Australian Government

Department of Defence
Science and Technology

The Operationalisation of Agent Transparency and Evidence for Its Impact on Key Human-Autonomy Teaming Variables

Adella Bhaskara

Aerospace Division
Defence Science and Technology Group

DST-Group-TR-3413

ABSTRACT

The growing interest in the use of autonomous systems for both military and commercial applications has been accompanied by a concomitant increase in research involving human-agent interaction. Transparency has been investigated as one factor that could improve human trust in, and appropriate reliance on, autonomous systems. This report provides a review of studies that have examined how the transparency of an autonomous system affects key variables such as operator performance, response time, situation awareness, perceived usability, and subjective workload. Theoretical frameworks that support transparency in autonomous systems including Lyons' models of transparency (2013) and the Situation Awareness-Based Agent Transparency (SAT) model (Chen et al., 2014) are also presented. Findings from these studies indicate that a certain amount of transparency seemed to improve operator performance, however too much transparency information could also decrease operator performance. Overall, the findings have not been clear-cut in terms of how much and what type of transparency information should be communicated to the operator. Future research should also examine the underlying mechanisms responsible for these transparency effects.

Produced by

Aerospace Division
Defence Science and Technology Group
506 Lorimer Street
Fishermans Bend VIC 3207

Telephone: 1300 333 362

October 2017
AR-016-986

APPROVED FOR PUBLIC RELEASE

The Operationalisation of Agent Transparency and Evidence for Its Impact on Key Human-Autonomy Teaming Variables

Executive Summary

The growing interest in the use of autonomous systems for both military and commercial applications has been accompanied by a concomitant increase in research involving human-agent interaction. Transparency has been identified as one factor that could improve human trust in, and appropriate reliance on, autonomous systems (Hancock et al., 2011). This report examines how transparency has been operationalised in the literature, and reviews evidence of the impact of transparency on key human-autonomy teaming in order to guide future research.

Transparency refers to an operator's awareness of an autonomous agent's actions, decisions, behaviour, and intention (e.g., Chen et al., 2014). Theoretical frameworks such as Lyons' models of transparency (2013) and the Situation Awareness-Based Agent Transparency (SAT) model (Chen et al., 2014) have been proposed to support transparency, respectively, in human-robot interaction and human-agent teaming. Specifically, these models provide guidance on what information should be communicated to the human to support the interaction between the human and the autonomous system.

There have been five studies that have manipulated different levels of transparency and investigated its effects on variables such as operator performance, response time, subjective workload, situation awareness, trust, and usability of the system (i.e., Mercado et al., 2015/2016; Selkowitz, Lakhmani, Larios, & Chen, 2016; Stowers et al., 2016; Wright, Chen, Barnes, & Boyce, 2015; Wright, Chen, Barnes, & Hancock, 2016a, 2016b). In general, these studies found that transparency information imparting information about an agent's reasoning improved operator performance; however, some studies found that additional transparency information actually worsened operator performance (Wright et al., 2015, 2016b). Transparency did not seem to affect operator response time and subjective workload (Mercado et al., 2015/2016; Selkowitz et al., 2016; Wright et al., 2015, 2016a, 2016b). Out of these five studies, only one study included a measure of situation awareness and found that the additional transparency information (i.e., predicted outcomes and agent's reasoning) improved operator situation awareness, but not when uncertainty information was also included (Selkowitz et al., 2016). These findings indicate that providing too much transparency information may overwhelm the operator. In terms of subjective trust and perceived usability, the results have been inconsistent. For example, Mercado et al. (2015/2016) found that subjective trust increased only when uncertainty information was given; however, Selkowitz et al. (2016) found that agent's reasoning, but not the additional uncertainty information, increased subjective trust. Finally, while Mercado et al. found that perceived usability scores increased when an agent's reasoning and uncertainty information were given, Stowers et al. (2016) found that the addition of uncertainty information actually lowered participants' scores on perceived usability.

In summary, although operator performance, situation awareness, perceived usability, and trust seemed to be affected by agent transparency, the results from past studies have not been clear-cut in terms of how much and what type of information should be included and communicated to the operator. Some of these studies suggest that while higher levels of transparency may improve some of the human-autonomy teaming variables, the highest transparency level did not always produce the best outcome. Future research should further investigate this issue (i.e., to specify the type of information that would be beneficial for the operator in a given context), and further examine the underlying mechanisms that could explain these transparency effects. Understanding these psychological processes is fundamental to designing an interface that supports transparency in human-autonomy teaming.

Author

Adella Bhaskara
Aerospace Division

Adella holds a Bachelor of Psychological Science (Honours) from the University of Queensland and a PhD in Applied Experimental Psychology from the University of Adelaide. She is a Visiting Research Fellow at the University of Adelaide and recently joined Defence Science and Technology Group to contribute to research on Human-Autonomy Teaming with the Human Factors group, in Aerospace Division.

This page is intentionally blank.

Contents

This page is intentionally blank.

1. Introduction

The purpose of this report is to provide a detailed analysis of the key literature relating to agent transparency in human-autonomy teaming. In addressing this aim, this report starts with a brief introduction to automation and autonomous systems, and why transparency is a relevant issue worth investigating (Section 1). This report then examines how transparency has been operationalised in the literature and discusses the theoretical frameworks—Lyons' models of transparency (2013) and the Situation Awareness-Based Agent Transparency (SAT) model (Chen et al., 2014)—that have been proposed to support transparency in human-autonomy teaming (Section 2). Following this, a critical review of the existing literature that has examined how transparency affects key human-autonomy teaming variables including operator performance, response time, situation awareness, perceived usability, and subjective workload is presented (Section 3). This review provides an overview of the experimental tasks, scenarios, and interfaces that were used in past studies; and aims to find commonalities and differences in how transparency has been implemented in an autonomous system and presented to the operator. This report ends by providing recommendations for future research on the basis of this review (Section 4).

1.1 Automation and Autonomous Systems

Automation has been used in many areas—such as manufacturing, health care, transportation, and aviation—to assist and enhance human performance (e.g., through an automated aid to reduce memory load), perform complex tasks that are beyond human capabilities (e.g., complex mathematical operations), and in hazardous environments where human safety could be compromised (e.g., mining; Wickens, Hollands, Banbury, & Parasuraman, 2013). One of the primary drivers of the use of automation is to reduce labour costs and increase productivity. Automated unmanned air vehicles, for example, cost less to fly and manufacture than manned airplanes (Cooke, Pringle, Pedersen, & Connor, 2006).

The degree of automation in a system can vary across a continuum, from the lowest level in which the human takes all responsibility to the highest level in which the system decides everything and acts autonomously (see Cummings, Bruni, Mercier, & Mitchell, 2007; Parasuraman, Sheridan, & Wickens, 2000; Ruff, Narayanan, & Draper, 2002; Sheridan & Verplank, 1978). Regardless of the different levels, higher levels of automation always imply more responsibility for the automated system, requiring less intervention by the human—and potentially resulting in less work for the human (Wickens et al., 2013).

Higher levels of automation are commonly referred to as *autonomous systems* and are often differentiated from the less complex, lower levels of automation. *Automation* refers to rule-based systems that have been programmed to achieve specific, pre-defined outcomes (e.g., modern programmable thermostats). *Autonomous systems*, however, have the ability to learn from the environment, have a degree of self-governance and self-directed behaviour, and are able to evolve and perform tasks or functions on their own. As a result, they are not directly predictable in their behaviour (e.g., self-learning robot that taught itself how to

walk; Hancock, 2017; Scharre, 2015). Additionally, as autonomous systems possess higher levels of independent intelligence, they have a greater capacity to make decisions in uncertain and unplanned circumstances (Schaefer, Chen, Szalma, & Hancock, 2016).

Agents can be implemented in both automation and autonomous systems. Agents are hardware or software-based computer systems possessing the following characteristics: (a) autonomy (they are able to operate without human intervention for a significant length), (b) social ability (they are able to interact and communicate with humans or other agents), (c) observability and reactivity (they are able to perceive the environment through sensors and react to it), (d) proactiveness (they are able to self-direct behaviour in anticipation of future events and to achieve mission goals; see Lakhmani, Abich, Barber, & Chen, 2016; also Chen & Barnes, 2014). Agents can vary in complexity—from simple reflex agents (e.g., thermostats) to autonomous agents (e.g., intelligent agents that can learn and evolve, Chen & Barnes, 2014).

The intelligence of an autonomous agent, however, is not as flexible or as robust as human intelligence in the capacity to understand patterns of behaviour, human intentions, implications, and ethical responsibilities (see Chen & Barnes, 2014). As tasks in autonomous system require higher levels of planning, judgment and decision making, and remote operations (also Cummings, Bruni, Mercier, & Mitchell, 2007; Goodrich & Cummings, 2015) the role of humans becomes more important. The involvement of human control in higher-level-knowledge-based behaviours (as opposed to lower-level-skill-based behaviours) is known as human supervisory control (Goodrich & Cummings, 2015).[1] More specifically, human supervisory control is the process by which the human and the system are inter-dependent: the human interacts with the agent, monitoring its actions, receiving feedback, and providing commands for future actions (Goodrich & Cummings, 2015; Wickens, Hollands, Banbury, & Parasuraman, 2013). The degree of human monitoring that is required also varies with the level of automation; for example, as the automation takes on more responsibility, the human requirement for monitoring increases (Parasuraman, 1987, as cited in Wickens et al., 2013). This human supervisory control is also referred to as human-agent teaming (Chen, Barnes, & Harper-Sciarini, 2011).

In the military aerospace domain, one of the primary goals of using an autonomous system is to minimise the number of operators, or to have a single operator supervising multiple unmanned assets (Shaw et al., 2010). However, managing multiple assets of differing constraints and capabilities (e.g., monitoring system performance, supporting frequent re-planning, and re-tasking in response to evolving mission needs and operational environment) increases cognitive workload and places significant demands on the operator's attentional resources. These could lead to (a) the operator losing situation awareness, (b) complacency or overreliance on the autonomous systems, or (c) skill degradation of the operator (Endsley & Kiris, 1995; Parasuraman et al., 2000; Parasuraman & Riley, 1997). Furthermore, due to the complexity of these autonomous systems, operators often have difficulty in understanding how the system works or why a certain decision has been made by an autonomous agent (Linegang et al., 2006). Consequently,

[1] Also known as human-on-the loop (Chen & Barnes, 2014), as compared to human-*in*-the loop.

operators sometimes question the accuracy and effectiveness of the agent's action; this in turn can decrease their trust in the autonomous agent (Lakhmani et al., 2016).

Implementing system *transparency* (e.g., Chen & Barnes, 2014) has been proposed as one method of addressing some of the problematic aspects associated with the interaction between humans and autonomous systems. In a transparent system, information regarding the autonomous agent's actions, decisions, behaviour, and intentions is communicated to the operator through an appropriate interface with the aim of improving trust in the system, performance, and situation awareness (e.g., Mercado et al., 2015; Selkowitz, Lakhmani, Chen, & Boyce, 2015; Wang, Jamieson, & Hollands, 2009). The next section presents definitions and models of transparency that have been proposed in the literature.

2. Transparency in Autonomous Systems

Transparency refers to an operator's awareness of an autonomous agent's behaviour, reliability, and intention. Specifically, transparency is about understanding why an autonomous agent behaved in a particular way (Kim & Hinds, 2006), understanding an agent's reliability (Wang et al., 2009), an agent's tendency for errors (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003), and an agent's intended action (Ososky, Sanders, Jentsch, Hancock, & Chen, 2014). Transparency has been further described as a means of sharing intent and awareness between the operator and the autonomous agent (Lyons, 2013). Essentially, the purpose of transparency is to facilitate an operator's comprehension of an autonomous agent's intent, performance, abilities, future plans, and reasoning process (Chen et al., 2014). Two models of transparency have been proposed in the literature: Lyon's (2013) models of transparency for human-robot interaction, and Chen et al.'s (2014) Situation Awareness–Based Agent Transparency Model.

2.1 Lyons' Models of Transparency (2013)

Lyons (2013) suggested several factors that are important to support effective human-robot interaction. These factors fall under two key aspects of transparency: the robot communicating to the operator about its knowledge of the system and its view of the world (robot-*to*-human), and the robot communicating its awareness about the state of the human operator (robot-*of*-human).

2.1.1 Robot-to-Human Transparency

In robot-to-human transparency, the information that it is crucial to convey to an operator can be described in terms of (a) an intentional model, (b) a task model, (c) an analytical model, and (d) an environmental model. According to the intentional model, it is important for the human operator to fully understand the design, purpose, and intent of the robotic system (and how these match with the operator's expectations). In other words, the operator needs to clearly understand why and for what purpose the robot was created,

and how the system seeks to perform these actions. For example, if the robot was to override a human directive, the operator should be aware that this could occur, and also understand why it occurred and when.

The task model relates to specific actions of the robot. Once the operator understands the design, purpose, and intent of the robotic system, he or she can begin to analyse the actions of the robot. According to the task model, the robot must communicate to the human an understanding of the task at hand, its intent in terms of what goals it is trying to accomplish, its progress in achieving those goals, as well as its capabilities and tendency for errors.

The analytical model is concerned with understanding the underlying analytical structure of the robot's decision process (i.e., how the robot is doing the analysis and how it makes decisions). Here, the robot needs to communicate to the human and share details about the rationale for its behaviour or system recommendations, as well as provide an understanding of the programming of the technology. According to Lyons (2013), such awareness will be useful in uncertain situations, as this enables humans to take over manual control when necessary. For example, knowing that a robot fuses information from satellite imagery and ground sensors (in order to detect the location of potential emergency zones) would be useful in cases where the ground sensor networks had been compromised.

Finally, the environmental model involves educating the operator about how the system senses information in the environment. More specifically, this involves communicating to the operator about the robot's understanding of the dynamics of its surrounding environments (e.g., for potential hostility and temporal constraints). According to Lyons (2013), this will help enhance operator situation awareness during uncertainty, and help calibrate the operator's reliance on the robot.

2.1.2 Robot-of-Human Transparency

While robot-*to*-human transparency focuses on the robot communicating to the operator about its knowledge of the system and its view of the world, robot-*of*-human transparency focuses on the robot communicating its awareness about factors relating to the human. Lyons (2013) robot-of-human transparency is explained through (a) the teamwork model and (b) the human state model. According to the teamwork model, both the human and the robot need to have a shared understanding of each other's role (i.e., about who is responsible for a given task or set of tasks), and of the level of autonomy that the system is operating under. Once a shared awareness has been established, the human state model explains that the robot needs to communicate an understanding of the humans' cognitive, emotional, and physical state. For instance, the robot needs to sense when the human is distressed or fatigued (e.g., through monitoring the human's cognitive workload) so that the robot could execute an alert to the human, and either recommend a higher level of autonomy or assume autonomous control if human limitations present a safety concern. In summary, Lyons' models of transparency suggest that for overall human-robot systems to function effectively, the human users need to understand information about the robot

(robot-*to*-human transparency), as well as having the robot communicate its awareness and understanding of human-centric factors (robot-*of*-human transparency).

2.2 Situation Awareness-Based Agent Transparency Model

Following Lyons' models of transparency (2013), Chen and colleagues (2014) have proposed a model of transparency called the Situation Awareness-Based Agent Transparency (SAT) model. Similar to Lyons' models of transparency, the SAT model provides guidance on what information should be communicated to human operators to promote agent transparency. The SAT model was influenced by: Endsley's (1995) theory of Situation Awareness (SA); the Beliefs, Desires, Intentions (BDI) Agent Framework (Rao & Georgeff, 1995); and the three factors considered fundamental to human automation trust (purpose, process, and performance; Lee & See, 2004). Before presenting a detailed overview of the SAT model, each of these theoretical frameworks will first be briefly discussed.

2.2.1 Theory of Situation Awareness

Endsley's (1995) theory of situation awareness (SA) involves three levels of awareness: (1) perception of elements in the environment, (2) comprehension of those elements, and (3) projection of their status in the near future. According to Endsley, the first step in achieving SA is to perceive the status, attributes, and dynamics of relevant elements in the environment. These elements could be perceiving aircraft, mountains, or warning lights (alongside their relevant characteristics such as colour, size, speed, and location). The second step in achieving SA is to comprehend or understand the situation by integrating various Level 1 SA elements in light of operator goals. For example, if a military pilot sees three enemy aircraft within a certain proximity of one another and in a certain geographical location, he or she must understand that this situation indicates certain things about the aircraft's objectives (Endsley, 1995). The third step in achieving SA is to project the future actions of the elements in the environment. For example, from observing behaviour patterns and the location of a threat aircraft, a fighter pilot should be able to project its likelihood of attack.

According to Endsley (1995), SA is built over time and the projection of future actions (Level 3 SA) is achieved through the knowledge of the status, attributes, and dynamics of the relevant elements (Level 1 SA) and the comprehension of the situation (Level 2 SA). In other words, SA goes beyond simply perceiving information about the environment, it also involves integrating and comprehending the meaning of that information, comparing it with operator goals, and projecting future states of the environment.

2.2.2 Purpose, Process, and Performance

The SAT model also incorporates *purpose, process,* and *performance* factors. These three factors have been identified as the antecedents for trust development in the context of human-agent interaction (Lee & Moray, 1992; Lee & See, 2004). As described in Lee and See's (2004) summary of the human-automation trust literature, *purpose* represents motives

or intent—it describes the extent to which the automation is being used according to the designer's intent. *Process* represents an understanding of the underlying qualities or characteristics that govern how the system behaves—it describes the extent to which the automation's algorithms are appropriate for a given task and situation and for achieving operator's goals. *Performance* represents the expectation of consistent, stable, and desirable performance or behaviour—it refers to the current and historical operation of the automation. On the basis of these three factors, under the SAT model Chen et al. (2014) argue that to promote agent transparency an operator should understand (a) the purpose of the automation or why it was developed, (b) the uses and limitations of the automation and whether it is capable of achieving the operator's goals in a given task, and (c) the reliability, predictability, and capability of the system.

2.2.3 Beliefs, Desires, Intentions

Finally, the development of the SAT model is also influenced by the Beliefs, Desires, Intentions (BDI) Agent Framework (Rao & Georgeff, 1995). BDI is a framework used to describe the behaviour of rational agents. *Beliefs* represent the information the agent has about the word; *desires* represent the motivational state of the system and information on the goals to be achieved; and *intentions* represent the desires that the agent has committed to achieving. According to the BDI framework, the agent's beliefs are updated based on its perception of the environment, communications with other agents or humans, and its inference mechanisms (also see Briggs & Scheutz, 2012). Once the agent's beliefs are updated, the agent continues with goal-selection (desires), and executing action plan (intentions). Accordingly, the SAT model states that to promote agent transparency, an operator should be informed about the agent's beliefs, desires, and intentions (Ososky et al., 2014).

2.2.4 The SAT Model: SAT Levels 1, 2, and 3

Following Endsley's (1995) theory of SA, the SAT model consists of three levels: SAT Levels 1, 2, and 3. The main difference between the SAT model and Endsley's SA model is that the SAT model focuses on operator's SA when there is an autonomous agent involved. When an autonomous agent is involved, the operator needs to have situation awareness of the autonomous agent, and according to Chen et al. (2014), this can be attained through agent transparency. Specifically, to support his or her SA, the operator needs to understand the agent's parameters, logic, and predicted outcomes (Ososky et al., 2014). The SAT model thus attempts to identify features of the environment and transparency requirements necessary to support operator's SA when an agent is involved—through the following three levels.

In the SAT model, the first level should provide the operator with basic information about the agent's current state, goals, intentions, and proposed actions. This information is presented to the operator through an interface. More specifically, Level 1 includes the agent's purpose (the current goal), process (the agent's intent, planning process, and the agent's current progress), current performance, and status (see Figure 1; Chen et al., 2014). The second level of the SAT model should provide the operator with information about the agent's reasoning process or rationale behind its actions or decisions. According to

Chen et al., such rationale can be displayed on an interface through a representation of "resource limitations, constraints/affordances (environmental, situational, vehicular, etc.), feasibility, risk, trade-offs between alternatives, and history of past performance" (p. 10). Finally, the third level of the SAT model should provide the operator with information regarding the agent's projection of the future state, such as predicted consequences and the likelihood of a plan's success or failure, and uncertainty associated with these projections. More specifically, the visualisation of Level 3 in the interface includes the projection to future states (e.g., expected outcomes, probabilities of success with a confidence interval) as well as limitations (e.g., reliability, likelihood of error, history of past performance; Chen et al., 2014).

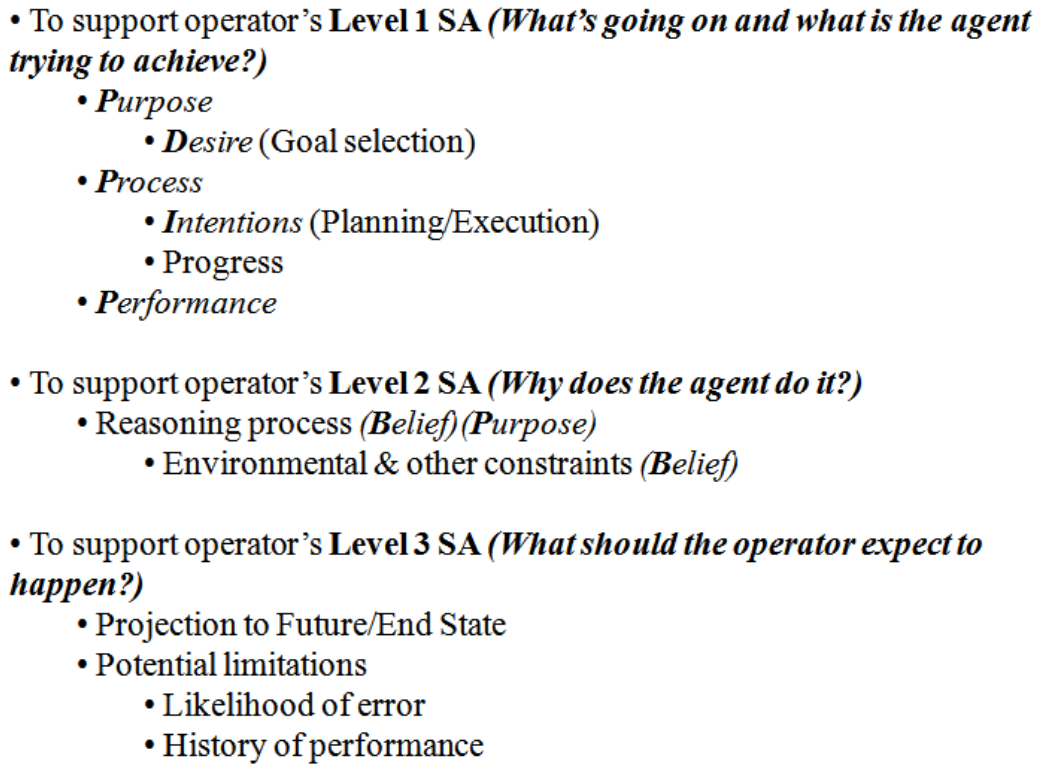

Figure 1. SA-based Agent Transparency model

Chen et al. (2014) stated that incorporating these three levels should improve operator SA of the autonomous agent, as they allow the operator to gain understanding of the agent's intent, reasoning process behind its action, projected outcomes, and uncertainty.[2] This would not only help the operator make informed decisions as to whether he or she should intervene, but should also lead to improving the operator's subjective trust as well as trust

[2] Although note that Chen et al. (2014) have also stated that the SAT model differs from Endsley's (1995) SA model, in that with the SAT model, the three levels may not be absolutely necessary to achieve transparency, and that the requirements to achieve system transparency are context dependent. For example, in a time-sensitive situation, the operator may only need to know the agent's proposed action (Level 1) and the projected outcome (Level 3) to make a decision.

calibration (i.e., proper reliance when the agent is correct [hit rates] and correct rejection when the agent is incorrect; Lee & See, 2004). A number of studies have investigated the effects of transparency in autonomous systems; these studies are reviewed in the next section.

3. Past Studies on Transparency in Autonomous Systems

This section contains a review of the existing literature concerning transparency in autonomous systems. To date, there have been five studies that have examined how transparency affects key human-autonomy teaming variables such as operator performance, response time, situation awareness, trust, perceived usability of the interface, and workload (i.e., Mercado et al. 2015/2016; Selkowitz et al., 2016; Stowers et al., 2016; Wright et al., 2015, 2016a, 2016b). This section is structured as follows. Section 3.1 starts by providing an overview of the experimental tasks, scenarios, and interfaces that were used in these studies. Section 3.2 then describes in more detail how transparency was investigated, operationalised, and implemented in these studies. Section 3.3 ends by presenting the results from these studies and discussing the extent to which transparency might facilitate human-autonomy teaming.

3.1 Overview of the Experimental Tasks

Different experimental scenarios, interfaces, and tasks were used in these five studies. In Mercado et al.'s (2015/2016) and Stowers et al.'s (2016) studies, participants took on the role of an operator supervising unmanned vehicles. Participants' task was to direct these vehicles to carry out missions while managing a commander's intent, as well as vehicle and environmental constraints. Similarly, in Wright et al.'s (2016a, 2016b) study, participants guided a convoy of a manned- and unmanned-vehicles through a simulated urban environment—with an agent assisting with the route planning task. Here, participants' task was to re-route the convoy as necessary according to events that occurred. In contrast, in Selkowitz et al.'s (2016) study, participants' task was to use the interface to gain information about an autonomous agent known as the Autonomous Squad Member (ASM), and its status, as they completed a route containing obstacles. Finally, in Wright et al.'s (2015) study, the experimental task involved route planning in a simulated urban environment. Specifically, participants had to direct a simulated dismounted soldier team to take an appropriate route as it moved from checkpoint to checkpoint. Participants' task was to ensure that the soldier team arrived at the final destination with the least required amount of resources possible.

In three of these five studies (i.e., Mercado et al., 2015/2016; Stowers et al., 2016: Wright et al. 2016), participants were required to make a decision and assess their decision based on the available information provided to them both through the environment and by the agent. Specifically, they were given a choice—between Plan A and Plan B—to follow or

reject the recommendation provided by the agent. When evaluating each plan, participants in Mercado et al.'s (2015/2016) and Stowers et al.'s (2016) studies had to take into account parameters such as speed, coverage, and capabilities of the vehicles suggested in each plan. Prior to receiving the mission objective, participants' task was to monitor unmanned vehicle positions and status, and received four intelligence messages containing either patrol report, updates on vehicle status, or commander intel messages (two of which were relevant to the task). In Wright et al.'s (2016a, 2016b) study, in addition to the route selection task, participants in their study had to (a) maintain communication with command (this included messages directed at other units which participants should disregard, and requests for information which required a response) and (b) maintain local security and detect threats.

In contrast, participants in the other two studies (i.e., Selkowitz et al., 2016; Wright et al., 2015) were not presented with two choices to make a decision. For example, participants' main task in Selkowitz et al.'s (2016) study was to monitor the interface (at varying levels of transparency) to gain information about the ASM and its status as they completed a route containing obstacles. During an obstacle encounter, the ASM would react by performing an action and participants' SA was then measured through SAGAT (Situation Awareness Global Assessment Technique) queries. While monitoring the interface, participants were also required to detect a target object that appeared in the environment. In Wright et al.'s (2015) study, participants' only task was to direct the soldier team to take an appropriate route as it moved from checkpoint to checkpoint. Participants were given a choice to select from three different routes to ensure that the soldier team arrived at the final destination using the least amount of resources.

3.2 Varying Levels of Transparency

Regardless of the differences in experimental scenarios, interfaces, and tasks, all of these five studies have manipulated transparency information at either three or four levels. Transparency information contains plans (or suggestions or explanations) made by an agent and communicated to the human operator through a computer interface. Transparency level increases as the amount of information presented to the operator increases (i.e., transparency information presented at higher levels would typically include transparency information presented at the lower level). However, it is important to note that the amount of transparency information provided at each level vary across studies. For example, some but not all of these studies used the SAT model to discriminate between transparency levels (i.e., Mercado et al. 2015/2016; Selkowitz et al., 2016; Stowers et al., 2016). Accordingly, these studies used different naming convention to denote different levels of transparency. For the purpose of this review, a more general term of low, medium, high, and very high transparency would therefore be used; and information presented at each level of transparency is further described below.

3.2.1 Low Transparency

This is the lowest level of transparency and in all five studies, the transparency information typically contained only basic plan or basic information. For example, in

Selkowitz et al.'s (2016) study, at low transparency, participants were only provided with basic information such as status, camera, map, resource indicators, and the Autonomous Squad Member location. Similarly in Wright et al.'s (2015) study, participants were only shown routes, which were coloured (either red, yellow, or green) to denote relative energy and/or time usage requirements needed to traverse that section of the route. When a choice of two plans was suggested by the agent, participants were shown a basic plan, with no reasoning as to why a particular plan was recommended (e.g., "A1 sector search on friendly boat" [Mercado et al., 2015/2016] and "Change to convoy path recommended" [Wright et al., 2016b]).

3.2.2 Medium Transparency

At medium transparency, all five studies added agent's reasoning in the transparency information. For example, in Mercado et al.'s (2015/2016) study, factors such as speed, coverage, capabilities, the environment, and the vehicle appropriateness—which would influence the agent's recommendation of Plan A, were presented in the interface through a text box, as well as through a sprocket (containing different wedge sizes and colours denoting each metric). In Stowers et al.'s (2016) study, participants were shown—through the Projected Plan Success tile—how the agent weighs each parameter (i.e., time, fuel/endurance, vehicle capability, and sensor coverage) according to its importance. Two statements regarding the intelligent agent's reasoning for Plan A and Plan B were also included (e.g., "Plan A: Hawk is the fastest UAV. Panther has pedestrian avoidance technology. Panther and Hawk are good for searching where visibility may be obscured"). Similarly, in Selkowitz et al.'s (2016), Wright et al.'s (2016b), and Wright et al.'s (2015) studies, (a) an icon depicting agent's reasoning (b) an explanation (e.g., "Change to convoy path recommended. Activity in area: Dense Fog"), and (c) text boxes and a bar graph, respectively, were added to their interfaces to indicate the reasoning behind the agents' decisions or suggestions.

3.2.3 High Transparency

At high transparency, additional information was presented through the computer interface. This information could relate to predicted outcomes, predicted consequences, uncertainty information, or additional reasoning of the agent's decision.[3] For example, the following studies included predicted outcomes or predicted consequences at high transparency (i.e., Selkowitz et al., 2016; Stowers et al., 2016; Wright et al., 2015). In Wright et al.'s (2015) study, a bar graph was added to the interface to show how specific route choices could impact the overall resource usage for mission completion. Likewise, in Selkowitz et al.'s (2016) study, predicted outcomes (of the agent's decisions and reasoning) were shown through an icon; and the severity of the predicted resources outcome was indicated by the number of red blocks displayed next to the icon (e.g., one block indicated low predicted resource usage, two blocks indicated moderate predicted resource usage, and three blocks indicated severe predicted resource usage). Similarly, in Stowers et al.'s (2016) study, the projected plan success of each plan (regarding time, fuel/endurance,

[3] Note that when two (instead of one) pieces of additional transparency information were presented to participants, the level of transparency increased to "very high transparency" (see Section 3.2.4).

vehicle capability, and sensor coverage) was displayed through the Projected Plan Success tile; and a projection statement was also added (i.e., "It is expected that Hawk will arrive the quickest, and Panther will move quickly on the base road").

In Mercado et al.'s (2015/2016) study, at high transparency, participants were told about the projection of uncertainty of the agent's recommendation. This information was shown through a transparent sprocket wedge, a transparent vehicle icon, and a bulleted statement in a text table (e.g., "It is uncertain how fog will affect speed"). Finally, in Wright et al.'s (2016a, 2016b) study, the high transparency was an added reasoning, where participants were additionally told when the agent had received the information (e.g., "Change to convoy path recommended. Activity in area: Dense Fog. Time of Report: 1 [h]"). Wright et al. indicated that while this information did not imply any confidence or uncertainty on the part of the agent, such additional information appeared to create ambiguity for the operator.

3.2.4 Very High Transparency

At very high transparency, further information was added on top of the additional information already provided in high transparency. This was investigated in two studies: Selkowitz et al. (2016) and Stowers et al. (2016) included both projected outcomes and uncertainty information at very high transparency. More specifically, in Selkowitz et al.'s study, participants were presented with basic information (low transparency), reasoning of the agent's decisions (medium transparency), predicted outcome (high transparency), and additional information regarding the uncertainty of the agent (very high transparency). The uncertainty of the agent was expressed through the use of semi-opaque blocks. For example, one solid red block and one semi-opaque red block indicate a certainty of low resource usage with the possibility of moderate resource usage. Uncertainty was also expressed through hazards and field events (displayed using icons surrounded by areas of effect). For example, if the exact location is uncertain, the icon is surrounded by a larger (cf. smaller), semi-opaque (cf. opaque) field to indicate that there might be a hazard or event in the general area.

Similarly, in Stowers et al.'s (2016) study, participants were presented with the basic plan, reasoning behind the agent's recommendation, projected plan success of each plan, and additional information regarding the uncertainty of the agent. Specifically, the uncertainty statement in their study contained two points: (a) the agent was uncertain about particular aspects of the tasking environment, and (b) the agent was making an assumption to deal with it (e.g., "It is uncertain how long the search will take. The agent assumes that the man will be found quickly").

3.3 Effects of Transparency on Autonomous Systems

These five studies have investigated the effects of transparency on a number of variables: performance, response time, situation awareness, trust, usability, and workload. The results of these studies will be discussed in turn below.

3.3.1 Performance

3.3.1.1 Performance Accuracy

Three studies (Mercado et al., 2015/2016; Stowers et al., 2016; Wright et al., 2016a, 2016b) have measured operator performance in terms of the operator's accuracy in accepting and rejecting the agent's recommendation—through hit rates, correct rejection rates, false alarm rates, and miss rates.

3.3.1.1.1 Hit Rates

Three studies have calculated operator hit rates (for correctly accepting the agent's recommendation) across the different levels of transparency conditions (i.e., Mercado et al., 2015/2016; Stowers et al., 2016; Wright et al., 2016b). Mercado et al. (2015/2016) found that proper agent usage rates were higher when participants were presented with agent's reasoning (i.e., contained in both medium and high transparency conditions), compared to when they were presented with only basic plan (low transparency). However, proper agent usage rates did not differ between the medium transparency condition (usage rate at 87%) and the high transparency condition (usage rate at 89%). In other words, the addition of reasoning information (medium transparency) increased proper use by 11% (cf. low transparency), but the addition of both reasoning and uncertainty information (high transparency) improved proper use only by an additional 2% (cf. medium transparency). These results suggest that the addition of reasoning information alone was sufficient to improve operator hit rates for correctly accepting the agent's recommendation.

The other two studies that have calculated hit rates either combined participants' hit rates with correct rejection rates (Wright et al., 2016b), or did not provide a full report of their results (Stowers et al., 2016). When hit rates were combined with correct rejection rates, Wright et al. (2016b) found that these combined rates were only marginally higher in the medium transparency condition (when agent's reasoning was present) than in the low transparency condition; and no difference was found between the medium transparency and the high transparency conditions. Between the three transparency conditions (i.e., medium transparency, high transparency, and very high transparency), Stowers et al. (2016) reported that the lowest and highest hit rates were found, respectively, in the medium transparency condition and the very high transparency condition. However, the statistical significance of these findings was not revealed, as the statistics and further details of these results were not reported in their study.

Taken together, all of these three studies agreed that the poorest hit rates were found in the lowest transparency condition. However, the evidence from some of these studies suggests that higher levels of transparency information (e.g., uncertainty information) may not be necessary to improve operator hit rates after a certain amount of transparency (i.e., agent's reasoning) has been provided (Mercado et al., 2015/2016; Wright et al., 2016b). In fact, the results from Mercado et al.'s (2015/2016) study indicate that providing agent's reasoning to operators (at medium transparency) was sufficient to increase the operator hit rates for correctly accepting the agent's recommendation. Although it was noted from Stowers et al.'s (2016) study that the highest hit rates were found when uncertainty

information was present, it was not clear from their study whether the presence of additional transparency information (i.e., projection plan success [high transparency], both projection plan success *and* uncertainty [very high transparency]) significantly improved operator hit rates over and above agent's reasoning alone (medium transparency), and if they significantly differed from each other.

3.3.1.1.2 Correct Rejection Rates

With the exception of Wright et al.'s (2016b) study, Mercado et al. (2015/2016) was the only other study that reported correct rejection rates. They found that the addition of agent's reasoning information (medium transparency) increased correct rejection rate significantly by 12% (cf. low transparency), but the addition of both reasoning and uncertainty information (high transparency) improved correct rejection rate even more by an additional 14% (cf. medium transparency). In short, participants were significantly more likely to correctly reject the agent's recommendation when they were presented with high transparency information (correct rejection rate at 81%) compared to when they were presented with medium transparency information (correct rejection rate at 67%). These results indicate that while presenting agent's reasoning to operators is important to improve their correct rejection rates, it was the additional uncertainty information that encouraged the operators to question and eventually reject the agent's recommendation when it was incorrect.

As mentioned in the previous section, correct rejection rates were also obtained in Wright et al.'s (2016b) study; however, in their study, correct rejection rates were analysed together with correct acceptance rates. Although they found that the rates in the medium transparency condition were higher than in the low transparency condition, the difference was marginal and no difference was found when these medium transparency and high transparency conditions were compared.

3.3.1.1.3 False Alarm Rates and Miss Rates

Wright et al. (2016b) was the only study that measured operator false alarm rates (where operators incorrectly accepted agent's recommendation). They found that participants in the medium transparency condition (when agent's reasoning was present) actually made significantly fewer incorrect acceptances than those in either the low transparency or the high transparency condition (when further reasoning was added). Although participants in the high transparency condition made fewer incorrect acceptances than participants in the low transparency condition, the difference was not significant. In other words, while the availability of agent reasoning helped reduce false alarm rates, providing *additional* transparency information in this case negated this effect. These results suggest that while access to agent reasoning in a decision-supporting agent can counter automation bias (i.e., false alarm), too much information could result in an out-of-the-loop situation and could increase complacent behaviour.

Wright et al. (2016b) was, again, the only study that reported operator miss rates where operators incorrectly rejected the correct agent's recommendation. No differences were

found between the three transparency conditions, suggesting that increased transparency information did not make any difference in terms of incorrect rejection rates.

3.3.1.2 General Performance

Wright et al. (2015) investigated the effect of transparency on operator performance (as defined by the total resources used in a route planning task). In Experiments 1 and 2, the route planning task involved participants directing a simulated dismounted soldier team to take an appropriate route; participants had to ensure that the team arrived at the final destination with at least the required amount of resources. Wright et al. found that the resource usage (in both Experiments 1 and 2) did not differ between the three transparency conditions. In Experiment 2, a robotic asset was added to the soldier team, and participants had to ensure that the team arrived at the final destination with sufficient resources of battery and fuel. They found that participants in the low transparency condition used significantly more fuel than those in either the medium transparency or the high transparency condition; however for battery usage, these significant differences were not found. For fuel usage, Wright et al. found that participants actually used significantly more fuel in the high transparency condition than those in the medium transparency condition, suggesting that the addition of predictive information in the high transparency condition may have actually hindered operator performance.

In summary, the pattern of results from the four studies discussed above suggests that although presenting agent's reasoning to participants may improve their performance, adding additional transparency information (in the form of projected outcomes or uncertainty information) did not reliably improve their performance further. The exception to this was the Mercado et al.'s (2015/2016) results, where they found that the additional uncertainty information increased participants' correct rejection rates significantly over the presence of agent's reasoning alone.

3.3.2 Response Time

Three studies have measured the effect of transparency on operator response time (Mercado et al., 2015/2016; Wright et al., 2015; Wright et al., 2016a, 2016b). In Mercado et al.'s (2015/2016) study, response time was defined as the time participants took to make a decision (whether to accept Plan A or reject and go with Plan B). Mercado et al. found that response time did not differ between the three transparency conditions. Similarly, Wright et al. (2016a, 2016b) also found that response time (the time taken between acknowledging the alert and selecting an appropriate route) did not differ between their three transparency conditions. Finally, Wright et al. (2015) measured, in Experiments 1 and 2, the time participants spent at each checkpoint as they guided a soldier team through their chosen route. Only in Experiment 2 did Wright et al. find that, as level-of-information increased, the time spent at each checkpoint also increased. Specifically, the time spent at each checkpoint in either the medium transparency or the high transparency condition was longer compared to the low transparency condition (when only basic information was presented), and no difference was found between the medium transparency condition (when agent's reasoning was added) and the high transparency condition (when predictive information was also present).

In summary, out of these three studies, only one study (Experiment 2 by Wright et al., 2015) indicated that operator response time may increase with additional transparency information). The other two studies (Mercado et al., 2015/2016; Wright et al., 2016a, 2016b) did not find evidence that the additional transparency information had any effect on operator response time.

3.3.3 Situation Awareness

Out of the five studies reviewed here, only one study included a measure of situation awareness (SA). In Selkowitz et al.'s (2016) study, participants were required to use the interface to gain information about the Autonomous Squad Member and simulated squad's status as they completed a route containing obstacles. During an obstacle encounter, the simulation would pause and participants' three levels of SA were assessed using SAGAT-style queries (Jones & Kaber, 2004) relating to the squad member's status (Level 1 SA), reasoning (Level 2 SA), and projected outcomes of its action and reasoning (Level 3 SA). Selkowitz et al. found that SAGAT query responses for Level 1 SA did not differ across the four transparency conditions utilised in their study (i.e., low transparency, medium transparency, high transparency, and very high transparency). However, for Level 2 SA, the next-highest transparency condition (i.e., high transparency) had the highest score and was significantly higher than the medium transparency condition (no other differences between conditions were observed). For Level 3 SA, again the high transparency condition had the highest score, and was significantly higher than either the low transparency condition or the medium transparency condition (no other differences were reported). These results suggest that the predicted outcomes information (high transparency) improved operator Level 2 SA and Level 3 SA more so than when the uncertainty information was added (very high transparency), supporting the evidence that providing too much information to operators is not necessarily a good thing (see Miller, 2014).

3.3.4 Subjective Trust

The effect of transparency on operator subjective trust has been investigated in three studies (i.e., Mercado et al., 2015/2016; Selkowitz et al., 2016; Wright et al., 2016a, 2016b). In Mercado et al.'s (2015/2016) study, subjective trust was measured using a modified Jian Trust Survey (Jian, Bisantz, & Drury, 2000). The modified Jian Trust Survey assessed participants' trust of the system on each of four stages of human information processing: (a) information acquisition, (b) information analysis, (c) decision and action selection, and (d) action implementation (Parasuraman et al., 2000). Specifically, participants answered 16 items and the aforementioned four stages were conceptualised in the scale as the following example: "The system is deceptive when… (a) gathering or filtering information, (b) integrating and displaying analysed information, (c) suggesting or making decisions, and (d) executing actions)." Participants rated each of these four stages on a 7-point Likert scale (1 [not at all], 4 [neutral], 7 [extremely]). In Mercado et al.'s study, only two subscales were analysed as their study only manipulated the display of information (trust during "information analysis") and performed "decision and action selection." The results from the "information analysis" subscale showed no significant differences between the three transparency conditions. Mercado et al. expected these results because although the

reliability of the agent's recommendation was not perfect, the information supporting agent transparency was always 100% accurate. However, the results from the "decision and action selection" subscale showed higher trust when high transparency information was presented compared to low transparency information (medium transparency condition did not differ from the other two conditions). These results suggest that operator trust in the agent's recommendation increased as the system became more transparent (i.e., when agent's reasoning and uncertainty were provided).

Selkowitz et al. (2016) also used the modified Trust in Automated Systems questionnaire (Jian et al., 2000) to measure history-based trust score. Selkowitz et al. defined history-based score as the ongoing, changing relationship of trust that is influenced by the operator's interaction with the agent. They found that the average of history-based trust score was significantly higher when high transparency information was presented (compared to both low transparency and medium transparency conditions); however, the very high transparency condition did not differ from the other conditions. According to Selkowitz et al.'s results, presenting agent's reasoning and the projected state of the system (the high transparency condition) increased trust, but not when uncertainty information was added (the very high transparency condition).

Finally, Wright et al. (2016a, 2016b) investigated the effect of transparency on operator trust and found that trust scores did not differ between the three transparency conditions. However, they found a slight curvilinear trend to the data (p = .046); they found that the medium transparency condition had the lowest trust scores (slightly lower than in the low transparency condition) and was significantly lower than the high transparency condition (which had the highest trust scores).

In summary, the trend from these three studies suggests that—up to a point—operator trust seems to increase as the level of transparency increases (e.g., Wright et al., 2016a, 2016b). However, there have been inconsistencies in the type of information presented at the high transparency condition (e.g., uncertainty vs. predicted outcomes; see Section 3.2.3), hence it is not clear which of the high transparency information was specifically responsible for improving operator trust. In Mercado et al. (2015/2016), providing information about agent's reasoning did not make a difference on operator trust (when compared to when only basic information was provided), however operator trust increased when agent's reasoning was also presented with uncertainty information. In contrast, the availability of predicted outcomes in Selkowitz et al.'s (2016) study significantly improved trust over and above agent's reasoning alone; however, when uncertainty information was added on top of predicted outcomes, participants' trust in the system waned slightly, resulting in no significant differences with the other transparency conditions.

3.3.5 Perceived Usability

Three studies have investigated the effect of transparency on operator perceived usability of the interface (i.e., Mercado et al., 2015/2016; Stowers et al., 2016; Wright et al., 2016b). The System Usability Scale (Brooke, 1996) was used in Mercado et al.'s (2015/2016) and Stowers et al.'s (2016) studies to measure users' overall feelings of usability (efficiency,

efficacy, and satisfaction) with the interface (e.g., "I think that I would like to use this system frequently"). This scale contained 10 items, each to be rated on a scale ranging from 1 (strongly disagree) to 5 (strongly agree). Mercado et al. found no difference in usability scores between the medium condition (when agent's reasoning was present) and the high transparency condition (when agent's uncertainty was added), although scores in these conditions were higher compared to the low transparency condition. However, between the three transparency conditions (i.e., the medium, high, and very high transparency), Stowers et al. found that while the high transparency condition (when projected plan success was added) had the highest perceived usability scores, the very high transparency condition (when uncertainty was further added) actually had the lowest perceived usability scores. Nonetheless, the significance of Stowers et al.'s results was not clear, as full results were not reported in their study. Finally, in Wright et al.'s (2016b) study, when agent's reasoning was present, the usability scores were significantly lower than when either only the basic plan was given or when the level of transparency information was greatest (when basic plan, agent's reasoning, and further reasoning were added).

So far, the results from these past three studies on perceived usability have not been consistent. While Mercado et al. (2015/2016) found that presenting agent's reasoning and agent's uncertainty were perceived to be useful, Stowers et al.'s (2016) study indicated that the additional of uncertainty information seemed to decrease participants' perceived usability. In contrast, Wright et al. (2016b) found that the presence of agent's reasoning actually lowered operator perceived usability compared to when this information was not present. These inconsistent findings might be due to differences in the interfaces used in these studies, and the way the different levels of transparency information were presented.

3.3.6 Subjective Workload

Subjective workload was measured in three studies (i.e., Mercado et al., 2016/2015; Selkowitz et al., 2016; and Wright et al, 2016a) using the National Air and Space Administration Task Load Index (NASA-TLX; Hart & Staveland, 1998). Mercado et al. (2015/2016) found that workload did not differ across the three different levels of transparency (i.e., low, medium, high transparency). Similarly, Selkowitz et al., (2016) did not find any differences on the mean global weighted workload across the four transparency level conditions (i.e., low, medium, high, very high transparency). In line with these two studies, Wright et al. (2016a) found that transparency had no significant effect on Global NASA-TLX scores.

3.3.7 Summary of Results

This section contained a review of five recent studies that have tested the effects of transparency on variables such as operator performance, response time, situation awareness, trust, perceived usability of the interface, and workload. Transparency does not appear to affect either operator response time or subjective workload, and certain forms of transparency information (i.e., agent reasoning) appears to improve operator performance (e.g., Mercado et al., 2015/2016). Further, in some studies, operators actually performed better when agent's reasoning was given without the addition of higher levels of transparency information (Wright et al., 2015, 2016a, 2016b). The exception to this was

Mercado et al.'s (2015/2016) results where they found higher performance (i.e., correct rejection rates) when uncertainty information was also presented in addition to information about agent's reasoning (compared to when only agent's reasoning was presented). For operator situation awareness, the addition of predicted outcomes improved operator Level 2 SA and Level 3 SA more so than when both predicted outcomes and uncertainty information were present.

In terms of subjective trust and perceived usability, the results have been inconsistent. For example, Mercado et al. (2015/2016) found that subjective trust was significantly higher when participants were given information about agent reasoning and uncertainty compared to when they were only given information about agent's reasoning. However, in Selkowitz et al.'s (2016) study, while agent's reasoning and projected outcomes increased trust significantly, trust was lowered when uncertainty information was added. Similarly for perceived usability, while some studies found that higher transparency information produced higher perceived usability (Mercado et al., 2015/2016), others found that the additional uncertainty information produced the lowest perceived usability scores (Stowers et al., 2016).

In summary, the general trends of results suggest that transparency does seem to affect operator performance, situation awareness, perceived usability, and trust. However, it is still unclear how much and what type of information should be given—as some of these studies suggest that while higher levels of transparency may improve some of these measures, the highest transparency level did not always produce the best outcome.

4. Summary and Directions for Future Research

Two models of transparency for human-robot interaction (Lyons, 2013) and human-agent teaming (the SAT model; Chen et al., 2014) have been proposed. Some of the studies reviewed here have used the SAT model to operationalise levels of transparency and investigate their effects on variables associated with human-agent teaming (i.e., Mercado et al. 2015/2016; Selkowitz et al., 2016; Stowers et al., 2016). Although the SAT model provides some guidelines for the types of information that could be included in each level, future research should further investigate exactly which type of information should be conveyed (and how much should be revealed) to the operator in a given situation. The findings of the studies reviewed in Section 3 suggest that appropriate care should be taken when presenting transparency information to avoid overwhelming the operator. Chen et al. (2014) argue that the requirements to achieve system transparency are context dependent, and that these three SAT levels may not be absolutely necessary to achieve transparency. For example, Chen et al. suggested that in a time-sensitive situation, an operator may only need to know the agent's proposed action (Level 1) and the projected outcome (Level 3) to make a decision (Chen et al.). However, this has not been tested; additionally, future research needs to investigate whether time pressure or additional workload would compromise the transparency effect.

The fundamental assumption underlying the research discussed in this review is that providing transparency information to the operator (regarding the agent's intent, performance, future plans, and reasoning process) will allow the operator to develop an accurate mental model of the system and its behaviour, leading to calibrated trust in (and more appropriate reliance on) the system—and ultimately leading to better operator SA and overall performance (Chen et al., 2014; Mercado et al., 2016). However, the studies that have been reviewed here have not directly tested whether agent transparency increased operator SA (except in Selkowitz et al., 2016's study; see Section 3), or whether agent transparency improved the operator's mental model. In order to provide greater insight into the psychological processes underlying operator performance, future research on agent transparency should directly measure operator SA and the accuracy of their mental models.

With the development of more complex systems, researchers have started to investigate factors—such as agent transparency—that could improve human trust and reliance on autonomous systems. An empirically supported model would be of great utility in guiding the design of an interface that supports transparency, and in testing the varying effect of different forms of transparency on human-autonomy team performance. The SAT model proposed by Chen et al. (2014) provides a good starting point; however, future research needs to tease apart exactly what type of transparency information would be beneficial for the operator, and in what contexts. Perhaps more importantly, future research needs to also investigate the underlying processes that explain *why* transparency information facilitated operator performance (e.g., by providing direct evidence that operator mental model and SA improved as a result of agent transparency). Understanding the underlying psychological processes is fundamental to designing an interface that supports transparency in human-autonomy teaming.

5. References

Briggs, G., & Scheutz, M. (2012). Multi-modal belief updates in multi-robot human-robot dialogue interactions. *Proceedings of 2012 symposium on linguistic and cognitive approaches to dialogue agents*, 67-72.

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189-194). London, UK: Taylor & Francis.

Chen, J. Y. C., & Barnes, M. J. (2014). Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems, 44,* 13-29. doi: 10.1109/THMS.2013.2293535

Chen, J. Y. C., Barnes, M. J., & Harper-Sciarini, M. (2011). Supervisory control of multiple robots: Human-performance issues and user-interface design. *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews, 41,* 435-454.

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

Cooke, N. J., Pringle, H. L., Pedersen, H. K., & Connor, O. (Eds.). (2006). *Human factors of remotely operated vehicles: Advances in human performance and cognitive engineering research* (Vol. 7). Amsterdam.

Cummings, M. L., Bruni, S., Mercier, S., & Mitchell, P. J. (2007). Automation architecture for single operator, multiple UAV command and control. *The International Command and Control Journal, 1,* 1-24.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*, 697-718. doi: 10.1016/S1071-5819(03)00038-7

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

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors: The Journal of Human Factors and Ergonomics Society, 37,* 381-394.

Goodrich, M. A., & Cummings, M. L. (2015). Human factors perspective on next generation unmanned aerial systems. In K. P. Valavanis & G. J. Vachtsevanos (Eds.), *Handbook of Unmanned Aerial Vehicles* (pp. 2405-2423). Netherlands: Springer.

Hancock, P. A. (2017). Imposing limits on autonomous systems. *Ergonomics, 60,* 284-291. doi:10.1080/00140139.2016.1190035

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of Human Factors and Ergonomics Society, 53,* 517-527. doi: 10.1177/0018720811417254

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

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

Jones, D. G., & Kaber, D. B. (2004). Situation awareness measurement and the situation awareness global assessment technique. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick (Eds.), *Handbook of human factors and ergonomics methods* (pp. 42.1-42.7). New York: CRC Press.

Kim, T., & Hinds, P. (2006). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN),* 80-85.

Lakhmani, S., Abich, J., Barber, D., & Chen, J. (2016). A proposed approach for determining the influence of multimodal robot-of-human transparency information on human-agent teams. *Proceedings of the 10th International Conference on Augmented Cognition, Canada,* 296-307. doi: 10.1007/978-3-319-39952-2_29

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35, 1243-1270. doi: 10.1080/00140139208967392

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of Human Factors and Ergonomics Society, 46,* 50-80.

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

Lyons, J. B. (2013). Being transparent about transparency: A model for human-robot interaction. *Proceedings of AAAI Spring Symposium on Trust and Autonomous Systems, Palo Alto,* 48-53.

Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barber, D., Procci, K., & Barnes, M. (2015). *Effects of agent transparency on multi-robot management effectiveness.* (ARL-TR-7466). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for multi-UxV management.

Human Factors: The Journal of Human Factors and Ergonomics Society, 58, 401-415. doi: 10.1177/0018720815621206

Miller, C. A. (2014). Delegation and transparency: Coordinating interactions so information exchange is no surprise. In R. Shumaker & S. Lackey (Eds.), *Virtual, augmented and mixed reality: Designing and developing virtual and augmented environments* (pp. 191-202). Berlin, Germany: Springer.

Ososky, S., Sanders, T., Jentsch, F., Hancock, P., & Chen, J. Y. C. (2014). Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. *Proceedings of SPIE Defense and Security on Unmanned Systems Technology XVI, United States, 90840E,* 1-12. doi: 10.1117/12.2050622

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors, 29,* 695-706.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of Human Factors and Ergonomics Society, 39,* 230-253.

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

Rao, A. S., & Georgeff, M. P. (1995). BDI agents: From theory to practice. *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95), San Francisco, USA,* 312-319.

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence, 11,* 335-351. doi: 10.1162/105474602760204264.

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors: The Journal of Human Factors and Ergonomics Society, 58,* 377-400. doi: 10.1177/0018720816634228

Scharre, P. D. (2015). The opportunity and challenge of autonomous systems. In A. P. Williams & P. D. Scharre (Eds.), *Autonomous systems: Issues for defence policymakers* (pp. 3-26). HQ SACT, Norfolk.

Selkowitz, A., Lakhmani, S., Chen, J. Y. C., & Boyce, M. (2015). The effects of agent transparency on human interaction with an autonomous robotic agent. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting, 59,* 806-810. doi: 10.1177/1541931215591246

Selkowitz, A. R., Lakhmani, S. G., Larios, C. N., & Chen, J. Y. C. (2016). Agent transparency and the autonomous squad member. *Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting, 60,* 1319-1323. doi: 10.1177/1541931213601305

Shaw, T. H., Emfield, A., Garcia, A., de Visser, E., Miller, C., Parasuraman, R., & Fern, L. (2010). Evaluating the benefits and potential costs of automation delegation for supervisory control of multiple UAVs. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, 54,* 1498-1502.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Cambridge, MA: Massachusetts Institute of Technology Man-Machine Systems Laboratory.

Stowers, K., Kasdaglis, N., Newton, O., Lakhmani, S., Wohleber, R., & Chen, J. (2016). Intelligent agent transparency: The design and evaluation of an interface to facilitate human and intelligent agent collaboration. *Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting, 60,* 1706-1710. doi: 10.1177/1541931213601392

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human Factors: The Journal of Human Factors and Ergonomics Society, 51,* 281-291. doi: 10.1177/0018720809338842

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (4th ed.). New Jersey: Pearson Education, Inc.

Wright, J. L., Chen, J. Y. C., Barnes, M. J., & Boyce, M. W. (2015). The effects of information level on human-agent interaction for route planning. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting, 59,* 811-815. doi: 10.1177/1541931215591247

Wright, J. L., Chen, J. Y. C., Barnes, M. J., & Hancock, P. A. (2016a). Agent reasoning transparency's effect on operator workload. *Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting, 60*, 249-253. doi: 10.1177/1541931213601057

Wright, J. L., Chen, J. Y. C., Barnes, M. J., & Hancock, P. A. (2016b). The effect of agent reasoning transparency on automation bias: An analysis of response performance. In S. Lackey & R. Shumaker (Eds.), *Virtual, Augmented and Mixed Reality: 8th International Conference, VAMR 2016 Held as Part of HCI International 2016 Toronto, Canada, July 17-22, 2016, Proceedings* (Vol. 9740, pp. 465-477). Switzerland: Springer International Publishing.

This page is intentionally blank.

DEFENCE SCIENCE AND TECHNOLOGY GROUP DOCUMENT CONTROL DATA			1. DLM/CAVEAT (OF DOCUMENT)
2. TITLE The Operationalisation of Agent Transparency and Evidence for Its Impact on Key Human-Autonomy Teaming Variables		3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED LIMITED RELEASE USE (U/L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)	
4. AUTHOR Adella Bhaskara		5. CORPORATE AUTHOR Defence Science and Technology Group 506 Lorimer Street Fishermans Bend VIC 3207	
6a. DST GROUP NUMBER DST-Group-TR-3413	6b. AR NUMBER AR-016-986	6c. TYPE OF REPORT Technical Report	7. DOCUMENT DATE October 2017
8. OBJECTIVE ID	9.TASK NUMBER 17/565	10.TASK SPONSOR SRI	
11. MSTC Aerospace Systems Effectiveness		12. STC Human Factors	
13. DOWNGRADING/DELIMITING INSTRUCTIONS		14. RELEASE AUTHORITY Group Leader, Aerospace Division	
15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT *Approved for public release.* OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111			
16. DELIBERATE ANNOUNCEMENT No Limitations.			
17. CITATION IN OTHER DOCUMENTS Yes			
18. RESEARCH LIBRARY THESAURUS Automation, Autonomous Agents, Human-Machine Interaction			

19. ABSTRACT

The growing interest in the use of autonomous systems for both military and commercial applications has been accompanied by a concomitant increase in research involving human-agent interaction. Transparency has been investigated as one factor that could improve human trust in, and appropriate reliance on, autonomous systems. This report provides a review of studies that have examined how the transparency of an autonomous system affects key variables such as operator performance, response time, situation awareness, perceived usability, and subjective workload. Theoretical frameworks that support transparency in autonomous systems including Lyons' models of transparency (2013) and the Situation Awareness-Based Agent Transparency (SAT) model (Chen et al., 2014) are also presented. Findings from these studies indicate that a certain amount of transparency seemed to improve operator performance, however too much transparency information could also decrease operator performance. Overall, the findings have not been clear-cut in terms of how much and what type of transparency information should be communicated to the operator. Future research should also examine the underlying mechanisms responsible for these transparency effects.