The Effects of Agent Transparency on Human Interaction with an Autonomous Robotic Agent

Anthony Selkowitz[1], Shan Lakhmani[1], Jessie Y.C. Chen[2], and Michael Boyce[2]
[1] Institute for Simulation and Training at the University of Central Florida; [2] Army Research Laboratory

We use the Situation awareness-based Agent Transparency model as a framework to design a user interface to support agent transparency. Participants were instructed to supervise an autonomous robotic agent as it traversed simulated urban environments. During this task, participants were exposed to one of three levels of information used to support agent transparency in the interface display. Our findings suggest that providing agent transparency information allows operators to properly calibrate trust without excess workload. Though, increased agent transparency information did not support operator situation awareness.

Autonomous robotic agents, for military operations, are becoming increasingly sophisticated and independent. As robotic autonomy increases, human understanding of the agent's behavior, reasoning, and outcome projections becomes paramount (Chen & Barnes, 2014). The Situation awareness-based Agent Transparency (SAT) model was developed to address what information an agent should communicate to its human team members in order to be more "transparent" to them (Chen et al., 2014). The primary goal for this study was to investigate the effects of integrating information, using the framework of the SAT model, into the display for a fully autonomous (robotic) squad member (ASM). The study examined if increased agent transparency influenced a supervisory operator's trust in the agent, perceived workload, and situation awareness of the mission environment.

There are many different definitions on what transparency is and how it should be implemented (Ososky, Sanders, Jentsch, Hancock, & Chen, 2014; Lyons, 2013). We define agent transparency as a property of an interface to communicate the intent, performance, future plans, and reasoning process of an agent to the user (Chen et al., 2014). Providing such "transparency information" in a clear and understandable presentation is theorized to aid the user when interacting with both automation and autonomous robotic agents (Chen et al., 2014; Kilgore & Voschell, 2014). When provided with this information, users more readily trust an autonomous agent after a failure (Wang, Jamieson, & Hollands, 2011). For the current study, the SAT model was used to inform the display of 'transparency information'.

The Situation awareness-based Agent Transparency (SAT) model is a foundation to design "transparent interfaces" (Chen et. al., 2014). The SAT model theorizes that presenting the user with information supporting operator situation awareness of the agent's intent, performance, future plans, and reasoning process will lead to an improved understanding of the robot's current actions, logic, and predictions. The SAT model is composed of three levels of information required to support operator SA of the system and its tasking environment. In Level 1 of the SAT model, the agent displays its current actions and plans. In Level 2, the agent displays its reasoning and the environmental constraints that it takes into account when performing actions. In Level 3, the agent displays the projected outcomes and uncertainty of its actions and reasoning. For the current experiment, Level 1 information was implemented by displaying the autonomous robot's current route and resource levels. The information displayed for Level 2 included the reasons behind the robot's route changes and current environmental constraints and affordances. The information displayed for Level 3 included the robot's projected resources and uncertainty information. Interfaces supporting agent transparency have been identified as a way to improve operator trust, situation awareness, and workload (Chen et al., 2014; Mercado et al., in preparation). These factors have been identified as critical variables in robot operator performance (Ososky et al., 2014).

When interacting with an autonomous robot, properly calibrated trust is critical to avoid performance decrements stemming from over-trust or under-trust (Parasuraman, Sheridan, & Wickens, 2000). For the current experiment, operator trust is defined as "[t]he attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee & See, 2004, p. 54.). Research has shown that there are two different concepts of trust in automated systems. Merritt and Ilgen (2008) argue for dispositional and history-based trust in human-automation interactions. Dispositional trust is based on the person's attitude toward automated systems without any interaction with the automated system in question. History-based trust refers to the person's trust after interacting with the automated system. We measured trust prior to interaction with the autonomous robot to use as a covariate for the post-interaction (history-based) trust, in the autonomous robot, analysis. One way that has been suggested to properly calibrate history-based trust in an autonomous robot is to increase the transparency of the interface (Chen et al., 2014; Ososky et al., 2014). Increasing the transparency of an interface is also theorized to increase the situation awareness of the operator of the autonomous robot.

Situation awareness (SA) refers to an individual's dynamic understanding of "what is going on" in a given system (Endsley, 1995). In this model, SA is comprised of three hierarchical levels: perception of elements within the environment, the comprehension of their meaning, and a projection of their status in the near future (Endsley, 1995). Robot operators must monitor the autonomous robotic agent's performance, maintain SA, and re-task if needed, so practical transparency information must be displayed to support the needs of awareness and control, while still maintaining the performance and cognitive benefits of automation (Miller, 2014). Chen et al.'s (2014) SAT Model uses the SA

framework to determine the information needed to maintain agent transparency.

Mental workload represents the cognitive resources demanded by a task that are needed to achieve a particular level of performance (Hart & Staveland, 1988). If task demand exceeds the operators' capacity, they enter a state of overload; consequently, their performance decreases (Hart & Staveland, 1988). This added workload can be avoided by shifting responsibilities to an automated system (Miller, 2014). While a more opaque system can improve the performance of the human-automation system, it can result in an increased potential for error and decreased knowledge, awareness, and control for the operator (Kilgore & Voshell, 2014; Miller, 2014). To counteract these detrimental effects, transparency information can be added to the system (Kilgore & Voshell, 2014; Miller, 2014). When implemented poorly, transparency information can obfuscate the autonomous robotic agent's behavior; even when implemented appropriately, increased transparency requires additional information in the system's interface, which may increase operator workload (Chen et al., 2014). Consequently, the challenge for increasing transparency in a human-autonomous robotics system is to implement it in a manner that keeps operators in the loop while minimizing additional workload.

Current Study

This experiment was designed to examine the influence that different levels of information designed to support agent transparency, as established by Chen et al.'s (2014) SAT model, have on operators' monitoring performance using the display for an autonomous robotic agent. The goal was to determine if the addition of reasoning and projection information to the interface would increase operators' trust in the autonomous robotic agent, raise operators' workload, and strengthen operators' situation awareness. Our hypotheses are as follows:

H1. In conditions with more transparency information available, operator trust will be greater.
H2. In conditions with more transparency information available, operators will report increased situation awareness
H3. In conditions with more transparency information available, operator workload will be greater.

Method

Participants

Forty-five individuals (M_{age}=21.04, SD_{age}=2.17) from the metropolitan area of Orlando, Florida participated in this experiment. There were 27 male, 17 female, and 1 undisclosed gender participants.

Design

A one-way, between subjects design, with three levels of transparency, was used for the current study. Level 1 viewed a display with only SAT model Level 1 information. Level 1+2 received SAT model Level 1 and Level 2 information. Level 1+2+3 observed a display with SAT model Level 1, Level 2, and Level 3 information. The dependent measures in the experiment included the operator's trust in the autonomous robotic agent, situation awareness, and workload.

Apparatus

The simulator was developed using C# and Net 4.5 framework and run using a standard personal computer desktop setup. An example of the simulator user interface can be seen in Figure 1.

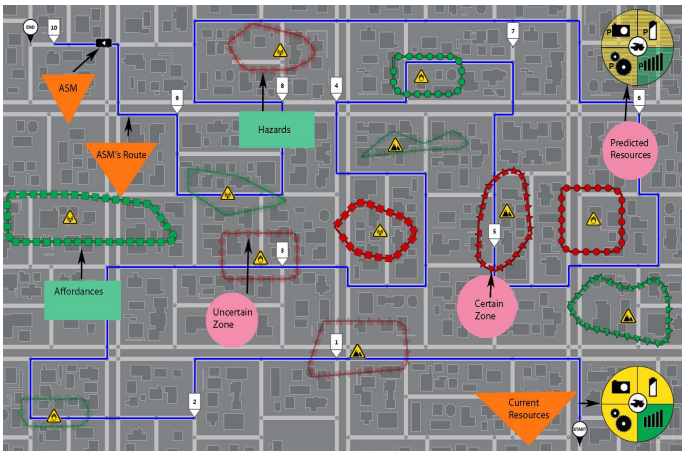

Figure 1. The SAT model-based information displayed in each condition. The triangles represent the Level 1 information. The squares represent the Level 2 information. The circles represent the level 3 information. The labels—triangles, rectangles, and circles— were not included in the experimental display.

Measures

Situation awareness

Situation awareness was measured using SAGAT, a freeze probe recall method (Endsley, 2000) that involves querying participants of their knowledge of SA elements during random freezes of the simulated environment (Stanton et al., 2013). This method was chosen because it elicits knowledge of task-specific elements, which needed when examining a monitoring task. Probes took the form of multiple choice questions about the ASM's resource levels, focusing on Level 1 SA. These items are listed below:

1. Which resources are currently green?
2. Which resources were last reduced?
3. When was the last time your current status icon changed?

Trust

Two trust measures were used in the current experiment. The Trust in Automated Systems scale was adapted from Jian, Bisantz, and Drury (2000) to be specifically about automated systems. The scale was adapted by replacing the word "system" with "automation". For the second trust measure, we further modified version the Trust in Automated Systems scale (Function Specific Trust in Automated Systems) to include the functions and capabilities of an autonomous robot as determined by Parasuraman, Sheridan, and Wickens' (2000) four classes of functions for the automation of information processing tasks: Gathering or Filtering Information;

Integrating and Displaying Analyzed Information; Suggesting or Making Decisions; Executing Actions. The purpose of using this scale was to disambiguate the participant's trust according to the functions of the autonomous robotic agent.

Workload

The NASA-Task Load Index (Hart & Staveland, 2000), a widely used assessment tool, was used to measure workload. The weighted scores were used to compute workload.

Procedure

Participant assignment to SAT display condition was randomized. Participants filled out the Trust in Automated Systems scale prior to monitoring the ASM. After completing the Trust in Automated Systems scale, the participant received training on the display elements of the ASM's display. Once the training was completed, the participant completed a scenario to familiarize them with the display and SAGAT style of prompts. After the familiarization scenario, the experimental scenario began. The participant was tasked to monitor the ASM', displayed on the map as a rectangular icon, as it moved to the next way point (see Figure 1). At the beginning of each waypoint, the ASM's route was revealed to the participant by way of a blue navigation line. Participants were informed that the ASM will take the most direct route possible, and reroute on its own accord; the navigation line reflected any route change. During the task, participants were instructed to answer the SAGAT prompts to the best of their ability. Once the ASM completed its route, the NASA-TLX was administered, and then the next route started. Overall, the participant monitored the ASM while it completed six routes. During each route, the SAGAT prompts were administered three times. Overall, the NASA-TLX was administered six times. Once all routes were completed, participants completed the Trust in Automated Systems scale and the Function Specific Trust in Automated Systems scale. Then the participants were debriefed and any questions that they had were answered.

Results

Trust

Two separate trust analyses were performed. The first trust analysis to be discussed is the 1-way Trust in Automated Systems (Post-interaction) by 3 SAT display (Level 1, Level 1+2, Level 1+2+3) Analysis of Covariance (ANCOVA) using score on the pre-interaction Trust in Automated Systems scale as a covariate, α=.05. The assumptions of ANCOVA were tested and none were violated. The pre-interaction trust describes participants' propensity to trust automated systems (Merritt & Ilgen, 2008). Trust in Automated Systems post-interaction scores were used as the dependent variable, and SAT display group was used as a between-subjects independent variable. Examination of the data suggested that trust scores were normally distributed and variances were homogenous. There was a significant effect for SAT display on the post-interaction Trust in Automated System score, $F(2, 41)=4.073$, $p<.05$, $\eta_p^2 = .165$ (Figure 2). Pairwise comparisons revealed that Level 1 (M=55.05, SD=10.36) had significantly lower post-interaction trust in automated systems than Level 1+2 (M=61.09, SD=10.18) $p<.05$. There was not a significant difference between Level 1+2+3 (M=57.27, SD=10.36) and the other conditions.

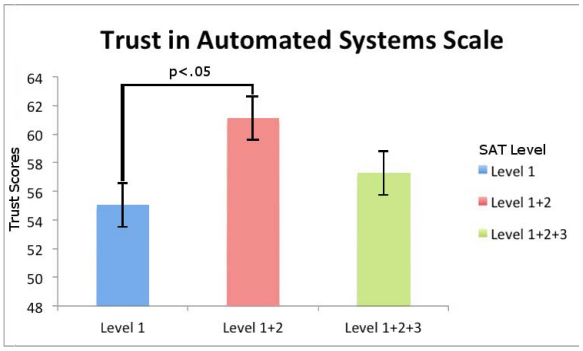

Figure 2. Trust in Automated Systems scores.

.

The second trust analysis examined the participants' trust using the Function Specific Trust in Automated Systems scale. Four separate 1 (Trust in automated system function) by 3 (SAT display group) ANCOVAs were performed using the pre-interaction Trust in Automated Systems scale as a covariate. The alpha level was reduced to .01 to account for running multiple comparisons. The dependent measure for the analyses was trust specific to the functions of automated systems. There were no significant differences in trust in the robot on the functions "Gathering or Filtering Information," "Suggesting or Making Decisions," and "Executing actions." Adjusting for pre-intervention test scores, trust specific to "Integrating and Displaying Analyzed Information" differed significantly by transparency information displayed, $F(2,41)= 5.274$, $p<.01$, $\eta_p^2 =.205$ (Figure 3). Examination of the data suggested that trust scores were normally distributed and variances were homogenous. Pairwise Bonferroni comparisons revealed significant differences between Level 1 (M=80.07, SD=19.94) and Level 1+2 (M=93.41, SD=19.94) at $p<.01$. No significant differences were observed between Level 1+2+3 (M=85.27, SD=19.94) and either Level 1 or Level 1+2.

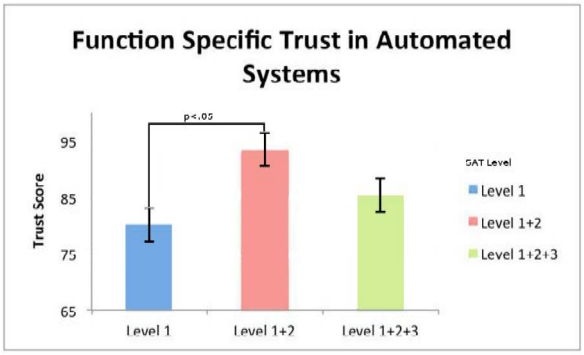

Figure 3. Function Specific (Integrating and Displaying Analyzed Information) Trust in Automated Systems scores.

Situation Awareness

Initial analysis of the data from the situation awareness metrics revealed that they were highly negatively skewed across all questions. Upon further examination, a trend was noticed in which Level 1 and Level 1+2 were negatively skewed and Level 1+2+3 was normally distributed

according to the Shapiro-Wilk test. In light of these findings, a 1 (Situation awareness prompt) by 3 (SAT display group: Level 1, Level 1+2, Level 1+2+3) independent samples Kruskal-Wallis test was the best course of action for analysis for each prompt. The following sections will examine significant results according to each question, non-significant results were obtained for the situation awareness prompt: "When was the last time your system icon changed?".

Transparency condition had a significant effect on the situation awareness prompt: "Which Resources were Last Reduced?" $H(2)=7.203$, $p<.05$. Pairwise comparisons with adjusted p-values revealed a significant difference ($p=.022$, $r=.40$) between Level 1+2 and Level 1+2+3 in which Level 1+2 (29.4) had a higher rank than Level 1+2+3 (16.6). There was no significant difference between Level1 and Level 1+2, or Level 1 and Level 1+2+3.

Another significant effect was noted for the situation awareness prompt: "Which Resources are Currently Green?" $H(2)=6.23$, $p<.05$. Pairwise comparisons, with adjusted p-values, indicated a trending significance ($p=.070$, $r=.34$) in which Level 1 (29.83) had higher rank than Level 1+2 (19.03). There was no significant difference on performance between Level 1 and Level 1+2+3, or Level 1+2 and Level 1+2+3.

Workload

An examination of the data found no violations of the assumptions of ANOVA. A 1 (weighted workload) by 3 (SAT display group: Level 1, Level 1+2, Level 1+2+3) between-subjects ANOVA was performed. No significant differences in weighted workload were observed according to SAT display group $F(2, 42)= .497$, $p=.612$, $\eta^2= .023$

Discussion

We use Chen et al.'s (2014) model of agent transparency to examine the impact of agent transparency information on operators' trust in the autonomous robotic agent, SA of the agent's display, and workload while using the system. The use of SAT levels allows us to examine differing amounts of agent transparency on a systematic level, starting with trust.

Trust

Adjusting for participants' dispositional trust in automation, participants exposed to information supporting ASM status and reasoning, Level 1+2, showed a notable change in self-reported trust. While the greater transparency in Level 1+2 aligns with Hypothesis 1, reported trust in Level 1+2+3 yields more ambiguous results. The displayed status and reasoning information communicated the trust cue of intent; the addition of prediction and uncertainty information may have subsequently informed operators of the limitations of the system as well, guiding them towards a more appropriate level of reliance (de Visser et al., 2014; Lee & See, 2004). The trust gained from the displayed reasoning information in Level 1+2+3 seems to have been counteracted by a loss of trust in the system following the display of uncertainty information. In contrast, Mercado et al. (in preparation) found that participants reported greater trust in an intelligent agent when the agent presented SAT Level 1+2+3 (with uncertainty as the Level 3 information), compared with SAT Level 1+2 and Level 1. However, the performance data showed that participants rejected the agent's incorrect recommendations more often in the Level 1+2+3 condition than in the other two conditions—indicating more effective trust calibration with increased levels of agent transparency. Uncertainty also influenced performance in a simulated automated driving task; participants who received uncertainty information demonstrated a faster time-to-take-over after automation failure (Helldin, Falkman, Riveiro, & Davidsson, 2013).

The Function Specific Trust in Automated Systems scale, rather than describing overall trust in automation, describes trust in light of specific automated tasks, centered on Parasuraman et al.'s (2000) delineation of classes of functions in automation. This approach distinguishes user trust of specific automation tasks within a larger system of automation. In the current system, the operator was tasked to observe the robotic agent's interface, understand its actions, and be ready to relay that information to others. The most relevant automation function class, which facilitated understanding of the agent's actions, was "Integrating and Displaying Analyzed Information." This difference between groups paralleled that of the overall system trust described above. The difference in trust between only Level 1 and Level 1+2, demonstrates the influence that reasoning and uncertainty information can have on operator trust in automated analysis systems. Human-Agent trust requires the belief that the trustee will accomplish the desired goal in a situation filled with uncertainty, and the introduction of uncertainty will cause operators to engage in compensatory action (Lee & See, 2004; Endsley, Bolte, & Jones, 2003). Participants who received uncertainty information in a simulated driving task demonstrated a lower, more appropriate trust in the system (Helldin et al., 2013). This calibration of trust, however, was not exhibited amongst the other classes of automation, suggesting that the expressed transparency information did not influence those classes of automation or that those tasks were not relevant and hence remained unobserved. Future automation tasks can be assessed using this Function Specific trust to ascertain a more complete understanding of the automation functions and capabilities of the robot being trusted.

Situation Awareness

The analysis for SA focused on the effects on the resources of the autonomous robotic agent, and consequently focused on three SA probes describing recent resource usage. Hypothesis 2 was not supported, but the score distribution suggested secondary issues contributed to this result. The high negative skew in Level 1 and Level 1+2, imply that the implementation of prediction and uncertainty transparency information decreased the accuracy of resource observation throughout Level 1+2+3, preventing the ceiling effect seen in the other groups. The addition of the predicted resource gauge may have split users' attention, so they tried to remember information from both gauges and failed or recalled information from the wrong gauge.

For the question "Which Resources were Last Reduced?," operators in Level 1+2+3 answered less accurately than their counterparts in Level 1+2. The aforementioned second gauge may have obfuscated the desired information, creating a situation where operators thought, incorrectly, that they knew when resources were reduced. The Kruskal-Wallis analysis detected a significant difference between groups for the "Which Resources are Currently Green?" question, but subsequent pairwise comparisons did not. While these results may have been hampered by a ceiling effect, especially when coupled with the trending significant difference between Level 1 and Level 1+2, the limited sample size restricts the conclusions that can be made. The results of "When was the last time your current status icon changed?" did not differ between groups, suggesting that the confusion over resource gauges was not as prominent with this feature. Overall, the SA probes attempting to measure resource display and use information may have been too easy, yielding distributions of scores frequently reaching the maximum. More probes in the future may remedy this issue.

Workload

Regardless of amount of transparency information displayed, operators reported a similar level of overall workload. This outcome is contrary to the expectations established by Hypothesis 3. This result, however, may be due to appropriate implementation of principles of ecological interface design (EID) (Burns & Hajdukiewicz, 2004, pp. 1-2). Using principles featured in EID, information supporting agent transparency can be displayed in a manner that visually represent objects and abstract principles in ways that reduce the need for higher order cognitive processes, through the use of visual representations such as lines, symbols, fields, and maps (Burns & Hajdukiewicz, 2004, pp. 4, 82-84; Kilgore & Voshell, 2014). The design of the interface may have allowed users to mostly intuit, rather than conventionally interpret, the information displayed to them, regardless of SAT information level displayed (Chen et al., 2014). Alternatively, given the similarity of SA scores between conditions, operators may not have processed the information they were given, though the trust scores suggest that the information was acknowledged.

Conclusion

The more granular examination of trust, focusing on different classes of automation function, allowed us to determine the kinds of tasks, visualized in the interface, that were influenced by the displayed transparency information during the task. This specific understanding of the operators' reaction to the visualizations of different system information in the interface, facilitates greater understanding of how the operator makes use of the system, which can drive future development and research. The addition of information, while influencing operator trust, did not influence workload, which suggests that the graphical implementation of this information, using principles from EID, allowed for more information without a concurrent increase in workload. This study shows a method for how transparency information can be implemented, using the SAT framework, in a way that allows for modulation of trust without excessive workload gains.

References

Burns, C. M., & Hajdukiewicz, J. (2013). *Ecological interface design*. CRC Press.

Chen, J.Y.C., & Barnes, M.J. (2014). Human-agent teaming for multi-robot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems, 44*(1), 13-29.

Chen, J. Y., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). *Situation Awareness-Based Agent Transparency* (No. ARL-TR-6905). Aberdeen Proving Ground MD: US Army Research Laboratory.

de Visser, E. J., Cohen, M., Freedy, A., & Parasuraman, R. (2014). A design methodology for trust cue calibration in cognitive agents. In *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments* (pp. 251-262). Springer International Publishing.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 32-64.

Endsley, M.R., Bolte B., Jones D.G. (2003) *Designing for situation awareness: an approach to human-centered design*. London: Taylor and Francis.

Jian, J.Y., Bisantz, A.M., & Drury, C.G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*(1), 53-71.

Jones, D.G. & Kaber, D.B., (2004). Situation awareness measurement and the situation awareness global assessment technique. In: Stanton, N., Hedge, A., Hendrick, H., Brookhuis, K., Salas, E. (Eds.), *Handbook of Human Factors and Ergonomics Methods.* CRC Press, Boca Raton, USA, pp. 419–427.

Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013, October). Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 210-217). ACM.

Kilgore, R., & Voshell, M. (2014). Increasing the transparency of unmanned systems: Applications of ecological interface design. In *Virtual, Augmented and Mixed Reality. Applications of Virtual and Augmented Reality* (pp. 378-389). Springer International Publishing.

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

Mercado, J., Rupp, M., Chen, J., Barnes, M., Barber, D., & Procci, K. (in preparation). Intelligent agent transparency in human-agent teaming for multi-UxV management.

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(2), 194-210.

Miller, C. A. (2014). Delegation and transparency: Coordinating interactions so information exchange is no surprise. In *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments* (pp. 191-202). Springer International Publishing.

Ososky, S., Sanders, T., Jentsch, F., Hancock, P., & Chen, J. Y. (2014, June). Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. In *SPIE Defense+ Security* (pp. 90840E-90840E). International Society for Optics and Photonics.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *30*(3), 286-297.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. International *Journal of Industrial Ergonomics, 39*(3), 490-500.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2011, September). The effects of design features on users' trust in and reliance on a combat identification system. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 55, No. 1, pp. 375-379). SAGE Publications.

Acknowledgments: This research was supported by the U.S. Department of Defense's Autonomy Research Pilot Initiative. The authors wish to thank MaryAnne Fields, Daniel Barber, Erica Valiente, Jonathan Harris, and Susan Hill for their contribution to this project.