Situation awareness-based agent transparency and human-autonomy teaming effectiveness

Jessie Y. C. Chen⁹, Shan G. Lakhmani¹, Kimberly Stowers¹, Anthony R. Selkowitz⁹, Julia L. Wright⁹ and Michael Barnes⁹

⁹U.S. Army Research Laboratory, Orlando, FL, USA; ¹Institute for Simulation and Training, University of Central Florida, Orlando, FL, USA

ABSTRACT

Effective collaboration between humans and agents depends on humans maintaining an appropriate understanding of and calibrated trust in the judgment of their agent counterparts. The Situation Awareness-based Agent Transparency (SAT) model was proposed to support human awareness in human–agent teams. As agents transition from tools to artificial teammates, an expansion of the model is necessary to support teamwork paradigms, which require bidirectional transparency. We propose that an updated model can better inform human–agent interaction in paradigms involving more advanced agent teammates. This paper describes the model’s use in three programmes of research, which exemplify the utility of the model in different contexts – an autonomous squad member, a mediator between a human and multiple subordinate robots, and a plan recommendation agent. Through this review, we show that the SAT model continues to be an effective tool for facilitating shared understanding and proper calibration of trust in human–agent teams.

Relevance to human factors/Relevance to ergonomics theory

Advances in machine intelligence have necessitated the prioritisation of transparent design for agents to be teamed with human partners. This paper argues for a model of transparency that emphasises teamwork, bidirectional communication, and operator-machine-situation awareness. These factors inform an approach to designing user interfaces for effective human–agent teaming.

Introduction

As autonomous systems become more intelligent and more capable of complex decision making (Shattuck 2015; Warner Norcross and Judd 2015), 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 (Chen and Barnes 2014). Indeed, Explainable AI (Gunning
a DARPA programme 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* (Defense Science Board 2016), 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 (15). Chen and her colleagues (2014) developed a model, *Situation Awareness-based Agent Transparency* (SAT), to organise the issues involved with supporting the human’s awareness of an agent’s: (1) current actions and plans, (2) reasoning process, and (3) outcome predictions. In this paper, results from three projects will be used to illustrate the utility of agent transparency for supporting human-autonomy teaming effectiveness. Additionally, the original SAT model is expanded to incorporate teamwork and bidirectional transparency to address the challenges in dynamic tasking environments.

The ‘agents’ discussed in this paper include virtual agents, which are software systems that can be deemed ‘intelligent’ (*autonomy at rest*; Defense Science Board 2016), and autonomous robots (*autonomy in motion*; Defense Science Board 2016), which are embodied and independent (Fong et al. 2005; Sycara and Sukthankar 2006). These agents are characterised by their autonomy, reactivity to the environment, and pursuit of goal directed behaviour (Russell and Norvig, 2009; Wooldridge and Jennings 1995). The focus of this paper is on these kinds of agents that interact with humans via mixed-initiative systems, systems wherein a human and an agent collaborate to make decisions (Allen, Guinn, and Horvitz 1999; Chen and Barnes 2014). Within the mixed-initiative framework, the agent can have several different tasks, and conduct each with a varying degree of autonomy, depending on the needs of each task and the needs of the human teammate (Kaupp, Makarenko, and Durrant-Whyte 2010). Agents designed to provide low level mixed-initiative interaction can report critical information as it arises or seek clarification in uncertain situations (Allen, Guinn, and Horvitz 1999). At higher levels, agents can be set to take initiative in predefined circumstances (e.g. when the human is overloaded or too slow to react), taking over the implementation of a plan – prompting the human for decisions when needed – or implementing pre-specified algorithms to achieve the objective, returning initiative to the human once the subtask is complete (Allen, Guinn, and Horvitz 1999; Chen and Barnes 2014; Parasuraman, Barnes, and Cosenzo 2007; Parasuraman and Miller 2004). At the highest levels of mixed-initiative interaction, the agent actively monitors the current task and uses information it has about its capabilities, the human’s capabilities, and the other demands on its resources to evaluate whether it should take initiative (Allen, Guinn, and Horvitz 1999). The human operator is frequently the final authority in the mixed-initiative framework, often performing in a supervisory capacity, especially in military environments due to the added complexity and safety considerations inherent in combat (Barnes, Chen, and Jentsch 2015; Chen and Barnes 2014; Goodrich 2013). This relationship is prone to various supervisory control issues including operator error rates, difficulty in maintaining situation awareness (SA), and inappropriate trust in the agent (Chen and Barnes 2014). Because poorly calibrated trust can result in catastrophic failures due to operator misuse or disuse of the automated systems, it is crucial that an agent’s behaviours, reasoning, and expected outcomes be transparent to the operator (Bitan and Meyer 2007; de Visser et al. 2014; Linegang et al. 2006; Parasuraman and Riley 1997; Seppelt and Lee 2007; Stanton, Young, and Walker 2007).
Situation awareness-based agent transparency

Lee and See (2004) identified three components of an agent whose transparency influences operators’ trust calibration: purpose, process, and performance (3Ps). Lee (2012) proposed that to increase transparency in machine agents, the system’s 3Ps and performance history should be displayed to the operator in a simplified form (e.g. integrated graphical displays) so the operator is not overwhelmed by too much information (Cook and Smallman 2008; Neyedli, Hollands, and Jamieson 2009; Selkowitz et al. 2017; Tufte 2001). Chen and associates (2014) leveraged Endsley’s model of SA (Endsley 1995), the BDI (Beliefs, Desires, Intentions) Agent Framework (Rao and Georgeff 1995), Lee’s 3Ps described above (2012), and other relevant previous work (Chen and Barnes 2012a, Chen and Barnes 2012b; Lyons and Havig 2014) to develop the SAT model (Figure 1). Drawing inspiration from Endsley’s (1995) three levels of SA – perception, comprehension, and projection – the SAT model describes the information that agents need to convey about their decision-making process in order to facilitate the shared understanding required to perform effective human–agent teamwork (Chen et al. 2014; Stubbs, Wettergreen, and Hinds 2007; Sycara & Sukthankar 2006). While SA is the cumulative result of three hierarchical phases, the SAT model is comprised of three independent levels, each of which describe the information an agent needs to convey in order to maintain a transparent interaction with the human (Chen et al. 2014; Endsley 1995).

At the first SAT level (SAT 1), the agent provides the operator with the basic information about the agent’s current state and goals, intentions, and plans. This level of

| Situation Awareness-based Agent Transparency |
|---------------------------------------------|
| **Level 1: Goals & Actions**                |
| *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. The original situation awareness-based agent transparency (SAT) model, adapted from Chen et al. (2014).
information assists the human’s perception of the agent’s current actions and plans. At the second level (SAT 2), the agent reveals its reasoning process as well as the constraints/affordances that the agent considers when planning its actions. In this way, SAT 2 supports the human’s comprehension of the agent’s current behaviours. At the third SAT level (SAT 3), the agent provides the operator with information regarding its projection of future states, predicted consequences, likelihood of success/failure, and any uncertainty associated with the aforementioned projections. Thus, SAT 3 information assists with the human’s projection of future outcomes. SAT 3 also encompasses transparency with regard to uncertainty, given that projection is predicated on many factors whose outcomes may not be precisely known (Bass, Baumgart, and Shepley 2013; Chen et al. 2014; Helldin 2014; Mercado et al. 2016; Meyer and Lee 2013). Prior research has shown the positive effects of agents conveying uncertainty on joint human–agent team performance effectiveness and operator perceptions of agent trustworthiness (Beller, Heesen, and Vollrath 2013; McGuirl and Sarter 2006; Mercado et al. 2016). The following section details how humans’ trust in an agent can be influenced by transparent interaction with that agent.

**Implications for operator trust**

Several empirical studies have investigated the effects of agent transparency on operator trust (Bass, Baumgart, and Shepley 2013; Helldin et al. 2014; Mercado et al. 2016; Zhou and Chen 2015). Mercado et al. (2016) found that when interacting with an intelligent planning agent, operator performance and trust in the agent increased as a function of agent transparency level. Similarly, Helldin et al. (2014) found that transparency in terms of sensor accuracy and uncertainty information improved both trust and operator task performance (correct target classifications using an automated classifier). However, there were costs for increased transparency in terms of decision latency and increased operator workload, which was not the case in the Mercado et al. (2016) study. In a recent study by Zhou and Chen (2015), system transparency (where displays detailed the underlying machine learning algorithms’ processes) improved operators’ ability to monitor the course of the agent as well as their acceptance of its outcomes.

Moreover, there is evidence that knowing a machine agent’s uncertainties improves human–machine team performance (Bass, Baumgart and Shepley 2013; Beller, Heesen and Vollrath 2013; Chen et al. 2016; Mercado et al. 2016). Transparency has been shown to have the (perhaps counterintuitive) quality of improving operators’ trust in less reliable autonomy by revealing situations where the agent has high levels of uncertainty, thus developing trust in the ability of the agent to know its limitations (Chen and Barnes 2014; Chen and Barnes 2015; Mercado et al. 2016). Furthermore, operator’s trust calibration can be enhanced by a decision-making process supported by transparent joint reasoning. For example, the agent may indicate that the current plan is based on considering constraints $x$, $y$, and $z$. Human operators, based on their understanding of the mission requirements and information not necessarily accessible to the agent (e.g. intelligence reports), may instruct the agent to ignore $z$ and focus only on the other two constraints. Without the agent’s explanation of its reasoning process, the human operators may not realise that the agent’s planning is suboptimal due to its consideration of an unimportant constraint. In other words, the SAT model strives to not only address operator trust and trust calibration but also to identify the information requirements needed to facilitate
effective joint decision making between the human and agents. The following section reviews the results of three projects based on the SAT model to further illustrate the effects of agent transparency on operator performance and trust calibration in human–agent teaming task environments.

**Programmes of research**

In this section, we describe three programmes of research that use the SAT model to inform the presentation of information from an agent to a human in various experimental paradigms. We will discuss empirical studies within each programme which investigate the effects of agent transparency on measures of human–agent teaming effectiveness, including shared understanding and proper trust calibration. These studies provided insights into the implementation of transparency in human–agent teams and provided us with the information that we used to update the model, which will be described following the current section. The paradigms evaluate different aspects of the efficacy of SAT-based transparency information: (1) simulated embodied agent (small ground robot), (2) an agent suggesting convoy route changes, and (3) an agent that helps operators manage multiple heterogeneous robots (e.g. planning courses of actions) for base defence.

**Autonomous squad member**

Current autonomy research programmes, such as the U.S. Department of Defense Autonomy Research Pilot Initiative (ARPI; Department of Defense 2013), have started to investigate some of the key human-autonomy teaming issues that have to be addressed in order for mixed-initiative teams to perform effectively in the real world, with all its complexities and unanticipated dynamics. One of the seven projects under the ARPI, the Autonomous Squad Member (ASM) project encompasses a suite of research efforts framed around a human’s interaction with a small ground robotic team member in a simulated dismounted infantry environment. The ASM is a robotic mule, carrying supplies while autonomously moving toward a rally point (Selkowitz et al. 2017). The ASM’s reasoning process uses information from the environment, outcomes of its past actions, current resource levels, and its understanding of the current state of its human teammates to determine the prioritisation of competing motivators – e.g. accomplish mission goals, maintain ability to achieve future goals (Gillespie et al. 2015). Based on that information, the ASM selects an action – e.g. move forward, reroute around an obstacle, seek cover – and predicts how that action will impact its resources (Selkowitz et al. 2017). One of the goals of the ASM project is to create a transparent agent – one that allows its human teammates to maintain an awareness of its current actions, plans, perceptions of the environment/squad, its reasoning, and its projected outcomes.

The agent’s interface conveys information about itself and its surroundings to support the human’s awareness. Of particular interest is the inclusion of the ASM’s perceived actions of its human teammates as part of the set of information it communicates to its teammates. The display of perceived squad action is an initial step in the implementation of bidirectional transparency between the human and the agent. The ASM gains information about the world and its human teammates through observation; it then
communicates that information, along with information about its reasoning process, to its teammates. The humans can then use the information they receive from the ASM to inform their subsequent actions.

Informed by the SAT model, the ASM’s user interface features an “at a glance” transparency module (upper left corner of Figure 2), where user-tested iconographic representations of the agent’s plans, motivator, and projected outcomes are used to promote transparent interaction with the agent (Selkowitz et al. 2016). Studies on the ASM’s user interface have investigated the effects of agent transparency on the teammate’s SA, trust in the ASM, and workload (Selkowitz et al. 2017).

In a recent human factors experiment, the participants monitored a virtual environment for threats while maintaining awareness of the ASM through its display (Selkowitz et al. 2017). Each participant performed the mission with four different interface configurations, with order of conditions counterbalanced between participants (Figure 2). The information presented in each interface configuration reflects the information that needs to be conveyed at each SAT level. In display condition SAT 1, the participants received information relevant the SAT level 1, including information about the ASM’s perception of its environment (Figure 2, items 9 and 10), its perception/understanding of its squad teammates’ current state (Figure 2, item 4), its current location (Figure 2, item 12), the current highest influence on its reasoning (Figure 2, item 1), its current resource levels (Figure 2, item 8), and current goal (Figure 2, item 12). In display condition SAT 1+2, where information relevant to both SAT level 1 and SAT level 2 are displayed, the ASM displayed the top ‘motivator’ (Figure 2, item 2) behind its current actions and plans (e.g. time or resources availability). Display condition SAT 1+2+3, where information relevant

Figure 2. The display for the autonomous squad member, with annotations.
to SAT level 1, SAT level 2, and the projection elements of SAT level 3 are displayed, further added information regarding both the ASM’s predictions (Figure 2, item 3) and its projection of future state (e.g. projected resource usage and time to reach its destination). Lastly, the condition SAT 1+2+3+U, where information from all three SAT levels are displayed, featured information from all of the previous levels in addition to the information regarding the ASM’s uncertainty of the information it is presenting (depicted in Figure 2, items 9 and 10, as the auras around the diamond/triangular icons). By examining these visual elements at such a granular level, we are able to specify which information best facilitates the human teammate’s needs.

In terms of SA related to the ASM (Jones and Kaber 2004), as measured by SAGAT-style queries, the human teammates had a significantly higher level of comprehension ($\eta^2 = .061$) of the reasoning of the ASMs’ actions (Level 2 SA) in the SAT 1+2+3 condition than in the SAT 1+2 condition ($\eta^2 = .066$). Similarly, they were better able to predict the ASM’s future state (Level 3 SA) in the SAT 1+2+3 condition than in the SAT 1 and SAT 1+2 conditions. There were no significant differences in the humans’ SAT 1 SA across conditions. These results suggest that an agent interface supporting all three levels of the SAT model facilitates humans’ comprehension of the agent’s reasoning and their projections of the ASM’s future state better than a SAT-based user interface supporting levels 1 and 2 alone. However, the additional uncertainty information available on the SAT 1+2+3+U user interface did not further benefit the humans’ performance, suggesting that the nuance provided by this additional information was not necessary to SA in the given task.

The human teammates had the greatest subjective trust in the ASM when it presented SAT 1+2+3 information, compared with other conditions ($\eta^2 = .097$). They reported significantly greater trust in the SAT 1+2+3 condition than in the SAT 1 and SAT 1+2 conditions. While there was a slight decline in subjective trust between SAT 1+2+3 and SAT 1+2+3+U, this difference was not significant. Results indicated no difference in workload incurred by participants across display conditions.

Overall results indicated that when humans were presented with increased SAT model-based information, they were able to use this information effectively and did not display significantly higher workload costs, which would typically be associated with adding additional information to displays. As the capabilities of the ASM increase, allowing for closer coordination between the human and the agent, bidirectional communication becomes more important. Not only will bidirectional communication between the human and the agent allow them both to be aware of the factors underlying each other’s decision-making process – and subsequently use that information in their own decision-making processes – but also allow them to correct erroneous factors the other may have. In order to support bidirectional transparency, control mechanisms will be added to the ASM ‘transparency module’ to enable human providing input (e.g. changing priority of constraints). The effectiveness of such a bidirectional module will be examined in a future human factors study.

**RoboLeader**

Another context in which aspects of the SAT model have been explored is in a mixed-initiative supervisory control task. An intelligent planning agent, RoboLeader, was developed to act as a mediator between an operator and a team of subordinate robots (Chen and
Figure 3. Operator control unit: user interface for convoy management and 360° tasking environment. OCU windows are (clockwise from the upper centre): (1) Map and Route Overview, (2) RoboLeader Communications Window, (3) Command Communications Window, (4) MGV Forward 180° Camera Feed, (5) MGV Rearward 180° Camera Feed, (6) UGV Forward Camera Feed and (7) UAV Camera Feed (Wright, Chen, Hancock, Yi, and Barnes, Forthcoming).

Previous research using the RoboLeader testbed has shown many benefits for the human–agent team, such as reducing operator’s perceived workload and improving operator’s task performance (Chen and Barnes 2012a; Chen and Barnes 2012b). However, when RoboLeader was at a high level of autonomy, humans displayed complacent behaviour in the form of automation bias (Wright et al. 2013). Automation bias has been described as a manifestation of inappropriate trust, i.e. overtrust (Parasuraman, Molloy, and Singh 1993), and can be encouraged by highly reliable systems, as well as operator inexperience and high workload (Chen and Barnes 2010; Lee and See 2004). This out-of-the-loop situation could have disastrous results when RoboLeader is not fully reliable, as the human operator would not be aware of potential mistakes.

The most recent RoboLeader study explored the effectiveness of agent reasoning transparency (SAT 2) on humans’ performance, specifically complacent behaviour (Wright, Chen, Hancock, Yi, and Barnes Forthcoming). In a series of scenarios designed to elicit high workload, the participants engaged in multiple tasks while guiding a convoy of
robotic vehicles through a simulated environment with the assistance of RoboLeader. Two experiments were conducted; in both experiments, agent reasoning transparency was manipulated via RoboLeader’s reports, which varied the amount of reasoning (i.e. no transparency, medium transparency, and high transparency) conveyed to the human teammate. The no transparency condition consisted of RoboLeader simply notifying the participant when a route change was recommended. In the medium transparency condition, RoboLeader notified the participant when a route change was recommended, and included the reason for the suggested change (e.g., Dense Fog ahead). The high transparency condition was the same as in the medium condition, but also included when the information was received upon which RoboLeader based its recommendation (e.g. Dense Fog ahead, 1 hour). The first experiment provided participants with information regarding their original route but no information regarding the revised route, while the second provided participants with information pertaining to both routes. In conjunction, these experiments allowed us to explore the effects of increasing the level of task environment information and how it affected the relationship between agent reasoning transparency and operator performance. Results showed that, when the participants had limited information about their environment (Experiment 1), reports that included RoboLeader reasoning (medium transparency) appeared to reduce complacent behaviour and improve performance on the operator’s route selection task, without increasing perceived workload. However, increasing the amount of RoboLeader reasoning conveyed in the report (high transparency), by adding time of report (which can be ambiguous and potentially difficult to interpret), negatively affected the participants’ task performance and complacent behaviour. In fact, results in the high transparency condition were similar to those evoked when no reasoning information was provided (no transparency). In both the no transparency and high transparency conditions, participants showed more complacent behaviour, lower task performance, and increased reported trust than those in the medium transparency condition.

When the participants had information regarding both route choices (Experiment 2), the effectiveness of increasing RoboLeader’s reasoning transparency appeared to be dependent upon individual difference factors such as (self-reported) complacency potential. While there was no difference in complacent behaviour when no reasoning information was provided (no transparency), individuals with a (reported) low potential for complacent behaviour performed better than their counterparts in both the medium and high transparency conditions. In fact, it appeared as though the presence of agent reasoning transparency did not benefit any individuals as much as it appeared to hurt performance in high complacency potential individuals while low complacency potential individuals appeared to be more resilient to the effects. This result may be due to the fact that the high transparency condition presented information (time of report) that could be potentially ambiguous and difficult for the participants to interpret. Therefore, the effects of high level of reasoning transparency may be different if the information is less ambiguous. In summary, these findings suggest that varying the amount of agent reasoning transparency provided, based on both the operator’s knowledge of the task environment and their individual differences, may be beneficial in mitigating complacent behaviour. The human’s knowledge of the task environment can vary over time, so effective personalisation would require bidirectional communication between the human and the agent.
Intelligent Multi-UxV planner with adaptive collaborative/control technologies (IMPACT)

In addition to assisting humans and managing robotic teams, agents are now able to function as planners that create and suggest strategies to follow in complex environments. This takes mixed-initiative decision making to a new level in which agents are not only performing actions and making recommendations, but also creating entire, detailed plans of activity. One such example can be found in 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 (Behymer et al. Forthcoming; Calhoun et al. 2017). This project involves the use of an agent that manages multiple unmanned vehicles in a military mission environment using a ‘play-calling’ paradigm (Fern and Shively 2009). ‘Plays,’ in this context, are templates from which an operator can choose to manage multiple heterogeneous unmanned vehicles (e.g. route reconnaissance using x aerial vehicles and y ground vehicles). One of the objectives of the IMPACT project is to develop an agent that can recommend a ‘play’ to the human operator.

In support of IMPACT’s objective to explore human–agent teaming, a laboratory study with a simulated agent that could be rapidly reconfigured (i.e. to convey different levels of transparency based on the SAT model) was undertaken with college students as participants. Specifically, the experimental design manipulations tested the impact of information transparency type and level on the participants’ decision-making performance. The agent played the role of ‘planner,’ generating and suggesting vehicle and route plans for the human; while the human (the experimental participant) acted as ‘decider,’ reviewing those plans and choosing which are implemented (Bruni et al. 2007). Transparency of the agent’s communication was manipulated in order to test its effect on the operator’s decision-making performance.

In the first study (Figure 4), completed by Mercado et al. (2016), three levels of transparency were implemented according to the SAT model reviewed in Figure 1: SAT 1 (basic plan information), SAT 1+2 (added reasoning information), and SAT 1+2+3 (added projection information, including uncertainty). Results indicated that operator task performance was highest and trust best calibrated (complying with the agent’s recommendations when it was correct and rejecting the recommendations when it was incorrect) when the highest level of transparency (SAT 1+2+3) was implemented. Additionally, this occurred without a detriment to response time or subjective workload. Finally, operators’ subjective trust in the agent’s ability to suggest and make decisions also increased with increased levels of agent transparency.

A replication of this study by Stowers et al. (Forthcoming) slight changes to the interface (Figure 5) examined three similar conditions, while delineating uncertainty from other projection information. The conditions implemented were SAT 1+2 (reasoning), SAT 1+2+3 (added projection), and SAT 1+2+3+U (added uncertainty). Results indicated that operator performance was once again highest when the highest level of transparency (SAT 1+2+3+U) was implemented. Additionally, this occurred without a detriment to response time or subjective workload. However, in this study, response time significantly increased with increased transparency, suggesting that operators were spending a longer time reading and processing information portrayed in the higher transparency conditions. Finally, operators’
Figure 4. Interface used to display transparency for study by Mercado and colleagues (2016). Transparency information is communicated through the use of text, vehicle characteristics on the map, and a pie chart.

Figure 5. Interface used to display transparency for study by Stowers and colleagues (2017). The interface was altered to pull transparency information to a separate area below the map, with text and a sliding bar scale representing the majority of transparency information offered. Vehicle characteristics on the map were also manipulated to correlate with such information.
subjective, state-based trust in the agent’s ability to analyse information increased when higher levels of transparency information were displayed.

These studies collectively suggest that transparency on the part of the agent benefits the human’s decision making and thus the overall human–agent team performance. As in a typical human–human team, this communication of relevant information can facilitate processing and decision making by aiding in the creation of a shared mental model between team members. What remains to be seen is how similar information on the part of the human (i.e. bidirectional transparency) can further facilitate the human–agent team performance.

The simulations used in Mercado et al. (2016) and Stowers et al.’s (Forthcoming) studies lend themselves to the inclusion of bidirectional communication. For example, button-clicking behaviour can aid bidirectional communication by allowing the human to give input to the agent. The agent can then update in response to such input, and give the human feedback. Indeed, it is possible that the present simulation, examined by college student participants, can be informed by the interfaces in the IMPACT system developed for use by trained military personnel (Calhoun et al. 2017). The IMPACT interfaces enable an operator to specify and revise parameters/constraints considered by the agent in developing a plan, as well as request more information from the agent on the rationale for one or more plans. A near-term plausible modification for this simulation is to add a chat box for the operator to input information to the agent, increasing the transparency of bidirectional communication along all three levels. Such a chat box would require either extensive training on the human’s part, or robust abilities to read natural language on the agent’s part. However, this additional communication to support transparency would be useful because it would allow the machine agent to learn from humans, an important goal in human–machine teams (Singh, Barry, and Liu 2004). This additional capability would also provide a mechanism by which the human and agent share in task completion (in this case, the shared task is ‘planning’).

The need for a dynamic SAT model

The review of the above three projects indicated the effectiveness of the SAT model to predict performance in a time constrained mission (ASM), the importance of unambiguous reasoning SAT information (RoboLeader), and the utility of all three SAT levels for planning using agent assistance, in particular, showing the benefits of informing the operator of agent uncertainties (IMPACT). The research also suggests that the original SAT model (Chen et al. 2014) may not be sufficient for complex military environments. SAT information must be dynamic, bidirectional and be sensitive to the operator’s role in shared decision making.

Since the SAT model was first proposed in 2014, progress in artificial intelligence and in the capabilities of autonomous systems has continued to advance. New systems are emerging that include more extensive machine learning capabilities than previously used (Koola, Ramachandran, and Vadakkeveedu 2016). The agents, in the aforementioned programmes of research, are also advancing in capabilities; they are able to make complex decisions and juggle priorities in environments characterised by continuous activity and change.
To capitalise on these advances in machine learning and their potential applications in mixed-initiative teaming, we have expanded the SAT model by incorporating aspects of teamwork transparency and bidirectional communications between humans and agents. The goal of this updated model is to better address joint action in human–agent teams, focusing on the bidirectional communication of information and the way each team member divvies up responsibilities in collaborative tasks. This updated model is most relevant to agents with the ability to learn from – or at the very least, understand – human input as part of their interaction, as such basic comprehension is a component of teamwork interaction (Barrett et al. 2012). The following section will review the theoretical underpinnings of human teaming research that ground these additions to the model. We will then present the revised model.

**Teamwork transparency and bidirectional communications**

Teams comprised solely of humans develop a shared cognitive understanding of their performance environment, the equipment needed to interface with it, the teammates with whom they work, and the patterns of interaction amongst those teammates (DeChurch and Mesmer-Magnus 2010; Mathieu et al. 2000). Having compatible knowledge structures for the aforementioned factors allows team members to predict each other’s actions, communicate with each other effectively, and coordinate their behaviour, thus increasing team performance (Cannon-Bowers, Salas, and Converse 1993; Sycara and Sukthankar 2006). Knowledge of one’s teammate – such as their skills or tendencies – allows team members to establish expectations from that teammate, while knowledge of the team’s interactions – such as roles and responsibilities, role interdependencies, and interaction patterns – allows team members to predict future team interactions (Mathieu et al. 2000). Knowledge of the shared understandings established in effective teams can be used to inform the design of interfaces supporting human–agent teaming (Mathieu et al. 2000; Sycara and Sukthankar 2006).

Lyons and associates (2017) suggest that transparency is comprised of a number of different dimensions (Lyons 2013). In one of these dimensions, the team dimension, the agent supports the human’s understanding of the division of labour between the human and the agent by conveying an understanding of which tasks fall under the human’s purview, which tasks are the agent’s responsibility, and which tasks require both humans’ and agents’ intervention (Bradshaw, Feltovich and Johnson, 2012; Bruni et al. 2007; Lyons 2013; Lyons et al. 2017). Accounting for such human–agent collaboration requires a more dynamic SAT model than that originally proposed by Chen and her associates (2014). In order to describe the transparency needed to support human–agent collaboration, the updated SAT model incorporates teamwork transparency and identifies the knowledge structures essential for facilitating team interaction; it also identifies the bidirectional transparency necessary to maintain both human and machine agents’ mutual understanding of their respective responsibilities.

In human–agent teaming, agent transparency is essential because it promotes the three major facets of human–agent interaction – mutual predictability of teammates, shared understanding, and the ability to redirect and adapt to one another (Lyons and Havig 2014; Sycara and Sukthankar 2006). To facilitate the creation and maintenance of a shared understanding between the human and machine agents, the machine must communicate
not only taskwork information (e.g. task procedures, likely scenarios, environmental constraints) but teamwork-related information as well (e.g. role interdependencies, interaction patterns, teammates’ abilities; Mathieu et al. 2000; Stubbs, Wettergreen, and Hinds 2007; Sycara and Sukthankar 2006). The addition of information specific to the interaction between the teammates allows each party to interpret cues from a compatible frame of reference, yielding more accurate predictions of future interactions and more flexible adaptation between the human and the agent in a mixed-initiative system (Mathieu et al. 2000; Stubbs, Wettergreen, and Hinds 2007; Sycara and Sukthankar 2006).

Teamwork, amongst humans or between humans and agents, requires a level of interdependence (Bradshaw, Feltovich, and Johnson 2012; Cannon-Bowers, Salas, and Converse 1993; Sycara and Sukthankar 2006). In mixed-initiative systems, a shared understanding of the division of labour can be established and negotiated by decomposing tasks into smaller sets of responsibilities and associating them with roles that can be dynamically assigned (Hayes and Scassellati 2013; Lyons 2013). This approach, coupled with the communication of intent, allows for mutual adaptation, where the agent can alter its role in the team and act autonomously to meet a goal (Chen and Barnes 2014; Sycara and Sukthankar 2006). While overt communication of roles, responsibilities, and the fulfilment thereof is expressed in the human teamwork literature (Cannon-Bowers, Salas, and Converse 1993; DeChurch and Mesmer-Magnus 2010), discussion of these factors in the context of human–agent teams has only recently been receiving increased attention (Hayes and Scassellati 2013; Lyons 2013; Sycara and Sukthankar 2006).

The incorporation of transparency with regard to teammate knowledge and teamwork interactions bolsters the original SAT model. Specifically, the updated model moves beyond characterising and framing transparency with regard to an agent’s communication about itself (i.e. current state, goals, reasoning, projections) and encompasses transparency as it applies to an agent’s understanding of its teammate’s roles and responsibilities and its interaction with its teammate as well. In a model incorporating these types of transparency, both humans and agents are responsible for maintaining transparency regarding their own tasks as well as their contributions to shared tasks, as illustrated in Figure 6. This bidirectional transparency allows humans and agent to take each other into account when making decisions, as previously suggested in the description of the implications of the Autonomous Squad Member. It would also allow them to use each other’s behaviour and reasoning as a resource in their own decision-making process, as described in the discussion of future research in the IMPACT project (Fong, Thorpe, and Baur 2003; Kaupp, Makarenko, and Durrant-Whyte 2010). In instances where agents use machine learning algorithms in their decision-making process, human behaviour and reasoning can be used to inform the agent’s model of the world; furthermore, if agents engage in behaviours supporting transparency, their human counterparts will know what information they need to provide to the agents in future iterations (Thomaz and Breazeal 2006). Incorporating human behaviour and understanding into the agent’s decision-making process more completely uses the capabilities of agents that include machine learning as part of their makeup.

Given the dynamic nature of the interaction between humans and agents in human–agent teams, we expanded the model to encompass the bidirectional, continuously iterating, and mutually interdependent interactions involved, thus creating the Dynamic SAT model (Figures 7 and 8). The human and agent interact throughout a mission: changing
Figure 6. Model of bidirectional, situation awareness-based agent transparency in human–agent teams. Both the human and the agent share their goals, reasoning, and projections to achieve their goals as a team. Both the human and the agent maintain transparency regarding their contributions to a shared task.

Figure 7. Agent decision making portion of the dynamic SAT model, including feedback and feed-forward loops. This conceptual model describes how changes in the agent's goals and actions, reasoning, or projections can influence the other two components.
plans, reacting to unforeseen events, and continually adapting to the environment. This interaction, coupled with the fact that the levels of the model are interrelated, make this model dynamic.

The Dynamic SAT model encompasses the agent’s decision-making process as well as the communication prerequisites needed to maintain the transparency of this process to both humans and agents. The three aspects of agent decision making that are highlighted in this model – Goals and Actions (encompassing the agent’s current goal, status, intent, plan, and action), Reasoning (the rationale behind the currently selected plan), and Projections (including both projection to future states and limitations) – influence each other through multiple feed-forward loops (Chen et al. 2014). As seen in Figure 7, a conceptual model of an agent’s decision-making process, the selection of a plan limits the possible outcomes that are projected to occur (Goals and Actions to Projections). Furthermore, changes in the plan affect the agent’s reasoning (e.g. constraints that need to be satisfied; Goals and Actions to Reasoning). Subsequently, any changes in reasoning yield a new projected outcome and influence any uncertainties concomitant with that new projected outcome (Reasoning to Projections).

The decision-making process described in this model can be illustrated using the example of a hypothetical, autonomous vehicle. If an autonomous vehicle has to travel to two points, a tree and a sign, and it selects a plan where it goes to the tree, then its predictions for the future include getting to the tree and exclude getting to the sign (Goals and Actions to Projections). So, by choosing to go to the tree, the autonomous vehicle fulfils its motivation to go to the tree and focuses on the constraints and affordances relevant to traveling to the tree (Goals and Actions to Reasoning). If the vehicle encounters an obstruction on the way to the tree, its projected outcome would change to incorporate the obstruction (Reasoning to Projections).

The feedback loop reacts in the opposite direction, as seen in Figure 7. The agent makes projections based on current information. As the agent gains experience, its understanding of its limitations (e.g. history of past performance, likelihood of error) informs its understanding of the affordances and constraints in its task environment, allowing the agent to alter its reasoning process to better reflect its updated understanding (Projections to Reasoning).
to Reasoning; Chen et al. 2014). Furthermore, the agent’s understanding of its constraints and affordances may suggest a different goal state as well as new planning approaches to achieving its objectives (Reasoning to Goals and Actions). Additionally, failure to achieve a projected outcome is noted by a mismatch between current progress and the achievement of the initial goal, which in turn causes the agent to re-plan (Projections to Goals and Actions). The re-planning, induced by the feedback loop, initiates the feed-forward loop, with a change in plan yielding changes in agent’s reasoning, which in turn results in new outcome predictions.

The hypothetical autonomous vehicle can also be used to describe the feedback loop. By the time the aforementioned vehicle has reached the tree, it has gained experience about its environment, its ability to manoeuvre in that environment, and how those factors influence its pursuit of its goal. Once the vehicle has achieved its current goal, it will switch to its next goal and start moving towards the sign. The vehicle will use its previous predictions coupled with the information it has gained from its previous actions to inform its rationale for future actions, such as avoiding the obstruction that previously got in its way (Projections to Reasoning). With an expanded knowledge of the environment and its capabilities in it, the vehicle can choose a plan to reach the sign that will avoid the obstruction that previously blocked it (Reasoning to Goals and Actions). The vehicle takes this new route and projects reaching the sign, but instead finds itself stuck in a mud puddle on the way. The vehicle identifies that its current state does not match its predictions and it has to re-plan (Projections to Goals and Actions). The vehicle switches its goal to getting itself out of the mud puddle. This new goal has affected the vehicle’s reasoning and projected outcomes, stimulating the feed-forward loop.

The feedback and feed-forward loops used by agents are informed by factors outside the agent, such as the environment. When a human is teamed with the agent, however, the human becomes one of the factors that influence the agent’s decision-making process. Unlike other factors, the human, as a teammate, can have a privileged loop into the agent’s decision-making process, with the potential to influence different aspects of the agent’s course of actions (Figure 8). Figure 8 describes not only the information about the agent decision-making process being conveyed to the human but also the human’s ability to communicate back to the agent about that process. The human is able to intervene at any time to provide input on the current plan, question the agent’s reasoning, or inform the projected outcome state, ensuring greater flexibility in the agent. For human intervention to be effective, however, the human has to be aware of the factors that the agent is taking into account and how that information is being used.

The dynamic nature of the agent’s decision-making process requires transparency when working with a human teammate. With information about the agent’s decision-making process, the human teammate can construct a compatible knowledge structure and make informed decisions. Similarly, the agent needs to understand the factors that its human teammate uses to make decisions. In the human–agent team, SA is bidirectional, so consequently SAT should be bidirectional. The human could not intervene effectively if the agent’s plan, its projected goal state, or its underlying reasoning process remain unknown; conversely, the agent could not optimise its plans unless it understood the human’s intent, constraints, and overall objective. Thus, two-way communication of reasoning information between humans and agents – bidirectional transparency – is important for effective performance in high functioning mixed-initiative teams. Establishing a
reciprocal relationship between human and agent teammates can also help mitigate issues concerning human bias. Smith, McCoy and Layton (1997) found that when a decision support system, used for flight planning, made a poor recommendation, 36% of participants who saw this recommendation before exploring on their own actually chose the system’s poor suggestion. In the first Roboleader study, when the agent only gave the human teammate a recommendation, the human exhibited complacent behaviour; however, the human teammates were less likely to exhibit complacent behaviour if they were given a recommendation coupled with information about the agent’s reasoning (Wright et al. Forthcoming). Conversely, too much information yielded nearly as much complacent behaviour as when the agent provided the recommendation only. In our model, bidirectional communication can be used to specify the human’s needs directly to the agent’s reasoning process. We posit that bidirectional communication between the human and agent can let a human–agent team reap the benefits of transparency, while mitigating the previously seen issue of too much information.

Figure 8 illustrates the actual communication between the human and the agent, showing where bidirectional communication within the more dynamic SAT model would occur. The agent can communicate information about its reasoning process to its human teammate. Conversely, humans can communicate relevant information about their decision-making processes, within the framework of the agent’s decision-making process. This mutual communication facilitates mixed-initiative interaction (Allen, Guinn, and Horvitz 1999). The extent to which agents can engage in collaborative decision making with humans can vary depending on how the mixed-initiative interaction is implemented (Allen, Guinn, and Horvitz 1999; Chen and Barnes 2014).

Ultimately, a well-designed ‘transparency module’ should not only present the agent’s description of itself (based on the three levels of SAT) and its understanding of the human’s task and their shared tasks, but also support effective ‘tweaking’ by the human. In other words, each SAT component should enable human’s fine tuning, either to correct the agent’s understanding or to provide instructions or information that the agent is not aware of. Of particular importance for human–agent teams is the effect of the dynamic SAT model on operator trust and reliance during dynamic and uncertain conditions. The human’s ability to ‘see’ into the agent’s intent, reasoning and projected outcomes is heavily dependent on the concomitant ability of the agent’s reasoning process to be predicated on its understanding of the human’s intention and world view. Trust between teammates, both synthetic and human, depends on mutual understanding of their respective roles, division of labour, and capabilities. Such bidirectional and dynamic communications should ensure the agent’s model of the tasking requirements are in sync with the human’s. Such systems should improve trust calibration by involving the human in transparent two-way communications.

Conclusions

The SAT model of transparency was developed as a model of transparency that focused on supporting the human’s awareness of the agent by emphasising the agent’s intent, reasoning, future plans, and the uncertainties associated with them. Three different research programmes – ASM, RoboLeader, and IMPACT – were navigated in order to explore how this paradigm influenced human–agent team performance, trust, SA, and
compliance. Using information gained from these studies, coupled with an understanding of new and future developments in agent technology, we have expanded the SAT model to better address the dynamic nature of new and upcoming human–agent teaming relationships. The expanded model emphasises the factors specific to the human–agent interaction, bidirectional communication, as well as the subsequent bidirectional transparency that is needed to maintain that interaction’s effectiveness. As agents become more complex and thus gain capability in mixed-initiative interactions, the need to address the explicit coordination between humans and agents grow. We propose that the expanded SAT model addresses this complexity and that future research into mixed-initiative human–agent teaming can use this model to inform their approach towards bidirectional transparency.

**Acknowledgments**

This work was supported by the U.S. Department of Defense Autonomy Research Pilot Initiative (Program Manager Dr DH Kim) and the U.S. Army Research Laboratory’s Human-Robot Interaction Program (Program Manager Dr Susan Hill). The authors would like to thank Ryan Wohleber, Paula Durlach, and Jack L. Hart for their assistance in the preparation of this manuscript.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Notes on contributors**

*Jessie Y. C. Chen* is a senior research psychologist with US Army Research Laboratory-Human Research and Engineering Directorate at the Field Element in Orlando, Florida. She earned her PhD in applied experimental and human factors psychology from the University of Central Florida.

*Shan G. Lakhmani* is a research assistant with the Institute of Simulation and Training at the University of Central Florida. He is currently pursuing his PhD in modelling and simulation.

*Kimberly Stowers* is a doctoral candidate in cognitive and human factors psychology at the University of Central Florida. She received an MS in modelling and simulation from the same university. Her research involves the study of attitudes, behaviours, and cognition in human–machine interaction.

*Anthony R. Selkowitz* is a post-doctoral fellow with Oak Ridge Associated Universities and US Army Research Laboratory-Human Research and Engineering Directorate at the Field Element in Orlando, Florida. He earned his PhD in applied experimental and human factors psychology from the University of Central Florida.

*Julia Wright* is a research psychologist with US Army Research Laboratory (ARL) – Human Research & Engineering Directorate at the field element in Orlando, Florida. Her research interest is bridging the intuitive gap between humans and technology, including human–agent teaming, human–technology interaction, visual attention, individual differences, and agent transparency.

*Michael Barnes* works for US Army Research Laboratory (ARL) where he works on a number of projects related to humans and autonomous systems. He currently serves as member of a NATO
committee related to his research efforts on autonomous systems and has co-authored over 90 articles on the human factors of military systems. In his present position, he is the ARL chief of the Ft. Huachuca, AZ field element where he supports cognitive research in intelligence analysis, unmanned aerial systems, and visualisation tools for risk management.

References

Allen, J. E., C. I. Guinn, and E. Horvitz. Sep./Oct. 1999. “Mixed-Initiative Interaction.” *IEEE Intelligent Systems and their Applications* 14(5): 14–23. doi: 10.1109/5254.796083.

Barnes, M. J., J. Y. C. Chen, and F. Jentsch. 2015. “Designing for Mixed-Initiative Interactions between Human and Autonomous Systems in Complex Environments.” 2015 *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. Hong Kong: IEEE. 1386–1390. doi: 10.1109/SMC.2015.246.

Barrett, S., P. Stone, S. Kraus, and A. Rosenfeld (2012). “Learning Teammate Models for ad hoc Teamwork.” In *AAMAS Adaptive Learning Agents (ALA) Workshop*, Valencia. 57–63.

Bass, E. J., L. A. Baumgart, and K. K. Shepley. 2013. “The Effect of Information Analysis Automation Display Content on Human Judgment Performance in Noisy Environments.” *Journal of Cognitive Engineering and Decision Making* 7(1): 49–65.

Behymer, K. J., M. J. Patzek, C. D. Rothwell, and H. A. Ruff. Forthcoming. Initial Evaluation of the Intelligent Multi-UXV Planner with Adaptive Collaborative/Control Technologies (IMPACT). Technical Report: AFRL-RH-WP-TR-2016-TBD, Wright-Patterson AFB, OH: Air Force Research Lab.

Beller, J., M. Heesen, and M. Vollrath. 2013. “Improving the Driver-Automation Interaction: An Approach Using Automation Uncertainty.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 56(6): 1130–1141. doi: 10.1177/0018720813482327.

Bruni, S., J. J. Marquez, A. Brzezinski, C. Nehmeand, Y. Boussemart. 2007, June. “Introducing a Human-Automation Collaboration Taxonomy (HACT) in Command and Control Decision-Support Systems.” In *Proceedings of the 12th International Command and Control Research and Technology Symposium*. Newport, RI.

Calhoun, G. L, H. A. Ruff, K. J. Behymer, and E. M. Mersch. 2017. “Operator-Autonomy Teaming Interfaces to Support Multi-Unmanned Vehicle Missions.” In *Advances in Human Factors in Robots and Unmanned Systems*, edited by P. Savage-Knepshield and J. Y. C. Chen, 113–126. Warsaw: Springer International Publishing.

Cannon-Bowers, J., E. Salas, and S. Converse. 1993. “Shared Mental Models in Expert Team Decision Making.” In *Individual and Group Decision Making: Current Issues*, edited by N. J. Castellan, 221–246. Hillsdale, NJ: Erlbaum.

Chen, J. Y. C., and M. J. Barnes. 2010. “Supervisory Control of Robots Using RoboLeader.” Proceedings of the Human Factors and Ergonomics Society Annual Meeting. San Francisco, CA: Sage. 1483–1487. doi: 10.1177/1541931105401927.

Chen, J. Y. C., and M. J. Barnes. 2012a. “Supervisory Control of Multiple Robots Effects of Imperfect Automation and Individual Differences.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54(2): 157–174. doi: 10.1177/001872081435843.

Chen, J. Y. C., and M. J. Barnes. 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., and M. J. Barnes. 2014. “Human-Agent Teaming for Multirobot Control: A Review of Human Factors Issues.” *IEEE Transactions on Human-Machine Systems* 44(1) 13–29. doi: 10.1109/THMS.2013.2293535.
Chen, J. Y. C., and M. J. Barnes. 2015. “Agent Transparency for Human-Agent Teaming Effectiveness.” 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Hong Kong: IEEE Press. 1381–1385. doi: 10.1109/SMC.2015.245.

Chen, J. Y. C., M. J. Barnes, A. R. Selkowitz, K. Stowers, S. G. Lakhmani, and N. Kasdaglis. 2016. “Human-Autonomy Teaming and Agent Transparency.” Companion Publication of the 21st International Conference on Intelligent User Interfaces. Sonoma, CA: ACM. 28–31. doi: 10.1145/2876456.2879479.

Chen, J. Y. C., K. Procci, M. Boyce, J. Wright, A. Garcia, and M. Barnes. 2014. Situation Awareness-Based Agent Transparency. Report No. ARL-TR-6905, Aberdeen Proving Ground, MD, U.S. Army Research Laboratory.

Cook, M. B., and H. S. Smallman. 2008. “Human Factors of the Confirmation Bias in Intelligence Analysis: Decision Support from Graphical Evidence Landscapes.” Human Factors: The Journal of the Human Factors and Ergonomics Society 50(5): 745–754. doi: 10.1518/001872095779049543.

de Visser, E. J., M. Cohen, A. Freedy, and R. Parasuraman. 2014. “A Design Methodology for Trust Cue Calibration in Cognitive Agents,” In Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments, edited by R. Shumaker, S.Lackey, 251–262. Cham, Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-07458-0_24.

DeChurch, L. A., and J. R. Mesmer-Magnus. 2010. “The Cognitive Underpinnings of Effective Teamwork: a Meta-Analysis.” Journal of Applied Psychology 95(1): 32–53. doi: 10.1037/a0017328.

Defense Science Board. 2016. Defense Science Board Summer Study on Autonomy, Washington, DC: Under Secretary of Defense.

Department of Defense. 2013. “Autonomy Research Pilot Initiative Web Feature.” Last modified June 14. http://www.acq.osd.mil/chieftechnologist/ari.html.

Endsley, M. R. 1995. "Toward a Theory of Situation Awareness in Dynamic Systems." Human Factors: The Journal of the Human Factors and Ergonomics Society 37(1): 32–64. doi: 10.1518/001872095779049543.

Fern, L., and R. J. Shively. 2009. “A Comparison of Varying Levels of Automation on the Supervisory Control of Multiple UASs.” Proceedings of AUVSI’s Unmanned Systems North America. New York, NY: Curran Associates Inc. 10–13. https://www.researchgate.net/profile/Lisa_Fern/publication/257645042_A_COMPARISON_OF_VARYING_LEVELS_OF_AUTOMATION_ON_THE_SUPERVISORY_CONTROL_OF_MULTIPLE_UASs/links/004635258b5a8ed6b7000000.pdf.

Fong, T. W., I. Nourbakhsh, C. Kunz, L. Flueckiger, J. Schreiner, R. Ambrose, R. Burridge, et al. 2005. “The Peer-to-Peer Human-Robot Interaction Project.” AIAA Space 2005. Long Beach, CA: American Institute of Aeronautics and Astronautics. doi: 10.2514/6.2005-6750.

Fong, T. W., C. Thorpe, and C. Baur. 2003. “Robot, Ask of Questions.” Robotics and Autonomous Systems 42(3): 235–243. doi: 10.1016/S0921-8890(02)00378-0.

Gillespie, K., M. Molineaux, M.W. Floyd, S. S. Wattam, and D. W. Aha. 2015. “Goal Reasoning for an Autonomous Squad Member,” Technical Report GT-IRIM-CR-2015-001, In Goal Reasoning: Papers from the ACS Workshop, edited by D. W. Aha, 52–67. Atlanta, GA: Georgia Institute of Technology, Institute for Robotics and Intelligent Machines.

Goodrich, M. A. 2013. “Multitasping and Multi-Robot Management.” In Oxford Handbook of Cognitive Engineering, edited by J. D. Lee and A. Kirklik, 379–394. New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780199757183.013.0025.

Gunning, D. 2016. Explainable Artificial Intelligence. http://www.darpa.mil/program/explainable-artificial-intelligence.

Hayes, B., and B. Scassellati. 2013. “Challenges in Shared-Environment Human-Robot Collaboration.” Paper presented at the 8th ACM/IEEE International Conference on Human-Robot Interaction, Workshop on Collaborative Manipulation. Tokyo, Japan, March 3-6. http://bradhayes.info/papers/hayes_challenges_HRI_collab_2013.pdf.

Helldin, T. 2014. “Transparency for Future Semi-Automated Systems.” PhD diss., Orebro University.
Helldin, T., U. Ohlander, G. Falkman, and M. Riveiro. 2014. “Transparency of Automated Combat Classification.” Engineering Psychology and Cognitive Ergonomics 22–33. doi: 10.1007/978-3-319-07515-0_3.

Jones, D. G., and Kaber, D. B. 2004. “Situation Awareness Measurement and the Situation Awareness Global Assessment Technique.” Handbook of Human Factors and Ergonomics Methods, edited by N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. W. Hendrick, 42–1-42-8. London: Taylor & Francis.

Kaupp, T., A. Makarenko, and H. Durrant-Whyte. 2010. “Human-Robot Communication for Collaborative Decision Making—A Probabilistic Approach.” Robotics and Autonomous Systems 58 (5): 444–456. doi: 10.1016/j.robot.2010.02.003.

Koola, P. M., S. Ramachandran, and K. Vadakkeveedu. (2016). “How Do We Train a Stone to Think? A Review of Machine Intelligence and its Implications.” Theoretical Issues in Ergonomics Science 17(2): 211–238. doi: 10.1080/1463922X.2015.1111462.

Lee, J. D. 2012. “Trust, Trustworthiness, and Trustability.” Presented at the Workshop on Human Machine Trust for Robust Autonomous Systems. Ocala, FL, (2012 Jan–Feb).

Lee, J. D., and K. A. See. 2004 “Trust in Automation: Designing for Appropriate Reliance.” Human Factors: The Journal of the Human Factors and Ergonomics Society 46(1): 50–80. doi: 10.1518/hfes.46.1.50.30392.

Linegang, M. P., H. A. Stoner, M. J. Patterson, B. D. Seppelt, J. D. Hoffman, Z. B. Crittendon, and J. D. Lee. 2006. “Human-Automation Collaboration in Dynamic Mission Planning: A Challenge Requiring an Ecological Approach.” Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 50(23): 2482–2486. doi: 10.1177/154193120605002304.

Lyons, J. B. 2013. “Being Transparent About Transparency: A Model For Human-Robot Interaction.” Technical Report SS-13-07. In Trust and Autonomous Systems: Papers from the AAAI Spring Symposium, D. Sofge, G. Kruijff, W.F. Lawless , pp. 48–53. doi: https://pdfs.semanticscholar.org/bf51/f5448981d985e082f2d9511d276ca6519944.pdf.

Lyons, J. B., and P. R. Havig. 2014. “Transparency in a Human-Machine Context: Interface Approaches for Fostering Shared Awareness/Intent.” In Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments, edited by R. Shumaker, S. Lackey, pp. 181–190. Cham, Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-07458-0_18.

Lyons, J. B., G. G. Sadler, K. Koltai, H. Battiste, N. T. Ho, L. C. Hoffmann, D. Smith, W. Johnson, and R. Shively. 2017. “Shaping Trust Through Transparent Design: Theoretical and Experimental Guidelines.” In Advances in Human Factors in Robots and Unmanned Systems, pp. 127–136, Cham, Switzerland: Springer International Publishing.

Mathieu, J. E., T. S. Heffner, G. F. Goodwin, E. Salas, and J. A. Cannon-Bowers. 2000. “The Influence of Shared Mental Models on Team Process and Performance.” Journal of Applied Psychology 85(2): 273–283. doi: 10.1037/0021-9010.85.2.273.

McGuirl, J. M., and N. B. Sarter. 2006. “Supporting Trust Calibration and the Effective Use of Decision Aids by Presenting Dynamic System Confidence Information.” Human Factors: The Journal of the Human Factors and Ergonomics Society 48(4): 656–665. doi: 10.1518/001872006779166334.

Mercado, J. E., M. A. Rupp, J. Y. C. Chen, M. J. Barnes, D. Barber, and K. Procci. 2016. “Intelligent Agent Transparency in Human-Agent Teaming for Multi-UxV Management.” Human Factors: The Journal of the Human Factors and Ergonomics Society 58.3(2016): 401–415. doi: 10.1177/0018720815621206.

Meyer, J., and J. D. Lee. 2013. “Trust, Reliance, and Compliance.” In The Oxford Handbook of Cognitive Engineering, edited by J. D. Lee and A. Kirlik, pp.109–124. New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780199757183.013.0007.

Neyedli, H. F., J. G. Hollands, and G. A. Jamieson. 2009. “Human Reliance on an Automated Combat ID System: Effects of Display Format.” Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Sage Publications) 53(4): 212–216. doi: 10.1177/154193120905300411.
Parasuraman, R., M. Barnes, and K. Cosenzo. 2007. “Adaptive Automation For Human-Robot Teaming in Future Command and Control Systems.” *International Journal of Command and Control* 1(2): 43–68.

Parasuraman, R., and C. A. Miller. 2004. “Trust and Etiquette in High-Criticality Automated Systems.” *Communications of the ACM* 47(4): 51–55. doi: 10.1145/975817.975844.

Parasuraman, R., R. Molloy, and I. L. Singh. 1993. “Performance Consequences of Automation-Induced ‘Complacency’.” *The International Journal of Aviation Psychology* 3(1): 1–23. doi: 10.1207/s15327108iap0301_1.

Parasuraman, R., and V. Riley. 1997. “Humans and Automation: Use, Misuse, Disuse, Abuse.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39(2): 230–253. doi: 10.1518/00187209778543886.

Rao, A. S., and M. P. Georgeff. 1995. “BDI Agents: From Theory to Practice.” *Proceedings of the First International Conference on Multiagent Systems*. Palo Alto, CA: Association for the Advancement of Artificial Intelligence Press, 312–319.

Russell, S., and P. Norvig. 2009. *Artificial Intelligence: A Modern Approach*. 3rd ed. Upper Saddle River, NJ: Prentice Hall Press.

Selkowitz, A. R., S. G. Lakhmani, C. N. Larios, and J. Y. C. Chen. 2016. “Agent Transparency and the Autonomous Squad Member.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60(1): 1319–1323. doi: 10.1177/1541931216601305

Selkowitz, A. R., C. N. Larios, S. G. Lakhmani, and J. Y. C. Chen. 2017. “Displaying Information to Support Transparency for Autonomous Platforms.” *Advances in Human Factors in Robots and Unmanned Systems*. Orlando, FL: Springer International Publishing, 161. doi: 10.1007/978-3-319-41959-6_14.

Seppelt, B. D., and J. D. Lee. 2007. “Making Adaptive Cruise Control (ACC) Limits Visible.” *International Journal of Human-Computer Studies* 65(3): 192–205. doi: 10.1016/j.ijhcs.2006.10.001.

Shattuck, L. G. 2015. “Transitioning to Autonomy: A Human Systems Integration Perspective.” Presented at Transitioning to Autonomy: Changes in the role of humans in air transportation, Moffett Field, CA, 10–12 March 2015. https://human-factors.arc.nasa.gov/workshop/autonomy/download/presentations/Shaddock%20.pdf.

Smith, P. J., C. E. McCoy, and C. Layton. 1997. “Brittleness in the design of cooperative problem-solving systems: The effects on user performance.” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 27(3): 360–371. doi: 10.1109/3468.568744.

Stanton, N. A., M. S. Young, and G. H. Walker. 2007. “The Psychology of Driving Automation: a Discussion with Professor Don Norman.” *International Journal of Vehicle Design* 45(3): 289–306. doi: 10.1504/IJVD.2007.014906.

Stowers, K., J. Y. C. Chen, N. Kasdaglis, O. Newton, M. Rupp, and M. Barnes. Forthcoming. *Effects of Situation Awareness-Based Agent Transparency Information on Human Agent Teaming for Multi-UxV Management*, Report No. ARL-TR-XX, Aberdeen Proving Ground, MD: Army Research Lab.

Stubbins, K., D. Wettergreen, and P. H. Hinds. 2007. “Autonomy and Common Ground in Human-Robot Interaction: A Field Study.” *IEEE Intelligent Systems* 22(2): 42–50. doi: 10.1109/MIS.2007.21.

Sycara, K., and G. Sukthankar. 2006. “Literature review of teamwork models.” Report No. CMU-RI-TR-06-50. Pittsburgh, PA: Robotics Institute, Carnegie Mellon University.

Thomaz, A. L., and C. Breazeal. 2006. “Transparency and Socially Guided Machine Learning.” Paper presented at the 5th International Conference on Development and Learning (ICDL), Bloomington, IN, May 31-June 3.

Tufte, E. R. 2001. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Warner Norcross and Judd. 2015. *Trucking, Mining Industries Blazing a Path to Vehicle Autonomy*. January 14. http://www.wnj.com/Publications/Trucking-Mining-Industries-Blazing-a-Path-to-Vehic.
Wooldridge, M., and N. R. Jennings. 1995. “Intelligent Agents: Theory and Practice.” The Knowledge Engineering Review 10(2): 115–152. doi: 10.1017/S0269888900008122.

Wright, J. L., J. Y. C. Chen, P. A. Hancock, I. Yi, and M. J. Barnes. Forthcoming. “Transparency of Agent Reasoning and its Effect on Automation-Induced Complacency.” Report No. ARL-TR-XX, Aberdeen Proving Ground, MD: Army Research Lab.

Wright, J. L., J. Y. C. Chen, S. A. Quinn, and M. J. Barnes. 2013. “The Effects of Level of Autonomy on Human-Agent Teaming for Multi-Robot Control and Local Security Maintenance ” Report No: ARL-TR-6724, Aberdeen Proving Ground, MD: Army Research Laboratory.

Zhou, J., and F. Chen. 2015. “Making Machine Learning Useable.” International Journal of Computational Science and Engineering 14(2): 91–109. doi: 10.1504/IJISTA.2015.074069.