

Insights into Human-Agent Teaming: Intelligent Agent Transparency and Uncertainty

Kimberly Stowers, Nicholas Kasdaglis, Michael Rupp, Jessie Chen,
Daniel Barber and Michael Barnes

Abstract This paper discusses two studies testing the effects of agent transparency in joint cognitive systems involving supervisory control and decision-making. Specifically, we examine the impact of agent transparency on operator performance (decision accuracy), response time, perceived workload, perceived usability of the agent, and operator trust in the agent. Transparency has a positive impact on operator performance, usability, and trust, yet the depiction of uncertainty has potentially negative effects on usability and trust. Guidelines and considerations for displaying transparency in joint cognitive systems are discussed.

Keywords Transparency · Human factors · Human-Machine interaction · Systems engineering · Supervisory control · Unmanned vehicles

K. Stowers · N. Kasdaglis · M. Rupp · D. Barber
Institute for Simulation and Training, 3100 Technology Parkway,
Orlando, FL 32826, USA
e-mail: kstowers@ist.ucf.edu

N. Kasdaglis
e-mail: nkasdagl@ist.ucf.edu

M. Rupp
e-mail: mrupp@ist.ucf.edu

D. Barber
e-mail: dbarber@ist.ucf.edu

J. Chen (✉)
U.S. Army Research Laboratory, Orlando, FL 32826, USA
e-mail: yun-sheng.c.chen.civ@mail.mil

M. Barnes
U.S. Army Research Laboratory, Ft. Huachuca, AZ 85613, USA
e-mail: michael.j.barnes.civ@mail.mil

1 Background

As warfighting environments become more complex, operators will increasingly collaborate with intelligent agents (IAs) to manage teams of robotic systems [1]. However, increased autonomy may come with a cost: operators may have difficulty understanding IA rationale during mixed-initiative decision-making [2]. Mixed initiative decision-making is not only a requirement for effective operations in war-fighting environments, but also an inherent behavior of all Joint Cognitive Systems [3]. Therefore, operational environments demand not only autonomy and flexibility, but also collaborative interaction between system cognitive actors (human operators and IAs) to reach optimized performance. As a whole, the overall system should support operators' recognition of the state of the world, anticipation of the consequences of state changes in the world, and appropriate adaptation of system means and goals. This can aid in operators' general performance, as well as management of abnormal system-wide events [4, 5]. To facilitate this support, an operator's display should be designed with resilient operations in mind [6]—that is, buffering, to absorb the information processing deficits of a human cognition; flexibility, rather than brittleness, to adapt to dynamic events; and appropriate trade-off mechanisms that solve conflicting goal and/or competing resource allocations.

Additionally, collaborative work between IAs and human operators in a system presupposes a priori roles for effective organizational automation. Generally, these roles may exist along a continuum of supervisory control [7, 8]. While both IAs and humans process, reason, and communicate, such processing must be explicit to the human in order to avoid confusion, instill trust, and structure action [9, 10]. Thus, if an operator is to make informed decisions, a system display must make explicit what the IA knows, does not know, reasons, and projects about its operation context and its goals.

1.1 Situation Awareness in Mixed Initiative Decision-Making

The construct of situation awareness (SA) formalizes human interaction within a given context [11, 12]. Whether conceptualized as a processes or product, SA explicates situational human cognition for decision-making, as it represents an operator's awareness of the immediate situation, comprehension of the situation, and prediction of future possibilities [13, 14]. The most commonly relied upon model parses SA into three levels [11]: perception of the situation elements, comprehension of these elements, and projection as it relates to the perceiver and situation elements in the future. IAs possess a similar computational ability for sensing, reasoning, and projecting about their environment.

In this sense, both humans and IAs are data driven and concept driven. That is, human-cognition and computational-cognition are each concerned with a means and an end to accomplish a purpose. For example, both perceive or sense the environment and are able to utilize data in addition to planning and modifying that data to accomplish a goal. Separate, both are entities whose performance may improve through an external intervention. However, a paradigm that accounts for collaborative and coordinated human-agent interaction would allow for a unified cognitive system that integrates human and IA cognitive processes and outcomes [15, 16].

1.2 Transparency and Supervisory Control of Intelligent Agents

In addition to information sharing between an operator and an IA, as well as coordination of both of their respective activities, it has been suggested that collaborative work must acknowledge that each part of a system possesses partial and overlapping information relevant to the fulfillment of the overall system purpose [16]. To benefit from that information, a collaborative work system must provide a means for a transparent field of view of each agent's unique perspective. In this regard, increasing *transparency* in IA interfaces can improve operator performance [9], provided it gives an understandable representation of the mission environment and constraints, and the IA's knowledge, intent, and limitations [17, 18].

Not only will IA transparency increase operator SA by giving insight into the IA's current action and intent, its relevant knowledge of the state of the world and situational constraints, but it will additionally engender trust between the IA and the operator, who must rely on the IA's reasoning and projections to make decisions [9, 10]. Transparency specifically facilitates appropriate calibration of trust for the operator. Such calibrated trust should lead to appropriate reliance on the IA [19]. As opposed to under-reliance (IA disuse) or over-reliance (IA misuse), which impede overall system effectiveness [20], establishing appropriate reliance can increase overall performance in the human-machine system [19].

1.3 Situation Awareness-Based Agent Transparency

In an effort to meet the above needs and guide the design of transparency in IA, Chen et al. [10] proposed the Situation awareness-based Agent Transparency (SAT) model. By applying underlying theoretical assumptions inherent to the understanding of both SA and agent transparency, this model can facilitate effective mixed-initiative decision-making (Fig. 1). The model functions as a corollary to the three levels of individual SA [11], yet is particularly relevant to the domain of human-IA teaming.

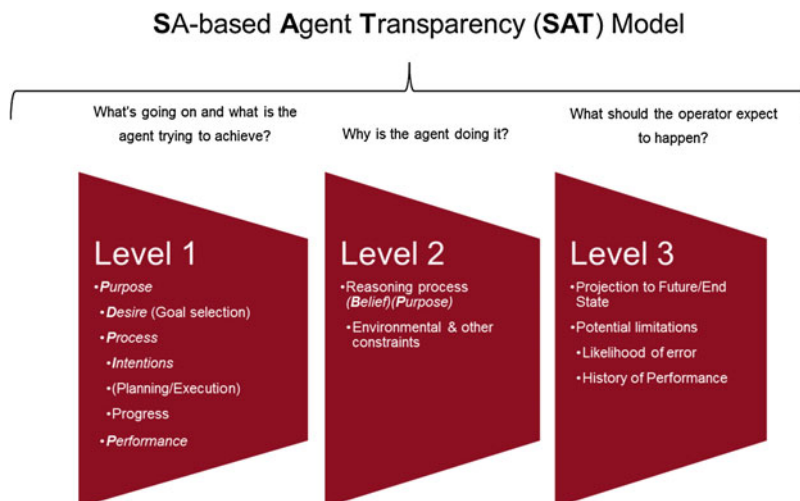


Fig. 1 SA-based agent transparency model diagram [10]

The SAT model provides a useful theoretical framework to guide design efforts of requisite IA display elements that support the operator's SA and facilitating appropriate trust calibration. While recognizing the danger in assigning human attributes to a computational agent, it is helpful to borrow such terms for clarity. Display requirements for human-supervisory control with IA thus correspond to the three levels of SA in humans. Each level seeks to provide the answer to three implicit questions of operator:

1. What is the agent trying to achieve?
2. Why is the agent doing it?
3. What should the operator expect to happen?

Implicit in the cognitive and computational process captured in Level 3 of the SAT model is the notion of uncertainty, specifically the fact that no future event can be absolutely known. Although an IA can make sense of the world, it does not necessarily know all parameters that may affect its actions. It is important for the IA to communicate this uncertainty as part of its interaction with the human for collaborative planning and decision-making. Thus, the IA must share its uncertainty concerning its reasoning and projections with the operator. For example, in order to make a suggestion, the IA often must "fill in the blank" regarding missing information—the IA must make an assumption. A transparent IA must then communicate the nature of that uncertainty and the assumption made by the IA to the operator.

This model is not solely a human model, nor is it a model only for the IA. Instead, it relates the IA cognitive process and products back to the human's supervisory purview. Level 1 communicates the IA's desires and intentions [21] as

they relate to its environmental, operational, and organizational context. As part of a goal-directed team, the IA examines its environment for data needed to algorithmically reason what actions are needed to achieve optimum system performance; such communicated reasoning is Level 2. Finally, the IA makes projection of the dynamic nature of the situation; Level 3 information provides the operator this insight.

1.4 Design of Transparency Displays for Heterogeneous UxV Management

The application of the above theoretical positions is particularly important in the study of multi-unmanned vehicle (UxV) management, where mixed-initiative decision making is integral to mission success. Increasingly, research is focusing on the development of IAs that can work with operators to manage teams of UxVs [1, 22]. One of those efforts is the Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT) project currently funded by the U. S. Department of Defense's Autonomy Research Pilot Initiative [23, 24]. IMPACT is investigating issues associated with human-machine interaction in military contexts [24], and flexible "play-calling," such as that done in football [25, 26]. Such "play-calling," whereby a person chooses from a set of options or plans in a "playbook," can be applied in many warfighting contexts where warfighters are frequently required to make diplomatic decisions based on a limited set of options. It may be particularly useful in UxV management [25].

As part of this effort, and to explore the SAT model's utility for UxV management, the SAT model served as a guideline in the design of two separate IA interfaces evaluated in two consecutive studies (Figs. 2 and 3, referred to as Interface 1 and Interface 2, respectively). These interfaces were adapted from the U. S. Air Force Research Laboratory-developed IMPACT/Fusion interface [22, 27], and were further developed to convey three different conditions of SAT for each interface. These conditions and descriptions of their corresponding graphical displays are given in Table 1.

2 Study Design and Implementation

We examined the above interfaces separately in a pair of consecutive studies [28, 29]. Both studies sought to test a series of predictions regarding whether the aforementioned implementation of the SAT model was successful in facilitating UxV management. Specifically, we wanted to examine the impact of information sharing on several performance parameters critical to the success of multi-UxV management. For example, while additional transparency can improve performance



Fig. 2 Interface 1: displaying information according to the SAT model

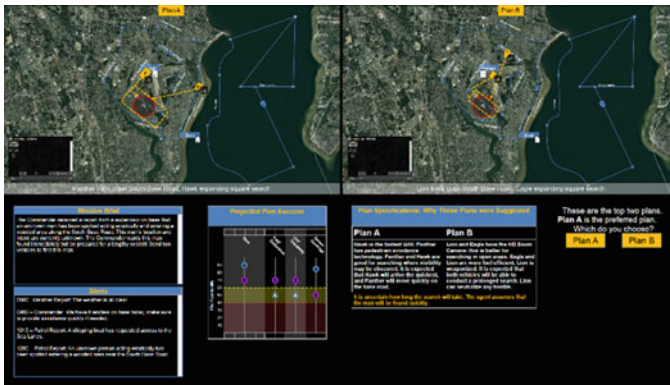


Fig. 3 Interface 2: displaying information according to the SAT model

Table 1 Manipulation of agent transparency using SAT levels to form experiment conditions. Each condition contains the components introduced in the prior condition

Interface 1		Interface 2	
SAT level	Display components	SAT level	Display components
L1	Map icons, plan details icon, and path show basic information	L1 + 2	Map icons, path, line graph, and text show basic information
L1 + 2	Pie graph and text add reasoning information to display	L1 + 2 + 3	Sliding points on line graph and extra text add reasoning and projection
L1 + 2 + 3	Opacity of sprocket pie graph varied and extra text add projections including uncertainty	L1 + 2 + 3 + U	Opacity of map icons and graph points varied, and extra text add assumptions and uncertainty

Table 2 Performance criteria according to IA and operator choice of plans

Performance criterion	Correct plan	IA suggestion	Operator choice
Proper IA use	A	A	A
Correct IA rejection	B	A	B

[10], it is important to consider the impact that this extra information has on response time and increased workload [30]. Furthermore, the usability of the interface may affect trust in machines [31]. Finally, as stated above, it has been suggested that additional transparency can improve trust on behalf of the human counterpart [9, 10].

To test these theoretical positions, we designed two studies, each with three conditions of transparency (see Table 1). During the corresponding experiments, participants took on the role of a UxV system operator whose task was to monitor and direct vehicles to carry out missions given to them by a simulated commander. Operators managed a team of six unmanned vehicles (UxVs): two unmanned aerial vehicles (UAVs), two unmanned ground vehicles (UGVs) and two unmanned surface vehicles (USVs), as well as an IA, which communicated plan options for completing the mission. To complete missions, operators needed to interpret their commander's intent, understand vehicle and environmental constraints, and ultimately decide whether to follow the IA's play-calling recommendation. The IA always suggested two options: Plan A as the most viable plan (which was its primary recommendation), and Plan B as the back-up plan. For 3 out of every 8 events, the IA's recommendation was incorrect due to information it did not have access to—updated commander's intent or other intelligence.

During each of these decisions, operators' performance (based on the criteria in Table 2), and response time were monitored by the simulation. After each block of events, we surveyed participants for information including their perceived workload, perceived interface usability, and their trust in the IA.

2.1 Study 1: Interface 1

Results from study 1 [28] indicated that proper IA use and correct rejection were both significantly greater when participants were presented with SAT L1 + 2 + 3 and L1 + 2 compared to L1. The greatest rates of proper IA use (when the IA's recommendation was correct) and correct rejection (when the IA's recommendation was incorrect) were found in L1 + 2 + 3, suggesting that operators were more likely to make correct decisions when presented with all three levels of SAT information. We found no significant differences for response time or workload, indicating that operators did not take longer to complete each decision nor did operators experience more effort as the amount of information to support agent transparency increased.

We analyzed operator trust in the IA after the first block of interactions, and examined it across two contexts: the IA's analysis of the information, and the IA's ability to suggest and make decisions. There were no significant differences across SAT level for trust in the IA's ability to analyze information. However, we found that operator's trust in the IA's ability to suggest and make decisions significantly increased as transparency increased. Specifically, participants felt the IA made decisions that were more accurate when presented with L1 + 2 + 3 as compared to L1 + 2 or L1. We also found a significant effect of SAT level on the perceived usability of the IA, where the IA was perceived to be the most usable when presented with L1 + 2 + 3.

While this study differentiated basic information, reasoning, and future projections according to the SAT model, we only examined uncertainty as part of SAT Level 3 information and not on its own. Thus, the role of uncertainty in affecting operator decision making remained unclear. Study 2 filled this gap by setting up different conditions whereby the final condition parsed out uncertainty from other Level 3 information (see Table 1).

2.2 *Study 2: Interface 2*

Results from study 2 [29] indicated that proper IA use and correct rejection were both significantly greater when SAT L1 + 2 + 3 + U was presented compared to L1 + 2. The greatest rates of proper IA use and correct rejection were found with L1 + 2 + 3 + U, suggesting that operators were more likely to make correct decisions when they were presented with all three levels of transparency, as well as uncertainty. As was the case in study 1, no significant difference was found for workload, indicating that operators did not experience more effort as the amount of information to support agent transparency increased. However, unlike study 1, there was a significant difference in response time between L1 + 2 and L1 + 2 + 3 + U, with L1 + 2 + 3 + U taking the longest for participants to complete. This was not unexpected, as an increase in information on the display should naturally take longer to process.

Contrary to study 1, in which we only analyzed trust after a single interaction with the interface, for study 2 we analyzed operator trust as it developed over time while also controlling for the effect of pre-existing implicit associations [32]. There was a significant difference across SAT level for trust in both the IA's ability to analyze information and the IA's ability to suggest and make decisions. Specifically, participants trusted the IA's ability to analyze information most when presented with L1 + 2 + 3 + U, while they trusted the IA's ability to suggest decisions most when presented with L1 + 2 + 3. We also found a significant effect of SAT level on the perceived usability of the IA, where the IA was perceived to be the most usable when displaying L1 + 2 + 3 and the least usable when displaying L1 + 2 + 3 + U. This perception of usability is somewhat consistent with the participants' trust in the IA's ability to make decisions, where their trust and

perceived usability peaked at L1 + 2 + 3 and tapered off when uncertainty was added to the interface. This finding adds further support to the idea that usability impacts trust [31]. It also raises several questions about the display of uncertainty [33], which will be discussed next.

3 Discussion

Overall, we found evidence supporting the use of the SAT model to improve operator performance, increase trust in the IA, and increase perceived usability of the system, while minimizing potential costs of workload. Displaying SAT L1 + 2 + 3 information provided the most benefits to operators' trust and perception of the agent's usability, while displaying L1 + 2 + 3 + U provided the most benefits to operators' performance. Due to these findings, we recommend that similar automated decision aid systems incorporate information displays that provide the operator with both information regarding the reasoning of each decision provided, as well as displays of possible future states and sources of potential uncertainties that might affect their decisions. Our results suggest a number of aspects that designers of human-machine systems should consider.

First, how can we utilize SAT-based displays to improve decision-making performance while minimizing the impact on response time? It makes sense that response time may increase when more information is presented. Increased response time may not always be problematic, but in time-critical tasks, milliseconds may make the difference between success and failure. In such contexts, it is important to design interfaces that communicate vital information to the user while minimizing the amount of processing required to make a decision. If we aim to make truly flexible machines that can adapt to the environment [6], we must also consider how this flexibility applies to the display of SAT-based information.

Next, how can we best display uncertainty in a way that is both useful and usable to the operator? Such optimization of the interface may have effects on not just usability, but trust and performance, as well. Our interfaces displayed uncertainty both graphically and in text, but we did not statistically differentiate the usability of each of these components. Further analysis of such component parts may yield information and best practices about the display of uncertainty. For example, it is possible that intuitive graphical representations will be perceived as more usable—and may even result in lower processing time—than textual representations. Furthermore, trust and perceived usability may change when the IA presents its uncertainty in different ways. For example, reporting percentages of certainty (e.g. “80 % probable”) may lead to drastically different perceptions and outcomes in the operator than more ambiguous graphical representations, and these potential differences must be considered [34].

Finally, are the results we found here generalizable? We argue that the outcomes examined here depend on the task and the context. This position is supported by prior research and theoretical discussions positing that both task and environment

influence overall human-machine performance [35]. The studies presented in this paper examined the management of UxVs in a military environment. As such, our findings are most applicable to similar mixed initiative decision-making tasks. What remains to be seen is the validity of these parameters in entirely different contexts.

Future studies should examine the display of SAT-based information in new contexts, and thus refine our understanding of the usefulness of agent transparency in human-machine interaction. Furthermore, future studies should more thoroughly examine the role of uncertainty as a key to achieving appropriate levels of transparency. While the display of uncertainty may have tradeoffs, it should not be eliminated from displays [34]. It is wholly necessary, as are the other facets of transparency, to the successful performance of overall human-machine systems.

Acknowledgments This research was supported by the U.S. Department of Defense Autonomy Research Pilot Initiative, under the Intelligent Multi-UxV Planner With Adaptive Collaborative/Control Technologies (IMPACT) project. We wish to thank Joseph Mercado, Katelyn Procci, Isacc Yi, Erica Valiente, Shan Lakhmani, and Jonathan Harris for their contribution to this project. We would also like to thank Gloria Calhoun and Mark Draper for their input.

References

1. Chen, J.Y.C., Barnes, M.J.: Human-agent teaming for multi-robot control: a review of human factors issues. *IEEE Trans. Hum. Mach. Syst.* **44**, 13–29 (2014)
2. Linegang, M.P., Stoner, H.A., Patterson, M.J., Seppelt, B.D., Hoffman, J.D., Crittendon, Z.B., Lee, J.D.: Human-automation collaboration in dynamic mission planning: a challenge requiring an ecological approach. In: *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, pp. 2482–2486. SAGE Publications, California (2006)
3. Hollnagel, E., Woods, D.D.: *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. CRC Press, Boca Raton (2005)
4. Boy, G.A.: *Cognitive Function Analysis*. Ablex Publishing Corporation, Stamford (1998)
5. Kasdaglis, N., Newton, O., Lakhmani, S.: system state awareness: a human centered design approach to awareness in a complex world. In: *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting*, pp. 305–309, SAGE Publications, California (2014)
6. Hollnagel, E., Woods, D., Leveson, N.: *Resilience Engineering: Concepts and Precepts*. Ashgate, Surrey (2006)
7. Boy, G.: Theories of human cognition: to better understand the co-adaptation of people and technology. In: Kiel, LD (ed.) *Knowledge Management, Organizational Intelligence and Learning, and Complexity*, vol. 3, Developed under the Auspices of the UNESCO, pp. 204–238. Eolss Publishers Co Ltd, Oxford (2009)
8. Parasuraman, R., Cosenzo, K.A., De, Visser, E.: Adaptive automation for human supervision of multiple uninhabited vehicles: effects on change detection, situation awareness, and mental workload. *Mil. Psychol.* **21**, 270–297 (2009)
9. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors* **46**, 50–80 (2004)
10. Chen, J.Y.C., Procci, K., Boyce, M., Wright, J., Garcia, A., Barnes, M.: Situation awareness-based agent transparency (ARL-TR-6905). Technical Report, US Army Research Laboratory, Aberdeen Proving Ground (2014)

11. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. *Hum. Factors* **37**, 32–64 (1995)
12. Smith, K., Hancock, P.: Situation awareness is adaptive, externally directed consciousness. *Hum. Factors* **37**, 137–148 (1995)
13. Endsley, M.R., Jones, D.G.: *Designing for Situation Awareness: An Approach to User-Centered Design*. CRC Press, Boca Raton (2012)
14. Stanton, N.A., Chambers, P.R., Piggott, J.: Situational awareness and safety. *Saf. Sci.* **39**, 189–204 (2001)
15. Woods, D.D.: Cognitive technologies: the design of joint human-machine cognitive systems. *AI Mag.* **6**, 86–92 (1985)
16. Hollnagel, E., Woods, D.D.: *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*. CRC Press, Boca Raton (2006)
17. Cook, M.B., Smallman, H.S.: Human factors of the confirmation bias in intelligence analysis: decision support from graphical evidence landscapes. *Hum. Factors* **50**, 745–754 (2008)
18. Neyedli, H.F., Hollands, J.G., Jamieson, G.A.: Beyond identity incorporating system reliability information into an automated combat identification system. *Hum. Factors* **53**, 338–355 (2011)
19. Gao, J., Lee, J.: Extending the decision field theory to model operator’s reliance on automation in supervisory control systems. *IEEE Trans. Syst. Man Cybern. A. Syst. Humans* **36**, 943–959 (2006)
20. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse. *Hum. Factors* **39**, 230–253 (1997)
21. Bratman, M.: *Intention, Plans, and Practical reason*. CSLI Publications, Stanford (1987)
22. Behymer, K.J., Mersch, E.M., Ruff, H.A., Calhoun, G.L., Spriggs, S.E.: unmanned vehicle plan comparison visualization for effective human-autonomy teaming. In: *Proceedings of 6th international conference on ahfe and the affiliated conference*, pp. 1022–1029. Elsevier B. V, Netherlands (2015)
23. U.S. Department of Defense—Research & Engineering Enterprise. *Autonomy Research Pilot Initiative*. <http://www.acq.osd.mil/chieftechologist/arpi.html>
24. Draper, M.: *Realizing autonomy via intelligent adaptive hybrid control: adaptable autonomy for achieving UxV RSTA team decision superiority—year 1 report*. US Air Force Research Laboratory, Dayton (in press)
25. Fern, L., Shively, R.J.: A comparison of varying levels of automation on the supervisory control of multiple UASs. In: *Proceedings of AUVSIs Unmanned Systems North America*, pp. 10–13. Curran Associates Inc., New York (2009)
26. Miller, C.A., Parasuraman, R.: Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. *Hum. Factors* **49**, 57–75 (2007)
27. Rowe, A., Spriggs, S., Hooper, D.: Fusion: a framework for human interaction with flexible-adaptive automation across multiple unmanned systems. In: *Proceedings of 18th symposium on aviation psychology*, pp. 464–469. Curran Associates Inc., New York (2015)
28. Mercado, J., Rupp, M., Chen, J., Barber, D., Procci, K., Barnes, M.: Intelligent agent transparency in human-agent teaming for multi-UxV management. *Human Factors* (in press)
29. Stowers, K., Chen, J.Y.C., Kasdaglis, N., Newton, O., Rupp, M., Barnes, M.: Effects of situation awareness-based agent transparency information on human agent teaming for multi-UxV management (in press)
30. Lyons, J.B., Havig, P.R.: Transparency in a human-machine context: approaches for fostering shared awareness/intent. In: Shumaker, R., Lackey, S. (eds.) *Virtual, Augmented and Mixed Reality: Designing and Developing Virtual and Augmented Environments*, pp. 181–190. Springer, Berlin (2014)
31. Wang, L., Jamieson, G.A., Hollands, J.G.: Trust and reliance on an automated combat identification system. *Hum. Factors* **51**, 281–291 (2009)
32. Merritt, S.M., Heimbaugh, H., LaChapell, J., Lee, D.: I trust it, but i don’t know why effects of implicit attitudes toward automation on trust in an automated system. *Hum. Factors* **55**, 520–534 (2012)

33. Stowers, K., Kasdaglis, N., Newton, O., Lakhmani, S., Wohleber, R., Chen, J.: Intelligent agent transparency: the design and evaluation of an interface to facilitate human and artificial agent collaboration. In: Proceedings of the Human Factors and Ergonomics 60th Annual Meeting (submitted)
34. Endsley, Mica R.: Designing for Situation Awareness: An Approach to User-Centered Design. CRC Press, Boca Raton (2011)
35. Stowers, K., Oglesby, J., Leyva, K., Iwig, C., Shimono, M., Hughes, A., Salas, E.: A framework to guide the assessment of human-machine systems. Human Factors (submitted)