Agent Transparency for Intelligent Target Identification in the Maritime Domain, and its impact on Operator Performance, Workload and Trust

Chan Gloria Wai Yee

This report is submitted in partial fulfilment of Master of Psychology (Organisational and Human Factors)

School of Psychology

University of Adelaide

December 2017

Wound Count: 5178 words (Literature Review); 7869 words (Article)

DECLARATION

This report contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge, this report contains no materials previously published except where due reference is made.

[redacted]

Chan Gloria Wai Yee

December 2017

Table of Contents

Acknowledgments

I would first like to express my sincere gratitude to my supervisor, Dr. Susan Cockshell for the continuous support of the research, for her patience, motivation and immense knowledge. Her door was always open whenever I had a question about my research. Her guidance and encouragement helped me in all the time of research and writing of this thesis.

I would also like to thank Dr. Kingsley Fletcher for his insightful comments and expertise that greatly assisted the research and writing, but also for his help in experiment design and data analysis. I would like to express my appreciation to Dr. Anna Ma-Wyatt for her guidance of research and I gratefully thank for her valuable comments on this thesis.

My sincere thank also goes to Ashley Arnold, who helped me in modifying the software for the purpose of this research. Without his precious support, it would not be possible to conduct this research. A special thanks to my workmates in DST Group Sebastian Tsui and Helen Geng for their support and company throughout the year, which brightened up my research time.

Literature Review on Agent Transparency for the Use of Uninhabited Vehicles

Uninhabited vehicles are considered to be vital assets for military and commercial operations with their reduced radar signatures, increased endurance over piloted airframes and, more importantly the ability to remove humans from immediate threats (Lewis, 2013). Although no human is required to be in the vehicles, a human operator is required to provide some guidance and to interpret and use the information from these vehicles. With a current focus of the military on reducing crewing, research is finding ways in which a single operator can manage multiple uninhabited vehicles as opposed to one uninhabited vehicle managed by multiple human operators (Cummings, Clare, & Hart, 2010). However, to do this effectively, a certain level of automation is required to assist a human operator to manage multiple uninhabited vehicles (Chen & Barnes, 2014).

With technology development over the last few decades, uninhabited vehicles are now highly automated and can perform a range of functions such as flying to a designated location without direct operator control (Arrabito et al., 2010). It is suggested that a single operator is able to manage around four to five uninhabited vehicles at a time with a significant amount of automation such as autopilot (Cummings, 2010). To achieve this, autonomous and sophisticated intelligent agents must be developed in order to support operators in the management of multiple uninhabited vehicles without becoming overloaded.

Intelligent agents potentially offer a huge degree of automation. However, to manage an uninhabited vehicle, the human and intelligent agent need to collaborate effectively in flight control, navigation, and mission and payload management (Cummings, Bruni, Mercier, & Mitchell, 2007). The role of the human operator in managing multiple uninhabited vehicles is more likely to be supervisory, or what has been termed ‘on the loop’, as opposed to ‘in the loop', which would allow the intelligent agent to potentially take on more tasking. How the human and agent will interact needs to be determined.

The role of the human operator while 'on the loop' is to appropriately evaluate decisions recommended by the intelligent agent, diagnose any problems and know when to reject an incorrect decision by the intelligent agent. Consequently, there is an increasing focus on the appropriate usage of intelligent agents so that human operators only rely on the intelligent agent when it is correct. Recent research has focused on supporting the operator to build a proper reliance on the intelligent agent to enable effective human-agent teaming (Chen & Barnes, 2014; Chen et al., 2014; Hoff & Bashir, 2015; Lee & See, 2004). In order to have appropriate usage and build proper reliance on the intelligent agent, it is suggested that human operators need to maintain an appropriate level of situation awareness about the agent's actions and its environment (Drury, Riek, & Rackliffe, 2006). Moreover, it has been suggested that the information specific to the purpose, process and performance of the intelligent agent is needed to allow the operator to have adequate 'human on the loop' performance (Lee & See, 2004).

Intelligent Agent

In artificial intelligence, an agent is defined as 'anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators' (Russell & Norvig, 2009, p. 34). Moreover, an agent that takes the best possible action in a situation is defined as an intelligent agent (Russell & Norvig, 2009). Generally speaking, an agent should also have autonomy over the choice of activity to achieve the goals (Russell & Norvig, 2009). The term 'intelligent agent' has been widely used interchangeably with other terms such as autonomous agent and rational agent; however, the term ‘intelligent agent’ is used for consistency in this paper.

Transparency in Intelligent Agent

Many intelligent agents have been used in different settings and for different purposes in recent years (Russell & Norvig, 2009). As the agents become more independent and sophisticated, it has been suggested that it is increasingly important for human operators to understand the agents’ behaviours, the reasoning process leading to those behaviours, and the predicted outcomes of those behaviours to enable the human operators to calibrate their trust in the agents appropriately and make informed decisions (Lee & See, 2004). Collectively these elements have been described as the transparency of the intelligent agent, and have been shown to have a significant influence on human trust and preference. Transparency is defined as the extent to which human operators’ can understand the intelligent agent’s ability, intent and situational constraints, which facilitates effective interactions between the human operator and the intelligent agent (Lyons, 2013). Transparency can prevent human operators from becoming overly dependent on the intelligent agent, and can assist operators to make informed decisions based on a clear understanding of the working mechanism of the intelligent agent (Fleischmann & Wallace, 2005).

Although agents being more transparent may be beneficial to effective human-agent teaming, there are some arguments against increasing agent transparency (Lyons, 2013). For example, it has been argued that increasing the transparency of an intelligent agent’s actions may overload the operator with too much information (Duggan, Banbury, Howes, Patrick, & Waldron, 2004). Increased agent transparency may encourage human operators to maintain an increased awareness of the observations, decisions and actions that the intelligent agent performs, thereby offsetting the any reduction in time or cognitive effort provided by the intelligent agent (Helldin, Ohlander, Falkman, & Riveiro, 2014). Miller and Parasuraman (2007) proposed a model to explain the relationship and tradeoff between competency, workload and unpredictability in human-agent teaming. Competency refers to the ability of the human-agent teaming to make correct decisions. Workload refers to the mental workload of the human operator when interacting with the agent. Unpredictability means the human operator's inability to understand what the agent will do. Operators can reduce workload by relying on the agent but this would increase unpredictability by reducing their awareness of the situation which relates to the agent's behaviours. Therefore, the role of the intelligent agent in human-agent teaming is not only to support human operators by saving their mental effort in the tasking environment, but also to maintain the right amount of awareness for achieving effective overall performance (Helldin et al., 2014).

Providing explanation of the Intelligent Agent's behaviours

While some argue that increasing agent transparency may overload human operators, others suggest that providing an explanation of how the intelligent agent arrives at a decision can enable operators to develop an appropriate reliance on the intelligent agent (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Paradis, Benaskeur, Oxenham, & Cutler, 2005). One study investigated the effect of different types of agent explanations on operators' understanding of the intelligent agent (Lim, Dey, & Avrahami, 2009). The study provided explanations of why the agent did or did not behave in a certain way, what the agent would do if an event occurs, and how to get the agent to do something in the current situation. Explaining the rationale behind the agent's actions was found to be the most effective way to improve the operators' understanding of the agent, to build a higher level of trust in the intelligent agent, and to increase the acceptance of the intelligent agent's action. Another study found that displaying the contextual information underlying the decision of an intelligent agent in an intelligence, surveillance, and reconnaissance task improved operator performance and enabled human operators to determine how much they should trust the intelligent agent's decision (Rovira, Cross, Leitch, & Bonaceto, 2014). Moreover, presenting the intelligent agent's reasoning process can enhance the involvement of human operators in the reasoning process. An explanation of the reasoning process can enable the human operator to understand why the intelligent agent makes a recommendation and allow the human operator to also apply their own knowledge and inference skills to the reasoning process (Herlocker, Konstan, & Riedl, 2000). Moreover, providing the reasoning behind the agent's actions can assist human operators to understand the strengths and limitations of the intelligent agent, to develop a better understanding of the intelligent agent's behaviours and to adopt a proper reliance on the intelligent agent (Herlocker et al., 2000).

Helldin (2014) conducted a study which examined the effects of increasing the transparency of the intelligent agent's recommendation in an automated target classification task. The conditions examined in the study were (1) without any intelligent agent support; (2) displaying intelligent agent's proposed decision; and (3) displaying intelligent agent's proposed decision and its reasoning. The study found that increasing the transparency improved performance, but also increased decision making time and workload. However, the authors noted that the amount of information presented in condition three was significantly more than in the other two conditions, and participants suggested that more training with the intelligent agent would help them to mak better and faster decisions.

Although incorporating transparency into the intelligent agent can be beneficial in relation to performance, the evidence also suggests the associated increase in information can require more time and effort to comprehend and, therefore, transparency may lead to a reduction in the decision quality for some operators (Ehrlich et al., 2011). For instance, a study showed that additional contextual information of the intelligent agent only improved operator performance in high task demand situations, while there was no performance improvement in the low task demand situations (Rovira et al., 2014). Therefore, more research is required to understand how to effectively present information to explain the intelligent agent's behaviour (Bunt, Lount, & Lauzon, 2012).

Description of Situation Awareness-based Agent Transparency (SAT) Model

Previous research shows that human operators criticise the effectiveness and accuracy of the intelligent agent's behaviours when the human operator has difficulty in understanding the agent's state (Linegang et al., 2006; Seppelt & Lee, 2007). Sarter and Woods (1995) identify the three most common challenges for human-agent teaming: understanding the agent's current state, comprehending the agent's intentions depending on its current behaviours, and projecting the future behaviours. According to prior research, an intelligent agent that provides some information on how it operates can improve human-agent teaming performance and facilitate appropriate trust (Seppelt & Lee, 2007; Wang, Jamieson, & Hollands, 2009). However, only presenting reasoning information on the intelligent agent's actions may not address all of the challenges for human-agent teaming. Chen et al. (2014) suggest the challenges identified in Sarter and Woods (1995) are closely related to Endsley (1995)'s Situation Awareness model: the perception of basic components, comprehension of the components' meaning, and projection of the future status. On this basis, Chen et al. (2014) developed the Situation Awareness-based Agent Transparency (SAT) model, which aims to address Sarter and Woods (1995)'s challenges through displaying transparency information to operators to support them in developing an accurate mental model of the agent. The SAT model is developed based on Endsley (1995)'s situation awareness model, Lee and See (2004)'s 3Ps (Purpose, Process and Performance) model and the Rao and Georgeff (1995)'s Belief-Desire-Intention (BDI) agent framework, which is discussed in detail below.

Building upon these theories, Chen et al. (2014) define agent transparency as 'the descriptive quality of an interface pertaining to its abilities to afford an operator's comprehension about an intelligent agent's intent, performance, future plans, and reasoning process' (p. 2). As shown in Table, Chen et al. (2014) propose that there are three levels of agent transparency in the SAT model that can enhance a human operator's situation awareness. The first level of the SAT model is to provide the human operator with the agent's current state, intentions, and the agent's proposed actions. The second level is to provide the human operator with the information about the rationale behind the agent's proposed actions and the constraints that the agent has considered in planning. The third level includes information about the agent's projected states, predicted consequences or events based on the agent's current status and conditions, and the likelihood of success or failure. The following paragraphs discuss how Endsley (1995)'s situation awareness model, Lee and See (2004)'s 3Ps (Purpose, Process and Performance) model and Rao and Georgeff (1995)'s Belief-Desire-Intention (BDI) agent framework are incorporated into the SAT model.

Table 1.

Situation Awareness-based Agent Transparency (SAT) Model (Chen et al., 2014)

Level 1	**Level 2**	**Level 3**
What's going on and what is the agent trying to achieve?	Why is the agent doing it?	What should the human operator expect to happen?

• Purpose ▪ Desire (Goal selection) • Process ▪ Intentions (Planning/ Execution) ▪ Progress • Performance	• Reasoning process (Belief/ Purpose) • Environmental and other constraints	• Projection to future/ End State • Potential limitations ▪ Likelihood of error ▪ History of performance

Situation awareness is the operator's state of knowledge of the dynamic environment (Endsley, 1995). Endsley (1995) proposes that situation awareness has three levels including the perception of basic components, comprehension of the components' meaning, and projection of the future status. Perception represents the basic level at which the operators perceive the information they need to know to achieve the goal. Comprehension means the human operators could integrate the perceived information with other information and interpret it accurately. Projection is the ability of human operators to make predictions based on the current situation. This model can guide the design of the intelligent agent to facilitate operators' acquisition of awareness about the agent's actions (Scholtz, Antonishek, & Young, 2005). Chen et al. (2014) has incorporated Endsley (1995)'s situation awareness theory into the SAT model, such that each level supports a level of the operator's situation awareness of the intelligent agent's actions (see Table 1). The higher levels of Endsley (1995)'s theory of situation awareness depend on the success of lower levels of situation awareness, such that an operator cannot project without being able to comprehend the meaning of the information. However, Chen et al. (2014) suggested that transparency is not a cumulative result over the levels. For instance, the operator may only be required to know the agent's actions (SAT Level 1) and the projected outcomes (SAT Level 3) to make a sufficiently informed decision in a time-critical situation.

The SAT model also incorporates Lee and See (2004)'s 3Ps model i.e. the system's Purpose, Process and Performance under Level 1 and Level 2 of the SAT model (Table 1). Lee and See (2004) suggest that these three components are critical in developing trust in the interaction between human and agent. When the 3P's information is communicated to the human operator, this increases operator trust in the intelligent agent by making it clearer to the human operator what the intelligent agent is doing (Lee & See, 2004). The information on Performance informs the human operator about what the intelligent agent is doing and its ability to achieve the operator's goals. The Process information informs the operators on how the intelligent agent is operating and the consistency of its actions. Information about Purpose conveys why the intelligent agent is operating in the way it is. Lee and See (2004) suggest Performance, Process and Purpose form the general basis of trust for the human operator.

Rao and Georgeff (1995) propose that the agent's belief, desire and intention (BDI) are the mental attitudes which represent the information, motivational and deliberative states of the intelligent agent. These attitudes form the reasoning process and drive the intelligent agent's behaviours and therefore it is important for the operators to understand the intelligent agent's BDI to achieve effective human-agent teaming. Belief is the information which the agent perceives from the situation. Desire is about what the agent wants to bring about. Intention is the desire that the agent has committed to achieve. The SAT model includes the agent's BDI components to support the operator's situation awareness of the intelligent agent in Level 1 and 2 as per Table 1.

Representing the three levels of SAT model in the design of an intelligent agent enables the human operator to understand the rationale behind the agent's actions and assist the human operator to make informed decisions (Chen et al., 2014). The aim of agent transparency is not to provide all of the system's capabilities, behaviours, and decision making rationale to the human operator, but to communicate the appropriate level of information to allow the operator to maintain adequate situation awareness of the intelligent agent's actions without becoming overloaded (Chen et al., 2014; Lee & See, 2004). Previous studies that have employed the SAT model are discussed in the following section to evaluate this proposed model.

Studies using the SAT model

A number of studies have examined the impact of agent transparency using the SAT model in various tasking environments, such as route planning and navigation of multiple uninhabited vehicles in human-robot teams (Mercado et al., 2016; Selkowitz, Lakhmani, & Chen, 2017; Selkowitz, Lakhmani, Larios, & Chen, 2016; Stowers et al., 2016; Stowers et al., 2017). In particular, Mercado et al. (2016) and Stowers et al. (2017) conducted two studies using different transparency interface designs to examine the utility of the SAT model in a route planning task for uninhibited vehicles management. Both studies aimed to meet the following goals: show information of all three SAT levels, maintain scalability in the display of agent transparency, and maintain ecological validity of the overall design (Stowers et al., 2016).

In both studies, participants were required to complete a number of missions by giving orders to the uninhabited vehicles through the intelligent agent. The intelligent agent provided the human operator with two plans of the actions that could be carried out by the uninhibited vehicles to complete the missions based on the commander's intent, vehicle capability and environmental constraints. Plan A was always the agent's primary recommendation and plan B was the secondary recommendation. Plan B was a better option in approximately one in three times, when the agent was incorrect due to changes in external information such as commander’s intent. Participants were required to select either Plan A or Plan B based on the information presented by the agent and the additional information that was given to them (Mercado et al., 2016; Stowers et al., 2017).

In the first study, Mercado et al. (2016) included Level 1, Level 1+2, and Level 1+2+3 based on the SAT model. The results suggest that human operators improved their performance without the cost of longer response time as agent transparency increase. The study also found no increase in subjective or objective workload when operators were presented with increased agent transparency. Moreover, the operators perceived the decision made by the agent as more trustworthy when presenting information on the intelligent agent’s intent, reasoning and projection (Mercado et al., 2016). However, the participants could only choose between the two recommended plans and could not modify the plans or reject both of them; therefore the study design forced the participants to choose one of the two plans where the participants might disagree with both plans. For instance, participants might choose one over the other as they perceived the chosen plan was relatively better than the other plan.

Building on the study by Mercado et al. (2016), Stowers et al. (2017) assessed the impact of agent transparency by using a different interface design and separating uncertainty from SAT Level 3 to further investigate the role of projection and uncertainty in building operator situation awareness of the intelligent agent. Uncertainty in the SAT model indicates that the agent may not know all of the factors that have an impact on its actions, and hence the future event could not be absolutely known. Therefore, the uncertainty information informs the human operators of the intelligent agent’s uncertainty and assumptions incorporated in its actions. Mercado et al. (2016) demonstrated that there are positive impacts of presenting additional information on intelligent agent's reasoning and projection compared to just presenting information on the intelligent agent's desires and intentions. Therefore, Stowers et al. (2017) investigated the impact of the following conditions: SAT Level 1+2, SAT Level 1+2+3 and SAT Level 1+2+3+ Uncertainty. The results showed that the human operators made more correct decisions without increasing their perceived workload when information on the intelligent agent's intent, reasoning, projection, and uncertainty are presented. However, operators took longer to make the decisions when the additional intelligent agent's uncertainty information is presented to them (i.e. SAT Level 1+2+3+ Uncertainty) compared to displaying the information of the agent's intent and reasoning (i.e. SAT Level 1+2). Moreover, operator perceived trust in both integrating information and decision making increased across transparency levels. In particular, participants perceived the intelligent agent to be most trustworthy in its *information integration* when presenting the information of the intelligent agent's intent, reasoning, projection and uncertainty (i.e. SAT Level 1+2+3+ Uncertainty), while to be most trustworthy in its *decision making* when presenting the information of the intelligent agent's intent, reasoning and projection (SAT Level 1+2+3). The perceived usability of the intelligent agent is consistent with the trust results in that participants perceived the intelligent agent to be most usable when the information on the *intelligent agent's intent, reasoning and projection* was displayed, and less usable when the information on the *intelligent agent's intent, reasoning, projection and uncertainty* was displayed.

Both studies suggest that when the agent is more transparent, operator performance improves without increasing their workload. However, because they also took more time, it is necessary to ensure there is no tradeoff between speed and accuracy. When more information is presented, further analysis in future research may yield information and best practices about the display of information, in particular the presentation of uncertainty information. The two studies presented the uncertainty information through text or graphics; however, a different representation of uncertainty, such as numerically or as a percentage, may impact operator perceived trust and usability. Moreover, Stowers et al. (2017) suggest that the impact of agent transparency depends on the task and environment, and future research is needed to investigate the impact of agent transparency in other contexts.

Trust in Intelligent Agent

An intelligent agent can be used in a number of tasks including acquiring and analysing information, making decisions, taking actions and monitoring other systems (Parasuraman, Sheridan, & Wickens, 2000). However, an intelligent agent may sometimes create errors when introduced into complicated situations, so it may not always be reliable in a military context (Hoff & Bashir, 2015). However, the operator may overly trust the intelligent agent and rely on its incorrect decisions which has the potential for serious consequences (Atoyan & Shahbazian, 2009). The willingness of the operator to rely on the intelligent agent in uncertain situations has been labelled as operator trust (Hoff & Bashir, 2015).

It is critical for the human operators to recognise when they should rely on the intelligent agent and when to override the intelligent agent (Lee & See, 2004; Schaefer, Chen, Szalma, & Hancock, 2016). When the intelligent agent makes mistakes in its decisions and the human operator overly trusts the intelligent agent, the operator may accept the agent’s incorrect decisions, which is a misuse of the intelligent agent. On the other hand, a human operator who has too little trust to the intelligent agent could disuse of the intelligent agent by ignoring the intelligent agent’s decision and forgo the potential benefits of using the intelligent agent such as improved performance and the saving of operator time and effort (Lee & See, 2004;

Parasuraman & Riley, 1997). Consequently, for effective human-agent teaming, the operators need to have an appropriate level of trust in the intelligent agent, also called calibrated trust. Calibrated trust has the potential to lead to better human-agent teaming performance, with lower operator workload and faster response time (Parasuraman & Riley, 1997). Calibrated trust occrus when the operator has an accurate mental model of the intelligent agent and depends on the intelligent agent within the agent's capabilities while being aware of its limitations. In this situation the operator can override the intelligent agent when it is outside of its capabilities (Lee & See, 2004). However, it may be difficult for the operators develop such calibrated trust.

Hoff and Bashir (2015) have systematically reviewed the empirical research on trust between human and automation and have identified that trust is affected by the human operator, the intelligent agent, and environmental factors. Hoff and Bashir (2015) further mapped these three factors to the three different layers of trust suggested by Marsh and Dibben (2003), which are dispositional trust, situational trust and learned trust. Dispositional trust is the individual variability in the tendency to trust the agent which may vary according to culture, age, gender and personality traits. Situational trust represents trust that varies due to the external environment such as the workload, perceived risk and task framing, and the internal context-dependent characteristics of the operator such as self-confidence and attentional capacity in a particular situation. Learned trust arises from the operators' evaluation of the intelligent agent's behaviours from their past experience or current interaction with the agent. Therefore, learned trust is affected by the operators' previous knowledge and the intelligent agent's performance. Transparency of the agent's capability could build the operators' learned trust and reduce the chances of the operators' misuse or disuse the intelligent agent (Hoff & Bashir, 2015).

Prior research has shown that human operators could develop an appropriate level of expectations of the agent's capability in achieving task goals when transparency information is displayed (Chen, Barnes, & Harper-Sciarini, 2011; Lee & See, 2004). Although greater transparency of an agent may facilitate an operator's trust calibration, calibrated trust has also been shown to be affected by the perceived workload and usability of the intelligent agent (Hoff & Bashir, 2015). Operators perceive the intelligent agent to be more usable and trustworthy when showing transparency information is displayed as the operator can easily form an accurate mental model of the intelligent agent. Without displaying transparency information, the operators are likely to perceive the intelligent agent to be less usable and trustworthy (Hoff & Bashir, 2015). Therefore, the information needs to be relevant and efficient to allow the operator to form an accurate mental model of the intelligent agent.

Accurate feedback on an agent's reliability could enable operators to build an appropriate level of trust and improve the human-agent teaming performance (Wang et al., 2009). Moreover, it has been suggested that operators are likely to have proper calibrated trust when the agent is transparent on its analytical, intentional and awareness-based parameters (Lyons, 2013). Given these findings, the SAT model provides a foundation for what information should be displayed to assist the operator in building the mental model of an intelligent agent while using it; specifically the agent's intent, reasoning and projection (Chen et al., 2014). Agents that have been designed based on the SAT model have shown that an operator's trust in intelligent agent increases as the agent transparency increases (Mercado et al., 2016; Selkowitz et al., 2017).

Workload in Transparency

One of the concerns with agent transparency is that the additional information presented in higher transparency levels may increase operator workload (Chen et al., 2014). Workload is described as ‘the cost of accomplishing mission requirements for the human operator’ (Hart, 2006, p. 904). When the operator performs a task with higher workload, it decreases the operator’s capability to do additional tasks (Cain, 2007). High operator workload may affect performance and situation awareness during the task, and lead to incorrect agent usage decisions (Chen & Barnes, 2012; Parasuraman & Riley, 1997). An increase in agent transparency may affect workload as it may require more cognitive effort to process the additional information (Lyons & Havig, 2014). However, Chen et al. (2011) suggest that increased agent transparency may reduce operator workload as the agent’s current state, rationale, and future state projections are directly presented to the operator. Therefore, agent transparency may potentially reduce the effort and time required to process this information. Nonetheless, Duggan et al. (2004) argue that increasing the transparency of an intelligent agent’s behaviours may overload the operators with too much information. Therefore, the challenge for agent transparency design is to implement the agent in a manner that allows the operator to be on the loop while minimizing the additional operator workload.

Research on the impact of agent transparency on workload has not produced consistent results. Helldin (2014) reports that additional transparency information improved operator performance at the cost of increasing workload, while Mercado et al. (2016) found increased transparency information did not increase the operator workload. In contrast, an increase in the transparency of uninhabited vehicle autonomy and functional capability has produced a reduction in workload and performance (Chen, Gonzalez, Campbell, & Coppin, 2014). The inconsistent findings suggest that the additional transparency information has the potential to have a positive or negative influence on operator workload, Therefore the additional information may need to be relevant and designed effectively, to assist in operators' decision making and may not increase the operator workload (Hoff & Bashir, 2015). Similarly, information visualisation techniques may assist the operators to understand the information and enhance their situation awareness (Robertson, Czerwinski, Fisher, & Lee, 2009). Hence, the way the transparency information is displayed is likely to have an impact on workload. The principles of Ecological Interface Design (EID), such as graphical displays and simplified text, offer an approach to display the additional information (Cook & Smallman, 2008; Neyedli, Hollands, & Jamieson, 2011).

Kilgore and Voshell (2014) reviewed the application of EID principles in presenting transparency information on uninhibited vehicles in the maritime domain and presented some design strategies, such as using more salient visual cues, to manage operator attention. An effective interface should assist operators to easily perceive and understand the critical information of the task and enable operators to execute effective strategies to drive the agent's behaviours (Kilgore & Voshell, 2014). The EID techniques use graphics to explicitly show abstract information which can increase the agent's transparency and observability. Representing critical and complex relationships may also improve usability by integrating different information. For instance, a second layer of information could be mapped as a graphical sub-element to direct operator attention, such as using the opacity of an icon to represent uncertainty (Kilgore & Voshell, 2014). The presentation of contextual information may also assist the human operator to overcome the cost of an imperfect intelligent agent without increasing the operator's workload (Rovira et al., 2014).

Future Research

Several studies have examined the effect of agent transparency on performance effectiveness, workload and trust in uninhabited vehicle management tasks (Chen & Barnes, 2015; Chen et al., 2014; Mercado et al., 2016; Stowers et al., 2016). All of these studies have found that performance improves when the agent is more transparent (Helldin, 2014; Mercado et al., 2016). However, greater agent transparency had an inconsistent effect on operator workload and response time in the studies (Chen et al., 2014; Helldin, 2014; Mercado et al., 2016). Studies using the SAT model, that is being transparent about the intelligent agent's intent, reasoning and projections, found improvements in performance without the costs of increasing workload and longer response time (Mercado et al., 2016; Stowers et al., 2017), suggesting this model holds greatest potential benefits in human-agent teaming.

Previous research in human-agent performance has focused on the route planning and navigation aspects of uninhabited vehicle management in a military context (Mercado et al., 2016; Stowers et al., 2016). There is limited research on the impact of agent transparency in other areas of uninhabited vehicle management such as mission and payload management, where incoming information from the sensors is monitored and analysed to meet mission requirements. Therefore, future research may build upon previous research with a new type of agent applying the SAT model (Stowers et al., 2017), such as a target identification agent. This future study could apply the SAT model to a new task, and investigate the impact of agent transparency on trust as well as identifying any possible trade-offs in performance with respect to response time and workload.

References

Arrabito, R., Ho, G., Lambert, A., Rutley, M., Keillor, J., Chiu, A., Au, H., Hou, M. (2010). *Human Factors Issues for Controlling Uninhabited Aerial Vehicles: Preliminary Findings in Support of the Canadian Forces Joint Unmanned Aerial Vehicle Surveillance Target Acquisition System Project*. (DRDC Toronto TR 2009-043). Toronto: Defence Research and Development Canada.

Atoyan, H., & Shahbazian, E. (2009). Analyses of the Concept of Trust in Information Fusion and Situation Assessment. In E. Shahbazian, G. Rogova, & M. J. DeWeert (Eds.), *Harbour Protection Through Data Fusion Technologies* (pp. 161-170). Dordrecht: Springer Netherlands.

Bunt, A., Lount, M., & Lauzon, C. (2012). Are explanations always important?: a study of deployed, low-cost intelligent interactive systems. *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, 169-178. doi:10.1145/2166966.2166996

Cain, B. (2007). *A review of the mental workload literature*. Toronto, Canada: Human System Integration Section, Defence Research And Development.

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

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*(1), 13-29. doi:10.1109/THMS.2013.2293535

Chen, J. Y. C., & Barnes, M. J. (2015). Agent Transparency for Human-Agent Teaming Effectiveness. *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 1381-1385. doi:10.1109/SMC.2015.245

Chen, J. Y. C., Barnes, M. J., & Harper-Sciarini, M. (2011). Supervisory Control of Multiple Robots: Human-Preformance Issues and User-Interface Design. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41*(4), 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). MD: Aberdeen Proving Ground.

Chen, T., Gonzalez, F., Campbell, D., & Coppin, G. (2014). Management of multiple heterogeneous UAVs using capability and autonomy visualisation: Theory, experiment and result. *2014 International Conference on Unmanned Aircraft Systems*, 200-210. doi:10.1109/ICUAS.2014.6842257

Cook, M. B., & Smallman, H. S. (2008). Human Factors of the Confirmation Bias in Intelligence Analysis: Decision Support From Graphical Evidence Landscapes. *Human Factors, 50*(5), 745-754. doi:doi:10.1518/001872008X354183

Cummings, M., Bruni, S., Mercier, S., & Mitchell, P. J. (2007). Automation Architecture for Single Operator Multiple UAV Command and Control. *The international C2 Journal, 1*(2), 1-24. doi:10.1.1.91.4878

Cummings, M., Clare, A., & Hart, C. (2010). The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling. *Human Factors, 52*(1), 17-27. doi:10.1177/0018720810368674

Drury, J. L., Riek, L., & Rackliffe, N. (2006). A decomposition of UAV-related situation awareness. *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 88-94. doi:10.1145/1121241.1121258

Duggan, G. B., Banbury, S., Howes, A., Patrick, J., & Waldron, S. M. (2004). Too Much, Too Little or Just Right: Designing Data Fusion for Situation Awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48*(3), 528-532. doi:10.1177/154193120404800354

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*(6), 697-718. doi:10.1016/S1071-5819(03)00038-7

Ehrlich, K., Kirk, S. E., Patterson, J., Rasmussen, J. C., Ross, S. I., & Gruen, D. M. (2011). Taking advice from intelligent systems: the double-edged sword of explanations. *Proceedings of the 16th international conference on Intelligent user interfaces*, 125-134. doi:10.1145/1943403.1943424

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, *37*(1), 32-64. doi:10.1518/001872095779049543

Fleischmann, K. R., & Wallace, W. A. (2005). A covenant with transparency: opening the black box of models. *Communications of the Adaptive Complex Enterprises*, *48*(5), 93-97. doi:10.1145/1060710.1060715

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904-908. doi:10.1177/154193120605000909

Helldin, T. (2014). Transparency for future semi-automated systems: effects of transparency on operator performance, workload and trust. (PhD dissertation), Orebro University, Orebro.

Helldin, T., Ohlander, U., Falkman, G., & Riveiro, M. (2014). Transparency of Automated Combat Classification. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 22-33). Cham: Springer International Publishing.

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). Explaining collaborative filtering recommendations. *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, 241-250. doi:10.1145/358916.358995

Hoff, K. A., & Bashir, M. (2015). Trust in Automation. *Human Factors, 57*(3), 407-434. doi:10.1177/0018720814547570

Kilgore, R., & Voshell, M. (2014). Increasing the Transparency of Unmanned Systems: Applications of Ecological Interface Design. In R. Shumaker & S. Lackey (Eds.), *Virtual, Augmented and Mixed Reality* (pp. 378-389). Cham: Springer International Publishing.

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

Lewis, M. (2013). Human Interaction With Multiple Remote Robots. *Reviews of Human Factors and Ergonomics*, *9*(1), 131-174. doi:10.1177/1557234x13506688

Lim, B. Y., Dey, A. K., & Avrahami, D. (2009). Why and why not explanations improve the intelligibility of context-aware intelligent systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2119-2128. doi:10.1145/1518701.1519023

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 Annual Meeting, 50*(23), 2482-2486. doi:10.1177/154193120605002304

Lyons, J. B. (2013). *Being transparent about transparency: A model for human-robot interaction*. Paper presented at the 2013 AAAI Spring Symposium Series, Stanford University, Stanford, CA.

Lyons, J. B., & Havig, P. R. (2014). Transparency in a Human-Machine Context: Approaches for Fostering Shared Awareness/Intent. In R. Shumaker & S. Lackey (Eds.), *Virtual, Augmented and Mixed Reality* (pp. 181-190). Cham: Springer International Publishing.

Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology, 37*(1), 465-498. doi:10.1002/aris.1440370111

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, 58*(3), 401-415. doi:10.1177/0018720815621206

Miller, C. A., & Parasuraman, A. (2007). Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors, 49*(1), 57-75. doi:10.1518/001872007779598037

Neyedli, H. F., Hollands, J. G., & Jamieson, G. A. (2011). Beyond Identity: Incorporating System Reliability Information Into an Automated Combat Identification System. *Human Factors, 53*(4), 338-355. doi:10.1177/0018720811413767

Paradis, S., Benaskeur, A., Oxenham, M., & Cutler, P. (2005). Threat evaluation and weapons allocation in network-centric warfare. *2005 7th International Conference on Information Fusion, 2,* 1078-1085. doi:10.1109/ICIF.2005.1591977

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, 39*(2), 230-253. doi:10.1518/001872097778543886

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*(3), 286-297. doi:10.1109/3468.844354

Rao, A. S., & Georgeff, M. P. (1995). BDI Agents: From Theory to Practice *Proceedings of the First International Conference on Mulitagent Systems* (pp. 312-319). Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

Robertson, G., Czerwinski, M., Fisher, D., & Lee, B. (2009). Selected Human Factors Issues in Information Visualization. *Reviews of Human Factors and Ergonomics, 5*(1), 41-81. doi:10.1518/155723409X448017

Rovira, E., Cross, A., Leitch, E., & Bonaceto, C. (2014). Displaying Contextual Information Reduces the Costs of Imperfect Decision Automation in Rapid Retasking of ISR Assets. *Human Factors, 56*(6), 1036-1049. doi:10.1177/0018720813519675

Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach, Global Edition* (3rd ed.). Upper Saddle River, NJ: Prentice Hall

Sarter, N. B., & Woods, D. D. (1995). How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control. *Human Factors, 37*(1), 5-19. doi:10.1518/001872095779049516

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. *Human Factors, 58*(3), 377-400. doi:10.1177/0018720816634228

Scholtz, J. C., Antonishek, B., & Young, J. D. (2005). Implementation of a situation awareness assessment tool for evaluation of human-robot interfaces. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 35*(4), 450-459. doi:10.1109/tsmca.2005.850589

Selkowitz, A. R., Lakhmani, S. G., & Chen, J. Y. C. (2017). Using Agent Transparency to Support Situation Awareness of the Autonomous Squad Member. *Cognitive Systems Research, 46,* 13-25. doi:10.1016/j.cogsys.2017.02.003

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 Annual Meeting, 60*(1), 1319-1323. doi:10.1177/1541931213601305

Seppelt, B. D., & Lee, J. D. (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

Stowers, K., Kasdaglis, N., Newton, O., Lakhmani, S., Wohleber, R., & Chen, J. Y. C. (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 Annual Meeting, 60*(1), 1706-1710. doi:10.1177/1541931213601392

Stowers, K., Kasdaglis, N., Rupp, M., Chen, J. Y. C., Barber, D., & Barnes, M. (2017). Insights into Human-Agent Teaming: Intelligent Agent Transparency and Uncertainty. In P. Savage-Knepshield & J. Chen (Eds.), *Advances in Human Factors in Robots and Unmanned Systems* (Vol. 499, pp. 149-160). Cham: Springer.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and Reliance on an Automated Combat Identification System. *Human Factors, 51*(3), 281-291. doi:10.1177/0018720809338842

Agent Transparency for Intelligent Target Identification in the Maritime Domain, and its impact on Operator Performance, Workload and Trust

Chan Gloria Wai Yee

University of Adelaide

ABSTRACT

Objective: To examine how increasing the transparency of an intelligent maritime target identification system impacts on operator performance, workload and trust in the intelligent agent.

Background: Previous research has shown that operator accuracy improves with increased transparency of an intelligent agent's decisions and recommendations. This can be at the cost of increased workload and response time, although this has not been found by all studies. Prior studies have predominately focussed on route planning and navigation, and it is unclear if the benefits of agent transparency would apply to other tasks such as target identification.

Method: Twenty seven participants were required to identify a number of tracks based on a set of identification criteria and the recommendation of an intelligent agent at three transparency levels in a repeated-measures design. The intelligent agent generated an identification recommendation for each track with different levels of transparency information displayed and participants were required to determine the identity of the track. For each transparency level, 70% of the recommendations made by the intelligent agent were correct, with incorrect recommendation due to additional information that the agent was not aware of, such as information from the ship's radar. Participants' identification accuracy and identification time were measured, and surveys on operator subjective workload and subjective trust in the intelligent agent were collected for each transparency level.

Results: The results indicated that increased transparency information improved the operators' sensitivity to the accuracy of the agent's decisions and produced a greater tendency to accept the agent’s decision. Increased agent transparency facilitated human-agent teaming without increasing workload or response time when correctly accepting the intelligent agent’s decision, but increased the response time when rejecting incorrect intelligent agent’s decisions. Participants also reported a higher level of trust when the intelligent agent was more transparent.

Conclusion: This study shows the ability of agent transparency to improve performance without increasing workload. Greater agent transparency is also beneficial in building operator trust in the agent.

Application: The current study can inform the design and use of uninhabited vehicles and intelligent agents in the maritime context for target identification. It also demonstrates that providing greater transparency of intelligent agents can improve human-agent teaming performance for a previously unstudied task and domain, and hence suggests broader applicability for the design of intelligent agents.

Agent Transparency for Intelligent Maritime Target Identification

Uninhabited vehicles are considered to be increasingly important for military and commercial operations (Lewis, 2013). Recent research suggests a single operator can manage multiple uninhabited vehicles (Cummings, Clare, & Hart, 2010). However, to do this effectively, an intelligent agent is required to assist a human operator (Chen & Barnes, 2014). Cummings, Bruni, Mercier, and Mitchell (2007) found a single operator is able to simultaneously manage around four to five uninhabited vehicles with a significant amount of automation. Much early research focused on the impact of different levels of automation in managing multiple uninhabited vehicles (Cummings et al., 2007; Cummings et al., 2010). However, more recent research has focused on supporting the operator to build a proper reliance of the intelligent agent in order to have effective human-agent teaming (Chen & Barnes, 2014; Chen et al., 2014; Hoff & Bashir, 2015; Lee & See, 2004). The aims of increasing automation in uninhabited vehicle management are to enhance human-agent performance through simplified operations, reduced operation costs, and lower human operator workload. Nonetheless, there are increasing concerns about the potential misuse and disuse of automation as the level of autonomy increases (Parasuraman, 1997). That is, can the operator appropriately evaluate the situation and reject advice when the intelligent agent's decision is incorrect? Research has suggested that influencing the transparency of the intelligent agent may improve operator trust and performance; and thereby create a proper reliance on the intelligent agent (Chen et al., 2014; Lee & See, 2004).

However, there are still several challenges that need to be addressed to achieve a proper reliance of the human operator on an intelligent agent (Chen & Barnes, 2014). Human operators may not understand the rationale made by the intelligent agent, and question the accuracy of an intelligent agent's decision (Linegang et al., 2006). Thus, the intelligent agent has the potential to increase operator workload if the operator needs to determine the rationale behind the agent's decisions (Chen et al., 2014). Lee and See (2004) suggest that presenting a human operator with information specific to the purpose, process and performance of the intelligent agent could enable the operator to have adequate 'human on the loop' performance. Therefore, Chen et al. (2014) propose the Situation Awareness-based Agent Transparency (SAT) model to support the operators' situation awareness (Endsley, 1995) of the intelligent agent regarding the agent's current understanding of the world, reasoning process, and projected outcomes.

Agent Transparency

In artificial intelligence, an agent is defined as 'anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators' (Russell & Norvig, 2009, p. 34). Moreover, an agent that takes the best possible action in a situation is defined as an intelligent agent (Russell & Norvig, 2009). The term has been widely used interchangeably with other terms such as autonomous agent and rational agent; however, the term 'intelligent agent' is used for consistency in this paper.

Previous research shows that human operators criticise the effectiveness and accuracy of an intelligent agent's behaviours when the human operator has difficulty in understanding the agent's state (Linegang et al., 2006; Seppelt & Lee, 2007). Sarter and Woods (1995) identify the three most common challenges for human-agent teaming: understanding the current agent's state, comprehending the agent's intentions, and projecting the future behaviours. According to prior research, an intelligent agent that provides some information on how it operates can improve human-agent task performance and facilitate appropriate trust (Seppelt

& Lee, 2007; Wang, Jamieson, & Hollands, 2009). Chen et al. (2014) suggest that the Sarter and Woods (1995)'s challenges are closely related to the Endsley (1995)'s Situation Awareness model: the perception of basic components, comprehension of the components' meaning, and projection of the future status. Chen et al. (2014) then developed the Situation Awareness-based Agent Transparency (SAT) model, which aims to address the Sarter and Woods (1995)'s challenges through displaying transparency information to operators to support them in developing an accurate mental model of the agent. The SAT model incorporates Endsley (1995)'s situation awareness model on how each SAT level could support each level of the operator's situation awareness on the agent, which are the perception of what the intelligent agent is doing, the reasoning of the intelligent agent's action and the projection of the intelligent agent's behaviours.

The SAT model has also incorporated Lee and See (2004)'s 3Ps (Purpose, Process and Performance) model. When the information about the intelligent agent's purpose, process and performance is communicated to the operator, greater trust is developed by clearly showing to the operator what the intelligent agent is doing (Lee & See, 2004). The Purpose information is about what the intelligent agent is trying to achieve. The Process information informs the operators about how the intelligent agent operates and the consistency of its actions. The information on Performance informs the human operator about what the intelligent agent is doing and its ability to achieve the operator's goals.

Moreover, Rao and Georgeff (1995) propose that an agent's beliefs, desires and intentions (BDI) are the mental attitudes which represent the information, motivational and deliberative states of the intelligent agent. These attitudes drive the intelligent agent's behaviours and therefore it is important for the operators to understand the intelligent agent's BDI to achieve effective human-agent teaming. As a result, the SAT model includes the agent's BDI components to support operator situation awareness of the intelligent agent.

As shown in Table 1, Chen et al. (2014) propose that there are three levels of agent transparency in the SAT model to support human operator situation awareness. The first level of the SAT model is to provide the human operator with the agent's current state, intentions, and the agent's proposed actions. The second level is to provide the human operator with information about the rationale behind the agent's proposed actions and the constraints that the agent has considered in planning. The third level includes information about the agent's projected states the predicted consequences of events based on the agent's current status and conditions, and the likelihood of success or failure. Incorporating the three levels of the SAT model into the displays of an intelligent agent interface enables the human operator to understand the rationale behind the agent's actions and assist the human operator to make informed decisions (Chen et al., 2014).

Table 1

Situation Awareness-based Agent Transparency (SAT) Model (Chen et al., 2014)

Level 1	**Level 2**	**Level 3**
What's going on and what is the agent trying to achieve? • Purpose ▪ Desire (Goal selection) • Process ▪ Intentions (Planning/ Execution) ▪ Progress • Performance	Why is the agent doing it? • Reasoning process (Belief/ Purpose) • Environmental and other constraints	What should the human operator expect to happen? • Projection to future/ End State • Potential limitations ▪ Likelihood of error ▪ History of performance

The aim of agent transparency is not to display all of the system's capabilities, behaviours, and decision making rationale to the human operator, but to communicate the appropriate level of information to allow the operator to maintain adequate situation awareness of the intelligent agent's actions without becoming overloaded (Chen et al., 2014; Lee & See, 2004). Some studies have examined the information that the intelligent agent should present to the human operators and suggested elements that may improve overall performance such as providing accurate feedback about the intelligent agent's reliability and providing an explanation of why the agent behaved in a certain way (Helldin, Ohlander, Falkman, & Riveiro, 2014; Lim, Dey, & Avrahami, 2009; Lyons, 2013; Wang et al., 2009). However, Chen et al. (2014) propose a three-level agent transparency model which identifies the essential information that should be shown to the operators to enable them to maintain proper situation awareness of the agent's action in the tasking environment without being overloaded.

Trust in Intelligent Agent

An intelligent agent has the potential to assist human operators to achieve better performance with lower workload (Parasuraman & Riley, 1997). However, these benefits may not be achieved without an appropriate level of trust (Lee & See, 2004). Lee and See (2004) define trust in automation as, 'the attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability' (p. 54). Trust is also described as an attitude towards automation which could affect reliance (Lee & See, 2004). People tend to rely on an intelligent agent that they trust and tend to reject an intelligent agent that they do not trust. Thus, trust guides the operators' reliance on an intelligent agent to overcome the cognitive complexity of managing an intelligent agent (Lee & See, 2004). If the operator over-trusts the intelligent agent, the operator becomes complacent and over-relies on the agent, which may result in a misuse of the intelligent agent. Overreliance on the agent may reduce the frequency with which operators monitor the agent, and therefore may reduce the operator's situation awareness and may result in detrimental consequences (Lee & See, 2004). On the contrary, if the operator under-trusts the agent, the operator may disuse the intelligent agent and undermine the potential benefits if offers. The misuse and disuse of the intelligent agent are influenced by how well the human operator matches the true capabilities of the intelligent agent to their trust in the intelligent agent, which determines whether the operators have a proper trust calibration (Lee & See, 2004). Calibrated trust occurs when the operator has an accurate mental model of the intelligent agent. In this situation the operator will depend on the intelligent agent when it is operating within the agent's capabilities, and also be aware of the agent's limitations and override the agent when it is outside of the agent's capabilities (Lee & See, 2004). Therefore, it is critical that operators have proper calibrated trust to avoid misuse and disuse of the intelligent agent and facilitate human-agent teaming (Lee & See, 2004).

Trust in an intelligent agent is a complex and multidimensional concept that is grounded on at least one of the intelligent agent's characteristics such as motives, intentions and actions (Hoff & Bashir, 2015). In order for operators to have a proper trust calibration, operators need to understand the intelligent agent's ability to achieve the intended goals so they can form an appropriate level of expectation of the agent's capability to achieve the goals (Lee & See, 2004). The information presented needs to be relevant and efficient to enable the operator to form an accurate mental model, which the operators can understand, to explain and predict the intelligent agent's actions and act accordingly. Otherwise, operators may perceive the intelligent agent to be less trustworthy and usable when the additional transparency information increases their workload (Hoff & Bashir, 2015). To reduce the misuse and disuse of the intelligent agent, a transparent agent should provide accurate and useful feedback to the operator (Hoff & Bashir, 2015). Wang et al. (2009) found that when an agent was transparent about its level of reliability this facilitated appropriate trust and improved operator performance. More specifically, an intelligent agent should provide its analytical, intentional and awareness-based parameters to the operator to build proper calibrated trust (Lyons, 2013). Chen et al. (2014) have further investigated what information should be included to build a proper calibrated trust in an intelligent agent and proposed the SAT model. Previous research that applied the SAT model to the design of displays for multiple uninhabited vehicles management has found that increasing the transparency levels improved both subjective and objective trust in the intelligent agent (Mercado et al., 2016; Selkowitz, Lakhmani, Larios, & Chen, 2016).

Workload

One concern with agent transparency is that the additional amount of information that needs to be presented as transparency levels increase may increase operator workload (Chen et al., 2014). Workload is described as 'the cost of accomplishing mission requirements for the human operator' (Hart, 2006, p. 904). When the operator performs a task with higher workload, it decreases the operator's capability to do additional tasks (Cain, 2007). It is suggested that operators are more likely to rely on the intelligent agent's decisions when they experience higher mental workload, which may lead to more incorrect agent usage decisions. One study showed that high operator workload reduced operator performance and situation awareness in managing multiple robots (Chen & Barnes, 2012). Therefore, an increase in operator workload may lead to incorrect agent usage decisions (Parasuraman & Riley, 1997). An increase in agent transparency may affect workload as it may require more cognitive efforts to process the additional information (Lyons & Havig, 2014). However, Chen, Barnes, and Harper-Sciarini (2011) suggested that increased agent transparency may reduce operator workload as the agent's current state, rationale, and future state projections are presented directly to the operator. Therefore, agent transparency may potentially reduce the time and processing effort required to acquire this information. Nonetheless, Duggan, Banbury, Howes, Patrick, and Waldron (2004) argue that operators found it difficult to process all the information displayed to them in a time-sensitive military situation, in which case increasing the transparency of an intelligent agent's behaviours may overload operators with too much information.

Research on the impact of agent transparency on workload has not produced consistent results. Mercado et al. (2016) found increasing transparency information did not increase operator workload, while Helldin (2014) reports that additional transparency information improved operator performance at the cost of increasing workload. In addition, an attempt to increase transparency by providing more direct and specific information about subsystem autonomy in group of heterogeneous uninhabited vehicles produced a reduction in workload and performance (Chen, Gonzalez, Campbell, & Coppin, 2014). However, additional information that is relevant and effectively designed does not increase the operator workload (Hoff & Bashir, 2015). To effectively display information, the display of the additional information needs to be in simplified form and meet ecological interface design principles such as using graphical displays and having simplified text (Cook & Smallman, 2008; Neyedli, Hollands, & Jamieson, 2011).

Individual differences

Prior studies have shown that video gamers perform better than non-gamers on different aspects of visual attention (Green & Bavelier, 2003; Green & Bavelier, 2006; Hubert-

Wallander, Green, & Bavelier, 2011). Gamers were found to be more flexible and efficient in distributing attention over space and time (Hubert-Wallander et al., 2011). Research has also found that frequent gamers had better performance and situation awareness, and had faster response time when managing an intelligent agent in a military context than the infrequent gamers (Chen & Barnes, 2012; Chen & Barnes, 2015). Thus, people with gaming experience may perform better on tasks that require rapid processing of visual information, multiple object tracking, and flexibility in attention allocation (Green & Bavelier, 2003; Green & Bavelier, 2006; Hubert-Wallander et al., 2011). Therefore, the current study was interested in the impact of gaming experience on operator performance, trust and workload when the agent is more transparent in a maritime target identification task.

Increasing the transparency of an intelligent agent may enhance the operators' ability to build a new and sound mental model. However, it is also suggested that the operators' previous mental models may also have an impact on human-agent teaming (Johnson-Laird, 1983). Thus, the current study was also interested in the operators' previous experience with programming and its impact on a maritime target identification task and examined the impact of programming and gaming experience across transparency levels.

Current Study

While increasing agent transparency has been shown to improve operator performance in uninhabited vehicle management, there are no consistent findings about the impact of transparency on operator workload and response time. Mercado et al. (2016) reported performance improved with no increase in workload or response time. However, Helldin (2014) found that greater transparency increased both operator workload and response time. Chen et al. (2014) found that a more transparent agent reduced operator workload and improved reaction time. Furthermore, prior research in human-agent performance has been focussed on route planning and navigation in a military context (Mercado et al., 2016; Stowers et al., 2016). The aim of this study was to build upon previous research through applying the SAT model to an intelligent agent that makes recommendations on the classification of contacts in a maritime surveillance task. The intelligent agent generated an identification recommendation based on the information from an uninhabited aerial vehicle (UAV), and the participants made the identification decision for each contact based on a set of identification criteria. The current study examined the impact of agent transparency on trust and investigated any possible trade-offs in performance with respect to response time and workload. Four hypotheses were generated:

1. Operator performance will improve as the level of agent transparency increased.
2. There will be no difference in response time across transparency level.
3. Operator trust in the agent will increase as the level of agent transparency increases.
4. There will be no difference in perceived workload across the transparency levels.

The current study also explored the impact of gaming and programming experience on operator performance, response time, trust and workload across transparency level.

Method

Participants

Participants were recruited from the Defence Science & Technology (DST) Group Edinburgh through an intranet daily news advertisement. A power analysis using GPower (Erdfelder, Faul, & Buchner, 1996) indicated that a sample of 27 participants would be able to detect expected effect sizes based on the previous research conducted by Mercado et al.

(2016) with 80% power using an ANOVA with a significance level of $p = .05$. Consequently, twenty seven staff (21 men, 6 women) aged between 21 and 56 ($M = 37$, $SD = 9.9$) participated in this study. Participation was voluntary and no incentive was given for participation.

Apparatus and Stimuli

A customized simulator was created to support the current study. The simulation software was run on an Intel I7 Workstation and the simulator interface was displayed on a 30inch Dell monitor with a resolution of 2560 x 1600 pixels. The simulator interface is shown in Figure 1 and consists of three main sections showing information received by different sources. The left-hand half of the interface was a simulated ship radar display on which ‘ownship’ was represented as a blue circle at the centre of the display and the positions of other tracks were represented by symbols at various ranges and bearing from ownship. The operator was required to select a track by clicking on a symbol using the mouse cursor. The radar display also contained a green circle which represented a range of 30 nautical miles from ownship, a horizontal blue band to the south of ownship which represented a shipping lane and a diagonal red line to the northwest of ownship which indicated the Australian exclusive economic zone. The bottom-right section of the interface displayed Automatic Identification System information about the selected track which comprised of vessel type, vessel name, port of registry, and the maximum and current speed. The top-right section of the interface showed the identification recommendation made by the intelligent agent for the selected track and allowed participants to select their assessment of the identity of the selected track.

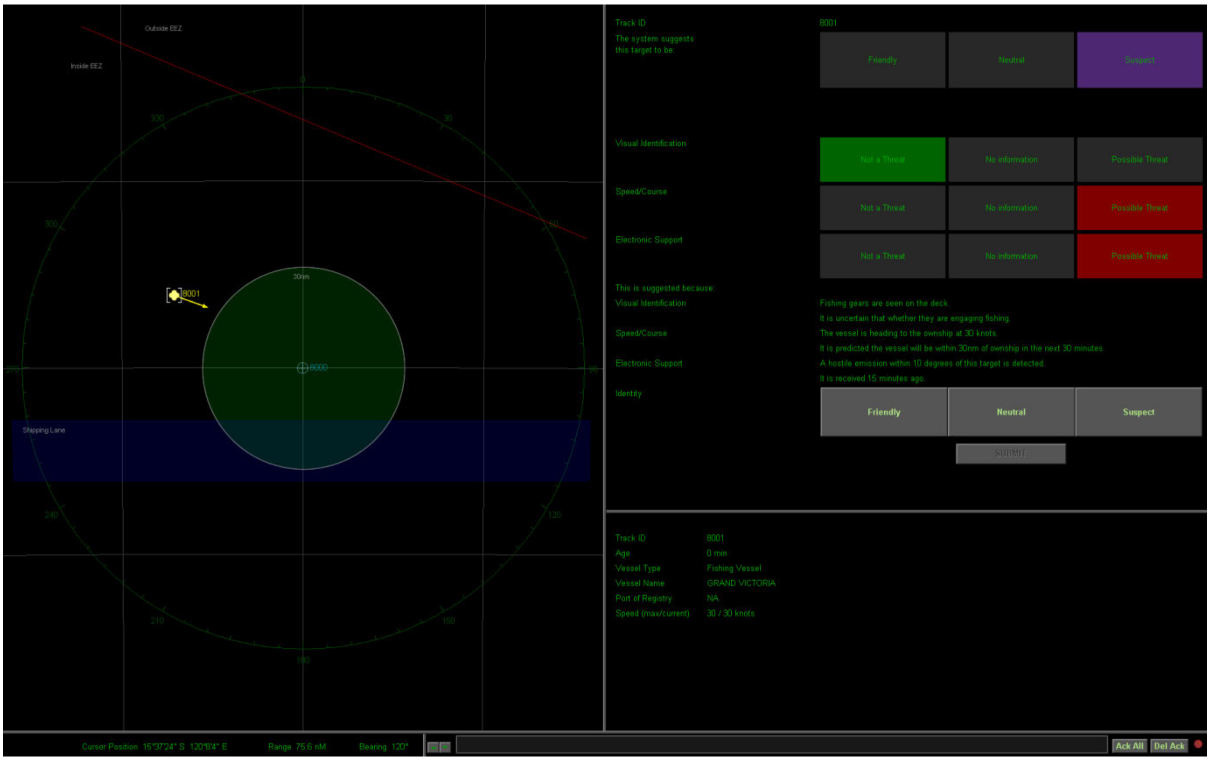


Figure 1. Example of the simulator screen with SAT Level 1+2+3 (See Appendix A for a full page simulator screen)

The intelligent agent section of the screen displayed different information depending on the level of agent transparency. The intelligent agent generated an identification recommendation based on the information received from the sensors on the UAV including the camera, radar and electronic support sensor. For all transparency levels, the intelligent agent's identification recommendation for the track was shown at the top of the section, and at the bottom of the section there were three grey identity boxes where the participants made the identification decision and a 'Submit' button for them to submit the identification decision. The participants made the identification decision of each track based on the three main identification attributes of visual identification, speed / course and electronic support. For the SAT Level 1 (basic information only) interface, there was a graphic representation showing whether the track was assessed to be a threat for each identification attribute (Figure 2). Green indicated ‘not a threat’, yellow meant ‘no information is available’, and red indicated ‘possible threat’. The SAT Level 1+2 (basic information and reasoning) interface, displayed text describing the reasoning behind the intelligent agent’s recommendation for each identification attributes in addition to the SAT Level 1 graphical display (Figure 3). For SAT Level 1+2+3 (basic information, reasoning and projection) interface, additional text was presented that provided projection information for each identification attribute, together with the graphical display and text on the reasoning information (Figure 4).

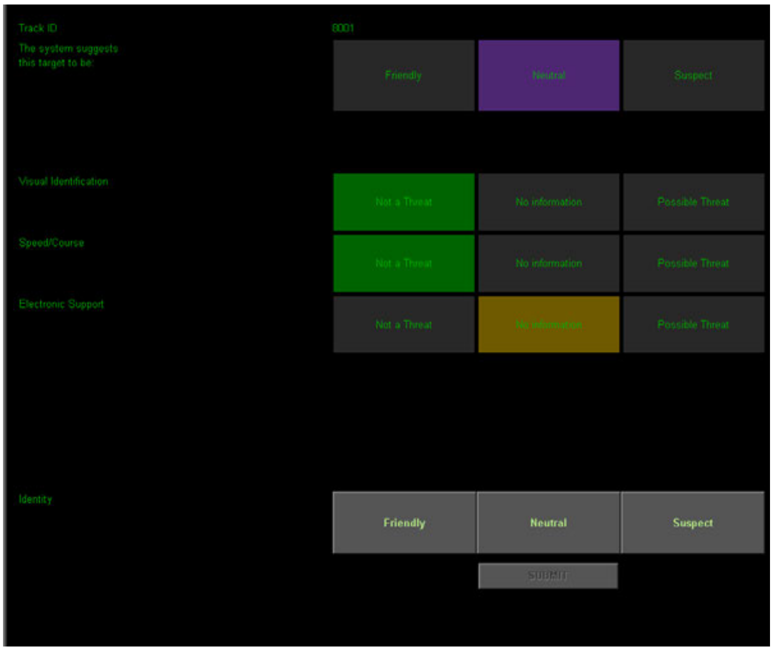


Figure 2. Example of the Intelligent Agent section with SAT Level 1 (basic information only).

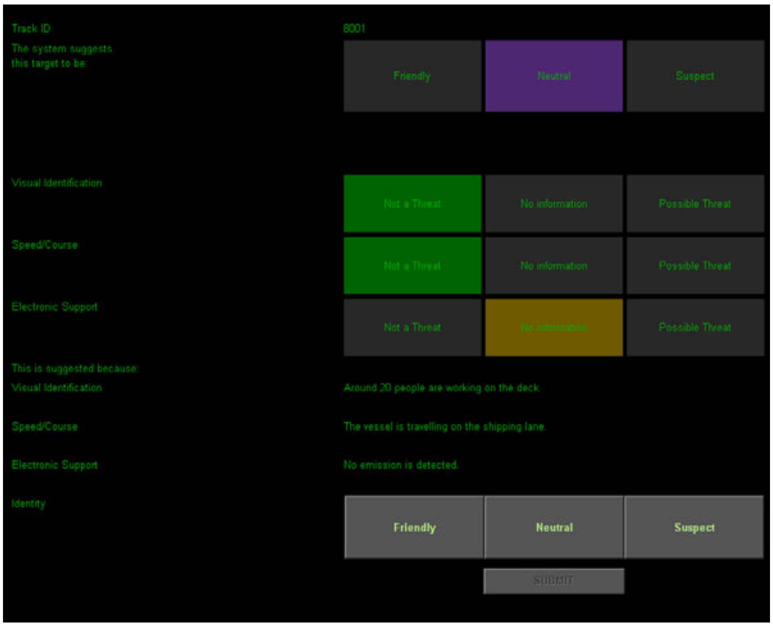


Figure 3. Example of the Intelligent Agent section with SAT Level 1+2 (basic information and reasoning).

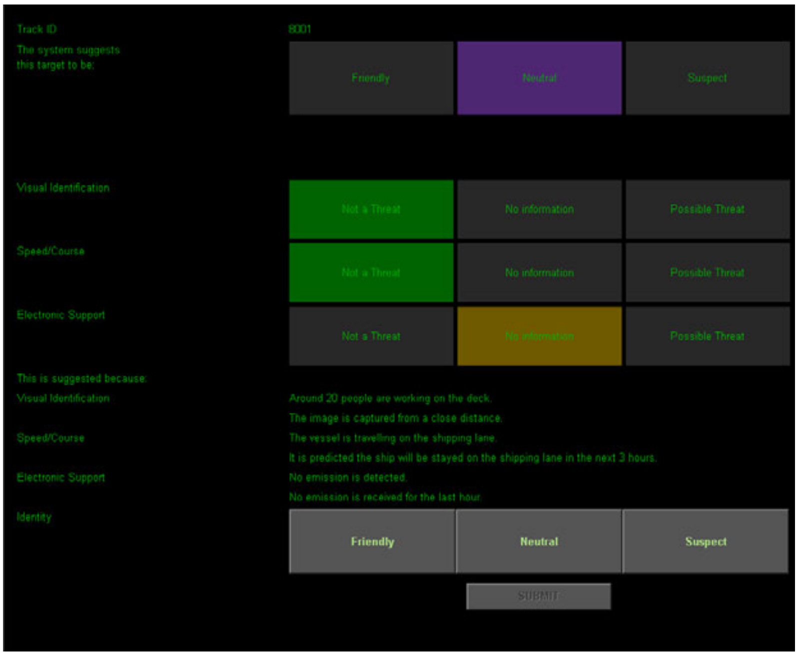


Figure 4. Example of the Intelligent Agent section with SAT Level 1+2+3 (basic information, reasoning and projection).

Design

A repeated-measures design was used in which participants were required to identify a number of tracks at each transparency level based on a set of identification criteria. The condition order (SAT Level 1, SAT Level 1+2, SAT Level 1+2+3) was counterbalanced using a Williams' randomization procedure (Williams, 1949). The participants completed three experimental sessions, and the conditions were in the order according to their participation sequence.

Each experimental session consisted of 22 tracks that needed to be identified, and the first two tracks in each session were considered as trials. For six out of the remaining twenty tracks in a session, the intelligent agent's recommendations were inaccurate due to additional information that was not available to the intelligent agent, that is, the information from the ship's radar and Automatic Identification System. The six tracks were pseudo randomised in each session. The choice of a reliability rate of 70% was based on Wickens and Dixon (2007)'s research which found that operator performance with automation reliability lower than 70% was worse than that with a lack of automation. Moreover, reliability that is too high may lead to over-reliance, while very low reliability may lead to under-reliance.

Measures

Participants' identification accuracy, identification time, workload, and trust in the intelligent agent were measured at each transparency level.

Operator performance. Identification accuracy was divided into measures based on Signal Detection Theory as shown in Table 2 (Chen et al., 2014). The correct acceptance rate was calculated as the proportion of trials where participants accepted a correct recommendation by the intelligent agent, while the correct rejection rate was calculated as the proportion of trials where participants rejected an incorrect recommendation by the intelligent agent. These measures of accuracy were translated into the Signal Detection Theory metrics of sensitivity and bias (Stanislaw & Todorov, 1999).

Table 2

Decision matrix by applying signal detection theory adapted from Chen et al. (2014)

Intelligent agent's recommendation	Participant's identification submission	Signal detection theory
Correct	Accept and submit correct	Hit (Correct acceptance)
Correct	Reject and submit incorrect	Miss
Incorrect	Reject and submit correct	Correct Rejection
Incorrect	Accept and submit incorrect	False Alarm
Incorrect	Reject and submit incorrect	Error

Response time. Response time for each track was defined as the time from when the participant clicked on the track to the time the participant submitted the identification of that track. Response times were then categorised based on the Signal Detection Theory and average response times were calculated for correct acceptance, correct rejection, miss, false alarm and error responses.

Workload. Workload was measured using a self-report questionnaire, the National Air and Space Administration Task Load Index (NASA-TLX) (Hart & Staveland, 1988) (Appendix B). NASA-TLX measures workload using six subscales of mental, physical, temporal demands, effort exerted, self-performance evaluation, and frustration felt during the task. Participants rated each subscale on a continuous scale ranged from 0 to 100, with lower scores representing lower workload and higher scores indicating higher workload in that subscale. The ratings of the subscales were equally weighted and averaged to create an estimate of overall workload (Hart, 2006). The Cronbach's alpha of the overall workload scale in Braarud (2001)'s study was 0.82.

Trust. Trust was measured using a modified version of an automation trust scale developed by Jian, Bisantz, and Drury (2000) (Appendix C). The questionnaire was modified by Mercado et al. (2015) to combine the scale with the four types of automation introduced by Parasuraman, Sheridan, and Wickens (2000). Parasuraman et al. (2000) identified four stages of information processing and suggested that each stage can be automated. The four stages are information acquisition (sensory processing), information analysis (perception), decision and action selection, and action implementation (response selection). Only trust of information analysis and decision and action selection were assessed in the current study as these were the two stages manipulated in this experiment. Participants were asked the trust questions for each stage of the information processing. Each question was scored on a 7-item Likert scale (1= not at all and 7= extremely). The Cronbach's alpha in Safar and Turner (2005)'s sample of the original version of the automation trust scale was strong ($\alpha = 0.93$).

Demographics. A demographic questionnaire collected information on the participant's age, gender, level of education, computer usage and computer / video gaming experience (Appendix D). For computer usage experience, participants were asked to choose from a list of the software programs they were capable of using without any help, and the number of languages they were capable of programming in. Participants in this study reported being capable of using at least five or more software programs. For computer programming experience, participants who were capable of programming at least in one language were categorised as 'programmer' and those who were incapable of programming in any languages were categorised as 'non-programmer'. In this sample, 77.8% of the participants were programmers, and 22.2% were non-programmers. For computer / video gaming experience, participants were asked to rate how often they played computer / video games. Participants who chose 'Daily' or 'Weekly' were classified as 'Gamer' and participants who chose the other options including 'Monthly', 'Less than once a month', 'I have played computer/video games in the past but not for many years' and 'Never' were classified as 'Non-gamer' based on Mercado et al. (2015)'s categorisation of gaming experience. In this sample, 55.6% of the participants were identified as gamers, and 44.4% were non-gamers.

Procedure

After the participants gave informed consent, they completed the demographic questionnaire. Participants then received training on the task, which took approximately 45 minutes. The training session consisted of a PowerPoint presentation that provided detailed instruction for performing the task followed by a simulation session to familiarise them with the target identification task and the user interface. Accuracy feedback was provided after each track was identified and participants could ask questions at any time during the training session. A booklet with all the materials they needed was given to them to refer to during the experiment. Participants were provided with a list of suspicious behaviours performed by vessels on the sea that identified them as 'Suspect' and a decision making tree to assist in making the identification decision. The training session was immediately followed by the experimental session. Participants were told that they were participating in a mission to protect Australian waters and their role was to identify whether vessels were friendly, neutral or suspect. Participants were instructed to be as quick and accurate as possible. No feedback was provided during the experimental sessions. Participants completed an experimental session for SAT Level 1, SAT Level 1+2 and SAT Level 1+2+3 in the counterbalanced order. After each experimental session, participants completed the workload and trust questionnaires. Each experimental session took approximately 20 minutes to complete.

Results

Analysis Approach

The effect of transparency level on performance, response time, workload and trust were examined using a series of univariate ANOVAS with planned comparisons (Tabachnick & Fidell, 2013). Performance was analysed as rates of correct acceptance (hit), correct rejection, false alarm and error (Table 2) based on Signal Detection Theory (Green & Swets, 1966). Miss rate was not analysed as it is the inverse of the correct acceptance (hit) rate according to the signal detection theory (Green & Swets, 1966). Sensitivity and response bias were also used to analyse the performance data based on the Signal Detection Theory (Stanislaw & Todorov, 1999). Sensitivity (*d'*) examined whether participants were sensitive to the accuracy of intelligent agent. Response bias (*c*) measured the tendency of participants to accept or reject the intelligent agent's recommendation. Response time was separately analysed by the Signal Detection Theory categories of correct acceptance, correct rejection, miss, false alarm and error. The first two tracks in each condition were treated as familiarisation trials and excluded from the performance and response time analysis. The effect of the programming and gaming experiences were examined using a series of mixed ANOVAs measures.

The significance level used in the analysis was $p < .05$. No correction was applied to examine differences between transparency levels in this study due to the small number of planned comparisons.

Operator Performance

Operator performance was analysed by the rates of correct acceptance, correct rejection, false alarm and error (Table 3). In particular, this study was interested in the correct acceptance and correct rejection rates across transparency levels, which indicated the correct usage of the intelligent agent (Figure 5). Note that correct acceptance reponses were from 14 tracks per session, with the other measures from the remaining 6 tracks per sessions, results in fewer responses and hence more variability for these measures.

Table 3

Means of Operator Performance by Transparency Level

Performance	M (*SD*)		
	Level 1	Level 1+2	Level 1+2+3
Correct Acceptance	83.86% (11.98%)	93.65% (7.76%)	97.62% (4.43%)
Correct Rejection	82.72% (18.19%)	75.93% (27.86%)	81.48% (27.48%)
False Alarm	14.20% (18.32%)	21.60% (25.66%)	15.43% (26.92%)
Error	3.09% (6.60%)	2.47% (6.03%)	3.09% (10.37%)

Analysis of correct acceptance rates revealed a significant main effect across transparency levels, F (2, 78) = 17.48, p <.001, partial η^2 = .40. There was a significant increase in correct acceptance rate from SAT Level 1 to SAT Level 1+2 with a mean difference of 9.79%, p = .009. The correct acceptance rate also significantly increased from SAT Level 1+2 to SAT Level 1+2+3 with a mean difference of 3.97%, p = .04. Correct acceptance rates in SAT Level 1+2+3 were significantly increased by 13.76% compared with SAT Level 1 (p < .001). The results revealed correct acceptance rates significantly increased with the increase of transparency levels.

Results for correct rejection rates revealed no significant difference in correct rejection between all three levels of agent transparency, F (2, 78) = .79, p = .46, partial η^2 = .029. Moreover, there was no significant difference found in the rate of false alarms, F (2, 78) =

1.02, p =.37, partial η^2 = .038, and in the rates of error, F (2, 78) = .097, p = .91, partial η^2 = .004, across transparency levels.

From Figure 5, the correct acceptance rates were greater than correct rejection rates in SAT Level 1+2 and SAT Level 1+2+3, while the correct acceptance rate and correct rejection rate in SAT Level 1 were not noticeably different. Therefore, post hoc t-tests were run to examine the differences of ratings between these two variables in each condition. The correct acceptance rates were significantly higher than correct rejection rates in SAT Level 1+2 and SAT Level 1+2+3, which were p =.005 and p =.004 respectively. However, no significant difference between the correct acceptance rate and correct rejection rate was found in SAT Level 1, p =.79.

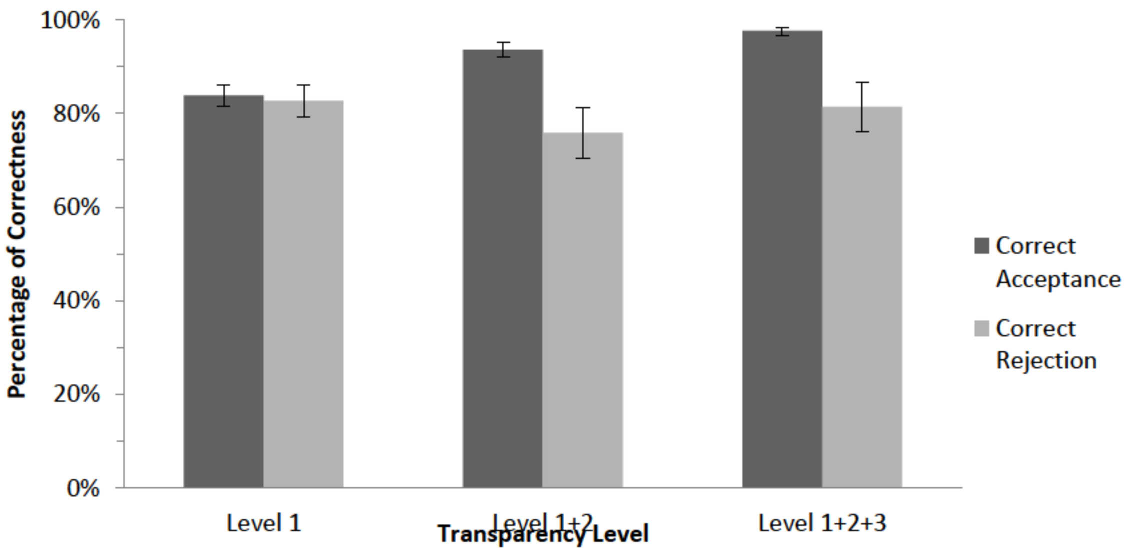


Figure 5. Correct Acceptance and Correct Rejection rates across transparency levels. Error bars indicate standard error of mean (SEM).

Signal Detection Analysis

Signal detection theory (SDT) was used to analyse sensitivity of the participants to the accuracy of the intelligent agent's accuracy and the tendency of the participants to accept or reject the agent's recommendation. Sensitivity (*d*') and response bias (*c*) were calculated from the difference between the hit (correct acceptance) and false alarm (incorrect acceptance) rates. When the hit or false alarm rates were zero or one, the data was corrected, adding 0.5 to both the number of hits and the number of false alarms and adding 1 to both the number of signal trials and the number of noise trials, before calculating the hit and false alarm rates (Stanislaw & Todorov, 1999).

Sensitivity (d'). Sensitivity increased with the transparency levels as can be seen from Figure 6. The results of a repeated-measures ANOVA on *d'* showed a significant effect across agent transparency levels, F (2, 78) = 5.97, $p = .005$, partial $\eta^2 = .19$. There was no significant difference in *d'* between SAT Level 1 (M = 1.99, SD = .14) and SAT Level 1+2 (M = 2.23, SD = .15), $p = .26$. However, *d'* in SAT Level 1+2+3 (M = 2.66, SD = .17) was significantly greater than that in SAT Level 1+2, p = .008, and in SAT Level 1 ($p = .006$).

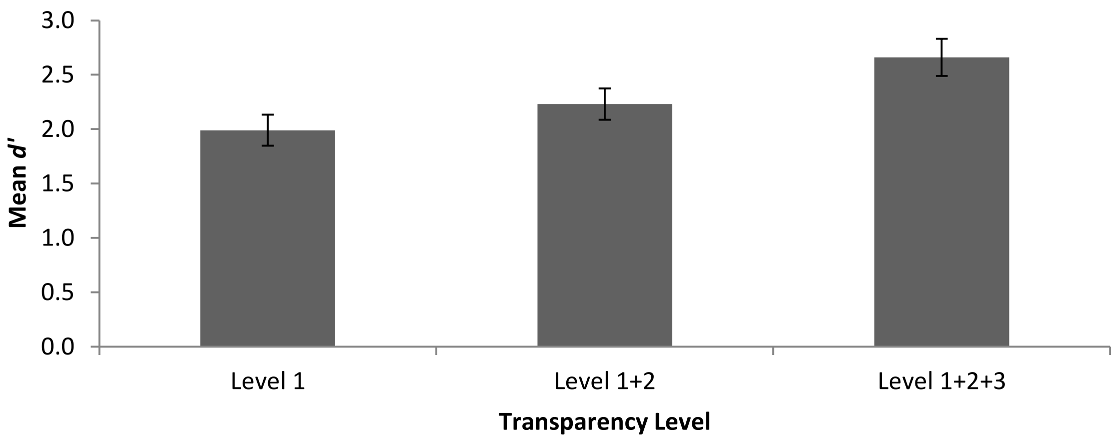


Figure 6. Sensitivity (*d'*) across transparency levels. Error bars indicate SEM.

Response bias (c). The mean response bias (*c*) in each transparency level is shown in Figure 7. The results of a repeated measures ANOVA on *c* revealed a significant difference across transparency levels, F (2, 78) = 7.26, $p = .002$, partial $\eta^2 = .22$. A significant decrease in *c* was found between SAT Level 1 (M = -.012, SD = .07) and both SAT Level 1+2 (M = -.33, SD = .09), $p = .009$, and SAT Level 1+2+3, $p = .002$. However, there was no significant change in *c* between SAT Level 1+2 and SAT Level 1+2+3 (M = -.34, SD = .07), $p = .90$. The results of *c* scores showed participants had an increased tendency to accept the agent's recommendation from SAT Level 1 to SAT Level 1+2, yet no further increase in the tendency of acceptance at SAT Level 1+2+3.

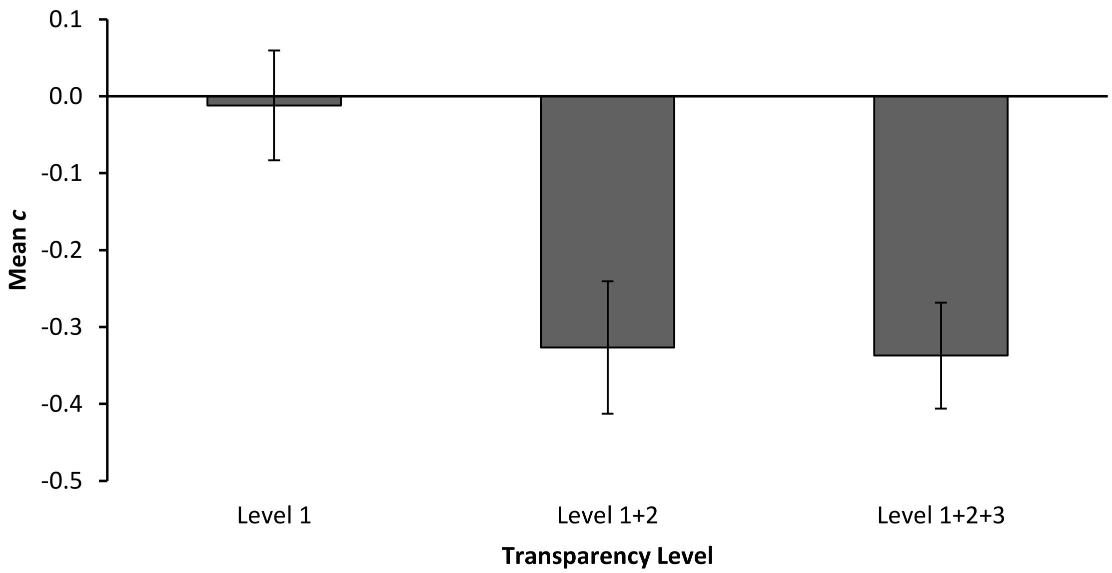


Figure 7. Response bias (*c*) across transparency levels. Error bars indicate SEM.

Response Time

Response time was analysed using the average response times for correct acceptance, correct rejection, miss, false alarm and error (Table 4). In this study, the average response times for correct acceptance and correct rejection were analysed as they indicated the time required for making a correct decision (Figure 8). The miss, false alarm and error had fewer responses and more variability in the data.

Table 4

Response Time Measures by Transparency Level

Response Time	M (*SD*)		
	Level 1	Level 1+2	Level 1+2+3
Correct Acceptance	24.83s (15.22s)	23.98s (12.06s)	27.41s (10.59s)
Correct Rejection	25.64s (11.76s)	33.07s (12.62s)	32.99s (13.01s)
Miss	53.32s (49.82s)	58.10s (34.29s)	39.14s (20.65s)
False Alarm	50.74s (48.84s)	46.93s (60.92s)	30.03s (13.41s)
Error	52.84s (45.28s)	46.86s (58.70s)	37.57s (17.48s)

No significant main effect in average response time for correct acceptance across all three levels of agent transparency was found, F (2, 78) = .90, p = .41, partial η^2 = .033. The results for response time for correct rejection revealed a significant main effect of agent transparency levels, F (1.5, 78) = 4.40, p = .017, partial η^2 = .145. Compared to the average response time for correct rejection rate in SAT Level 1, that in SAT Level 1+2 was 7.43s significantly longer (p = .024) and that in SAT Level 1+2+3 was 7.35s significantly longer (p = .001). No significant difference was found between SAT Level 1+2 and SAT Level 1+2+3 with a mean difference of 0.08s, p = .98. Moreover, no significant main effects were found in the average response times for miss, false alarm and error across transparency level (p >.05). Note that the miss, false alarm and error had fewer responses and more variability in the data.

From Figure 8, the average response times for correct rejection were greater than those for correct acceptance in SAT Level 1+2 and SAT Level 1+2+3. Therefore, post hoc t-tests were run to examine the differences of ratings between these two variables in each condition. The average response times for correct rejection were significantly higher than those for correct acceptance in SAT Level 1+2 and SAT Level 1+2+3, which were p <.001 and p =.004 respectively. However, no significant difference between the average response time for correct rejection and correct acceptance was found in SAT Level 1, p =.71.

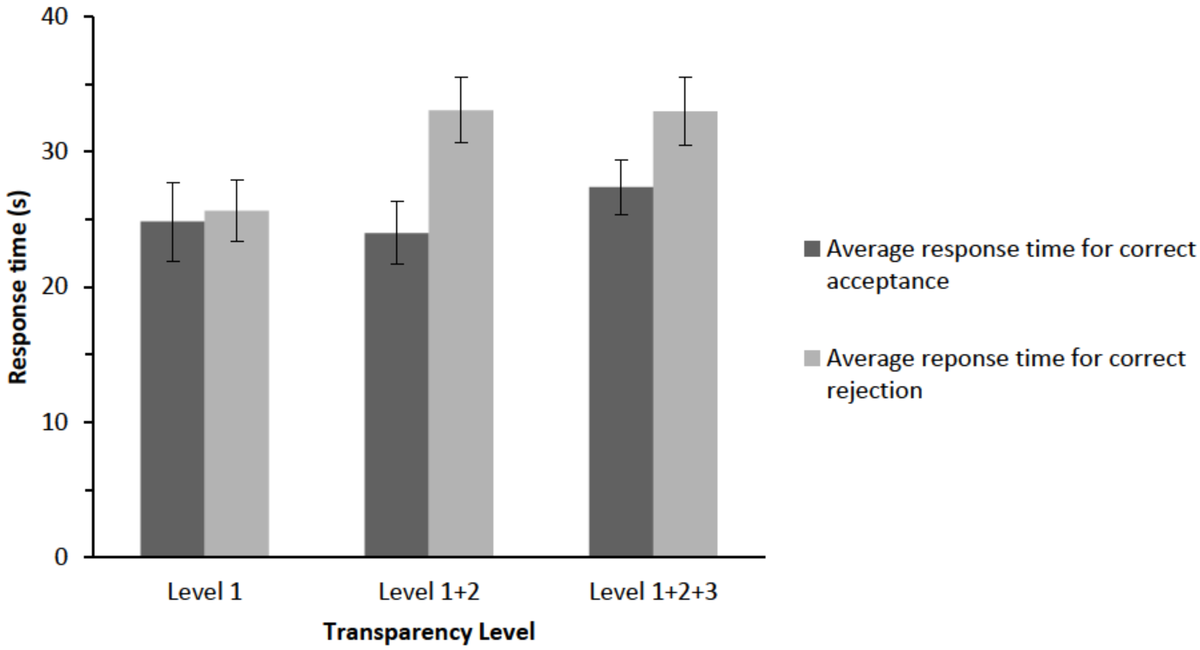


Figure 8. Average response times for correct acceptance and correct rejection across transparency levels. Error bars indicate SEM.

Workload

The scores of the six TLX subscales were combined and averaged to generate a global workload score. The global workload score was the highest in SAT Level 1 (M = 38.33, SD =14.55), subsequently reduced in SAT Level 1+2 (M = 35.93, SD =13.23) and SAT Level 1+2+3 (M = 34.88, SD =14.78). However, analysis revealed that there was no significant difference in global workload across transparency levels, F (2, 78) = .80, p =.46, partial η^2 = .030. ANOVA measures were also used to examine any differences among the six TLX subscales. The results revealed a significant decrease in the Performance workload subscale across agent transparency levels, F (2, 78) = 3.24, p = .047, partial η^2 = .11, indicating that participants perceived their performance improved with increased transparency level. No significant main effects across transparency levels were found in other workload subscales (p > .05).

Trust

Perceived trust was separated into trust of the agent's information analysis and trust of the agent's decision making. The mean scores of each scale across SAT Level are shown in Figure 9. Two separate within-subjects ANOVAs on the Information Analysis and Decision and Action Selection Trust subscales were conducted.

The results for the Information Analysis subscale showed a significant agent transparency level effect, F (2, 78) = 5.93, p = .005, partial η^2 = .19. No significant difference was revealed between SAT Level 1 (M = 4.76, SD = .96) and SAT Level 1+2 (M = 4.98, SD = .93), p = .17. However, trust for the Information Analysis subscale in SAT Level 1+2+3 (M = 5.22, SD = .99) was significantly greater than in both SAT Level 1(p = .002) and SAT Level 1+2 (p = .043). The results showed that the trust in the agent's ability to integrate and display information increased as transparency level increased.

No significant transparency level effect in trust for the Decision and Action Selection subscale was found, F (2, 78) = 2.00, p =.15, partial η^2 = .07. However, the results demonstrated a trend that trust in the agent's ability to suggest or make decisions increased as transparency level increased (Figure 9). Trust for the Decision and Action Selection subscale was the lowest in SAT Level 1 (M = 4.22, SD = .93), subsequently increased in SAT Level 1+2 (M = 4.41, SD = .85) and SAT Level 1+2+3 (M = 4.53, SD = 1.02).

From Figure 9, the trust for the Information Analysis subscale was greater than the trust for Decision and Action Selection subscale within each condition. Therefore, post hoc t-tests were run to examine the differences of ratings between these two subscales in each condition. The trust in Information Analysis subscale were significantly higher than the trust for Decision and Action Selection subscale in Level 1, Level 1+2 and Level 1+2+3, which were p =.001, p <.001 and p <.001 respectively.

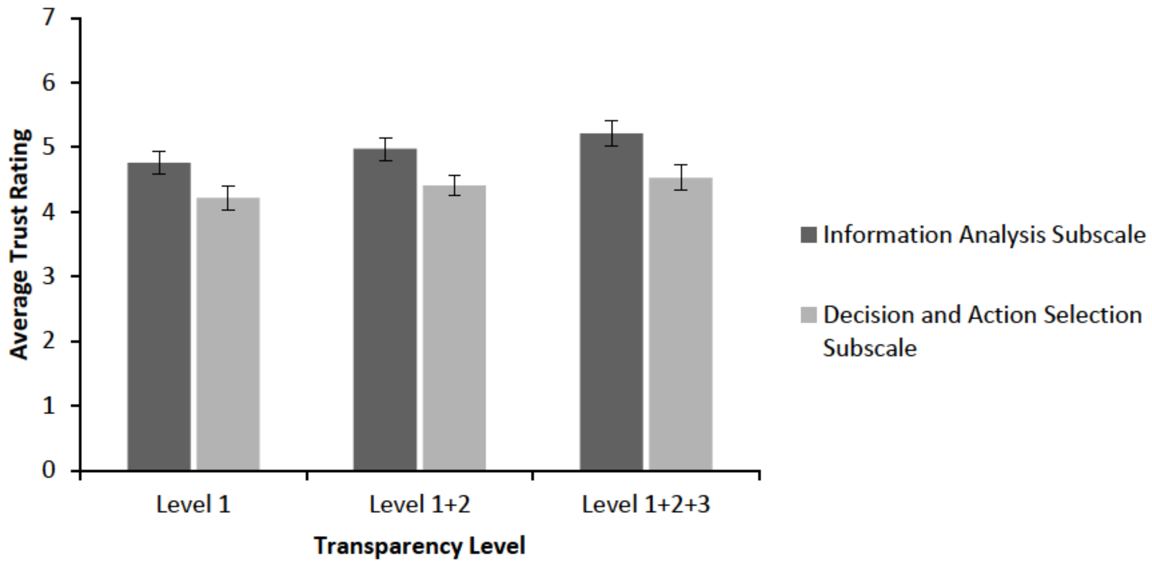


Figure 9. Trust subscales by transparency levels. Error bars indicate SEM.

Individual Differences

Results on individual differences in gaming experience and computer programming were evaluated across transparency levels. Participants were categorised as being gamers or non-gamers and as being programmers or non-programmers. No significant differences were found on operator performance, response times, trust and workload across transparency levels between gamers and non-gamers (p > .05). There were also no significant differences in any independent variable across transparency levels between programmers and non-programmers

(p > .05). Note that over 70% of the participants were categorised as programmers, which results in different sample sizes between the two groups.

Discussion

This study investigated the impact of introducing agent transparency to a maritime target identification task. The Situation Awareness-Based Agent Transparency (SAT) model was used to provide a foundation for what information was to be displayed to support human-agent teaming. The current study examined whether a more transparent agent could support an operator in identifying a contact in the maritime context and how the transparency information affected operator performance, trust and workload.

Operator Performance

Greater transparency enabled participants to accept more correct decisions made by the intelligent agent; however transparency was not shown to be beneficial when the intelligent agent's decisions were incorrect. Signal Detection Theory (Green & Swets, 1966; Stanislaw & Todorov, 1999) measures of sensitivity and response bias showed that participants were more sensitive to the accuracy of the intelligent agent's decisions, but also more likely to accept the intelligent agent's recommendations when the agent was more transparent. This suggests that presenting the agent's intents, reasoning and projection can improve operator performance by improving the human operators' ability to distinguish the correct and incorrect intelligent agent's recommendation, but also increase the tendency for operators to rely on the intelligent agent.

One explanation for the increase in the tendency to accept the intelligent agent's decision is that displaying transparency information enabled the operators to understand the agent's decision making rationale; therefore, operators are more likely to rely on and accept the agent's recommendations. Alternatively, Mercado et al. (2015) suggests that operators may more rely on the intelligent agent when the mental demand increases due to the increased amount of information. However, in the present study increasing transparency information did not increase operator subjective workload. This suggests that the increased tendency to accept the intelligent agent's decisions across transparency level was not caused by changes in workload.

When the agent was more transparent, participants were better able to discriminate the accuracy of the intelligent agent's decisions yet were more likely to rely on the intelligent agent's decisions. This finding partially supports Hypothesis 1 that operator performance would improve with increased transparency. It is inconsistent with previous studies (Helldin, 2014; Mercado et al., 2016; Selkowitz et al., 2016; Stowers et al., 2016) which found that a more transparent agent improved operator performance both when the intelligent agent's decisions were correct and incorrect, and Finger and Bisantz (2000)'s study that presenting information of uncertainty supported operator decision making. The current finding suggests that the transparency information may help the operators' decisions in accepting the correct intelligent agent's decisions; however, the information about the agent's intent, reasoning and projection may not help the operators in making decisions when the intelligent agent's decisions are incorrect. It may be because operators might perceive they need to have a stronger argument before rejecting the intelligent agent's incorrect decisions when the intelligent agent's intent, reasoning and projection are presented to them. Decision making research suggests that new evidence that supports the present belief has a greater impact on beliefs than that which does not support it (Ross & Lepper, 1980). Future research could examine how the operators arrive at their decisions when the intelligent agent's decision is correct in comparison with that when the intelligent agent's decision is incorrect.

In the current study, participants did not take longer to accept the correct intelligent agent's decisions when the intelligent agent was more transparent yet more time was required to reject the incorrect intelligent agent's decisions as transparency increased. It partially supports Hypothesis 2 that there would be no significant difference in response time across transparency levels, and is partially consistent with the prior research (Helldin, 2014; Mercado et al., 2016). Mercado et al. (2016)'s study showed that greater agent transparency improved performance without increasing response time, while Helldin (2014)'s research reported that the improved performance also increased the time required to make a decision. One explanation of the current study's finding is that operators might spend more time and be more careful when evaluating additional transparency information when they need to override the intelligent agent's incorrect decisions. The current study showed the participants were quicker in accepting the correct intelligent agent's decisions. This may be because the additional transparency information enabled them to better understand the intelligent agent's correct decision, which compensated for the time required to read the additional information.

The decision to use an automation reliability level of 70% (Wickens & Dixon, 2007) meant that only 30% of the intelligent agent's recommendation were incorrect in this study. Thus, only six out of twenty trials could be a correct rejection, false alarm, or error in each session. This resulted in a small number of data points and large variability in the error, false alarm and correct rejection rate, which may have led to the lack of significant results. More data points may enable a better understanding of the impact of increasing the transparency on operator performance when the intelligent agent's decisions are incorrect. Overall, the present study suggests that increasing transparency information improved the operators' sensitivity in the intelligent agent's decisions by enabling them to correctly accept the intelligent agent's recommendation faster, but also caused an increase in the time taken to correctly reject the intelligent agent's recommendation.

Trust in the Intelligent Agent

The performance data showed participants improved their performance across the transparency levels when the intelligent agent's decisions were correct and had no significant improvement when the intelligent agent's decisions were incorrect. It suggests greater transparency reduces the disuse decisions, which occurred when operators reject the capabilities of the intelligent agent and refuse to accept the intelligent agent's decisions. However, the absence of difference in correct rejection rates across transparency levels suggests that increasing transparency may not reduce the misuse decisions, which arise when operators become complacent and overly rely on the intelligent agent's decisions. Therefore, greater transparency may assist operators not to under-trust the agent by reducing the disuse decisions; however, it may not help operators in preventing over-trust. Moreover, the Signal Detection Theory analysis revealed that the sensitivity increased in relation to the transparency levels. It indicates participants were able to better discriminate the accuracy of the intelligent agent's decisions with the increased information on the agents' intent, reasoning and projection.

Additionally, the Signal Detection Theory analysis showed that the operators were better able to discriminate the correct intelligent agent's decisions from the incorrect intelligent agent's decisions and were more reliant on the intelligent agent when intelligent agent's decisions were correct. Considering the performance data and Signal Detection Theory analysis as the objective trust measure, it indicates that greater transparency on the agent's intents, reasoning and projection lead to greater operator trust when the intelligent agent's decisions are correct.

However, the objective trust measure may only partially assess the operators' trust in the intelligent agent. Participants might not trust the intelligent agent and ignore the agent's recommendations; however, they might still be able to make the correct decisions manually with more transparency information being available (Parasuraman & Riley, 1997). Consequently, the subjective trust measure was examined to provide further insight into the participants' trust in the intelligent agent. When the intelligent agent was more transparent, participants reported greater trust in the intelligent agent ability to integrate and display information, while there was no significant difference in participants' trust in the intelligent agent's ability in making decisions. The opposite was found with previous finding (Mercado et al., 2016), which found participants' trust in an intelligent agent's decision making ability increased while trust in the intelligent agent's ability to integrate and display information showed no significant difference with increased transparency levels. Together with the objective trust data, the operators in the current study increased their trust in the intelligent agent and relied more on the intelligent agent when the agent was more transparent in its intention and behaviours. This supports Hypothesis 3 that operator trust would increase with increased transparency information. One explanation of the increase in trust is that the operators might have a better understanding of the capability and limitations of the more transparent agent, and how it arrives at its decisions. Overall, greater transparency caused the participants to be more reliant on the agent when the intelligent agent's decisions are correct, however, participants did not show more reliance on the intelligent agent when the intelligent agent's decisions were incorrect. Therefore, the current study did not find that participants had more calibrated trust with increased transparency information, which is inconsistent with a previous finding (Mercado et al., 2016). An explanation of the trust result is that the decision making required to reject the intelligent agent’s recommendation might be difficult due to the complexity of the target identification task. Future research may examine the operator decision making strategies when the intelligent agent’s recommendations are correct and incorrect. Moreover, trust is a complex construct and is influenced by the human operator, the intelligent agent and environmental factors. The current finding shows that each level of agent transparency impacts on the operator trust to a different extent. For instance, operator trust of the intelligent agent’s ability to integrate and display information significantly increased from SAT Level 1+2 to SAT Level 1+2+3; however, no difference was found between SAT Level 1 and SAT Level 1+2. Future research may investigate how the additional information in each level of agent transparency impacts operator trust.

Furthermore, participants were sensitive to the experiment’s manipulation of reliability as reflected in the results of subjective trust measures. Participants perceived the intelligent agent as being more trustworthy in displaying analysed information than in suggesting or making decisions for all transparency levels. The agent was designed to be accurate in analysing the information all the time in this experiment, while the decision made by the intelligent agent was only accurate for 70% of the time for all transparency levels. This is supported by Wang et al. (2009) which found that disclosing the reliability level of the intelligent agent positively influenced the operator in having an appropriate level of trust. Thus, it indicates that participants recognised the reliability of the intelligent agent in the current study and were able to differentiate their trust in different elements of the intelligent agent.

Operator Workload

In contrast to Helldin (2014)'s finding which suggested that increasing transparency information could increase the demand on operators' information processing capacity, the current study showed that presenting information on agent's intent, reasoning and projection enabled the agent to be more transparent without increasing operator workload. Therefore, the result is consistent with Mercado et al. (2016)'s finding and supports Hypothesis 4. It may be because the information helped the operator to understand the rationale behind the agent's decision and therefore it reduced the operator's mental effort. Additionally, participants perceived that their performance improved when the agent was more transparent, which is consistent with the performance data. Overall, a more transparent agent improved operator performance without increasing workload, which is consistent with the other findings (T. Chen et al., 2014; Mercado et al., 2016).

Limitation

Participants took longer to correctly reject than to correctly accept the intelligent agent's recommendation with greater transparency; however, the correct rejection rates were significantly lower than the correct acceptance rates when the agent was more transparent. It is possible that this may be due to an increase in cognitive effort when the intelligent agent's recommendation was incorrect; however, operators did not report any change in workload with different levels of agent transparency. It is possible that an objective measure of workload such as eye movement and pupillary responses may be more sensitive to changes (Buettner, 2013). While some research has suggested objective workload measures may not be correlated (Crabtree, Bateman, & Acton, 1984), Mercado et al. (2016) found consistent results on increasing agent transparency with the subjective workload measure (Mercado et al., 2016). Therefore, using other workload measures may provide more insight into the longer response time for correct rejection.

In addition, civilian participants were recruited for this study who had little or no prior experience with maritime contact identification. However, the interface used in the current study was designed for civilians based on the Ecological Interface Design (Neyedli et al., 2011; Vicente & Rasmussen, 1992). For instance, simplified text was used in the interface to allow the naïve participants to easily understand the content of the task. Moreover, 25 out of 27 participants had already completed tertiary education, and over 70% of the participants reported they were able to program in at least one language. The result might not be generalisable to the general and Defence population. Nonetheless, Helldin (2014) recruited Defence participants in the study and found improvement on performance when the agent was more transparent.

Future Research

The current study focused on the impact of increasing agent transparency in a target identification task but uncertainty information was not included in SAT Level 3 (intelligent agent's projection). This study showed displaying the intelligent agent's intents, reasoning and projection showed to improve operator performance. Future research could examine the impact of including uncertainty in agent transparency in the projection information. Uncertainty might be critical for decision making in target identification. Disclosing the uncertainty information has been shown to reduce operators' attempts to make a final identification without an increase in workload or time required (Riveiro, Helldin, Falkman, & Lebram, 2014). Moreover, Selkowitz, Lakhmani, and Chen (2017) have separated uncertainty from SAT Level 3 in a route planning task, and found that operators were more cautious in trusting the agent to make decision and execute actions when the uncertainty information was presented (Selkowitz et al., 2017). Moreover, operators have been shown to be less likely to adopt risky behaviours with the uncertainty information (Andre & Cutler, 1998). Therefore, follow-up research could consider uncertainty separately from SAT Level 3 and investigate the impact of displaying information of intelligent agent's projection and uncertainty respectively on operator performance, trust and workload in the target identification task.

In the current study, participants spent more time correctly rejecting an agent's recommendation across transparency level, while they did not spend longer time correctly accepting the recommendation when transparency increased. It is unclear what the participants' decision making was strategy in the present study. Human operators, in particular military operators, heavily rely on their own subjective experience and decision making strategies to make a decision (Roux & van Vuuren, 2007). When presenting information about the agent's uncertainty, some operators may prefer to examine all possible options in relation to the worst case scenarios, while some operators may evaluate the options as to their expected outcomes (Roux & van Vuuren, 2007). Therefore, future research could explore participants' decision making strategy to gain a better understanding of the participants' approach to the task including how they arrive at their decision in accepting or rejecting the agent's recommendations. Moreover, a baseline condition, which is a condition without the intelligent agent's recommendation could be included in a future study to further understand how the participants utilise the intelligent agent.

The current study examined the impact of increasing transparency on operator performance in terms of accuracy and sensitivity. Future research could investigate how an increase in agent transparency affects the operators' choice of target identity. Previous research has shown that operators are more likely to choose suspect identities in target identification as they may consider the worst case scenario when the uncertainty information is presented (Riveiro et al., 2014). Therefore, future research could investigate the influence of transparency level on the choice of target identity.

The current study used simple graphic and textual representations to display transparency information for civilian participants. Future research could examine other interface designs to display the three transparency levels information for experts based on the Ecological Interface Design and investigate their impact on operator performance, trust and workload (Neyedli et al., 2011; Vicente & Rasmussen, 1992).

Conclusions

The current study has broadened the research of agent transparency by showing that greater transparency of the intelligent agent's decision enhances operators' ability to assess the accuracy of an intelligent agent's decisions for a target identification task in the maritime domain. Moreover, increasing agent transparency did not increase operator workload or increase the time taken to accept correct intelligent agent's decisions, yet it caused operators to spend more time rejecting incorrect intelligent agent's decisions. Displaying more transparency information was beneficial in terms of building operator trust in the intelligent agent as operators perceived the intelligent agent to be more trustworthy and were more reliant on the intelligent agent's correct decisions. This finding may facilitate the design and use of intelligent agents for uninhabited vehicles management in the maritime domain. Future research could investigate other interface designs for displaying agent transparency information to experienced military operators.

Acknowledgments

I would first like to express my sincere gratitude to my supervisor, Dr. Susan Cockshell for the continuous support of the research, for her patience, motivation and immense knowledge. Her door was always open whenever I had a question about my research. Her guidance and encouragement helped me in all the time of research and writing of this thesis.

I would also like to thank Dr. Kingsley Fletcher for his insightful comments and expertise that greatly assisted the research and writing, but also for his help in experiment design and data analysis. I would like to express my appreciation to Dr. Anna Ma-Wyatt for her guidance of research and I gratefully thank for her valuable comments on this thesis.

My sincere thank also goes to Ashley Arnold, who helped me in modifying the software for the purpose of this research. Without his precious support, it would not be possible to conduct this research. A special thanks to my workmates in DST Group Sebastian Tsui and Helen Geng for their support and company throughout the year, which brightened up my research time.

Key Points

- Greater agent transparency improved operator ability to assess the accuracy of an intelligent agent’s decisions while also causing a higher tendency to follow the agent’s decision in an intelligent maritime target identification system.
- Displaying information on the intelligent agent’s intention, reasoning and projections facilitated the human-agent teaming without the cost of increased workload or a longer time to accepting correct intelligent agent’s decisions, while increasing the time taken to reject incorrect intelligent agent’s decisions.
- Increasing the agent transparency information enabled operators to build a higher level of trust and be more reliant on the correct intelligent agent’s decisions.

References

Andre, A. D., & Cutler, H. A. (1998). Displaying Uncertainty in Advanced Navigation Systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 42*(1), 31-35. doi:10.1177/154193129804200108

Braarud, P. Ø. (2001). Subjective Task Complexity and Subjective Workload: Criterion Validity for Complex Team Tasks. *International Journal of Cognitive Ergonomics, 5*(3), 261-273. doi:10.1207/S15327566IJCE0503_7

Buettner, R. (2013). Cognitive Workload of Humans Using Artificial Intelligence Systems: Towards Objective Measurement Applying Eye-Tracking Technology. In I. J. Timm & M. Thimm (Eds.), *KI 2013: Advances in Artificial Intelligence* (pp. 37-48). Berlin, Heidelberg: Springer.

Cain, B. (2007). *A review of the mental workload literature*. Toronto, Canada: Human System Integration Section, Defence Research And Development.

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

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*(1), 13-29. doi:10.1109/THMS.2013.2293535

Chen, J. Y. C., & Barnes, M. J. (2015). Agent Transparency for Human-Agent Teaming Effectiveness. *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 1381-1385. doi:10.1109/SMC.2015.245

Chen, J. Y. C., Barnes, M. J., & Harper-Sciarini, M. (2011). Supervisory Control of Multiple Robots: Human-Preformance Issues and User-Interface Design. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41*(4), 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). MD: Aberdeen Proving Ground.

Chen, T., Gonzalez, F., Campbell, D., & Coppin, G. (2014). Management of multiple heterogeneous UAVs using capability and autonomy visualisation: Theory, experiment and result. *2014 International Conference on Unmanned Aircraft Systems*, 200-210. doi:10.1109/ICUAS.2014.6842257

Cook, M. B., & Smallman, H. S. (2008). Human Factors of the Confirmation Bias in Intelligence Analysis: Decision Support From Graphical Evidence Landscapes. *Human Factors, 50*(5), 745-754. doi:doi:10.1518/001872008X354183

Crabtree, M. S., Bateman, R. P., & Acton, W. H. (1984). Benefits of Using Objective and Subjective Workload Measures. *Proceedings of the Human Factors Society Annual Meeting, 28*(11), 950-953. doi:10.1177/154193128402801103

Cummings, M., Bruni, S., Mercier, S., & Mitchell, P. J. (2007). Automation Architecture for Single Operator Multiple UAV Command and Control. *The international C2 Journal, 1*(2), 1-24. doi:10.1.1.91.4878

Cummings, M., Clare, A., & Hart, C. (2010). The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling. *Human Factors, 52*(1), 17-27. doi:10.1177/0018720810368674

Duggan, G. B., Banbury, S., Howes, A., Patrick, J., & Waldron, S. M. (2004). Too Much, Too Little or Just Right: Designing Data Fusion for Situation Awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48*(3), 528-532. doi:10.1177/154193120404800354

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 32-64. doi:10.1518/001872095779049543

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28*(1), 1-11. doi:10.3758/bf03203630

Finger, R., & Bisantz, A. M. (2000). Utilizing Graphical Formats to Convey Uncertainty in a Decision Making Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 44*(1), 13-16. doi:10.1177/154193120004400104

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(6939), 534-537. doi:10.1038/nature01647

Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: the case of action video game players. *Cognition, 101*(1), 217-245. doi:10.1016/j.cognition.2005.10.004

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904-908. doi:10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology, 52*, 139-183. doi:10.1016/S0166-4115(08)62386-9

Helldin, T. (2014). Transparency for future semi-automated systems: effects of transparency on operator performance, workload and trust. (PhD dissertation), Orebro University, Orebro.

Helldin, T., Ohlander, U., Falkman, G., & Riveiro, M. (2014). Transparency of Automated Combat Classification. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 22-33). Cham: Springer International Publishing.

Hoff, K. A., & Bashir, M. (2015). Trust in Automation. *Human Factors, 57*(3), 407-434. doi:10.1177/0018720814547570

Hubert-Wallander, B., Green, C. S., & Bavelier, D. (2011). Stretching the limits of visual attention: the case of action video games. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(2), 222-230. doi:10.1002/wcs.116

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics, 4*(1), 53-71. doi:10.1207/S15327566IJCE0401_04

Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness (6 ed.). New York: Harvard University Press.

Kulesza, T., Stumpf, S., Burnett, M., & Kwan, I. (2012). Tell Me More? The Effects of Mental Model Soundness on Personalizing an Intelligent Agent. In J. A. Konstan, E. H. Chi, & K. Höök (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1-10). New York: ACM.

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

Lewis, M. (2013). Human Interaction With Multiple Remote Robots. *Reviews of Human Factors and Ergonomics, 9*(1), 131-174. doi:10.1177/1557234x13506688

Lim, B. Y., Dey, A. K., & Avrahami, D. (2009). Why and why not explanations improve the intelligibility of context-aware intelligent systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 2119-2128. doi:10.1145/1518701.1519023

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 Annual Meeting, 50*(23), 2482-2486. doi:10.1177/154193120605002304

Lyons, J. B. (2013). *Being transparent about transparency: A model for human-robot interaction*. Paper presented at the 2013 AAAI Spring Symposium Series, Stanford University, Stanford, CA.

Lyons, J. B., & Havig, P. R. (2014). Transparency in a Human-Machine Context: Approaches for Fostering Shared Awareness/Intent. In R. Shumaker & S. Lackey (Eds.), *Virtual, Augmented and Mixed Reality* (pp. 181-190). Cham: Springer International Publishing.

Mercado, J. E., Rupp, M., Chen, J. Y. C., Barber, D., Procci, K., & Barnes, M. (2015). *Effects of agent transparency on multi-robot management effectiveness*. (No. 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, 58*(3), 401-415. doi:10.1177/0018720815621206

Neyedli, H. F., Hollands, J. G., & Jamieson, G. A. (2011). Beyond Identity: Incorporating System Reliability Information Into an Automated Combat Identification System. *Human Factors, 53*(4), 338-355. doi:10.1177/0018720811413767

Orosy-Fildes, C., & Allan, R. W. (1989). Psychology of Computer Use: XII. Videogame Play: Human Reaction Time to Visual Stimuli. *Perceptual and Motor Skills, 69*(1), 243-247. doi:10.2466/pms.1989.69.1.243

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, 39*(2), 230-253. doi:10.1518/001872097778543886

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*(3), 286-297. doi:10.1109/3468.844354

Rao, A. S., & Georgeff, M. P. (1995). BDI Agents: From Theory to Practice *Proceedings of the First International Conference on Mulitagent Systems* (pp. 312-319). Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

Riveiro, M., Helldin, T., Falkman, G., & Lebram, M. (2014). Effects of visualizing uncertainty on decision-making in a target identification scenario. *Computers & Graphics, 41*, 84-98. doi:10.1016/j.cag.2014.02.006

Ross, L., & Lepper, M. R. (1980). The perseverance of beliefs: Empirical and normative considerations. In R. A. Shweder & D. Fiske (Eds.), *Fallible Judgment in Behavioral Research* (pp. 17-36). San Francisco: Jossey-Bass.

Roux, J., & van Vuuren, J. (2007). Threat evaluation and weapon assignment decision support: A review of the state of the art. *ORiON, 23*(2), 151-187. doi:10.5784/23-2-54

Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach, Global Edition* (3rd ed.). Upper Saddle River, NJ: Prentice Hall

Safar, J. A., & Turner, C. W. (2005). Validation of a Two Factor Structure for System Trust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*(3), 497-501. doi:doi:10.1177/154193120504900360

Sarter, N. B., & Woods, D. D. (1995). How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control. *Human Factors, 37*(1), 5-19. doi:10.1518/001872095779049516

Selkowitz, A. R., Lakhmani, S. G., & Chen, J. Y. C. (2017). Using Agent Transparency to Support Situation Awareness of the Autonomous Squad Member. *Cognitive Systems Research, 46*, 13-25. doi:10.1016/j.cogsys.2017.02.003

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 Annual Meeting, 60*(1), 1319-1323. doi:10.1177/1541931213601305

Seppelt, B. D., & Lee, J. D. (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

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149. doi:10.3758/bf03207704

Stowers, K., Kasdaglis, N., Newton, O., Lakhmani, S., Wohleber, R., & Chen, J. Y. C. (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 Annual Meeting, 60*(1), 1706-1710. doi:10.1177/1541931213601392

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston: Pearson Education.

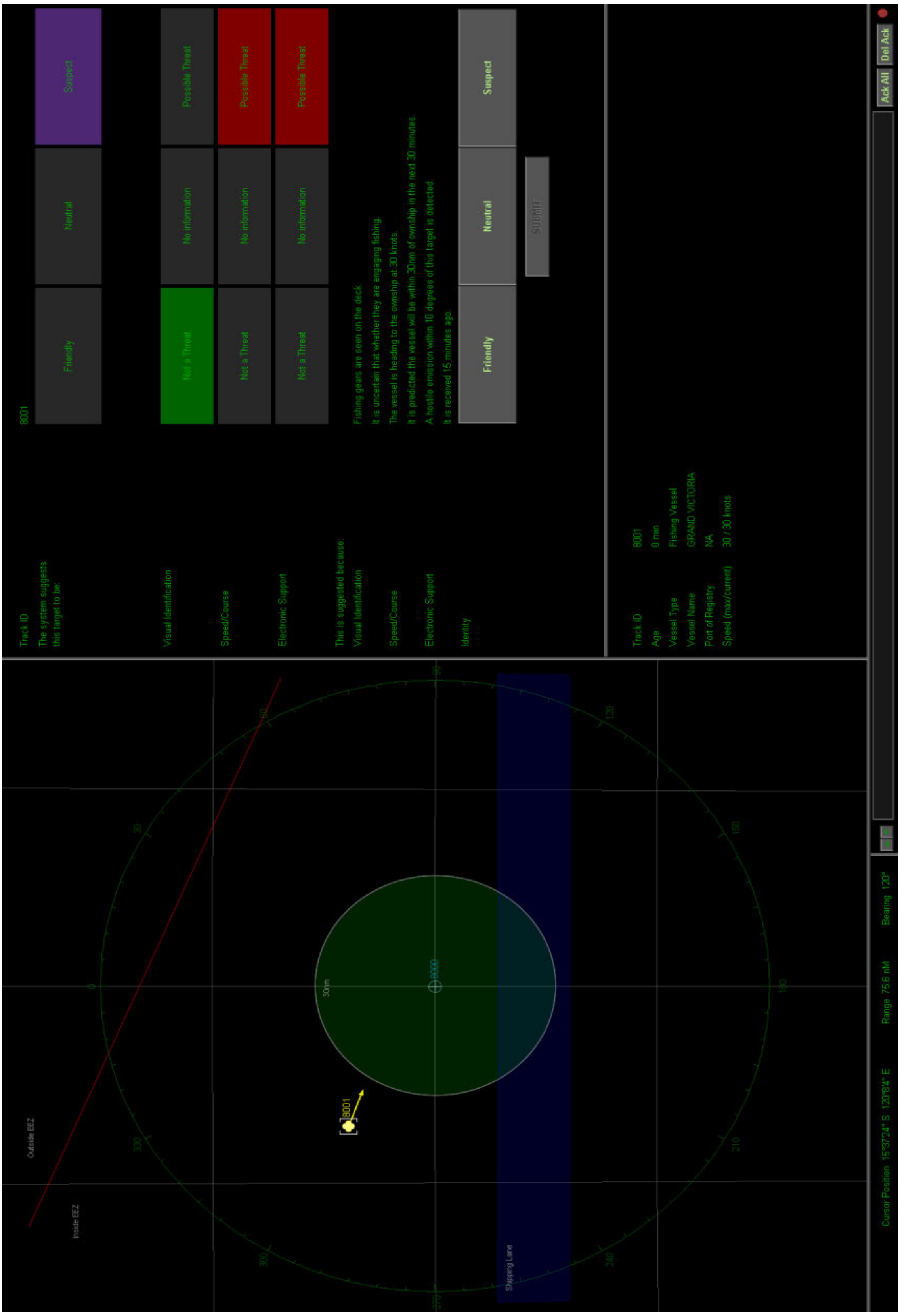
Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on systems, man, and cybernetics, 22*(4), 589-606.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and Reliance on an Automated Combat Identification System. *Human Factors, 51*(3), 281-291. doi:10.1177/0018720809338842

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: a synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8*(3), 201-212. doi:10.1080/14639220500370105

Williams, E. (1949). Experimental Designs Balanced for the Estimation of Residual Effects of Treatments. *Australian Journal of Chemistry, 2*(2), 149-168. doi:10.1071/CH9490149

Appendix A


Track ID
8001
The system suggests this target to be:
Friendly
Neutral
Suspect
Visual Identification
Not a Threat
No information
Possible Threat
Speed/Course
Not a Threat
No information
Possible Threat
Electronic Support
Not a Threat
No information
Possible Threat
This is suggested because:
Visual Identification
Fishing gears are seen on the deck.
It is uncertain that whether they are engaging fishing.
Speed/Course
The vessel is heading to the ownship at 30 knots.
It is predicted the vessel will be within 30nm of ownship in the next 30 minutes.
Electronic Support
A hostile emission within 10 degrees of this target is detected.
It is received 15 minutes ago.
Identity
Friendly
Neutral
Suspect
SUBMIT
Track ID
8001
Age
0 min
Vessel Type
Fishing Vessel
Vessel Name
GRAND VICTORIA
Port of Registry
NA
Speed (max/current)
30 / 30 knots
Ack All
Del Ack
Outside EEZ
Inside EEZ
Shipping Lane
30nm
8001
Cursor Position 15°37'24" S 120°8'4" E
Range 75.6 nM
Bearing 120°

Appendix B

NASA Task Load Index (TLX)

Participant No: ______ Session:____ Condition:____ Scenario:____ Date:_________

Please mark the point with a cross (x) on each rating scale that matches your experience during the last time period.

Mental Demand
How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was task easy or demanding, simple or complex, exacting or forgiving?

Low High

Physical Demand
How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low High

Time Demand
How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Low High

Effort
How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low High

Performance
How successful do you think you were in accomplishing the goals of the task set by the analyst (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Good Poor

Frustration Level
How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Low High

Appendix C

Trust Survey

Participant No: ________ Session:____ Condition:____ Scenario: ____ Date:___________

For each of the following items and situations, circle the number which best describe your feeling or your feeling or your impression based on the system you just used. For each item, consider the following situations:

- When the system is integrating information, generating predictive displays, and/or presenting its analysis.
- When the system is making decisions and/or selecting actions.

1. The system is deceptive.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

2. The system behaves in an underhanded manner.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

3. I am suspicious of the system’s intent, action, or output.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

4. I am wary of the system.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

5. The system's action will have a harmful or injurious outcome.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

6. I am confident in the system.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

7. The system provides security.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

8. The system has integrity.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

9. The system is dependable.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

10. The system is reliable.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

11. I can trust the system.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

12. I am familiar with the system.

	Not at all			Neutral			Extremely
Integrating and Displaying Analysed Information	1	2	3	4	5	6	7
Suggesting or Making Decisions	1	2	3	4	5	6	7

Appendix D

Demographics Questionnaire

Participant No: ________ Date: ___________

Please circle the corresponding response.

Your Age: ________

Your Gender:
Male Female

What is the highest level of education you have completed?

a) Less than Year 12 or equivalent
b) Year 12 or equivalent
c) Vocational Qualification / Associate Diploma / Advanced Diploma
d) Bachelor degree / Bachelor degree Honours
e) Postgraduate degree (including postgraduate diploma, Master and Doctorate)

Which of the following software programs are you capable to use without any help?
(Please tick on one or more boxes)

- ☐ Word processing (e.g. Word)
- ☐ Spreadsheet (e.g. Excel)
- ☐ Presentation software (e.g. PowerPoint)
- ☐ Databases (e.g. Access)
- ☐ Graphic / Movie editing software (e.g. Photoshop, iMovie)
- ☐ Internet (e.g. Internet Explorer, Chrome)
- ☐ Email (e.g. Outlook)
- ☐ Others: ____
- ☐ None

How many languages are you capable to program in (e.g. Java, JavaScript, or C)?

a) None

b) 1-3

c) 4-6

d) 7-9

e) 10 or more

How often do you play computer/video games?

a) Daily

b) Weekly

c) Monthly

d) Less than once a month

e) I have played computer / video games in the past but not for many years

f) Never

The Journal of the Human Factors and Ergonomics Society

Manuscript Preparation

- *Human Factors* manuscripts should be prepared according to **editorial style and ethical guidelines** of the *Sixth Edition of the Publication Manual of the American Psychological Association* (APA, 750 First St., NE, Washington, DC 20002; 800/374-2721).
- All text must be double-spaced with 1-inch margins, and must contain page numbers. Other formatting instructions for text, tables, figures, and references, are included in the *Publication Manual*.
- Exceptions to the APA *Publication Manual* are as follows:

1. **Use a structured abstract.** Prepare a structured abstract of no more than 250 words, with information arranged under the following subheadings (include the subheadings in your abstract), with each subheading beginning on a new line. We recognize that these categories may be a bit awkward for review papers or papers that use nontraditional methodologies, such as modeling or naturalistic observation, but we encourage the authors to do their best to adapt to this structure.
    1. Objective
    2. Background
    3. Method
    4. Results
    5. Conclusion
    6. Application (nontheoretical works)—A statement that reflects the practical impact of this work to a broad audience.

    View examples of structured abstracts at (empirical article and review article)

2. **Footnotes are not permitted.** Such notes should be incorporated into the text.
3. Add line numbering to the entire manuscript, starting with line 1 for the title of the submission. Line numbering aids the reviewers when commenting on the manuscript.
4. **Place all figures and tables (with captions) within the manuscript where first mentioned in the text.** If accepted, figures, tables, and captions will be placed at end of manuscript according to the APA *Publication Manual*. Guidelines for figures are explained on the SAGE Figure Guidelines page. Recommendations for presenting data in text, tables, and figures is available in a *Human Factors* article, "Guidelines for Presenting Quantitative Data in HFES Publications" (Gillan, Wickens, Carswell, & Hollands, 1998).
    Please indicate in your cover letter whether any of your figures must contain color.

Authors may be responsible for paying the costs for color. HFES will notify the author of such costs.

5. **Each manuscript should contain the following components, in the following order:**
   1. **Title page,** which contains:
      1. Title (25 words maximum)
      2. Each author's name and affiliation (institution, city, state, country) — OMIT IF REQUESTING A DOUBLE-BLIND REVIEW
      3. Running head
      4. Manuscript type
      5. Exact word count of text (not including title page, abstract, biographies), and references
      6. Acknowledgments (including contact information for corresponding author). If applicable, list funding sources and other pertinent disclosures. If no such acknowledgments are present in the initial submission, HFES will assume that no disclosures are necessary.
   2. **Abstract page,** which contains:
      1. Structured abstract
      2. Up to 5 keywords (exclude words that already appear in the title). View the current list of keywords. The importance of keywords to authors finding your article, and tips for choosing keywords, can be found at SAGE Publications.
      3. Précis: a 50-word description (in 1–3 sentences) of the manuscript, which will appear in the Table of Contents below the title and authorship information
   3. **Main body of paper.**
      Please note that all manuscripts must contain an explicit and clear discussion of the study's practical implications. (If applicable, state explicit design recommendations or principles).
      When reporting results, authors should follow the guidelines in the Publication Manual of the American Psychological Association. Authors are strongly encouraged to include measures of effect size (e.g., partial eta-squared) and variability (e.g., standard of mean, confidence intervals), and include standard error bars on data plots, as applicable to the study.
   4. **Key points:** A list of key points in bullet form, inserted prior to the References list

5. **References** (in APA style of hanging indent)
6. **Biographies:** For each author, indicate the current affiliation and highest degree obtained (field, year obtained, institution).

6. Authors are strongly encouraged to provide supplemental materials that would facilitate replication of the studies. Such materials would be available on-line at the journal's website. Examples include data, instructions, stimuli, algorithms, and questionnaires.