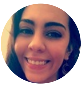
Intelligent Agent Transparency: The Design and Evaluation of an Interface to Facilitate Human and Intelligent...

Conference Paper · September 2016

DOI: 10.1177/1541931213601392

Kimberly Stowers
University of Central Florida

Nicholas Kasdaglis
Florida Institute of Technology



Olivia B. Newton
University of Central Florida

Ryan Wohleber
University of Central Florida

Intelligent Agent Transparency: The Design and Evaluation of an Interface to Facilitate Human and Intelligent Agent Collaboration

Kimberly Stowers[1], Nicholas Kasdaglis[2], Olivia Newton[1], Shan Lakhmani[1], Ryan Wohleber[1], Jessie Chen[3]
[1]University of Central Florida, [2]Florida Institute of Technology, [3]U.S. Army Research Laboratory

We evaluated the usability and utility of an unmanned vehicle management interface that was developed based on the Situation awareness–based Agent Transparency model. We sought to examine the effect of increasing levels of agent transparency on operator task performance and perceived usability of the agent. Usability and utility were assessed through flash testing, a focus group, and experimental testing. While usability appeared to decrease with the portrayal of uncertainty, operator performance and reliance on key parts of the interface increased. Implications and next steps are discussed.

Unmanned vehicles (UxVs) are an increasingly important tool for military and commercial operations (Lewis, 2013). A current goal in UxV operations is to allow a single operator to manage multiple UxVs (Cummings, Clare, & Hart, 2010). This can be done by allowing an intelligent agent (IA) to manage the UxVs for a human operator (Chen & Barnes, 2014). Such an approach to UxV management is under investigation through many efforts, including the Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT) project funded by the U.S. Department of Defense's Autonomy Research Pilot Initiative (Draper, 2013).

There are numerous challenges to proper reliance on an IA. Human operators may find it challenging to maintain situation awareness (SA) due to difficulty keeping up with the overwhelming amount of information the IA must obtain and manipulate to facilitate its management of a multi-UxV system. Additionally, human operators may be unable to ascertain the logic behind decisions made by the IA or the impact those decisions might have on the future states of the mission (Linegang et al., 2006). These difficulties may lead to degraded performance and an inability to complete vital tasks (Chen & Barnes, 2012a).

In order to ensure that operators maintain SA and adequate "human on the loop" performance (Maybury, 2012), it is necessary to present the human operator with information specific to the purpose, process, and performance of the automated system (Lee & See, 2004). In line with this approach, we examined transparency as a tool to facilitate appropriate trust in the IA. We focused on the display of transparency information on an IA interface that was used in mission planning scenarios. Here we detail our approach to conceptualizing agent transparency, our implementation of transparency, and the impact of transparency information on the usability of an IA interface.

AGENT TRANSPARENCY

As conceptualized by Endsley (1995), SA is characterized by three distinct levels: Level 1, *perception*; Level 2, *comprehension*; and Level 3, *projection*. Chen and colleagues (2014) proposed the SA-based Agent Transparency (SAT) to support these three levels of SA while accounting for unique requirements that arise from human-agent teaming. The SAT model formalized IA information processing along three levels of abstraction: SAT Level 1, *current actions and plans*; SAT Level 2, *reasoning*; SAT Level 3, *projection*. The first SAT level (L1) is characterized by information regarding the agent's state and goals relating to its environment. The second SAT level (L2) includes information about the agent's reasoning process (e.g., constraints and affordances that the agent considers). The third SAT level (L3) involves information regarding the agent's prediction of future outcomes, as well as uncertainty (U; e.g., environmental and temporal) that the agent is dealing with.

The present effort sought to use the SAT model to inform the design of an IA interface for UxV management. The aim of this effort was to understand how transparency might (1) improve operators' maintenance of SA and (2) facilitate appropriate trust and reliance on the IA as a decision support system. Generally, the SAT model informed implementation of transparency in the following ways. Under L1 of the SAT model, the interface informed the operator of the IA's current actions and plans. With L2 of the SAT model, the interface informed the operator of its reasoning regarding these plans. L3 of the SAT stipulates that a system provide the operator with projected outcomes of the IA's suggestions. Accordingly, the present interface displayed projections, or expected future states of the world based on the various options for action that the IA presented.

L3 also stipulates that the interface communicate the *uncertainty* with which these projections are made. This notion of uncertainty is inherent to sensing, reasoning and projection. For example, knowledge of the state of the world is likely never fully known by any sensing or perceiving agent. Therefore, suggestions by the IA require contention with the notion of uncertainty. The IA must make assumptions about what it believes to be true (or will be true of its environment, constraints, and resources). As an important consideration in the reasoning process of the IA and its ability to provide projections based on possible courses of action, the SAT model (Chen et al., 2014) suggests, under L3, that these uncertainties and assumptions be made explicit to the human operator. As such, this was considered a requisite design consideration in our work.

EXPERIMENTAL TASK AND DESIGN

Before explaining the design of our interface, it is first important to clearly identify and detail the task in which the interface was used. In line with the larger IMPACT project, our task utilized a play-calling scenario involving multiple UxVs performing military defense (Behymer, Mersch, Ruff, Calhoun, & Spriggs, 2015). The idea of play-calling is based on Miller and colleagues' (2005) playbook approach to adaptive automation wherein plans of action are based on fixed templates of UxV selection and maneuvers, called "plays." These plays have the advantage of being quick to call, commonly understood by the operator and automation, and consistent in different situations. In the present tasking, operators were given missions, which were described through several separate messages containing intelligence and information about the commander's intent at the beginning of each scenario. The IA then presented the operator with two "plays" or plans for completing the mission successfully, and the operator's task was to select the better plan based on the information received.

For our experiment, we intended to test the effect of the interface's transparency (based on the SAT model) on the operators' performance and perception of the IA. Our design closely matched a similar study on agent transparency by Mercado and colleagues (Mercado, Rupp, Chen, Barnes, Barber, & Procci, 2016). Their study utilized three conditions containing the following levels of SAT information, respectively: L1, L1+2, and L1+2+3. Their effort only examined uncertainty within L3 as described above. We also created three conditions, but with an additional focus on uncertainty as a separate variable: L1+2, L1+2+3, and L1+2+3+U. Each of these conditions were represented by increasingly transparent interfaces, as discussed next.

INTERFACE DESIGN AND DEVELOPMENT

The above task was embedded in an adapted version of the IMPACT Fusion interface (Rowe, Spriggs, & Hooper, 2015). In adapting this interface, we sought to meet the following goals: (1) display all three SAT levels, (2) maintain scalability in the display of SAT, and (3) maintain ecological validity of the overall interface. First, SAT levels needed to be represented in a way that was directly perceivable (Gibson, 1977) and actionable (Kasdaglis & Deaton, 2012; Kasdaglis, Newton, Lakhmani, 2014) so that the operator had a sense of what was possible and what could be done (Norman, 2002). Next, the display's communication of the IA's status and plans, reasoning, and projections had to be scalable in order to handle an increase in complexity (e.g., additional UxVs and mission dependency factors). Finally, since IMPACT's testbed is intended for highly trained military operators, concerns of eventual operators had to be weighed against the available experimental participant pool of mostly non-military affiliated college students. We compromised between the needs of the two populations by simplifying the interface enough to be usable by the student population, while also keeping a reasonable representation of factors that would be presented in a real-world environment.

Display of Transparency Information

The IMPACT testbed and the interface we adapted from it are modular in nature, allowing us to display the IA's status and plans, reasoning, and projection in a scalable and ecologically valid way. Careful consideration led to the final inclusion of three key modules or tiles in our display: a *Plan Maps* tile, a *Projected Plan Success* tile, and a *Plan Specifications* tile. These tiles were necessarily different from each other in the way they presented information, as our goal was to make the SAT levels as salient as possible through multiple means of display. Other tiles included in the display were the mission briefs, alerts, and decision tiles. These provided information about the state of the world, as opposed to SAT information, so they are not discussed here. The three key modules will be explored in more detail below.

Plan Maps. The centerpiece to our interface, like the testbed it is adapted from, is the map of the environment in which the UxVs are operating. Our interface presents two maps, one for each plan the IA suggests to the operator (Figure 1). Maps are a necessary part of any play-calling scenario involving UxVs as it allows the operator to see where UxVs are located and where they are going. Maps are also a useful backdrop for embedding important information about UxVs and we thus sought to utilize it with that in mind.

We sought to display L2 and L3 SAT (reasoning and projections – including uncertainty) on the map according to the corresponding conditions (L1+2, L1+2+3, and L1+2+3+U). L2 information was displayed by the size of UxVs, with larger UxVs being faster. L3 information was displayed as an icon attached to each UxV to depict the number of time units it was from its goal location. Uncertainty information was displayed through changes in opacity of the UxVs themselves, where less opaque vehicles were considered to have uncertain capabilities or time projections. These cues are meant to convey the meaning of the data they represent without the need to rely on existing mental models (Kosara, Healey, Interrante, Laidlaw, & Ware, 2003).

Figure 1: Plan Maps tile for condition 3. A map for each plan is included, with SAT levels 1, 2, 3, and uncertainty information displayed.

Projected Plan Success. In addition to the map, we determined it was important to present all levels of SAT information in one scalable graphic, which we titled Projected Plan Success (Figure 2). To do this, we used a plan comparison tool developed by U.S. Air Force researchers (Behymer et al., 2015). This graphic shows the evaluation of both plans (A and B) on four different parameters: time, coverage, fuel endurance, and general capability. Circles representing each plan slide up and down scales corresponding to each parameter. Plan acceptability is rated on this sliding scale; the higher a circle is on the scale, the more acceptable the plan is for meeting that parameter.

The information presented on the Projected Plan Success tile increases with each condition in order to display increasing amounts of SAT information. In condition L1+2, this tile shows how heavily the IA is weighing each parameter as a representation of the IA's recommended plan and reasoning. This is done by ordering the parameters from left to right, with decreasing widths for less important ones. In condition L1+2+3, the tile additionally portrays L3 SAT information by showing the projected plan success of each plan for all four parameters. In condition L1+2+3+U, uncertain plans were represented with hollow circles instead of filled ones.

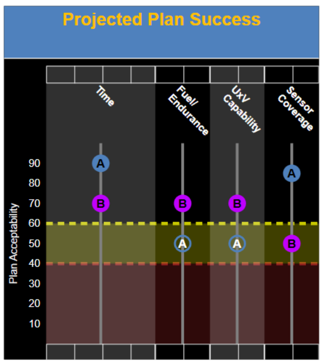


Figure 2: Projected Plan Success tile for condition 3. Hollow circles represent uncertain parameters for each plan.

Plan Specifications. The Plan Specifications tile displays three levels of SAT via systematically written text (Figure 3). We adapted the use of this tile from Mercado and colleagues' work (Mercado et al., 2016). In our adaptation, information was added in a structured way for each condition to enhance operator ability to read the information (Johnson, 2014). For condition L1+2, two statements regarding the IA's reasoning were included for each plan. For condition L1+2+3, a projection statement was added which started with "It is expected that…." Finally, for condition L1+2+3+U, an uncertainty statement specified two points: (1) the IA was uncertain about certain aspects of the tasking environment, and (2) the IA was making an assumption to deal with it.

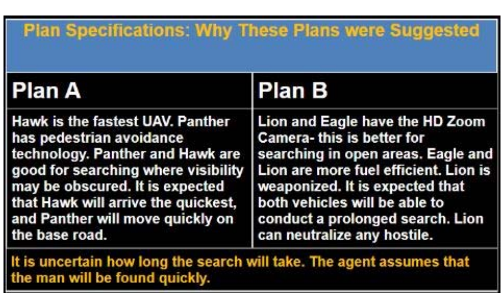


Figure 3: Plan Specifications tile for condition 3. The orange text at the bottom represents uncertain parameters for each plan.

INTERFACE EVALUATION

To test our implementation of transparency information, we evaluated our interface in two ways: (1) expert assessment and (2) experimental assessment. For the expert assessment, a convenience sample of four human factors specialists were asked to evaluate the design of the interface, the presentation of information in each tile, and how these elements could be used together to inform a decision-making task. This evaluation consisted of flash testing (Boag, 2010), open-ended text questions, and a focus group discussion. Images and storyboards were used for this evaluation. For experimental assessment, a sample of 53 young adults used our simulation while we monitored their performance and collected self-reported data. The results from both of these assessments are presented below.

Expert Assessment

Four human factors specialists were given a flash test of the interface components (for an explanation of this method, see Boag, 2010), and a brief introduction to the overall interface, after which the four specialists evaluated each tile individually. The specialists were invited to discuss how interface elements communicated the relevant information and potential areas of confusion.

From the flash testing and focus group outcomes, we revised the design of several elements. For example, various elements were difficult for the specialists to interpret including variation in icon letter sizes, wording, and, phraseology as well as the meaning of icon coloring and size. Training was adjusted to include checks of understanding to address these interpretation issues. Further, colors and contrast were enhanced and increased (respectively) to better distinguish them from background elements. We also used the specialists' assessments to inform the development of the training that was to be presented to experiment participants.

Experimental Evaluation

After suggestions from the expert assessment were implemented, 53 young adults were recruited for an experimental evaluation of our interface. As stated previously, we tested three conditions: L1+2, L1+2+3, and L1+2+3+U. Conditions were counterbalanced in a within-subjects design, with each participant receiving several play-calling scenarios per condition. Prior to the experimental task, participants received approximately 45 minutes of training to familiarize themselves with the interface and its components. This included both a narrated PowerPoint presentation and several training missions with the simulator itself.

During the simulation, several performance-related measures were collected, including correct responses, response time, subjective workload, trust, and subjective SA. In addition to performance-related measures, we collected information on participants' perceived usability of the interface and self-reported interface utility for each major part of the interface. The full results regarding participant performance using this interface are reported elsewhere (Stowers, Newton, Rupp, Kasdaglis, Chen, & Barnes, in progress). However, here we will focus only on performance as it relates to usability.

Performance. For each event, neither plan A or B were considered "wrong." However, one of the plans was always more optimal than the other and thus considered the "correct" response. Based on this guideline, total correct responses for each condition were calculated as a ratio ranging from 0 to 1 and compared using a repeated measures ANOVA. A significant main effect of transparency was found, $F(2, 104) = 10.31$, $p < .001$, $\eta_p^2 = .17$; performance was highest in condition L1+2+3+U ($M = .76$, $SD = .17$) and lowest in condition L1+2 ($M = .62$, $SD = .20$).

Usability. To assess usability for each interface condition, we conducted a repeated-measures ANOVA on responses to the System Usability Scale (Brooke, 1996). Results revealed that condition L1+2+3 had the highest perceived usability ($M = 72.86$, $SD = 15.82$) while condition L1+2+3+U had the lowest ($M = 68.44$, $SD = 18.20$), $F(2, 94) = 3.65$, $p = .03$, $\eta_p^2 = .07$. This is an interesting contrast to the performance data. Although participants indicated that the interface in condition L1+2+3+U had the poorest usability, performance with that interface was the highest.

Interface Utility. Interface utility was measured on a 7-point Likert scale with the question "How much did you rely on "X", where "X" was replaced by each key module of the interface – the Plan Maps tile, the Projected Plan Success tile, and the Plan Specifications tile. Participant responses were analyzed using a repeated-measures ANOVA. A significant effect of transparency was found for use of the Projected Plan Success tile, $F(2, 104) = 32.64$, $p < .001$, $\eta_p^2 = .386$. Participants claimed to use the Projected Plan Success tile most in condition L1+2+3 ($M = 5.10$, $SD = 1.22$), and least in condition L1+2 ($M = 3.38$. $SD = 1.72$). No significant differences were found in regards to the other tiles. Across conditions, participants claimed to rely most heavily on the Plan Specifications tile.

DISCUSSION

We designed and evaluated the user interface of an IA for multi-UxV management based on the SAT model (Chen et al., 2014). The findings from our expert assessment and experimental evaluation together provided insights for the effective display of transparency information. Experimental evaluation showed that, while operator performance increased with transparency level, perceived usability did not. There may be several explanations for this difference in outcomes. The most straightforward explanation is that the interface design in condition L1+2+3+U is simply less usable than that in condition L1+2+3 because of the way uncertainty information was displayed in condition L1+2+3+U. However, success using the interface in condition L1+2+3+U suggests that objective usability may not be the only explanation. To speculate further on objective usability, follow up efforts will need to conduct more targeted usability testing. For now, it is worth considering other explanations.

Another possibility is that operators did not like the increase in task-load despite the utility of the added information. This interpretation is supported by the findings of Yeh and Wickens (1988) who also saw a decrease in perceived usability with increased task-load despite increased or no difference in performance.

If task-load is the source of the usability and performance discrepancy, one solution may lie in the development of an adaptive interface that can adjust task-load based on physiological indicators of workload and fatigue. Such an interface could optimize performance at the expense of perceived usability unless the operator was overloaded or fatigued to help maintain optimal reliance on automated aids.

Finally, it is possible that condition L1+2+3+U was the least aesthetically pleasing. Aesthetics have often been linked to perceived usability such that more beautiful interfaces tend to receive higher perceived usability scores despite having identical layout and menu structures to the less beautiful interfaces with which they are compared (c.f. Tractinsky, Katz, & Ikar, 2000). It is possible that the additional information in condition 3 made the interface seem cluttered, causing users to see the interface as less attractive and usable.

This is an interesting consideration as it may explain both the increase in performance and decrease in perceived usability. Since condition L1+2+3+U had a large amount of text in comparison to the other two conditions, and participants indicated they were relying more on this than the other interface modules, they may have felt frustration with the amount of reading required to evaluate plan options, or they may have thought it did not look as attractive. Yet they still may have benefitted from the supplementary transparency information provided in the tile. If aesthetics is the source of the low usability scores, it will be necessary to consider more aesthetically pleasing ways to incorporate additional information into the interface while still enhancing performance.

Regardless of the explanation for the discrepancy, there is evidence in previous work of a dissociation between usability and performance that supports the idea that usability and performance do not always coexist. For example, Sondregger and Sauer (2010) found that what appears to be most usable was linked to poor performance, whereas Dillon (2002) found that poor perceived interface design was linked to poor performance (Dillon, 2002). This dissociation means that developers may be faced with a difficult choice between keeping operators content and optimizing performance. This fact would also be important to consider should operators be permitted to control the level of transparency afforded to them by the interface. Operators may be inclined to decrease the level of transparency at the expense of performance.

A potential solution to the dissociation between usability and performance may lie in the development of an adaptive graphical display that effectively represents textual information that appears often, thus reducing the amount of text that needs to be read.

CONCLUSION

Several concerns were raised during the expert assessment session, many of which were addressed by changes to the interface, but some of which were addressed through increased training. Experimental results did not necessarily corroborate certain complaints raised during the focus group, but instead offered supplemental information, thus highlighting the need to consider usability alongside performance. While expert assessment helped us to improve our displays prior to implementation, the experimental data shows that there may still be areas that can be fine-tuned, such as communication of information on the Plan Maps and on the Projected Plan Success tile. While the understanding of this information may be trainable, especially in expert populations, the more intuitive these areas of the interface are, the easier it will be to train operators and achieve improvements in mission success.

Future research should examine in more detail the dissociation between performance and usability, and explore how this may relate to the display of uncertainty in interfaces. Research should also investigate the utility of adaptive levels of transparency as a possible solution to tailoring interfaces for timeliness. Recent work has identified interface adaptiveness as a necessary feature of successful automation implementation (Gutzwiller et al., 2015) and interface design in general (Preece, Sharp, & Rogers, 2015). It may be essential for successful human-machine teams, especially as increases in automation technology lead to interfaces that are more complex.

ACKNOWLEDGMENT

This research was supported by the DoD Autonomy Research Pilot Initiative, under the Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT) project. The authors wish to thank Michael Rupp, Jonathan Harris, and Michael Barnes for their contribution to the design of these interfaces. The authors would also like to thank Mark Draper and Gloria Calhoun for their input.

REFERENCES

Behymer, K. J., Mersch, E. M., Ruff, H. A., Calhoun, G. L., & Spriggs, S. E. (2015). Unmanned vehicle plan comparison visualizations for effective human-autonomy teaming. *Procedia Manufacturing*, *3*, 1022-1029.

Boag, P. (2010, October 29) Making user testing happen [Blog post]. Retrieved from https://boagworld.com/usability/testing-tools/

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4-7.

Chen, J. Y. C. & Barnes, M. J. (2012a). Supervisory control of multiple robots: Effects of imperfect automation and individual differences. *Human Factors, 54*(2), 157-174. doi: 10.1177/0018720811435843

Chen, J. Y. C. & Barnes, M. J. (2012b). Supervisory control of multiple robots in dynamic tasking environments. *Ergonomics, 55*(9), 1043-1058. doi: 10.1080/00140139.2012.689013

Chen, J. Y. C., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). Situation awareness-based agent transparency (ARL-TR-6905). Aberdeen Proving Ground: Army Research Laboratory.

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.

Cummings, M. L., Clare, A., & Hart, C. (2010). The role of human-automation consensus in multiple unmanned vehicle scheduling. *Human Factors (52)1,* 17-27.

Dillon, A. (2002). Beyond usability: process, outcome and affect in human-computer interactions. *Canadian Journal of Library and Information Science.*

Draper, M, H. (2013). Realizing autonomy via intelligent adaptive hybrid control: Adaptable autonomy for achieving UxV RSTA team decision superiority. *OSD ASD (R&E) Autonomy Research Pilot Initiative (ARPI).* Volume 1.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32-64.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward and ecological psychology* (pp. 67-82). Hillsdale, NJ: Lawrence Erlbaum.

Goodrich, M. A. (2010). On maximizing fan-out: towards controlling multiple unmanned vehicles. In M. Barnes & F. Jentsch, (Eds.), *Human–robot interactions in future military operations* (pp. 375-395). Surrey, UK: Ashgate.

Grote, G. (2009). *Management of uncertainty: Theory and application in the design of systems and organizations.* Springer.

Gutzwiller, R. S., Lange, D. S., Reeder, J., Morris, R. L., & Rodas, O. (2015). Human-computer collaboration in adaptive supervisory control and function allocation of autonomous system teams. In *Virtual, Augmented and Mixed Reality* (pp. 447-456). Springer.

Johnson, J. (2014). *Designing with the Mind in Mind* (2nd ed.). Waltham, MA: Morgan Kaufman.

Kasdaglis, N., & Deaton, J. (2012). The art of managing information in the national airspace architecture. In J. B. Strother, J. M. Ulijn & Z. Fazal (Eds.), *Information overload; an international challenge for professional engineers and technical communicators* (pp. 195). Hoboken, NJ: Wiley.

Kasdaglis, N., Newton, O., & Lakhmani, S. (2014). System state awareness: A human centered design approach to awareness in a complex world. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*(1), 305-309.

Kosara, R., Healey, C. G., Interrante, V., Laidlaw, D. H., & Ware, C. (2003). User studies: Why, how, and when?. *IEEE Computer Graphics and Applications 4*, 20-25.

Lee, J. D. & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80. doi: 10.1518/hfes.46.1.50_30392

Lewis, M. (2013). Human interaction with multiple remote robots. *Reviews of Human Factors and Ergonomics*, *9*(1), 121-174. doi: 10.1177/1557234X13506688

Linegang, M., Stoner, H. A., Patterson, M. J., Seppelt, B. D., Hoffman, J. D., Crittendon, Z. B., & Lee, J. D. (2006). Human-automation collaboration in dynamic mission planning: A challenge requiring an ecological approach. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, 2482-2486. doi: 10.1177/154193120605002304

MacLean, A., Young, R. M., Bellotti, V. M., & Moran, T. P. (1991). Questions, options, and criteria: Elements of design space analysis. *Human–computer Interaction, 6*(3-4), 201-250.

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

Miller, C., Funk, H., Wu, P., Goldman, R., Meisner, J., & Chapman, M. (2005, September). The Playbook™ Approach to Adaptive Automation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 49*(1) 15-19.

Murphy, R. R. & Burke, J. L. (2010). The safe human-robot ratio. In M. J. Barnes & F. G. Jentsch (Eds.), *Human-robot interactions in future military operations* (pp. 31-49). Surrey, UK: Ashgate Publishing.

Norman, D. A. (2002). *The design of everyday things* New York, NY: Basic books.

Preece, J., Sharp, H., & Rogers, Y. (2015). *Interaction Design-beyond human-computer interaction*. Chichester, UK: John Wiley & Sons.

Rowe, A., Spriggs, S., & Hooper, D. (2015, May). Fusion: a framework for human interaction with flexible-adaptive automation across multiple unmanned systems. In *Proceedings 18th Symposium on Aviation Psychology.* Dayton, OH.

Sonderegger, A., & Sauer, J. (2010). The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied ergonomics*, *41*(3), 403-410.

Stowers K., Newton O., Rupp, M., Kasdaglis N., Chen J.Y.C., Barnes M. (in progress). The Importance of Agent Transparency in Collaborative Decision-Making among Human Agent Teams.

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with computers*, *13*(2), 127-145.

Yeh, Y. Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors*, *30*(1), 111-120.

UNCLASSIFIED//APPROVED FOR PUBLIC RELEASE, DISTRIBUTION IS UNLIMITED