Agent Transparency and Human-Autonomy Teaming Effectiveness in Multi-Robot Management Contexts

Jessie Y. C. Chen, Kimberly Stowers, Ryan Wohleber, Michael Barnes
US Army Research Laboratory, Aberdeen Proving Ground, MD
UNITED STATES OF AMERICA

yun-sheng.c.chen.civ@mail.mil; kim.stwrs@gmail.com; ryan.wohleber@gmail.com;
michael.j.barnes.civ@mail.mil

ABSTRACT

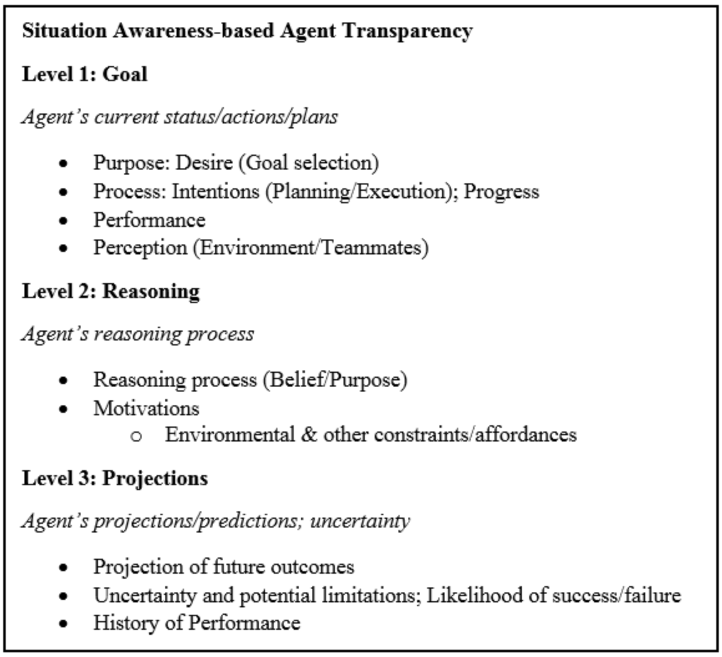
As intelligent agents (IAs) become more sophisticated, it is crucial to develop user interfaces that make the IA's intent, reasoning and predicted future states transparent to the human operator. Chen et al. leveraged Endsley's model of situation awareness (SA) and developed the Situation awareness-based Agent Transparency (SAT) model to organize the issues involved with supporting the human's awareness of an agent's: current actions and plans (L1), reasoning process (L2), and outcome predictions (L3). Projection uncertainty (U) may also be communicated. The current paper presents two human factors studies that investigated the effects of agent transparency on the effectiveness of human-agent teaming in military multi-robot management contexts. The first study investigated effects of levels of agent transparency; the second study focused on effects of agent communication framing (agent's style of communication). The user interfaces employed in the experiments were based on the SAT model. Overall, the results showed that perceptions of the IA (its aptitude and usefulness) increased with agent transparency. Trust survey results showed that the complimentary IA was rated as less trustworthy compared with the critical IA, especially in low transparency conditions. Participants tended to agree more with the critical IA when it was more opaque than when it was more transparent; on the other hand, participants' agreements with the complimentary IA did not differ regardless of the IA's transparency.

1.0 INTRODUCTION

As autonomous systems become more intelligent and more capable of complex decision making [1][2], it becomes increasingly difficult for humans to understand the reasoning process behind the system's output. However, such understanding is crucial for both parties to collaborate effectively, particularly in dynamic environments [3][4]. Indeed, *Explainable AI* [5], a DARPA program announced in August 2016, underscores the importance of making intelligent systems more understandable to the human. The U.S. Defense Science Board recently published a report, *Summer Study on Autonomy* [6], in which the Board identified six barriers to human trust in autonomous systems, with 'low observability, predictability, directability, and auditability' as well as 'low mutual understanding of common goals' being among the key issues (p. 15).

1.1 Situation awareness-based Agent Transparency (SAT) Framework

Chen et al. [3][7] leveraged Endsley's model of situation awareness (SA) [8][9] and developed the Situation awareness-based Agent Transparency (SAT) model, which mirrors the three levels (perception, comprehension, and projection) of Endsley's model (Figure 1). At the first SAT level (L1), the operator is provided with the basic information about the agent's current state and goals, intentions, and proposed actions (*perception* of the environment from the agent's perspective). At the second level (L2), the operator is provided with information about the agent's reasoning process behind those actions and the constraints/affordances that the agent considers when planning those actions (*comprehension*). At the third level (L3), the operator is provided with information regarding the agent's projection of future states, predicted consequences, likelihood of success/failure, and any uncertainty associated with the aforementioned projections (*projection*). Understanding the effects of uncertainty on the third level of SAT was considered of particular importance, because projection is predicated on many factors whose outcomes are never known precisely [7][10]-[13]. Prior research has shown that agents conveying uncertainty can improve the joint human-agent team performance and are also perceived as more trustworthy [10][14][15].



Situation Awareness-based Agent Transparency

Level 1: Goal

Agent's current status/actions/plans

- Purpose: Desire (Goal selection)
- Process: Intentions (Planning/Execution); Progress
- Performance
- Perception (Environment/Teammates)

Level 2: Reasoning

Agent's reasoning process

- Reasoning process (Belief/Purpose)
- Motivations
  - Environmental & other constraints/affordances

Level 3: Projections

Agent's projections/predictions; uncertainty

- Projection of future outcomes
- Uncertainty and potential limitations; Likelihood of success/failure
- History of Performance

Figure 1: Situation awareness-based Agent Transparency (SAT) Framework

1.2 Impact

Successful operational employment of autonomous systems requires command and control agility. Indeed, agility in tactical decision-making, mission management, and control is one of the key attributes for enabling heterogeneous unmanned vehicle (UxV) teams to successfully manage the 'fog of war' with its inherent complex, ambiguous, and time-challenged conditions. Mission effectiveness relies on rapid identification and management of uncertainties that can disrupt an autonomous team's ability to complete complex operations. Increasingly, research and development efforts are focusing on developing intelligent agents (IAs) that can work with human operators on managing the UxV teams [3][16]. One of those efforts is the Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT) project, which is currently funded by the ARPI. IMPACT is investigating issues associated with flexible play-calling [17][18], global cooperative control and local adaptive/reactive capability, and bi-directional human-autonomy interaction in military mission contexts [19]. However, in order for the IA to collaborate with human operators, the IA must be able to communicate its current and future states and receive commands to alter them when necessary [3]. Furthermore, knowledge of the IA's state of uncertainty has proven to be a crucial factor in the ability of humans to interact with automated systems [4][7][10]-[13][20]. In particular, it is important for operators to understand not only the locus of the uncertainty, but also to understand its source [21].

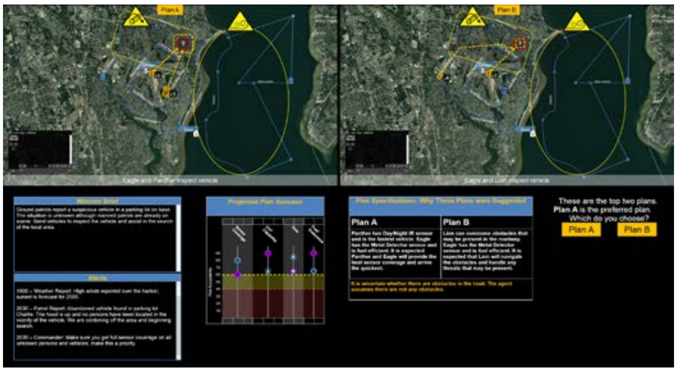
1.3 Current Studies

Two studies conducted under the IMPACT project are summarized in this paper. The first study investigated effects of levels of agent transparency; the second study focused on effects of agent communication framing (agent's style of communication). In both studies, participants completed blocks of mission vignettes, in which they were instructed to choose one of two plans provided by an IA, which always recommended an optimal plan (A) and a back-up plan (B). Three out of eight times, plan B was in fact better than plan A, based on updated intelligence or commander's instructions provided to the participants. The participant's task was to select the best plan based on current mission requirements. The user interfaces employed in the experiments were based on the SAT model.

2.0 STUDY 1

2.1 Methodology

This experiment utilized a within-subjects design with agent transparency as the independent variable. A previous study using the same UI from IMPACT tested the utility of Level 1 transparency and found that it was most useful when combined with higher levels of transparency [10]. Thus, transparency was only tested at three higher levels: (a) Level 1+2: containing reasoning information, (b) Level 1+2+3: containing reasoning and projection information *without* uncertainty, and (c) Level 1+2+3+U containing reasoning and projection *with* uncertainty. The UI varied per condition by showing corresponding pieces of SAT-level information on the map, in text, and on a sliding bar scale (Figure 2).


Plan A
Plan B
Eagle and Panther inspect vehicle
Eagle and Lion inspect vehicle
Mission Brief
Alerts
Projected Plan Success
Plan Specifications: Why These Plans were Suggested
Plan A
Plan B

Mission Brief

Ground patrols report a suspicious vehicle in a parking lot on base. The situation is unknown although manned patrols are already on scene. Send vehicles to inspect the vehicle and assist in the search of the local area.

Alerts

1900 – Weather Report: High winds reported over the harbor; sunset is forecast for 2000.

2030 – Patrol Report: Abandoned vehicle found in parking lot Charlie. The hood is up and no persons have been located in the vicinity of the vehicle. We are cordoning off the area and beginning search.

2030 – Commander: Make sure you get full sensor coverage on all unknown persons and vehicles, make this a priority.

Plan Specifications: Why These Plans were Suggested

Plan A	Plan B
Panther has Day/Night IR sensor and is the fastest vehicle. Eagle has the Metal Detector sensor and is fuel-efficient. It is expected Panther and Eagle will provide the best sensor coverage and arrive the quickest.	Lion can overcome obstacles that may be present in the roadway. Eagle has the Metal Detector sensor and is fuel efficient. It is expected that Lion will navigate the obstacles and handle any threats that may be present.

It is uncertain whether there are obstacles in the road. The agent assumes there are not any obstacles.

These are the top two plans. **Plan A** is the preferred plan. Which do you choose?

Plan A Plan B

Figure 2: IMPACT Study 1

Fifty-three students from an American university were recruited for this experiment. The average age of participants was 21.7 years (*sd* = 3.6). Prior to beginning the experiment, participants completed several individual difference surveys, including an Implicit Association Test (IAT) to track any bias they may have against technology. They then received about 1 hour of training. The subsequent experiment was divided into 3 blocks of 8 missions, with conditions counter-balanced. Using the interface presented in Figure 2, participants were presented with two plans to complete each mission, and were instructed to choose just one of them to move forward. For each set of plans, the agent always recommended one plan over the other, but the human always made the final decision. Reliability of the agent was held constant such that it was right in 5 out of every 8 scenarios. Information regarding mission objectives and commander's instructions were given before each set of decision options were presented, while information regarding specific mission parameters and vehicle capabilities were given on the decision screen (Figure 2). Each experimental block (including corresponding surveys) took participants about 30 minutes to complete.

After each block, participants completed surveys, including a measure of workload on an interval scale from 1 (low) to 10 (high) (NASA-TLX [22]), and a trust survey using a 7-point Likert scale (modified Trust in Automated Systems survey with two subscales [23][24]). Performance was measured throughout by tracking total correct responses, "proper use" (instances when the human agreed with the IA and answered correctly) and "correct rejections" (instances when the human disagreed with the IA and answered correctly). This performance measure was adapted from signal detection theory [25], but signal detection analyses were not completed.

2.2 Results

2.2.1 Operator Performance - Total Correct Response

A repeated-measures ANOVA showed significance, indicating an increase in operator performance (Total Correct Response, which combines Proper Use and Correct Rejections; Figure 3) with transparency level, $F(2, 104) = 10.31$, $p < .001$, $\eta_p^2 = .17$. Pairwise comparisons showed a significant difference between Level 1+2 and Level 1+2+3 ($p < .05$), as well as between Level 1+2 and Level 1+2+3+U ($p < .001$). Total correct responses were highest in the Level 1+2+3+U SAT condition and lowest in the Level 1+2 SAT condition.

2.2.2 Operator Performance - Proper Use and Correct Rejections

To examine exactly what areas of performance were most impacted, a repeated-measures MANOVA of proper use and correct rejections was performed. A significant multivariate effect indicated an increase in performance with transparency across both sub-measures: $F(4, 206) = 6.05$, $p < .001$, $\eta_p^2 = .11$. A univariate effect was found for proper use, $F(2, 104) = 10.66$, $p < .001$, $\eta_p^2 = .17$. Pairwise comparisons with Bonferroni correction indicated a significant difference between Level 1+2 and Level 1+2+3 ($p < .01$), as well as between Level 1+2 and Level 1+2+3+U ($p < .001$). As can be seen in Figure 3, both proper use and correction rejections were highest in the Level 1+2+3+U condition.

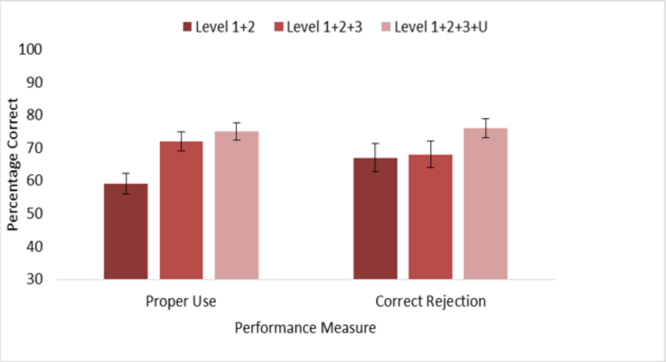


Figure 3. Percentage correct for both IA proper use and IA correct rejection rates.

2.2.3 Operator's Perceived Workload

In order to examine operator workload, we conducted a repeated-measures ANOVA on the total NASA-TLX scores, as well as a repeated measures MANOVA on the 6 unweighted subscales of the NASA-TLX. The effect of total workload was not significant, $F(2, 104) = .86$, $p = .43$, $\eta_p^2 = .02$. The effect of the subscales was also not significant, $F(12, 198) = .87$, $p = .58$, $\eta_p^2 = .05$.

2.2.4 Operator's Perceived Trust in the IA

Individual differences in IAT scores were covaried with participant trust scores in order to account for pre-existing biases that may affect trust. Trust was thus examined using a repeated-measures MANCOVA across the two subscales of our modified trust survey. A significant multivariate effect was found, $F(4, 202) = 2.61$, $p < .05$, $\eta_p^2 = .05$. Additionally, significant main effects of both subscales were found: (a) for integrating and diplaying information – $F(2, 102) = 3.48$, $p < .05$, $\eta_p^2 = .06$, and (b) for suggested decisions – $F(2, 102) =$

4.08, $p < .05$, $\eta_p^2 = .07$. Pairwise comparisons with Bonferroni correction indicated a significant difference between Level 1+2 and Level 1+2+3 in *suggesting decisions* ($p < .05$; Figure 4).

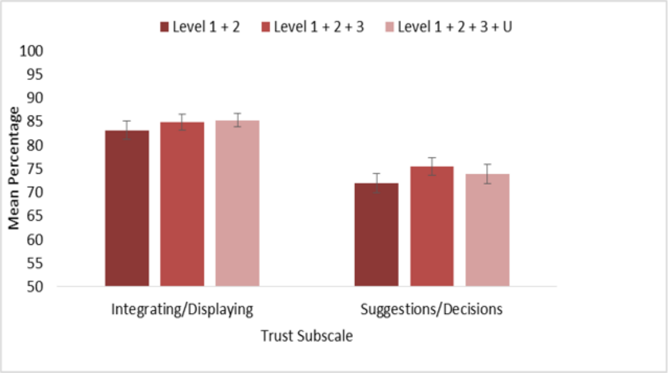


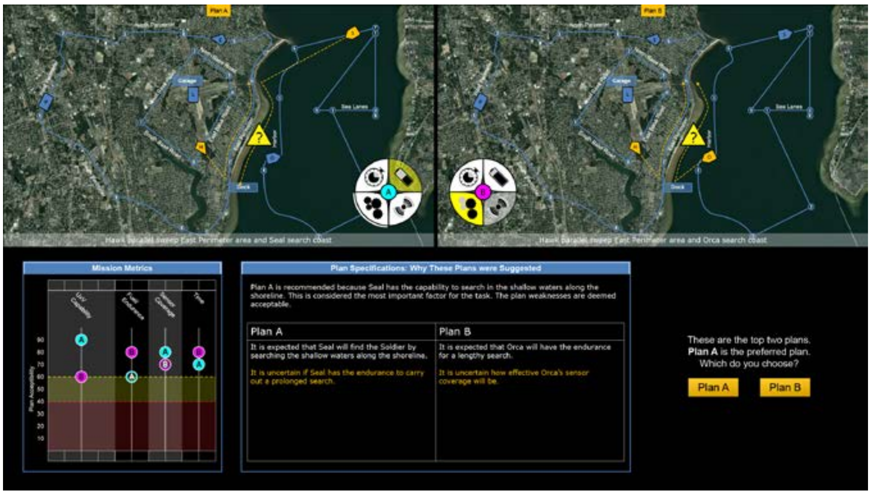
Figure 4. Trust in integrating and displaying information, as well as suggesting decisions.

3.0 STUDY 2

The second study focuses on the human factors aspects of the agent's transparency and communication framing in the context of human-autonomy teaming via IMPACT technologies and capabilities. The overall goal of the this study was to understand the interaction between level of agent transparency communication, according to the SAT model [3], and the agent's framing of communication. In line with Dzindolet et al.'s findings [21], we expect trust and evaluation of the agent to be higher with a high transparency interface than with a low transparency interface. When the agent is more transparent, and critical of the participant's plan decisions (critical framing), it will be perceived better and trusted more than a complimentary agent as it highlights reasons for error. On the other hand, we expect that a more opaque yet complimentary agent will increase trust in the agent by helping to counteract the overweighing of faults.

3.1 Methodology

Twenty-nine students from an American university were recruited for cash payment. Data were analyzed for 26 (17 men, 12 women, $M_{\text{age}} = 20.03$, $SD_{\text{age}} = 2.09$). Three were omitted from analysis due to technical issues. This experiment involved a 2x2 mixed design with agent transparency as the within-subjects independent variable and communication framing as the between-subjects independent variable. Agent transparency was tested at two levels: (a) L1+2: containing reasoning information, and (b) L1+2+3+U containing reasoning and projection *with* projection uncertainty information. Communication framing was tested as two contrasting attitudes from the agent: (a) *Critical:* highlighting a parameter of the chosen plan that is not satisfied, and (b) *Complimentary:* highlighting a parameter of the chosen plan that is optimal. The user interface varied per condition by showing corresponding pieces of SAT-level information on a map display, in text, and on a sliding bar scale (Figure 5). Prior to the experimental trials, participants received about 1 hour of training. The experiment was divided into 2 blocks of 8 missions. Transparency order and communication framing were counterbalanced within sets of four participants, within which the scenarios where the agent's recommendations were correct and incorrect were held constant. The choice of correct and incorrect scenarios was randomized for each set but kept the 5 correct and 3 incorrect ratio described below.

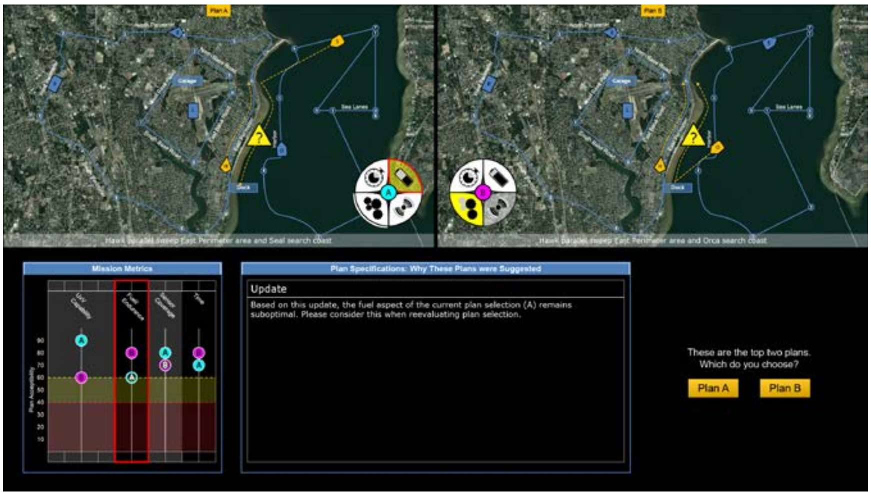

Plan A
Plan B
Hawk parallel sweep East Perimeter area and Seal search coast
Hawk parallel sweep East Perimeter area and Orca search coast
Mission Metrics
UxV Capability
Fuel Endurance
Sensor Coverage
Time
Plan Acceptability
90
80
70
60
50
40
30
20
10
Plan Specifications: Why These Plans were Suggested
Plan A
Plan B

Plan A is recommended because Seal has the capability to search in the shallow waters along the shoreline. This is considered the most important factor for the task. The plan weaknesses are deemed acceptable.

Plan A	Plan B
It is expected that Seal will find the Soldier by searching the shallow waters along the shoreline.	It is expected that Orca will have the endurance for a lengthy search.
It is uncertain if Seal has the endurance to carry out a prolonged search.	It is uncertain how effective Orca's sensor coverage will be.

These are the top two plans.
Plan A is the preferred plan.
Which do you choose?

(a)


Plan A
Plan B
Hawk parallel sweep East Perimeter area and Seal search coast
Hawk parallel sweep East Perimeter area and Orca search coast
Mission Metrics
UxV Capability
Fuel Endurance
Sensor Coverage
Time
Plan Acceptability
90
80
70
60
50
40
30
20
10
Plan Specifications: Why These Plans were Suggested
Plan A
Plan B

Update

Based on this update, the fuel aspect of the current plan selection (A) remains suboptimal. Please consider this when reevaluating plan selection.

These are the top two plans.
Which do you choose?

(b)

Figure 5. Multiple-heterogeneous UxV control mission plan decision (a) and mission update decision (b) screenshots.

Using the interface presented in Figure 5a, participants were presented with two plan options to complete each mission. Both plans were viable and were balanced in their fulfillment of the mission needs, but optimal plans prioritized the commander's intent and addressed more mission goals and needs. The agent always recommended one plan over the other, and the participant's task was to select the best plan based on current mission requirements. After the participant made the initial selection, an update informed the participant and agent of an aspect of the scenario which had evolved after the initial plans were generated (Figure 5b). Once the update was acknowledged, the agent re-assessed the plans and informed the participant of a parameter that had either *remained or become suboptimal* (criticizing condition) or had *remained or become optimal* (complimentary condition). The participant confirmed or changed the plan based on the new information. While the transparency manipulation applied throughout the entire interaction, the communication framing manipulation only applied post-update. Reliability of the agent was held constant such that its initial recommendation was correct for 5 of every 8 scenarios. The agent's post-update assessment was correct in all scenarios (though no actual recommendation was made for the post-update plan decision). Participants assessed their trust and perception of the agent after each block of scenarios. Within each block, participants also filled out items for usefulness and reliance following each of the eight initial mission plan decisions.

3.2 Results

To investigate the influence of both transparency and framing, as well as to account for order effects, a series of 2 (low & high transparency) x 2 (complimentary & critical framing) x 2 (transparency block order) mixed model ANOVAs were run with transparency as the within-subjects variable. We reported here only main effects and interactions for transparency and framing; including order in the analysis is important to account for order effects, but is not of specific interest here. Correlations were then run to help with interpretation. In light of relatively low statistical power, we reported trends with a *p*-value greater than .05 but less than .10.

3.2.1 Operator Performance

There were no significant task performances differences. However, for agreement with the agent, there was a main effect for transparency, *F*(1,26) = 4.28, *p* = .049, η^2_p = .141, as well as an interaction between transparency and framing, *F*(1,26) = 5.35, *p* = .029, η^2_p = .171. Agreement with the complimentary agent was consistent between transparency conditions. In contrast, agreement with the critical agent was higher in the low transparency automation condition than in the high transparency condition (Figure 6).

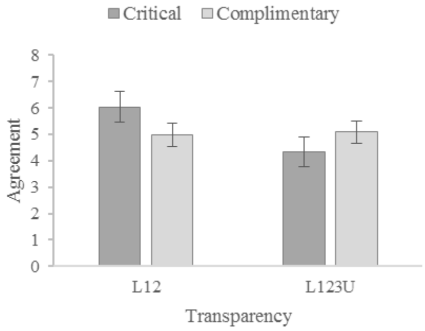


Figure 6. Mean number of scenarios per block participants agreed with the agent.

3.2.2 Operator Trust in Agent

Trust results were mixed. Trust Survey 1 [26] revealed no significant findings, but a main effect trend for framing, *F*(1,26) = 3.97, *p* = .057, η^2_p = .132. Participants were more trusting when the agent was critical (Figure 7). With regard to trust in the agent's ability to integrate and display analyzed information, Trust Survey 2 [23] revealed a significant main effect for transparency, *F*(1,26) = 6.5, *p* = .017, η^2_p = .200. There were no other significant findings, but a trend suggests the possibility of an interaction between transparency and framing were there more statistical power, *F*(1,26) = 3.4, *p* = .077, η^2_p = .116. Participants were relatively distrustful of the low transparency complimentary agent (Figure 8).

3.2.3 Perceptions of Agent

There were main effects of both transparency, *F*(1,26) = 4.9, *p* = .036, η^2_p = .16, and a non-significant interaction trend between the variables, *F*(1,26) = 3.9, *p* = .058, η^2_p = .13, for perceptions of agent aptitude with regard to integrating and displaying information [27]. Regarding perceptions of agent aptitude in suggesting plan decisions, there were also main effects for transparency, *F*(1,26) = 5.4, *p* = .028, η^2_p = .17, and framing, *F*(1,26) = 4.8, *p* = .038, η^2_p = .16, but an interaction was not significant *p* > .10. In both cases, participants perceived the agent to be more apt when transparency was high. Perceptions of agent aptitude when integrating and displaying information were higher when the agent framed the update plan critically (Figure 9). For suggesting decisions, the interaction was driven by the difference between the low transparency complimentary condition and the other three conditions. There were no significant (or trends of) differences between condition in perceived automation reliability (*p* > .10).

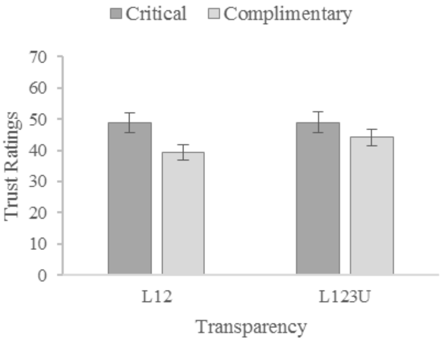


Figure 7. Trust Survey 1 ratings

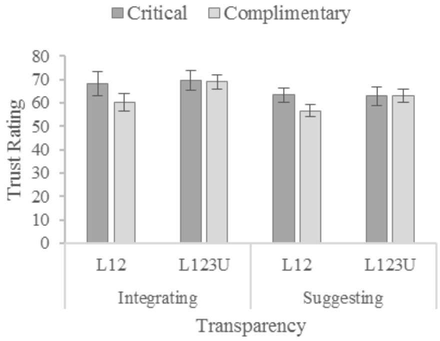


Figure 8. Trust Survey 2 ratings

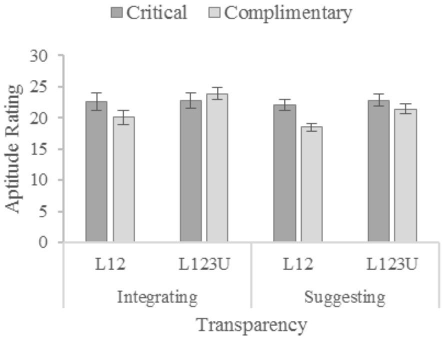


Figure 9. Automation aptitude ratings

3.2.4 Post Scenario Ratings

There was a strong main effect of transparency, $F(1,26) = 49.59$, $p < .001$, $\eta^2_p = .656$, and of framing $F(1,26) = 11.87$, $p = .002$, $\eta^2_p = .314$, for ratings of automation interface usefulness; participants found the high transparency and critical agent configurations to be more useful, respectively (Figure 10). For reliance, there were also main effects for transparency, $F(1,26) = 46.82$, $p < .001$, $\eta^2_p = .643$, and framing, $F(1,26) = 12.56$, $p = .002$, $\eta^2_p = .326$. Participants reported more reliance when transparency was high and the update assessment was critical in nature (Figure 11).

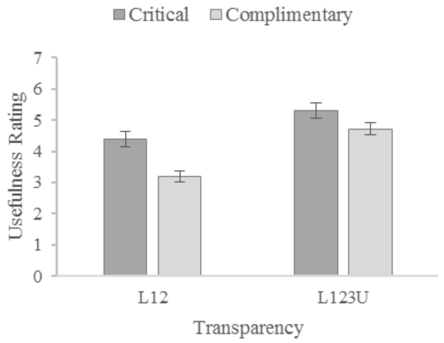


Figure 10. Ratings of usefulness of the automation's transparency information interface.

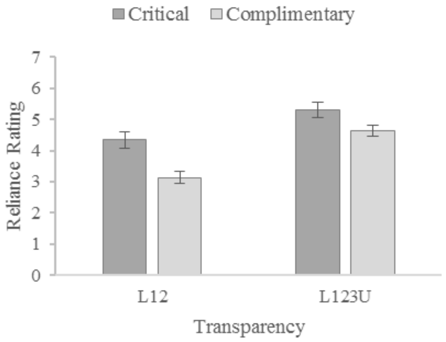


Figure 11. Reliance on the automation's transparency information interface.

4.0 CONCLUSIONS

Results from these studies completed as part of the IMPACT program yielded several insights to the utility of transparency, as well as best practices for implementing information transparency as part of an intelligent agent's interface. Primarily, it was found that agent transparency, as operationalized and implemented according to Chen et al.'s SAT model, is useful for improving performance in complex decision making such as that done in multi-UxV management tasks. Additionally, this performance is increased without a cost to workload. However, it should be noted that response time does increase for a few seconds, which may or may not create an issue depending on the mission environments. Overall, the results showed that perceptions of the IA (its aptitude and usefulness) increased with agent transparency. Trust survey results showed that the complimentary IA was rated as less trustworthy compared with the critical IA, especially in low transparency conditions. Participants tended to agree more with the critical IA when it was more opaque than when it was more transparent; on the other hand, participants' agreements with the complimentary IA did not differ regardless of the IA's transparency.

Additional research is needed which explores human-agent teaming in a more bidirectional manner. Future studies will examine human-agent teaming with bidirectional communications to evaluate the utility of SAT-based interfaces in a dynamic manner [4]. This can be used to inform the design of field research being done with finalized intelligent agents that are capable of behaving independently. Additional efforts will also focus on developing a repository of user interface design elements (e.g., visualizations) to support SAT-based interfaces.

5.0 REFERENCES

[1] Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence.* Knopf, New York.

[2] Husain, A. (2017). *The Sentient Machine: The Coming Age of Artificial Intelligence.* Simon & Schuster, New York.

[3] 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,* 13-29.

[4] Chen, J., Lakhmani, S., Stowers, K., Selkowitz, A., Wright, J., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomic Science, 19,* 259-282.

[5] Gunning, D. (2016). Explainable Artificial Intelligence. http://www.darpa.mil/program/explainable-artificial-intelligence.

[6] Defense Science Board. (2016). Defense Science Board Summer Study on Autonomy. *Washington, D.C.: Under Secretary of Defense.*

[7] J. Y. C. Chen, K. Procci, M. Boyce, J. Wright, A. Garcia, & M. Barnes. (2014). *Situation awareness-based Agent Transparency* (Technical Report: ARL-TR-6905). Aberdeen Proving Ground MD: US Army Research Laboratory.

[8] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37,* 32–64.

[9] Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *J. Cog. Eng. & Decision Making, 9,* 4–32.

[10] Mercado, J., Rupp, M., Chen, J., Barber, D., Procci, K., & Barnes, M. (2016). Intelligent agent transparency in human-agent teaming for multi-UxV management. *Human Factors, 58,* 401–415.

[11] Bass, E. J., Baumgart, L. A., & Shepley, K. K. (2013). The effect of information analysis automation display content on human judgment performance in noisy environments. *J. Cog. Eng. and Decision Making, 7,* 49–65

[12] Helldin, T. (2014). *Transparency for Future Semi-Automated Systems* (Doctoral Dissertation). Orebro University.

[13] Meyer, J., & Lee, J. (2013). Trust, reliance, compliance. In J. Lee and A. Kirlik (Eds.), *The Oxford Handbook of Cognitive Engineering*. Oxford, UK: Oxford University Press.

[14] McGuirl J. M., & Sarter, N. B. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Human Factors*, *48,* 656–665.

[15] Beller, J. Heesen, M., & Vollrath, M. (2013). Improving the driver–automation interaction: An approach using automation uncertainty. *Human Factors*, *55,* 1130–1141.

[16] Calhoun G.L., Ruff, H.A., Behymer, K.J., & Frost, E.M. (2018) Human-autonomy teaming interface design considerations for multi-unmanned vehicle control, *Theoretical Issues in Ergonomics Science, 19*, 321-352.

[17] Fern, L., & Shively, R. J. (2009). A comparison of varying levels of automation on the supervisory control of multiple UASs. *Proc. AUVSIs Unmanned Systems North America* (pp. 10-13).

[18] Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors*, *49,* 57-75.

[19] Draper, M. et al. (2018). Realizing autonomy via intelligent adaptive hybrid control: Adaptable autonomy for achieving UxV RSTA team decision superiority (also known as Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT)) (Tech Rep. AFRL-RH-WP-TR-2018-0005). Wright-Patterson Air Force Base, OH: U.S. Air Force Research Laboratory.

[20] Wickens, C., & Dixon, S. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8,* 201-212.

[21] Dzindolet, M., Peterson, S., Pomranky, R., Pierce, L., & Beck, H. (2003). The role of trust in automation reliance," *Int. J. Human-Computer Studies, 58,* 697-718.

[22] Hart, S., & Staveland, L. (1988). Development of NASA TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: Elsevier.

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

[24] Parasuraman, R. Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Trans. Syst., Man, Cybern A, Syst., Hum., 30,* 286–297.

[25] Tanner W. P. Jr, & Swets, J.A. (1954). A decision-making theory of visual detection. *Psychological review, 61,* 401-409.

[26] Lyons, J. B., Koltai, K. S., Ho, N. T., Johnson, W. B., Smith, D. E., and Shively, R. J. (2016). Engineering trust in complex automated systems. *Ergon. Des. Q. Hum. Factors Appl., 24,* 13–17.

[27] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q., 13,* 319,