal teams as an analog for future human-robot teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*(1), 1553-1557. doi: 10.1177/1071181312561309

Rasmussen, J. (1979). *On the structure of knowledge -- A morphology of mental models in a man-machine system context*. (Riso-M-2192). Roskilde, Denmark: Riso National Laboratory.

Sharkey, N. (2008). Grounds for discrimination: Autonomous robot weapons. *RUSI Defense Systems*, 86-89. Retrieved from RUSI website: http://www.rusi.org/downloads/assets/23sharkey.pdf

Sheridan, T. B. (2002). *Humans and automation* (Vol. 3). Santa Monica, CA: John Wiley & Sons, Inc.

Sims, V. K., Chin, M. G., Sushil, D. J., Barber, D. J., Ballion, T., Clark, B. R., . . . Finkelstein, N. (2005). *Anthropomorphism of robotic forms: A response to affordances*. Paper presented at the Human Factors and Ergonomics Society Annual Meeting, Orlando, FL.

Singer, P. W. (2009). *Wired for war: The robotics revolution and conflict in the 21st century*. New York, NY: The Penguin Press..

Wiegmann, D., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science, 2*(4), 352-367. doi: 10.1080/14639220110110306

Acknowledgments

The research reported in this document/presentation was performed in connection with Contract Number W911NF-10-2-0016 with the U.S. Army Research Laboratory. The views and conclusions contained in this document/presentation 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 heron.