2016

Transparency in human-agent teaming and its effect on complacent behavior

Julia Wright
University of Central Florida

TRANSPARENCY IN HUMAN-AGENT TEAMING AND ITS EFFECT ON COMPLACENT BEHAVIOR

by

JULIA L. WRIGHT

B.S. Grand Valley State University, 2011
M.S. University of Central Florida, 2013

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2016

Major Professor: Peter A. Hancock

ABSTRACT

This study examined how transparency of an intelligent agent's reasoning affected complacent behavior in a route selection task in a simulated environment. Also examined was how the information available to the operator affected those results.

In two experiments, participants supervised a three-vehicle convoy as it traversed a simulated environment and re-routed the convoy when needed with the assistance of an intelligent agent, RoboLeader. Participants were randomly assigned to an Agent Reasoning Transparency condition. Participants received communications from a commander confirming either the presence or absence of activity in the area. They also received information regarding potential events along their route via icons that appeared on a map displaying the convoy route and surrounding area. Participants in Experiment 1 (low information setting) received information about their current route only; they did not receive any information about the suggested alternate route. Participants in Experiment 2 (high information setting) received information about both their current route and the agent recommended an alternative route.

In the first experiment, access to agent reasoning was found to be an effective deterrent to complacent behavior when the operator has limited information about their task environment. However, the addition of information that created ambiguity for the operator encouraged complacency, resulting in reduced performance and poorer trust calibration. Agent reasoning did not increase response time or workload and appeared to have improved performance on the secondary task. These findings align with studies that have shown ambiguous information can increase workload and encourage complacency,

as such, caution should be exercised when considering how transparent to make agent reasoning and what information should be included.

In the second experiment, access to agent reasoning was found to have little effect on complacent behavior when the operator had complete information about the task environment. However, the addition of information that created ambiguity for the operator appeared to encourage complacency, as indicated by reduced performance and shorter decision times. Agent reasoning transparency did not increase overall workload, and operators reported higher satisfaction with their performance and reduced mental demand. Access to agent reasoning did not improve operators' secondary task performance, situation awareness, or operator trust. However, when agent reasoning transparency included ambiguous information complacent behavior was again encouraged. Unlike the first experiment, there were notable differences in complacent behavior, performance, operator trust, and situation awareness due to individual difference factors. As such, these findings would suggest that when the operator has complete information regarding their task environment, access to agent reasoning may be beneficial, but not dramatically so. However, individual difference factors will greatly influence performance outcomes.

The amount of information the operator has regarding the task environment has a profound effect on the proper use of the agent. Increased environmental information resulted in more rejections of the agent recommendation regardless of the transparency of agent reasoning. The addition of agent reasoning transparency appeared to be effective at keeping the operator engaged, while complacent behavior appeared to be encouraged both when agent reasoning was either not transparent or so transparent as to become ambiguous.

Even so, operators reported lower trust and usability for the agent than when environmental information was limited. Situation awareness (SA2) scores were also higher in the high information environment when agent reasoning was either not transparent or so transparent as to become ambiguous, compared to the low information environment. However, when a moderate amount of agent reasoning was available to the operator, the amount of information available to the operator had no effect on the operators' complacent behavior, subjective trust, or SA. These findings indicate that some negative outcomes resulting from the incongruous transparency of agent reasoning may be mitigated by increasing the information the operator has regarding the task environment.

This work is dedicated to everyone who sets their own dreams aside while supporting others' dreams, and to my daughter Brittany, who helped me realize why my dreams are important too.

ACKNOWLEDGMENTS

This work was supported in part by the U. S. Army Research Laboratory. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the U. S. Army Research Laboratory or the U. S. Government.

I was supported by a Presidential Fellowship from the University of Central Florida. I am grateful to Dr. Peter A. Hancock and the members of the fellowship committee for their support and encouragement.

I would like to thank the members of my committee: Dr. Florian G. Jentsch, Dr. James L. Szalma, Dr. Peter A. Hancock, for serving as my academic advisor throughout my graduate career, and Dr. Jessie Y. C. Chen, for serving as my Army Research Laboratory mentor. Thank you for your patience and guidance; I've learned so much under your tutelage, most importantly how much more I have yet to learn. I would also like to thank Dr. Daniel Barber, Jonathon Harris, Isacc Yi, and Olivia Newton for their contributions to this work.

I would especially like to thank the members of my cohort, whose encouragement, support, and kindness made a challenging process not only tolerable but downright enjoyable. You helped me stay focused when I became distracted, cheered me when I despaired and listened when I needed to vent. I will forever cherish your friendship and camaraderie, particularly Gaby Hancock, Ashley Hughes, and most of all my 'school husband', Tim 'Stroop' White. We are doctors now, baby!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS/ACRONYMS

ANOVA	Analysis of Variance
ART	Agent Reasoning Transparency
ART1	Agent Reasoning Transparency Level 1
ART2	Agent Reasoning Transparency Level 2
ART3	Agent Reasoning Transparency Level 3
AT	Agent Transparency
C.I.	Confidence Interval
CP	Complacency Potential
CPRS	Complacency Potential Rating Scale
DoD	Department of Defense
DT	Decision Time
EXP	Experiment
EXP1	Experiment 1
EXP2	Experiment 2
FA	False Alarm
FC	Fixation Count
FD	Fixation Duration
Frust	Frustration
ID	Individual Differences
IED	Improvised Explosive Device
IR	Infrared

LOA	Level of Assistance
LOR	Level of Reasoning
NASA-TLX	National Aeronautics and Space Agency Task Load Index
Max	Maximum
MD	Mental Demand
Mdn	Median
Min	Minimum
OOTL	Out of the Loop
PAC	Perceived Attentional Control
PDia	Pupil Diameter
Perf	Performance
PhyD	Physical Demand
RED	Remote Eyetracking Device
RL	RoboLeader
RSPAN	Reading Span
RT	Response Time
SA	Situation Awareness
SA1	Situation Awareness Level 1
SA2	Situation Awareness Level 2
SA3	Situation Awareness Level 3
SAT	Situation awareness-based Agent Transparency
SE	Standard Error

SD	Standard Deviation
SMI	Sensomotoric Instrument
SOT	Spatial Orientation Test
SV	Spatial Visualization
TD	Temporal Demand
TOR	Time of Report
WMC	Working Memory Capacity

CHAPTER 1: INTRODUCTION

A soldier on the battlefield is often required to conduct multiple concurrent tasks. These include demands such as maintaining local security, identifying and assessing threats and maintaining situation awareness (SA) (Barnes, Chen, Jentsch, & Haas, 2006; Chen, Durlach, Sloan, & Bowens, 2008; Hancock & Szalma, 2008). Employing robotic assets to assist in these respective duties permits the soldier to manage such multiple tasks as they increase in complexity. However, contemporary research shows that one operator managing multiple robotic assets suffers from performance decrement, reduced SA, and increased workload (Chen et al., 2008; Lewis, 2013; Wang, Lewis, Velagapudi, Scerri, & Sycara, 2009). In response to these concerns, an intelligent agent, "RoboLeader" (RL), has been developed to assist in the management of a team of supportive robots (Chen, Barnes, & Qu, 2010). Studies on this technology have indicated that using an intelligent agent as the mediator of the robotic team helps to improve operators' performance, SA, and decrease associated workload (Chen & Joyner, 2009; Chen & Terrence, 2009). However, in a recent RoboLeader study (Wright et al., 2013), operator performance degraded at the highest level of agent assistance. This might be due to the occurrence of automation-induced complacency (see Parasuraman, Molloy, & Singh, 1993; Parasuraman, Sheridan, & Wickens, 2000). Whether this behavior was due to premature cognitive commitment (Langer, 1989), some other complacent behavior, such as automation bias, or if the operator understood they had insufficient knowledge to appropriately override the automation remained unclear. What is clear is that there is still much to learn about human performance issues associated with human-agent teaming.

In the realm of human-automation interaction, a current topic of investigation is the quality of the interaction between the human operator and automated systems, specifically, how the operators' understanding of the system's actions affect their performance, and what qualities are contained within the automated system that might enhance this interaction. This area of interest is referred to as 'transparency', but presently there is no consensus on exactly how it should be defined. Transparency has been described both as something the automation provides (whether by design or by behavior) (Cramer, et al., 2008; Cuevas, Fiore, Caldwell, & Strater, 2007; Kim & Hinds, 2006), and as the understanding or knowledge an operator has regarding the systems performance, behavior, or internal state (Cheverst et al., 2005; Cring & Lenfestey, 2009; Jameson, Baldes, Bauer, & Kroner 2004). Regardless of the definition used, it is agreed that the lack of transparency within human-automation interaction negatively impacts operator performance.

Appropriate levels of transparency between the human and the agent must be present to enhance the effectiveness of the interaction. However, no quantitative method for defining and assessing agent transparency yet exists. To address this need, a model of Situation awareness-based Agent Transparency (SAT; Chen et al., 2014) has been developed. This model defines agent transparency (AT) 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*." The SAT model has levels that approximately correspond to Endsley's (1995) situation awareness (SA) model. However, it also incorporates Lee and See's (2004) "three P's" (i.e., purpose, process, and performance) for human-agent trust development, the Beliefs, Desires, Intentions (BDI) Agent Framework (Rao & Georgeff, 1995), as well as findings from recent studies (Chen & Barnes, 2012a, 2012b; Chen & Barnes, 2014; Cring & Lenfestey, 2009; Lyons & Havig, 2014).

This dissertation proposes to investigate performance associated with human-agent teaming as it pertains to appropriate agent transparency. This evaluation is set within the framework of the SAT model. Current Department of Defense (DoD) research (Mercado et al., in press; Boyce et al., in press) has explored the relationship between access to agent reasoning and decision-making, within the framework of the SAT model, in static single-task conditions. The present research investigated such factors but used a dynamic, multi-tasking simulation that emulated a real-world military environment.

Thus, this research addresses the following questions:

1. How does increased knowledge of the task environment affect an operator's performance, situation awareness, workload, and trust in an agent, such as RoboLeader?
2. How does increased access to the agent's reasoning affect operator's performance, situation awareness, workload, and trust?
3. Does increased knowledge of the task environment reduce the need for access to agent reasoning?
4. Does increased access to agent reasoning reduce complacency?
5. Do individuals with low attentional control, low spatial ability, or low working memory capacity have an increased need for access to agent reasoning?

Accrued results are expected to elucidate how the operators' knowledge of the task environment interacts with their understanding of agent reasoning to create 'transparency,' as well as how increased access to the reasoning behind automation 'decisions' affects a human operators' ability to interact effectively with said automation. While insufficient transparency may hinder operator trust in the automation, too much transparency may also have detrimental effects on operator performance, situation awareness, and decision-making by encouraging complacent behavior. This work will also investigate how individual difference factors influence the human-agent relationship in terms of transparency, and the subsequent effect on related human performance issues.

The findings of this work are expected to expand the current understanding as to how agent transparency exerts influence; identify the role operator knowledge and access to agent reasoning have in the interaction that creates the emergent construct 'transparency,' and how transparency should be evaluated from both a system and an operator's standpoints. These findings may serve to benefit human-automation research directly, but also promises to inform system designers, as well as guide operator training and performance evaluations.

CHAPTER 2: LITERATURE REVIEW

What is Automation?

References to automation date as far back as Ancient Greece, with tales of the god Hephaestus creating automatons to work in his smithy (Graves, 1960, p. 150). Automatons represent mechanical devices that, once started, complete a predetermined function, movement sequence or series of movement sequences on their own (Koetsier, 2001). Homer's description of self-opening doors in the Iliad, Plato's praise for Daedalus' moving sculptures (Automatones, n.d.), and Ktesibios' water-powered automata, including the first cuckoo clock, all appear in writings of the period that pre-date the birth of Christ (Pollard & Reid, 2007, p. 132). Hellenistic automata were designed as both tools and toys, meant to empower but primarily to entertain (Automatones, n.d.).

Early medieval automata were almost exclusively for entertainment. Eighth and ninth century Bagdad boasted many such automata, i.e., wind-powered statues, artificial animals, even a programmable flute player (Koetsier, 2001). In thirteenth century Italy, Count Robert II's reconstruction of the castle at Hesdin included plans for an elaborate park, containing many examples of mechanized fountains, sundials, and animals (Truitt, 2010). However, the earliest conception of automata as tools would not reoccur until the Renaissance.

Early Renaissance automata did continue to serve mostly as entertainment. However, the transition to from toy to tool had begun. Mechanical clocks and carillon appeared in the 1300's (Koetsier, 2001). It could reasonably be argued that Turriano's mechanical monk (circa 1560) is an example of a tool rather than a toy, an 'automatic prayer machine' (King, 2002), since the predominant religious view of that era held that prayers yielded tangible results, and in fact, the church ran a lively business in which people paid to have masses said for their beloved dead. Advancement in clockwork inspired early philosophers (Koetsier, 2001), as evidenced by Dr. John Dee's Wheeling Beetle (circa 1547) (Coovert & Thompson, 2014; Hancock, 2009), and in the seventeenth century Descartes' '*The World'* was published (Descartes, 1998/1664). The notion of the 'body as a machine' seemed to captivate artisans, and automata began to emulate life in great detail, as evidenced by de Gennes' eating peacock (Ancient Toys, 1887) and later, Vaucanson's digesting duck (Riskin, 2003) and Merlin's Silver Swan (The Silver Swan, n.d.). However, the 'body as a machine' idea also appears to have spurred an intuitive leap, in terms of automation development. Here, the emphasis in automation design began to evolve from mere entertainment to much more practical concerns.

In the eighteenth century, the development of automation for practical purposes increased dramatically, both in quantity and diversity. One of the most accepted of current definitions of automation is "*the use of electronic or mechanical devices to replace human labor*" (Sheridan, 2002). Although the term 'automation' was not coined until the 1950's (see Sheridan, 2002, p. 9), examples of automated systems (as per Sheridan's definition) appear as far back as the 18th century (Bennett, 1996; Sheridan, 2002). During the era now known as the Industrial Revolution, automated systems were developed to resolve labor-intensive tasks, such as spinning yarn or cleaning cotton, thus increasing production rates while decreasing manual labor requirements (Bennett, 1996; Sheridan, 2002, p. 10). The invention (and refinement) of the steam engine created a portable power source for such new machinery. Modern automation was developed as a method to free humans from performing tasks for which they are ill-suited, and in the process, their role changed from that of the laborer to periodic 'supervisor' (Fitts et al., 1951).

While automation initially was intended for tasks that were tedious, repetitive, or requiring vigilance to perform precisely and consistently, over time it became the panacea that allowed humans access to new dimensions of operation (e.g., complex mathematical computations), as well as enable humans to work in unsuitable environments (e.g., gathering soil samples inside volcanic craters). Today automation encompasses not only physical workload but mental workload as well, including such aspects as decision making (Sheridan, 2002). As systems continue to become more autonomous the role of humans in such human-automation systems is shifting once again, from that of periodic supervisor, where they still need to oversee and occasionally intervene in the process, to a level of pure administration, in which they oversee multiple systems and do not (typically) become involved in the direct operations of any.

Early automation was little more than mechanization of specific tasks, with the human operator still very involved in the production process, so much so that at times he appeared to be part of the machine itself (Figure 1). Once the notion of having machines perform the work of humans took hold, the next step in development was to have the machines perform the work in the same manner as the human. The eighteenth century saw the appearance of the feedback control mechanism and the development of various forms of governors, which allowed automatic adjustments to processes (Bennett, 1996). The Analytical Engine, the first programmable calculating machine, was designed by Babbage in the early nineteenth century. Babbage's machine programming used punch cards to direct the machine in specifying and performing a series of tasks. However, Babbage never created a completed machine, and the programmable calculator was forgotten until the 1940s when Zuse built his version (Koetsier, 2001). As advancements in automation were made, the ability to have the automation carry out multiple, successive tasks became a reality.

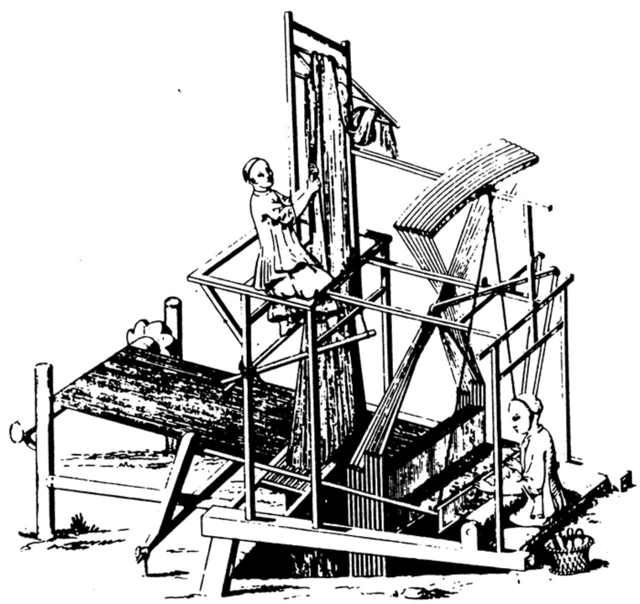


Figure 1. Chinese drawloom, circa Han Dynasty. The loom mechanized the task of raising and lowering the warp, creating space for the passage of the weft. The 'draw boy' keeps the strings separated while the weaver (seated) operates the loom (Koetsier, 2001).

Once automated systems began to imitate how humans work, both in the ability to perform successive tasks and the ability to alter task sequence, a new problem arose. Specifically, this was the problem of which tasks should be assigned to the automation and which to the human operator. Researchers have struggled to identify an optimal method for determining task assignment (machine or man) for many years (Chapanis, 1965; Fitts et al., 1951; Sheridan, 2006). Often they resort to some type of 'Men Are Better At, Machines Are Better At' (MABA-MABA; Fitts et al., 1951) function allocation list meant to equate a degree or level of automation with some perceived justification based on system performance (e.g., Endsley & Kaber, 1999; Parasuraman, Sheridan, & Wickens, 2000; Sheridan, 2000). One example of an MABA-MABA list is the 'Ten Levels of Automation of Decision and Action Selection' model (Table 1; Parasuraman et al., 2000), which defines automation as varying along a continuum of levels, with each level specifying which responsibilities are assigned to the human and which to the automation. While the lowest levels have the human maintaining authority and executing all actions, at each successive level the automation increasingly becomes more autonomous. At the highest level, the automation is completely autonomous, making decisions and carrying out actions without human input or approval. Thus, as the automation level increases, the responsibilities of the human operator decrease, until at the highest level of automation the human no longer has a role (see Table 1).

Table 1. Levels of automation of decision and action selection (Parasuraman et al., 2000).

HIGH	10. The computer decides everything, acts autonomously, ignoring the human.
	9. informs the human only if it, the computer, decides to
	8. informs the human only if asked, or
	7. executes automatically, then necessarily informs the human, and
	6. allows the human a restricted time to veto before automatic execution, or
	5. executes that suggestion if the human approves, or
	4. suggests one alternative
	3. narrows the selection down to a few, or
	2. The computer offers a complete set of decision/action alternatives, or
LOW	1. The computer offers no assistance: human must take all decisions and actions.

Automation at any level can have both beneficial and detrimental effects on human performance (Parasuraman & Mouloua, 1996; Parasuraman & Riley, 1997). As automation levels increase, increased automation-induced complacency and reduced situation awareness become particularly problematic, as the human operator falls ‘out of the loop’ (Endsley, 1996; Parasuraman, Molloy, & Singh, 1993). While the MABA-MABA lists have demonstrated some utility in system design, they have many shortcomings (Bainbridge, 1983; Dekker & Woods, 2002; Fuld, 2000). This has led some researchers to conclude that their use should be abandoned in favor of a more human-centered design that would stress the importance of the human-machine relationship becoming more synergistic, rather than a dichotomous assignment of specific tasks (Boy, 2014; Dekker & Woods, 2002; Hancock & Chignell, 1993; Marras & Hancock, 2014).

Issues in Automation

Automation-induced Complacency

Complacency has been defined as “*self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state*” (Billings, Lauber, Funkhouser, Lyman, & Huff, 1976). According to this definition, the human operator adopts the assumption that all is as it should be [with the automation], even without evidence that this is true. They thus become less diligent in their supervision of the automation. Initially considered to be a result of boredom, more recent studies have indicated that complacency and boredom are distinct and separate constructs (Parasuraman et al., 1993). Automation-induced complacency is thought to occur when conditions are such that the operator’s trait complacency combines with task conditions that favor such complacent behavior, typically in multitasking environments when an operator must divide their attention across multiple tasks. Therefore, when discussing complacency, a distinction must be made between the propensity for complacency and actual complacent behavior.

Complacency potential is a trait of the user and evidenced in their attitude towards (i.e., trust in, reliance upon and confidence in) automation (Parasuraman, Molloy, & Singh, 1993). Langer (1989) described this concept as a form of premature cognitive commitment. It is an attitude based on prior exposure and reinforced when following encounters are similar to the first (i.e., routine, repetition). Once such attitudes are formed, they become the basis for future actions, often without further thought or analysis. The Complacency Potential Rating Scale (Pop & Stearman, in review; Singh, Molloy, & Parasuraman, 1993) has been shown to be effective in distinguishing between complacency potential and more generalized attitudes towards automation. It is used here to assess participants' trait attitude towards automation.

Complacent behavior occurs when factors create conditions that favor inaction (or continued repetitive action) on the part of the operator. Complacent behavior may be expressed in many ways, e.g., failing to follow all steps in set procedures, or overload condition causing the operator to attend to one task while (erroneously) entrusting the less than perfectly reliable automation to carry out another (Parasuraman et al., 1993). Complacent behavior could also be described as a manifestation of inappropriate trust, particularly overtrust. Operator inexperience, high workload, and consistently reliable systems encourage such overtrust, resulting in more complacent behavior (Chen & Barnes, 2010; Lee & See, 2004; Parasuraman, Molloy, & Singh, 1993). Complacent behavior is operationalized here as accepting RoboLeader's route suggestion when it is not correct.

Situation Awareness

Situation awareness (SA) was first conceptualized during World War I (Gilson, 1995). It has been a contentious topic ever since, as it tends to be a highly subjective construct which researchers have yet to agree on how to define and operationalize (see Dekker, Hummerdal, & Smith, 2010; Flach, 1995; Gilson, 1995; Sarter & Woods, 1991; Smith & Hancock, 1995). In spite of this debate, there does appear to be some consensus that SA is a useful construct that appears to hold utility for both researchers and designers (Gilson, 1995; Parasuraman, Sheridan, & Wickens, 2008; Stanton, Chambers, & Piggott, 2001). Three predominant theories of situation awareness appear in the literature. Respectively, these are a reflective process-driven model based on Russian psychology (Bedny & Meister, 1999), an embedded world model (Smith & Hancock, 1995), and a three-level model for assessing SA (Endsley, 1995). These models describe different aspects of SA, and while none of the theories alone operationalize and quantify SA completely, the combination of the three appears to address all of the various aspects of SA (i.e., individual cognitive processes, interaction with the environment, and final output/assessment). While the two former models focus on the process of acquiring and maintaining SA, the latter focusses on SA assessment (Stanton, Chambers, & Piggott, 2001).

The reflective process-driven model proposed by Bedny and Meister (1999) focuses on internal mental activity and how the cognitive functions interact to form an understanding of situations and events. They refer to the overall process as 'operative reflection', and describe it as *"[Operative reflection] provides dynamic orientation in a* *situation, the opportunity to reflect not only on the present but the past and future, as well as not only actual but potential features of situations. This dynamic reflection contains logical-conceptual, imaginative, conscious and unconscious components. Based on these, individuals developmental models of external events" (p. 71).* This is not a model of SA per se, but activity theory (Bedny & Meister, 1999). The authors proposed that SA is included in the model, distributed over several 'blocks' and created via the interaction of these blocks, and they contend that is a more comprehensive manner in which to address SA than the other theories. However, there appear to be a number of intuitive leaps embedded in their model that are not addressed. Thus the blocks they say constitute SA are not connected in the model, so how these blocks interact remains unspecified. This theory does not offer a viable construct of SA, so as such its utility as an experimental tool is limited.

The embedded world theory of SA proposed by Smith and Hancock (1995) emphasizes the dynamic nature of the ongoing interaction between an individual and their environment, and defines SA as *"adaptive, externally directed consciousness."* This is later explained in more detail as *"[SA is] the invariant [that] codifies the information that the environment may make available, the knowledge the agent requires to assess that information, and the action the knowledge will direct the agent to take to attain its goals"* (Smith & Hancock, 1995, p. 141). Here, SA is an ever-changing emergent property of the interaction between the environment and an individual's consciousness rather than a measurable artifice of an individual's consciousness. While this definition and description of SA captures the essence of what SA is, it has limited utility as an experimental construct, mainly due to its breadth. Interestingly, this theory appears to compliment and complete the product model of SA (Endsley, 1995).

The product model of SA takes care to distinguish state situation awareness from the process of acquiring SA, thus also distinguishing itself from the former theories. Here situation awareness (SA) is defined as "*the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future"* (Endsley, 1988, 1995). This model describes SA as something contained within the individual, separate from yet influenced by individual differences, as well as a function of system design (environment) (see Hancock & Diaz, 2002). This model has been operationalized into 'levels,' with each level distinct from the others, yet having a cumulative nature (e.g., in that level 3 SA cannot be attained without first achieving level 1 SA). However, to what degree each level must be achieved before the next is attainable is not well defined, and this may well be variable dependent on task complexity. This model of SA is used here to quantify how well access to information and reasoning support the participant during mission completion.

Although we attempt to assess SA at a single point in time, SA is not acquired instantly but developed over time (Endsley, 1995). Time is often a critical aspect of SA, both in understanding when an event will occur in the future as well as assessing how relevant information is to the current state. Time is particularly impactful on levels 2 and 3 SA (see Endsley, 1995) as these incorporate understanding of the past to present state awareness for comprehension and projection of future states. Temporal understanding can be critical in dynamic environments, where the operator may have to change strategies in order to maintain situation awareness (Endsley, 1988; 1995). Here, knowledge of when a piece of information was received is manipulated in order to introduce uncertainty regarding the relevance of that information. I hypothesize that such uncertainty will be negatively related to the operator's SA.

At each increasing level of automation, the operator becomes more removed from the inner loop of control as their role changes from actor to supervisor. This distance eventually creates an 'out-of-the-loop' condition that leads to reduced operator SA (Chen & Barnes, 2010; Chen & Joyner, 2009; Endsley, 1995; Parasuraman et al., 1993). To avoid this loss of SA, an intermediate level of automation has been recommended to keep the operator engaged. Endsley and Kiris (1995) found this to be partially effective, whereas as the level of automation increased operators' level 1 SA improved, but their level 2 SA did not. This finding indicated the increase in the level of automation encourages a change from active engagement with the automation to a more passive engagement, which can result in reduced understanding. This threatens task effectiveness when comprehension and problem-solving are crucial. In this work, the agent's level of automation is kept at an intermediate level of autonomy, in order to control the effects of information and reasoning with varying automation influences.

Autonomy

Unlike automated systems, which follows scripts in which all possible courses of action have already been determined, autonomous systems exercise a degree of choice regarding their actions. They do this using information gathered rather than relying exclusively on information supplied at the design stage (Russell & Norvig, 2003). The

Department of Defense (DoD) defines autonomy as "*a capability (or a set of capabilities) that enables a particular action of a system to be automatic or, within programmed boundaries, 'self-governing'*" (Murphy & Shields, 2012). This definition advances the terms, 'automatic' as a trait of an automatic (scripted) system, whereas 'self-governing' is necessarily a trait of an autonomous system. Such confusion is commonplace in discussions of automatic and autonomous systems. To disambiguate such terms, we look to Parasuraman et al.'s (2000) model, which defines automation in regards to two particular aspects of human information processing (see also Manzey, Reichenbach, & Onnasch, 2012). First, how thoroughly the automation supports the four stages of human information processing (information acquisition, information analysis, decision and action selection, and action implementation, see Figure 2). Secondly, how involved the human is in the information processing (and subsequent action was taken).

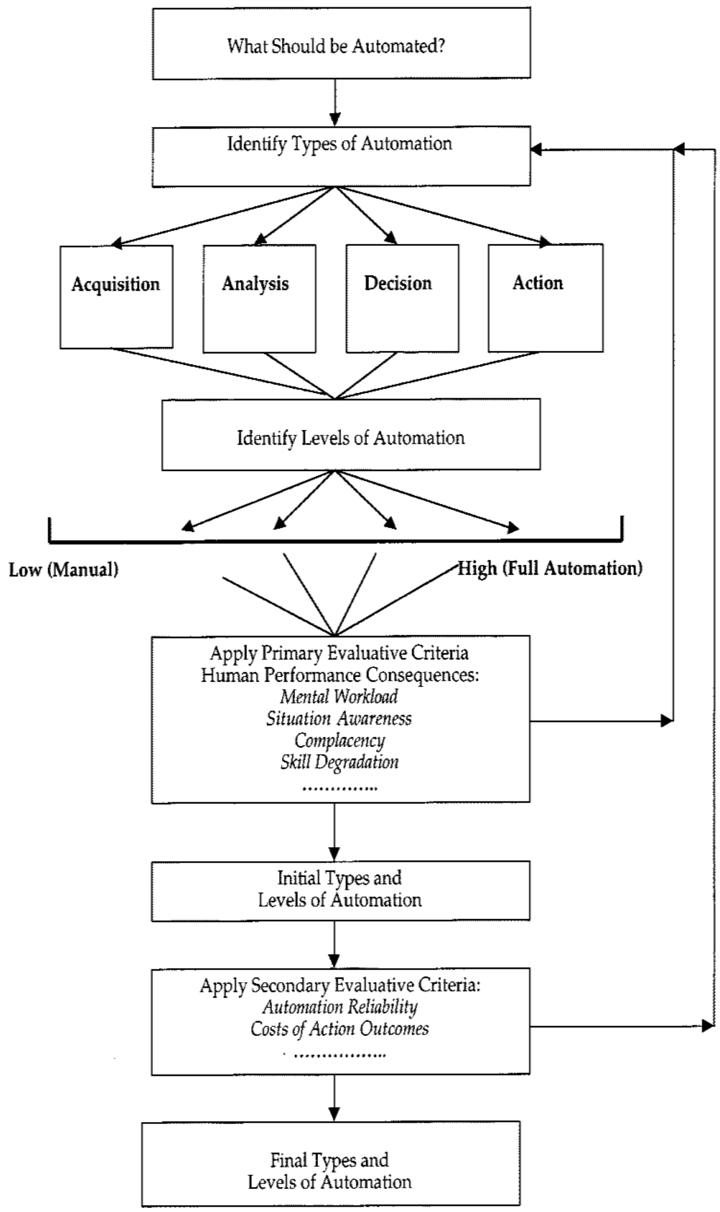


Figure 2. Flowchart depicting the application of Parasuraman et al.'s (2000) automation model. For each of the four stages of human information processing (information acquisition, information analysis, decision and action selection, and action implementation) a level of automation is selected, ranging from low to high. Subsequent evaluations determine if the assignment was appropriate, and allow for adjustments as needed.

The first aspect is assessed within each level of automation (Table 1). This ranges from simple 'detect and react' scenarios to more advanced 'analyze inputs, select appropriate action, and execute selected action' decisions. The second aspect is delineated by each successive level of automation; system autonomy is increasing while human involvement is decreasing until a point is reached where the system even decides whether to inform the human as to its actions. As such, the levels of automation encompass autonomy, particularly in levels 5 (concurrence) and higher, as these levels incorporate a dynamic, self-governing aspect to automation's behavior. The focus here is on the decision aspect of autonomy; specifically, the shared decision space between the human operator and the autonomous agent. Consequently, the present focus is on level 5, or concurrence automation.

Agents

What is an agent?

According to the Merriam-Webster dictionary, an agent is simply '*1. One who acts or can act; 2. One that acts or exerts power on the behalf of another*' (Agent, n.d.). According to the first definition, an agent acts. Whether this action is self-directed or at the behest of another is not addressed. In the second definition, an agent is not necessarily independent, but rather a respondent. The agent acts in the place of, or at the direction of, another. The agent acts, but does not necessarily understand when there is a need to take action, or decide what action to take, or instigate their action, or even assess the result of actions once they are completed. One example of such an agent would be the proxies manning phones at an auction. They act solely at the direction of the anonymous bidder.

The proxy does not necessarily understand if the bids are wise, or choose the amount of the next bid, or know the value of the item being bid upon, or understand the overall impact of the purchase on the bidder's finances. Such an agent is not an independent actor, but rather an interface for another. Chignell, Hancock, and Loewenthal (1989) defined an intelligent interface as *"an intelligent entity meditating between two or more interacting agents who possess an incomplete understanding of each other's knowledge and/or form of communication" (p.2)*. From this example, many agents could be viewed as forms of intelligent interfaces.

The definition of agent used in the computer/artificial intelligence realm is somewhat different. An agent is capable of perceiving its environment through sensors (e.g., eyes, ears, cameras, proximity switches), and of affecting its environment through actuators (e.g., hands, motors) (Russell & Norvig, 2003). This definition does not address the agent's independence (or lack thereof). However, one thing is made clear by these definitions of agents, that is, agency does not equate autonomy.

Russell and Norvig (2003) introduced the idea of independence in their intelligent agent definition by including the idea that an agent interacts with their environment; they sense their environment and then act upon said environment. An intelligent agent can be human, robot, or even a disembodied entity, such as a software computer program, so long as it is capable of detecting the environment through some sort of input (e.g., hands, eyes, sensors, network packets) and then affecting the environment through some kind of output or actuator (e.g., hands, actuators, information display, network packets). Not only can these intelligent agents be independent, but they can also be rational. That is, they interact with their environment in order to achieve a specific goal and measure their success according to specific performance criteria.

Autonomous Agents

Autonomous agents create general knowledge about their environment based on their sensory inputs and experience. The more an autonomous agent learns, the more the agent can rely on this subsequent experience and knowledge to form decisions rather than its original programming (Russell & Norvig, 2003). Franklin and Graesser (1997) defined an autonomous agent as *"a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future" (p.4).* An autonomous system can operate independently of a human operator; updating its work objectives as environmental circumstances change. Also, it can anticipate and deliberate upon outcomes, and execute subsequent actions as required. It has everything it needs to carry out its directive successfully, except for an 'operating force' (Schulte, Meitinger, & Onken, 2009), which is the only part of the work system that pursues the complete work objective, and is the role of the human supervisor. The operating force is the highest authority in the work system and the part of the system capable of defining the work objective. While an autonomous system can modify the work objective if needed, it cannot define the initial work objective. As in the Clockwork Universe Theory (Descartes, 1649); while the clock, once started, may run perfectly, it cannot start itself. Autonomous agents begin their interaction with the environment with pre-programmed knowledge concerning their work objectives. However as they gain experience with their environment that knowledge is updated and revised. The more the agent relies on experience and learned information rather than prior programming to achieve successful performance, the more autonomous that agent becomes. Regardless of the level of autonomy the agent achieves, human interaction will be required at some point, if only in the beginning. While agents working autonomously to achieve human objectives is a future idea, what is currently commonplace is not agents working independently of humans, but humans working alongside agents.

Human-Agent Teaming

Humans Supervising Teams

There are many examples of humans supervising teams of robotic entities, both in the military and civilian arenas. Military battlefield applications include casualty extraction, IED detection/disposal, reconnaissance, and surveillance. Civilian applications include search and rescue, firefighting, and space exploration (Chen & Barnes, 2014). In these complex and dynamic applications, the current state of technology requires a many-to-one supervision model, with multiple human operators overseeing a single robotic entity. Such team numbers increase as the complexity of the robot's tasks and environment increase (Murphy & Burke, 2010). As operator team size grows it tends to become unwieldy, so the development of autonomous systems that can assist the human operators and eventually reduce the number of these human operators becomes necessary. Hence, the move to a one-to-many model is encouraged. Development of systems, such as the mixed-initiative system that can assist human operators to oversee teams of robots is the first step towards achieving this goal.

Mixed-Initiative Systems

Mixed-initiative systems incorporate elements of both adaptive (level of automation is changeable by the system; Parasuraman et al., 2000) and adjustable (level of automation is changeable by an external operator or system; Bradshaw et al., 2003) automation. This allows the human and an agent to work in concert, each with authority to make decisions (Goodrich, 2010). Although both have the ability to make changes, the human operator is the ultimate authority in mixed-initiative systems. Mixed-initiative systems are effective in keeping the human supervisor in the loop, reducing operator workload, and thus increasing the number of subordinate robots that the operator can direct (Barnes et al., 2014; Chen & Barnes, 2010; Chen & Barnes, 2014). While mixed-initiative systems have been lauded as the most flexible system for supervisory control (Calhoun, Ruff, Draper, & Wright, 2011), these systems are also particularly susceptible to mode confusion (Goodrich, 2010; Sarter, 2008). Mode confusion is when the operator believes the automation is in a different mode than it currently is, and as a result, their responses to the automation prove inappropriate (Joshi, Miller, & Heimdahl, 2003). An example of a mixed-initiative system developed by the DOD to investigate such human-agent teaming issues is RoboLeader.

RoboLeader

An intelligent agent, RoboLeader, was developed to simplify interactions between a human supervisor and a robotic team (Chen, Barnes, & Qu, 2010). The human supervisor interacts with the RoboLeader, which interprets the supervisor’s goals and then commands a team of lower capability robots through route planning and convoy management. This allows the human to focus on high-level decisions regarding convoy management, freeing their attention for other tasks such as maintaining security and communications.

RoboLeader Findings

The addition of an intelligent agent to manage the robotic team brings unique problems. While the operator benefits from reduced workload, findings indicate they did not always improve performance and SA. Chen, Barnes and Qu (2010) found no difference in target detection rate and accuracy between the Baseline and RoboLeader-aided conditions, although there was an improvement in mission completion time. Similar findings were reported by Wright and colleagues (2013), such that increasing RoboLeader's level of assistance (LOA) did not always improve SA or task performance. Indeed, in some cases (i.e., high spatial ability individuals), performance in the highest LOA decreased. Effectively conveying information to the supervising operator in a manner that allows them to assimilate the information and stay engaged in their supervisory task becomes challenging when the agent is handling multiple complex tasks (Kilgore & Voshell, 2014). Transparency of the agent's intent and reasoning may encourage the operator to stay involved and in-the-loop. The effects of such a manipulation have yet to be tested.

Trust in Automation

The amount a user trusts an autonomous system directly affects their willingness to use it, as well as their performance and how they respond to unexpected scenarios (Lee & See, 2004). The higher the level of autonomy of the system, the more important the level of information the system supplies becomes in fostering trust in the human operator (Wang, Jamieson, & Hollands, 2009). If the information is not presented in a manner familiar to an operator, this reduces automation transparency (Kim & Hinds, 2006). The present dissertation investigates how the appropriate level of information and the preferred manner in which the information is displayed affects performance and trust in the route-planning agent.

There are two major types of trust: dispositional trust and history-based trust (Merritt & Ilgen, 2008). Dispositional trust is a stable trait describing someone's feelings about something before any actual encounter. In the present instance, such trust refers to how one feels about working with a remote monitoring and communications system. Dispositional trust is generated by exposure to a wide variety of sources, primarily social influences such as media and literature, and as such it can vary widely between individuals (Hancock, Billings, & Schaefer, 2011; Schaeffer, Hancock, & Chen, 2015). Dispositional trust in automation varies along many demographic divides (e.g., age, ethnicity, gender, education). However, studies exploring this have had only mixed results (Ho et al., 2005, Merrit & Ilgen, 2008). The present dissertation does not include assessments of dispositional trust, as it has not been shown to be predictive of performance in prior intelligent agent studies using university student participants (Mercado et al., in press; Wright et al., 2013). However, the relationship between dispositional trust in automation and task performance still remains to be explained.

In contrast to dispositional trust, history-based trust is developed from direct interaction with systems and is composed of multiple factors. These can include a human's abilities and effectiveness with the system, the system's behavior, and reliability, as well as environmental factors such as risk and uncertainty (Hancock et al., 2011; Masters, Miles, D'Souze, & Orr, 2004; Schaefer et al., 2014). As an individual's experience working with a particular system grows, they calibrate their trust to an appropriate level (Fallon et al., 2010). When improperly calibrated, this trust is expressed as either insufficient, where the operator does not rely on the automation sufficiently, versus excessive, where the operator relies too much on the automation (Lee & See, 2004; Parasuraman, Molloy, & Singh, 1993). The present research explores the effect of access to agent reasoning on history-based trust. History-based trust is evaluated using the Usability and Trust Survey (Appendix I; Chen & Barnes, 2012a). Objectively, history-based trust is operationalized as the time to accept or reject RL route selections, with higher trust being reflected in shorter selection times.

Transparency and Level of Reasoning

To be transparent means to be easy to be perceived or detected (Transparent, n.d.). Within the human-automation research community, there is presently no consensus as to exactly how transparency should be defined. Transparency has been described both as something the automation provides (whether by design or behavior) (Cramer et al., 2008; Cuevas, Fiore, Caldwell, & Strater, 2007; Kim & Hinds, 2006), as well as the understanding or knowledge an operator has regarding the systems behavior (Cheverst et al., 2005; Cring & Lenfestey, 2009; Jameson, Baldes, Bauer, & Kroner 2004). When referring to automation or automated systems, early constructs of transparency focused on explaining the system's behavior in an effort to foster trust. When users do not understand the rationale behind a system's recommendations, they begin to question the accuracy and effectiveness of that system (Linegang et al., 2006). As the users' understanding of the rationale behind a systems' behavior grows, the more accurate the users' calibration of their trust and reliance (Lee & See, 2004; Lyons, 2013; Mercado et al., under review). The more autonomous that a system becomes, the more important transparency becomes as a factor in user understanding and trust (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Kim & Hinds, 2006). A recent definition of agent transparency, *"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"* (Chen, Procci, Boyce, Wright, Garcia, & Barnes, 2014) expands on earlier constructs by extending the idea of agent transparency beyond simply explaining the agents' behavior and fostering user trust, but also facilitating the operator's comprehension and SA.

SAT Model

The SA-based Agent Transparency model (SAT; Chen et al., 2014; Figure 4) describes knowledge of what is happening in the environment and the agent's goals as supporting the operators' Level 1 SA (i.e., what is the agent trying to do), understanding the agent's reasoning process as supporting the operators' Level 2 SA (i.e., why does the agent do it), and providing future projections, likelihood of success, and uncertainty information as supporting the operators' Level 3 SA (i.e., what should happen; Endsley, 1995). When the operator has knowledge of the agents' intent, and understands the agents' reasoning, as well as anticipating likely outcomes based on the information, the operator can accurately calibrate their trust level (Lee & See, 2004). This is particularly important in evolving environments, where the operator's goals may not always coincide with the agents' goals (Linegang et al., 2006).

Figure 3. The SAT Model, illustrating how agent transparency is defined at each level (Chen et al., 2014).

When environmental information or the agent's reasoning is not available to the operator, the operator has no motivation to participate in decision-making, thus encouraging a human-out-of-the-loop isolation (Parasuraman, Sheridan, & Wickens, 2000; Wickens, 1994). This can be mistaken for automation-induced complacency (Parasuraman, Molloy, & Singh, 1993). The present dissertation investigated how the operator's knowledge of the current state of the environment, access to agent reasoning, and uncertainty affects decision-making ability, as measured via the route selection task. Research has indicated that the addition of information concerning uncertainty at a high level of agent transparency can improve operator performance in a decision-making task (Mercado et al., under review). However, that study used a single task in a static environment. The notion that excessive 'transparency' could result in the opposite effect than that originally intended (i.e., to enhance the human operators' performance and situation awareness while reducing cognitive workload) is a concern that is examined (Miller, 2014; Ososky, Sanders, Jentsch, Hancock, & Chen, 2014). The present dissertation explored this position as it relates to the addition of information to convey uncertainty. It is hypothesized that the addition of such information in a dynamic, multitasking environment will have a detrimental effect on task performance. Such effects do not impact all operators equally, so several individual difference factors relating to task performance will also be examined.

The Role of Individual Differences in Human-Agent Teaming

Within human-automation interaction and human-agent teaming research, several individual difference factors have been discussed as being impactful on operator performance. Such research has indicated that people with higher perceived attentional control (PAC) are more efficient in allocating attention, and are less susceptible to performance degradation in a multitasking environment than those with low PAC (Chen & Joyner, 2009; Derryberry & Reed, 2002; Rubinstein, Meyer, & Evans, 2001). There are also differential effects due to inherent spatial ability (SpA). These have been found on teleoperation tasks, robotic operations, and target detection tasks (Chen et al., 2010; Chen et al., 2008; Lathan & Tracey, 2002). Differences in working memory capacity (WMC) have also been shown to affect performance in multi-robot supervision tasks (Ahmed et al., 2013). Here I examine the differential effects of PAC, SpA, and WMC on multitasking performance, operator SA, and perceived workload. Complacency Potential (Singh, Molloy, & Parasuraman, 1993) affects an individual's ability to monitor automation adequately and to detect automation failure. Thus, complacency potential is also examined as a mediating factor in route selection. I examine each of these factors (PAC, SpA, and WMC) and how they influence performance in this work.

Attentional Control

Attentional Control, also known as endogenous attention, represents a person's ability to control what they attend to and what they ignore (Posner, 1980; Posner & Petersen, 1989). People with higher attentional control are more effective at switching between tasks and focusing attention than persons with lower attentional control. They are also better at threat disengagement, which is returning their attention to less threatening stimuli after diverting it to a threat (Derryberry & Reed, 2002). Previous RoboLeader studies have found links between individual differences in such attentional control, and system reliability, and associated cognitive workload. For example, in a simulated gunnery task using an aided target recognition software system, reliability was found to have interactive effects with attentional control (Chen & Terrence, 2009). False alarm prone (FAP) alerts negatively impacted those with high perceived attentional control (PAC) individual's more than miss prone (MP) alerts. However, low PAC individuals were more negatively affected by the miss prone alerts than the false alarm prone alerts. This illustrated differences in attentional control: high PAC individuals were able to switch their attention more readily, so the FAP alert encouraged them to focus needlessly on the area and search for the (reported) target. However, low PAC individuals were not able to switch their attention as readily and so were more dependent on alerts in general. As such, the performance of the latter group was more affected by the MP alerts than the FAP alerts.

Individual differences in workload (as indicated by pupil diameter) were also attributed to PAC (Wright et al., 2013). For high PAC individuals', workload steadily decreased as the level of assistance (LOA) increased. Low PAC individuals showed no such difference in workload across LOA conditions. Consistent workload, regardless of LOA, is only beneficial if the workload is relatively low; in this case, the low PAC individuals had very high workload in all LOA. As such, the low PAC individuals gained no benefit from increasing automated assistance.

In the present work, persons with high PAC are expected to make better use of additional information than their low PAC peers. As the level of information increases, the performance of high PAC individuals' on the route planning task is hypothesized to improve, while the low PAC individuals are expected not to benefit to the same degree from the additional information.

Spatial Ability

Spatial ability represents the capacity to navigate or manipulate objects in a three-dimensional environment (Eliot & Stumpf, 1987). Spatial ability is a basic dimension of human intelligence and comprises a domain of abilities rather than a single skill (Lathan & Tracey, 2002). Spatial ability correlates highly with general intelligence (Lohman, 1996). However, it has not been found to be predictive of overall academic performance. Rather, it appears to be predictive in several creative and task-specific domains, such as higher-order mathematics and engineering aptitude (Lohman, 1996). Spatial ability in the form of mental rotation has also been found to be a mediator of performance on spatial working memory tasks (Christie, Cook, Ward, Tata, Sutherland, Sutherland, & Saucier, 2013).

Physiological as well as performance-based evidence shows that object-based (visualization) and egocentric perspective-based measures of spatial abilities rely on different processing systems (Zacks et al., 1999). How many distinct spatial processing factors exist and how each should be characterized remains unclear (Lathan & Tracey, 2002). As understanding further develops as to the spatial ability construct and what abilities fall under this domain, taxonomies of its factors and sub-factors will assumedly continue to evolve (Carroll, 1993; Lohman, 1988).

Spatial visualization (SV) is the ability to manipulate visual patterns (Carroll, 1993), or to rotate objects mentally in space (Hegarty & Waller, 2005; Lohman, 1988). Evaluations of SV do not incorporate speed of manipulation but do include difficulty in the complexity of the manipulation (i.e., rotating, twisting, inverting). The caveat of SV is that these manipulations are "in space", i.e., without reference to a framed reference. This factor, as evaluated using the Cube Comparison Test (Ekstrom et al., 1976; Thurstone, 1951), has been found to be predictive of performance on target detection tasks (Chen & Joyner, 2009; Chen & Terrance, 2009; Fincannon, 2013).

Spatial Orientation (SO) is the ability to image objects from different perspectives (Kozhevnikov & Hegarty, 2001; Lohman, 1988), whether from an egocentric or an exocentric perspective (Gugerty & Brooks, 2004). This factor, evaluated using the Spatial Orientation Test (Gugerty & Brooks, 2004) has been found to have differential influences on target detection capabilities, as well as operators' situation awareness (Fincannon, 2013; Wright et al., 2013). The Spatial Orientation Test requires participants to conduct a navigation task, coordinating their egocentric view of the world with an external framed reference. In such environments, participants have forward and rearward views from their vehicle, and need to coordinate that information with what is displayed on a provided map. Successful integration of this information is expected to lead to improved SA scores and better performance on route planning tasks.

Working Memory Capacity

Working Memory (WM) refers to a part of the memory system responsible for comprehension, reasoning, planning, and implementing behaviors (Cowan, 2008). In Baddeley's (2000) model of working memory, there are four components; a visuospatial sketchpad (analog/spatial memory), a phonological loop (linguistic memory), the central executive (attentional control), and the episodic buffer, a "*limited capacity temporary storage system that is capable of integrating information from a variety of sources*" (*p. 421).* In general, the capacity of working memory is limited to seven ±2 items (Miller, 1956). Working memory capacity (WMC) is not only limited in the number of items that can be retained but time sensitive also (Melton, 1963). The more items to be remembered, the shorter the duration those items will stay in working memory.

Research has indicated that working memory capacity is correlated with an individual's attentional control (Engle, Kane, & Tuholski, 1999), in that high WM individuals allocate their attention differently than low WM individuals (Bleckley, Durso, Crutchfield, Engle, & Khanna, 2003). Here working memory capacity is evaluated as a covariate for assessing individual differences in performance due to PAC and SpA, since previous studies have indicated WMC correlates with these factors (Engle et al., 1999).

Working memory is directly related to an individual's situation awareness, particularly level 2 (comprehension) and level 3 (projecting future states), and in multitasking environments where multiple tasks compete for limited resources (Endsley, 1995; Wickens & Holland, 2000). Here, WMC is treated as an individual difference factor when evaluating performance differences on SA measures

Eye Tracking Measures to Consider

It has been asserted that underlying cognitive activities can be reliably inferred from eye tracking metrics (Beatty, 1982; Jacob & Karn, 2003). In an earlier RoboLeader study (Wright et al., 2013), eye-tracking metrics proved useful in evaluating differences in workload that subjective measures of workload did not reveal. This work incorporates two visual measures, 1) fixation count and 2) pupil diameter, as objective measures of cognitive workload.

Fixation Count (FixC)

When an individual focuses on a specific location, it is known as a 'fixation.' During fixations, the eye is not completely still but makes low-velocity movements, such as drifts and adjustments to maintain focus. Commercial eye tracking systems typically detect fixations using dispersion algorithms, identifying data points close enough together over a specified period of time as a fixation (see Figure 4). Such fixations typically last between 200-300 ms, but may last up to several seconds (Holmqvist et al., 2011). The number of fixations has been shown to correlate positively with search difficulty (Ehmke & Wilson, 2007) and encoding memory (Tatler, Gilchrist, & Land, 2005). Fixation count has been correlated negatively with search efficiency and mental workload (Goldberg &

Kotval, 1999; Van Orden, Jung, & Makeig, 2000). Fixation Count (FixC) is useful both as a within-participant and between-participants measure.

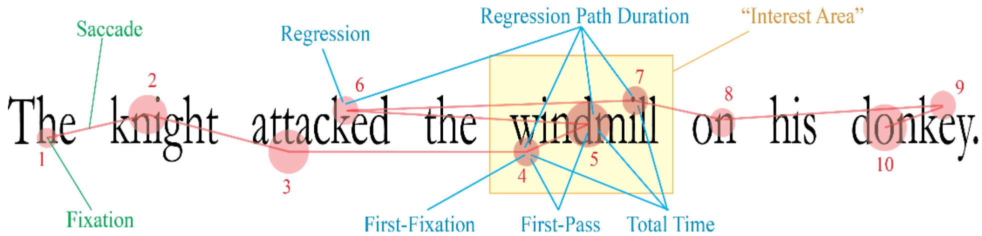


Figure 4. Illustration of common eye tracking metrics. The length of time spent in an area of a specified size is denoted by colored circles (longer time in the area indicated by larger circles). These circles denote fixations.

Pupil Diameter (PDia)

The size of the pupil opening is a result of changes in the iris. This is a function controlled by the autonomic nervous system and sensitive to both external (e.g., ambient light, distance to stimulus, viewing angle) and internal (e.g., emotion) factors. Pupil diameter is measured by imposing an ellipse over the pupil area and measuring the length of the vertical and horizontal axes. Of the two axes, the horizontal has proved to be less sensitive to artifacts due to the eyelid occlusion/closure (Holmqvist et al., 2011). The diameter of the pupil increases as mental workload and interest increase (Beatty, 1980; Iqbal, Zheng, & Bailey, 2004; Kang et al., 2009; Peavler, 1974; Van Orden et al., 2000; Van Orden, Limbert, Makeig, & Jung, 2001). Pupil Diameter (PDia) is useful both as a within-participant and between-participants measure.

Research Objective

The present research investigates how agent transparency, within the context of human-agent teaming, influences operator performance and behavior in a dynamic, multitasking environment. In two experiments, the effect of increased access to agent reasoning is evaluated within two contexts. Experiment one is a low environmental information environment. Experiment two concerns a high information environment. Each experiment had participants' complete three missions at a specific level of agent transparency. Results were compared between subjects to evaluate how the difference in transparency affected operator performance, workload, trust, SA, and complacent behavior. The two experiments' findings were compared to evaluate how differences in available information affected operators' preferred level of transparency.

The present results are expected to elucidate how the operators' knowledge of the environment interacts with their understanding of agent reasoning to create 'transparency,' as well as how increased access to the reasoning behind automation 'decisions' effects a human operators' ability to interact effectively with said automation. Too little transparency may hinder human trust in the automation. However, too much may have similarly detrimental effects on operator performance, situation awareness, and decision-making, thus encouraging complacent behavior. In addition, this work investigated how several individual difference factors of common interest within the human-automation interaction community influence the human-agent relationship in terms of agent transparency, and the subsequent effect on the related human performance issues.

Primary Hypotheses:

Based on the review of the literature, this proposal posits the following hypotheses:

1. Overall, increased access to agent reasoning will improve task performance and operator SA, and reduce complacent behavior.

2. However, this improvement will be mitigated by increasing the level of information the operator has available about the task environment.
3. Increasing the transparency of the agent's reasoning to include information that could create uncertainty for the operator will improve performance when information is sparse, but degrade performance when the operator has more information about the task environment.

Table 2. The anticipated pattern of findings. An up arrow (↑) indicates an increase in performance on the measure from the next lower level condition, while a down arrow (↓) indicates a decrease.

DV Measure	EXP 1 ART 2 → ART 1	EXP 1 ART 3 → ART 2	EXP 2 ART 2 → ART 1	EXP 2 ART 3 → ART 2	Comparison: EXP 2 → EXP 1 ART 1	Comparison: EXP 2 → EXP 1 ART 2	Comparison: EXP 2 → EXP 1 ART 3
Route Selection Task (RS)							
Correct accepts and rejects	↑	↓	↑	↓	↓	↑	↓
Decision Time	↑	↓	↑	↑	↑	↑	↑
Target Detection Task (TD)							
Targets Detected	↓	↓	↓	↓	↓	↓	↓
False Alarms	↓	↓	↑	↑	↑	↑	↑
Complacent Behavior							
Incorrect Acceptances	↓	↑	↓	↑	↓	↓	↑
Situation Awareness Scores							
SA Level 1 Queries (Perception)	↑	↑	↑	↓	↓	↓	↓
SA Level 2 Queries (Comprehension)	↑	↑	↑	↑	↑	↑	↓
SA Level 3 Queries (Projection)	↑	↓	↑	↓	↓	↓	↓
Workload (Global NASA-TLX)	↑	↑	↑	↑	↑	↑	↑
Operator Trust	↑	↓	↑	↓	↑	↑	↓

CHAPTER 3: EXPERIMENT 1

Study Overview

Experiment 1 investigated how access to agent reasoning affected the human operator's decision-making, task performance, situation awareness, and complacent behavior in a multitasking environment when limited environmental information is available. The participants' role was to supervise a convoy of vehicles as it progressed through an urban environment, maintaining communications with their command and identifying potential threats along the way. They were provided with a map of the area with a predetermined route marked. Icons referring to events in the area appeared on the map, some of which affected the convoy's route. Information received from command could contradict or update the information provided on the map. When approaching such an area, RoboLeader suggested altering the route, and the participant either accepted or rejected the suggestion. The level of the agent reasoning transparency (ART) behind RoboLeader's (RL) recommendation as to the appropriate route to continue the convoy's progress was manipulated between participants, varying from simple notifications to reports including recency of report. Each participant completed three missions at a specific ART. As the convoy progressed through the simulated environment, the participants' maintained communication with 'command'; receiving incoming messages and responding when appropriate (SA probes). While overseeing the convoy's progress, the participants also conducted a target detection task by monitoring the vehicles' camera feed and identifying potential threats in their environment. The number of threats was held constant across routes.

Hypotheses:

Based on the review of the literature, this proposal posits the following hypotheses:

It was hypothesized that access to agent reasoning would reduce complacent behavior, improve task performance, and increase trust in the agent—but only to a degree, beyond which increased access to agent reasoning would negatively impact performance, increase complacent behavior, and reduce trust in the agent (i.e., ART1 < ART2 > ART3). This hypothesis recapitulates an inverted [extended] U-shaped function often observed in operators in stressful conditions (Hancock and Warm, 1989; Yerkes and Dodson, 1908). Decision time was also examined as a facet of performance, and as such was expected to increase as access to agent reasoning increased: ART1 < ART2 < ART3. Although RL's messages were slightly longer in ARTs 2 and 3 than in ART1, additional time was not expected to be required for reading the messages. Participants were expected to take longer to process the information and reach their decision, resulting in a longer decision time. Shorter response times may indicate less deliberation on the part of the operator before accepting or rejecting the agent recommendation. This could mean either positive automation bias or reduced task difficulty.

> Hypothesis 1: Access to agent reasoning will reduce incorrect acceptances, ART1 > ART2, and increased transparency of agent reasoning will increase incorrect acceptances, ART2 < ART3. When agent reasoning is not available, incorrect acceptances will be greater than when agent reasoning is present, ART1 > ART2+3.

Hypothesis 2: Access to agent reasoning will improve performance (number of correct rejections and acceptances) on the route selection task, ART1 < ART2, and increased transparency of agent reasoning will reduce performance on the route selection task, ART2 > ART3. When agent reasoning is not available, performance will be lower than when agent reasoning is present, ART1 < ART2+3.

Hypothesis 3: Access to agent reasoning will increase operator trust in the agent, ART1 < ART2, and increased transparency of agent reasoning will decrease operator trust in the agent, ART2 > ART3.

It is hypothesized that increasing agent reasoning transparency will, in turn, increase the operators' workload. Typically, increased automation assistance reduces operator workload, as the operator is able to offload a portion of their duties to the automation. However, in the case of agent reasoning transparency, the amount of information the operator must process increases as the agent reasoning becomes more transparent. It is expected that this increased mental demand will be reflected in the workload measures.

Hypothesis 4: Access to agent reasoning will increase operator workload, ART1 < ART2; and increased transparency of agent reasoning will increase operator workload, ART2 < ART3. When agent reasoning is not available, workload will be lower than when agent reasoning is present, ART1 < ART2+3.

It is hypothesized that agent reasoning transparency will support the operators' situation awareness (SA). Access to the agent reasoning will help the operator better comprehend how objects/events in the task environment affect their mission, thus informing their task of monitoring the environment surrounding the convoy and making them cognizant of potential risks. This understanding will also enable them to make more accurate projections regarding the future safety of their convoy. However, the addition of information that appears ambiguous to the operator will have a detrimental effect on their ability to correctly project future status.

> Hypothesis 5: Access to agent reasoning will improve SA scores, and increased transparency of agent reasoning will improve SA1 and SA2 scores, but will reduce SA3 scores;
>
> - SA1: ART1 < ART2, ART2 < ART3;
> - SA2: ART1 < ART2, ART2 < ART3;
> - SA3: ART1 < ART2, ART2 > ART3.

It is hypothesized that increasing agent reasoning transparency will reduce performance on the target detection task. The increased mental demand on the operator will affect their ability to monitor the environment for threats effectively. However, access to agent reasoning will allow operators' to maintain higher selection criteria, resulting in fewer FAs.

> Hypothesis 6: Access to agent reasoning will reduce the number of targets detected and the number of FAs, ART1 > ART2, and increased

transparency of agent reasoning will again result in fewer targets detected and fewer FAs, ART2 > ART3.

The effects of individual differences in complacency potential, perceived attentional control, spatial ability, and working memory capacity on the operator's task performance, trust, and SA were also investigated.

Hypothesis 7: High CPRS individuals will have fewer correct rejects on the route planning task than Low CPRS individuals.

Hypothesis 8: High CPRS individuals will have higher scores on the usability and trust survey than Low CPRS individuals.

Hypothesis 9: High CPRS individuals will have lower SA scores than Low CPRS individuals.

Hypothesis 10: Individual differences, such as SpA and PAC, will have differential effects on the operator's performance on the route selection task and their ability to maintain SA.

Hypothesis 11: High WMC individuals will have more correct rejects and higher SA2 and SA3 scores than Low WMC individuals.

Table 3. The anticipated patterns of findings (hypotheses) for Experiment 1. Information in the columns indicates the expected score or performance change between agent reasoning transparency conditions (i.e., ART1, ART2, and ART3).

DV Measure		Add Transparency	Increase Transparency
Route Selection Task (RS)	Correct accepts and rejects	ART 1 < ART 2	ART 2 > ART 3
	Decision Time	ART 1 < ART 2	ART 2 > ART 3
Target Detection Task (TD)	Targets Detected	ART 1 > ART 2	ART 2 > ART 3
	False Alarms	ART 1 > ART 2	ART 2 > ART 3
Complacent Behavior	Incorrect Acceptances	ART 1 > ART 2	ART 2 < ART 3
Situation Awareness Scores	SA1 Queries (Perception)	ART 1 < ART 2	ART 2 < ART 3
	SA2 Queries (Comprehension)	ART 1 < ART 2	ART 2 < ART 3
	SA3 Queries (Projection)	ART 1 < ART 2	ART 2 > ART 3
Workload	Global NASA-TLX	ART 1 < ART 2	ART 2 < ART 3
Trust	Incorrect Rejections	ART 1 > ART 2	ART 2 < ART 3
	Usability and Trust Survey	ART 1 < ART 2	ART 2 > ART 3

Task Environment

Simulation Scenario. The human operator supervised a three-vehicle convoy as it proceeded through an urban environment, following a predetermined route, on a reconnaissance mission. An intelligent agent, RoboLeader, managed convoy behavior and route planning. As the convoy progressed, events (e.g., threats present, environmental hazards/obstacles) occurred that might necessitate altering the convoy's route. The agent suggested a potential route revision, and the operator would have to accept or reject the suggestion.

Each vehicle (a UAV, an UGV, and an MGV) was equipped with an indirect camera feed, displaying the environment below (i.e., the UAV), forward of (i.e., the UGV), or surrounding the vehicle (i.e., the MGV). The MGV is equipped with 360° indirect-vision capability, which the U.S. Army is currently developing, and the operator assessed via two 180° camera displays, one forward-view, and one rearward view (see Figure 5). In addition to the convoy supervisory duties, operators were required to maintain local security around the convoy via the vehicles' indirect-vision camera feeds by identifying threats present in the immediate vicinity. Operators monitored communications from command, responding when appropriate.

RoboLeader. To frame the task in current-day capabilities, RoboLeader was dependent on incoming information from a variety of sources upon which to base its recommendations, and as such, its environment was partially observable. It was possible for the operator to have more up-to-date information regarding the environment than RoboLeader, thus creating the justification for overriding RL's suggestions when they were inappropriate. Each route-planning recommendation was independent of the others, making the environment (from RoboLeader's Point Of View) episodic and stochastic.

Methodology

Experimental Participants

Seventy-six participants (between the ages of 18 and 40) were recruited from the Institute for Simulation and Training's and the Psychology Departments' SONA Systems. Participants received their choice of compensation: either cash payment ($15/hr) or Sona Credit at the rate of 1 credit/hour. Sixteen potential participants were excused or dismissed from the study; nine were dismissed early due to equipment malfunctions, one withdrew during training claiming they did not have time to participate, three fell asleep during their session and were dismissed, two could not pass the training assessments and were dismissed, and one did not pass the color vision screening test and was dismissed. Those who were determined to be ineligible or withdrew from the experiment received payment for the amount of time they participated, with a minimum of one hours' pay. Sixty participants (26 males, 33 females, 1 unreported, Min_{age} = 18 years, Max_{age} = 32 years, M_{age} = 21.4 years) successfully completed the experiment, and their data was used in the analysis.

Experimental Apparatus

Simulator. The Mixed Initiative Experimental (MIX) Testbed, used in the Chen et al. (2010) RoboLeader experiment, was modified and used in this experiment (Figure 5). The MIX Testbed is a distributed simulation environment for investigation into how unmanned systems are used and how automation affects human operator performance. This platform includes a camera payload and supports multiple levels of automation. Users can send mission plans or teleoperate the platform with a computer mouse while being provided a video feed from the camera payload. Typical tasks include reconnaissance and surveillance. RoboLeader has the capability of collecting information from subordinate robots with limited autonomy (e.g., with the capability of collision avoidance and self-guidance to reach target locations), making tactical decisions, and coordinating the robots by issuing commands, waypoints, or motion trajectories (Chen et al., 2010).

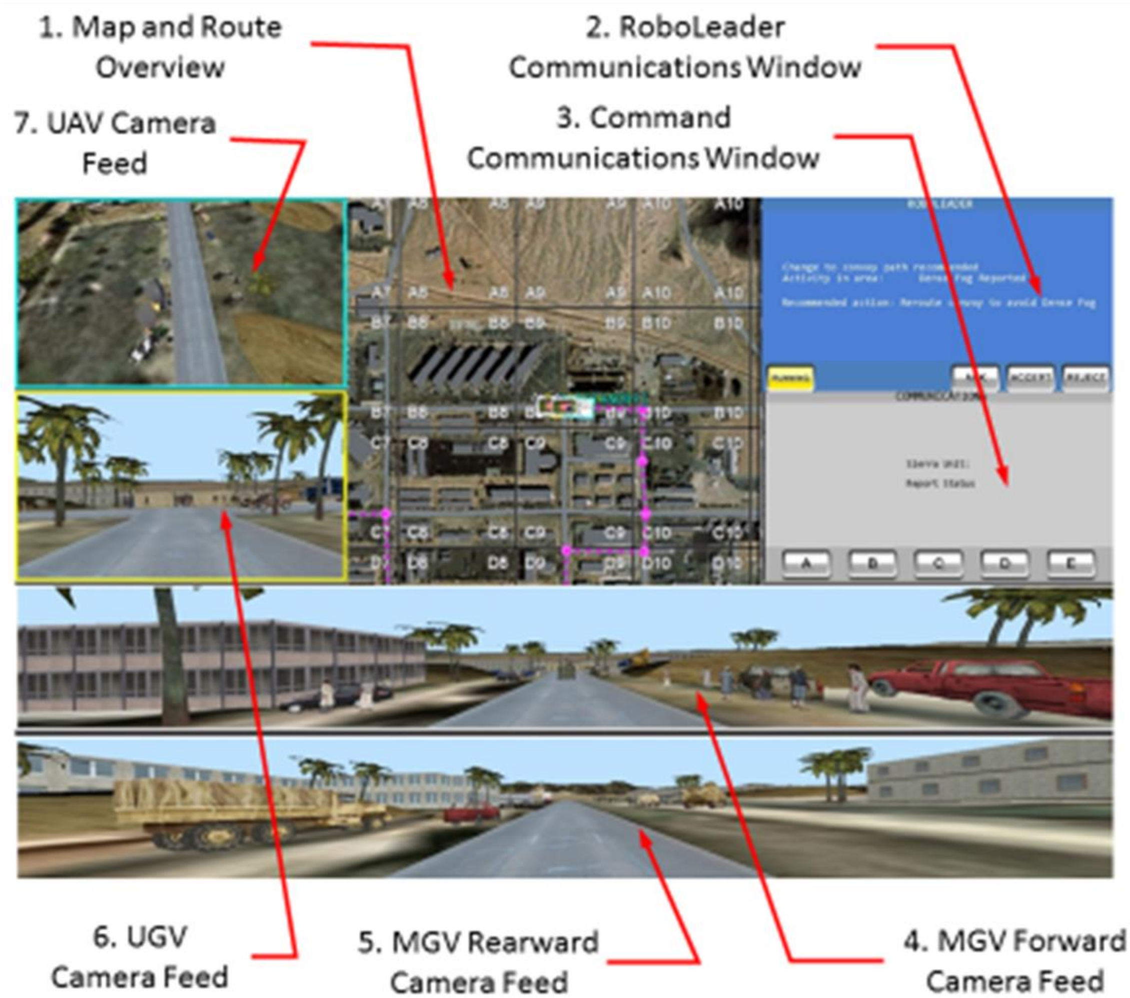


Figure 5. Operator Control Unit: User interface for convoy management and 360° tasking environment. OCU windows are (clockwise from the upper center): 1. Map and Route Overview, 2. RoboLeader Communications Window, 3. Command Communications Window, 4. MGV Forward 180° Camera Feed, 5. MGV Rearward 180° Camera Feed, 6. UGV Forward Camera Feed and 7. UAV Camera Feed.

Eye Tracker. A Sensomotoric Instrument (SMI) Remote Eyetracking Device (RED) was used to collect eye movement data. The SMI RED system uses an infrared (IR) camera-based tracking system and allows completely non-contact operation. Eye and head movements, which can be observed at approximately 0.03º of spatial resolution and sampled at the rate of 120 Hz, along with measurement reliability data, was logged in real time and synchronized with performance data from other systems. Only the participants' eye gaze coordinates were measured and recorded; no video of the participants' eyes and faces were recorded. The system was individually calibrated for each participant prior to each mission.

Surveys and Tests

Demographics. A demographics questionnaire was administered at the beginning of the training session (Appendix B). Information on participant's age, gender, handedness, and video gaming experience was collected.

Ishihara Color Vision Test. An Ishihara Color Vision Test (with 9 test plates) was administered via PowerPoint presentation. Since the RoboLeader OCU employs several colors to display the plans for the robots, normal color vision is required to interact effectively with the system. Participants who incorrectly identified a slide were given the opportunity to try again once; if on their second chance they could not correctly identify the slide that was counted as a miss. Participants who incorrectly identified 2 or more slides were dismissed.

Individual Difference Factors. Descriptive statistics pertaining to individual differences (ID) measure scores are listed here. ID results were dichotomized into

High/Low groups via median split of all scores. These groups were then evaluated within each ART for compliance with required analytic assumptions. Finally, for each ID measure, these groups were assessed across ARTs to ensure that 1) that the high and low groups were distinct from one another, and 2) that high/low group membership was consistent between ARTs.

Perceived Attentional Control. A questionnaire on Attentional Control (Derryberry & Reed, 2002) was used to evaluate participants' perceived attentional control (Appendix C). The Attentional Control survey consists of 21 items, measures attention focus and shifting, and has been shown to have good internal reliability (α = .88). Scoring range is 21 – 84 points, with higher scores indicating greater attentional control. High/Low group membership was determined by median split of all participants' scores (Min_{PAC} = 41.0, Max_{PAC} = 74.0, Mdn_{PAC} = 61.0, M_{PAC} = 60.5, SD_{PAC} = 7.5; PAC_{LOW} N = 28, PAC_{HIGH} N = 32).

Cube Comparison Test. Two aspects of spatial ability were assessed, spatial visualization (SV) and spatial orientation (SOT). The Cube Comparison Test (SV; see Appendix D; Ekstrom, French, & Harman, 1976) measures an individual's ability to manipulate objects mentally in 3D space. Participants have 3 minutes per part to compare 21 pairs of 6-sided cubes and determine if the rotated cubes are the same or different. Each part was scored using the formula: $\left[\left(\frac{\#attempted}{21}\right)\left(\frac{\#correct}{\#answered}\right)\right] * 100$, where attempted items included both answered and skipped items, answered items included any item where an answer was supplied (whether correct or incorrect), and skipped items were items that were not answered, but were followed by at least one answered item. This scoring method not only converts the score to a scale 1 – 100, it also addresses both speed and accuracy, while quantifying the impact of skips on overall score. Each part was scored using this formula, then the scores from both parts averaged to give the participants' overall score. High/Low group membership was determined by median split of all participants' scores (Min_{SV} = 0.234, Max_{SV} = 0.95, Mdn_{SV} = 0.60, M_{SV} = 0.61, SD_{SV} = 0.18, SV_{LOW} N = 30, SV_{HIGH} N = 30).

Spatial Orientation Test. The Spatial Orientation Test (SOT) measures an individual's ability to orient themselves in a 3D world (SOT; see Appendix E; Gugerty & Brooks, 2004). It is a computerized test consisting of a brief training segment and 32 test questions, which score is based on both accuracy and response time. Individual scores are calculated by dividing average response time by total number correct. Higher performance is indicated by lower scores. High/Low group membership was determined by median split of all participants' scores (Min_{SOT} = 3.97, Max_{SOT} = 39.32, Mdn_{SOT} = 12.72, M_{SOT} = 14.15, SD_{SOT} = 8.41, SOT_{LOW} N = 27, SOT_{HIGH} N = 33).

Complacency Potential Rating Scale. The updated Complacency Potential Rating Scale (CPRS; Pop & Stearman, in review; Singh, Molloy, & Parasuraman, 1993; Appendix F) measures an individual's attitude towards automation and automated devices, and has been shown to have high internal consistency ($r > .98$) and test-retest reliability ($r = .90$). The CPRS has 20 items, 4 of which are filler items, and each item is scored from 1 (strongly agree) to 5 (strongly disagree). Several items are negatively worded and are reverse-scored in the final tally. CPRS scores range from 16 (low complacency potential) to 80 (high complacency potential). The developers suggest classifying participants as either low or high complacency potential using the median split of the CPRS scores. High/Low group membership was determined by median split of all participants' scores (Min_{CPRS} = 28.0, Max_{CPRS} = 49.0, Mdn_{CPRS} = 39.5, M_{CPRS} = 39.9, $CPRS_{LOW}$ N = 30, $CPRS_{HIGH}$ N = 30).

NASA-TLX. Participants' perceived workload was evaluated with the computerized version of the NASA-TLX questionnaire, which uses a pairwise comparison weighting procedure (Hart & Staveland, 1988; Appendix G). The NASA-TLX is a self-reported questionnaire of perceived demands in six areas: mental, physical, temporal, effort (mental and physical), frustration, and performance. Participants evaluated their perceived workload level in each of these areas on 10-point scales, as well as completed pairwise comparisons for each subscale.

RSPAN. Verbal working memory capacity was assessed using the automated reading span task (RSPAN; Redick et al., 2012; Unsworth, Heitz, Shrock, & Engle, 2005; Appendix H), which has shown to have high internal (partial score α = .86) and test-retest (α = .82) reliability. Participants were shown a sentence, and they determined if the sentence made sense as written (e.g., "Andy was stopped by the policeman because he crossed the yellow heaven"). They indicated whether the sentence made sense (YES) or not (NO) by hitting either the F key (YES) or the J key (NO) on their keyboard. Participants were given feedback how they are performed on this task and were instructed to keep their performance above 80%. After evaluating the sentence, they were shown a letter to be recalled later. At the end of each set, participants were prompted to recall the letters in the proper order. Sentence-letter set sizes varied between 3 and 6 items, and each participant received 3 sets of each set size, for a total of 54 sentence-letter sets. Participants who scored lower than 80% on the sentence verification task were dismissed from the study. Working memory capacity was evaluated by using the participants' total letter set score (sum of all perfectly recalled letter sets), with higher numbers indicating greater working memory capacity, (Min_{RSPAN} = 5.0, Max_{RSPAN} = 51.0, Mdn_{RSPAN} = 32.5, M_{RSPAN} = 31.3, SD_{RSPAN} = 11.1). High/Low group membership was determined by median split of all participants' scores, $\text{RSPAN}_{\text{LOW}}$ N = 30, $\text{RSPAN}_{\text{HIGH}}$ N = 30.

Situation Awareness (SA). Participants' SA was evaluated by periodic queries via the communications panel. Queries were designed to assess the users level 1 (perception), level 2 (reasoning/comprehension) or level 3 (projection) SA (Appendix K). Queries were in the form of multiple choice questions, and participants responded by clicking the button on the communications panel that corresponded with the correct answer.

Usability and Trust Survey. Participant trust in the system was evaluated using the Usability and Trust Survey (Chen & Barnes, 2012a; Appendix I). The survey consisted of 20 questions rated on a scale of 1 to 7, with a scoring range of 20 – 140 points. Negative questions such as "The RoboLeader display was confusing" were reverse coded (a score of 7=1, 6=2, etc.). Positive questions such as "The RoboLeader system is dependable" and "I can trust the RoboLeader system" were regularly coded, with the sums of the positive and negative questions combined to create a global score. Higher scores indicate higher trust and usability of the system.

Procedure

After being briefed on the purpose of the study and signing the informed consent form, participants completed the demographics questionnaire, the working memory capacity test (RSPAN), and the Ishihara Color Vision Test. Those that successfully completed the RSPAN task with a sentence comprehension score of 80% or higher and those that passed the color vision test continued with the study, those that did not pass were debriefed and dismissed. Then participants completed the Attentional Control Survey, the Cube Comparisons test, the Spatial Orientation test, and the Complacency Potential Rating Scale.

Participants then received training and practice on their tasks for the experimental session. Training was self-paced and delivered by PowerPoint® slides. Participants were trained on the elements of the OCU, identifying map icons and their meanings, steps for completing various tasks, and completed several mini-exercises for practice. The training session lasted approximately 1.5 hours. Before proceeding to the experimental session, participants had to demonstrate that they could recall all icons and their meanings, as well as perform all tasks without any help. Participants who scored too low on the assessments were allowed to review the information again. If the participant had still not reached 90% proficiency after additional training, the participant was paid for the time they spent in the experiment up to that point and dismissed.

Participants were given a 5-minute break between the practice session and the experimental session. The experimental session lasted about 2 hours. Each experimental session had three missions, each lasting approximately 30 minutes. Participants were randomly assigned to an Agent Reasoning Transparency (ART) condition (ART1, ART2, or ART3), which were counterbalanced across participants to ensure an equal N in each condition. Each experimental session had three missions, each lasting approximately 30 minutes. The mission order was counterbalanced across participants to avoid order effects. At the beginning of each mission, the eye tracker was calibrated using the 9-point calibration setting.

During the missions, participants guided a convoy of three vehicles (their own MGV, a UAV, and a UGV) through a simulated urban environment, moving along a predetermined route. As the convoy proceeded through the environment, events occurred which necessitated altering the route. Each mission had six events that affected the predetermined route, and each event caused RoboLeader to suggest a route revision. Events and their associated area of influence were displayed on the map with icons. RoboLeader suggested a potential route revision, and the operator either accepted the suggestion or rejected it and kept the convoy on its original path. The participant was given no information regarding the safety of the alternate route, only the understanding that the suggested routes were at least as safe as the original route. Two of RoboLeader's route change suggestions [per mission] were inappropriate, and the participant needed to reject the suggestion correctly. Once RoboLeader suggested a route, there was a limited amount of time (15 seconds) for the participant to acknowledge the suggested change. When the participant acknowledged RL's suggestion, the simulation paused until the participant either agreed with or rejected RL's suggestion. If time expired before the participant acknowledged, RoboLeader automatically continued convoy movement along the original route.

Participants viewed communications from RoboLeader via a text feed in the upper right-hand corner of the OCU. When they received a message from RoboLeader, they were to click the acknowledge button on the RoboLeader communication window. Messages from RoboLeader varied based on ART. In ART1, messages from RoboLeader were simple notifications that RoboLeader is suggesting a route revision, corresponding to one of the map icons. In ART2, RL messages were notifications that provided specific details about the warning denoted by the map icons. In ART3, RL messages were the same as in ART2. However, the messages also included information as to how long ago the information was received (e.g., 1 hour, 4 hours, 6 hours). Transcripts of RoboLeader messages for each ART are in Appendix J.

The participant maintained communication with their ‘command’ via a text feed directly below RoboLeader’s communication window. Participants viewed messages from command, not all of which were directed to the participant. Each mission contained 12 information updates from command, two of which would result in the need to override RoboLeader’s route recommendation. Communications included messages directed at other units, which the participant should disregard. These messages were intended to create ‘noise,' as well as maintain a consistent rate for incoming messages (one message from either source approximately every 30 seconds). Transcripts of command messages for each mission are in Appendix L. In all conditions, command would also request information from the operator (SA queries). Requests for information required a response from the participant, which they did by selecting the appropriate response in the communication window on the OCU. Each mission contained 18 requests for information, and these were used to assess the participants SA. Transcripts of SA queries for each mission are in Appendix K.

Simultaneously, the participants had to maintain local security surrounding their own MGV by monitoring the MGV and UGV indirect-vision displays and detect targets in the immediate environment. Once a hostile target was detected, the participants identified the target by clicking on the target using the mouse. Mouse clicks in the camera feed windows produced a camera shutter sound, so the participant had verification that they did successfully click in the window. However, they did not receive feedback regarding their performance on the target detection task. There were civilians and friendly dismounted soldiers in the simulated environment to increase the visual noise present in the target detection tasks.

After completing all three missions, participants assessed their perceived workload using the NASA-TLX, and their trust in RL's suggestions by completing the Usability and Trust Survey. Participants were then debriefed, and any questions they had were answered by the experimenter.

Pilot testing was performed to determine whether the amount of proposed training was sufficient for participants to understand fully and correctly complete the assigned experimental tasks. Changes to training length and content were made as deemed necessary by pilot participants' performance.

Experimental Design

The study was a between-subjects experiment. Independent variables were Agent Reasoning Transparency and Individual Difference factors. Dependent measures were route selection task score, decision time, target detection task scores, workload, SA, and trust scores.

Independent Measures

The between-subjects factor was Agent Reasoning Transparency (ART). There were three levels of agent reasoning: low, medium, and high. ART was manipulated via RoboLeader messages. In the low ART condition (ART1), the agent recommended a course of action but otherwise offered no insight as to the reasoning behind the recommendation. In the medium ART condition (ART2), the agent recommended a course of action and gave the reason behind this recommendation. In the high ART condition (ART3), the same information as in ART2 was conveyed, and the recency of the information supporting the recommendation was also included. The time information was chosen as an example of the sort of information a system designer may provide to the user, but the user may not be certain how to incorporate this information into their decision. In this way, the effect of the type of agent reasoning information could be assessed as well as the availability. Participants completed three missions in their assigned ART.

Dependent Measures

Route Selection Task Measures:

- Performance Score: Participants were scored on whether they correctly accepted or rejected RoboLeader's route selection, and those scores summed across all missions. The score range for this score is 0 (no correct rejects or accepts) to 18 (correctly accepted or rejected all RoboLeader suggestions).
- Incorrect Acceptances: Twice each mission RoboLeader made a suggestion that should have been correctly rejected. Incorrect acceptances of these suggestions were indicative of complacent behavior; the participant scored 1 point for each

incorrect acceptance, and these were summed across all missions. The score range for this measure is 0 – 6, with higher scores indicating more complacent behavior and lower scores indicating less.

- Incorrect Rejections: Four times each mission RoboLeader made a suggestion that should have been correctly accepted. Incorrect rejections of these suggestions were indicative of low trust and/or poor SA; the participant scored 1 point for each incorrect rejection, and these were summed across all missions. The score range for this measure is 0 – 12, with higher scores indicating more distrustful behavior and lower scores indicating less.
- Decision Time: Decision time was averaged across missions. Decision Time was quantified as the time between agent alert and participant route selection. Reduced decision time when ART was available or increased (compared to decision time in the notification only condition) could indicate overwork, resulting in complacent behavior.

Target Detection Task Measures:

- Targets correctly detected (percentage): Number of targets correctly identified was expected to decrease in overwork conditions.
- Number of False Alarms: Number of false alarms was expected to increase as ART increases.
- In addition to Hits and FAs, two signal detection theory measures were also used to assess participant performance on the target detection task:

- d' – A measure of sensitivity to target. Values near 0 indicate correct detection probability near chance while higher values indicate increased discernibility of targets and participant sensitivity to targets.
- Beta (β) – The likelihood ratio, a measure of response bias. Higher values of Beta indicate a more conservative response bias.

Situation Awareness Scores: Each mission contained 18 SA queries, 6 for each of the three SA levels. SA queries were designed to assess the participants' SA at a specific SA level (i.e., SA1 – level 1 SA, perception; SA2 – level 2 SA, reasoning, comprehension; SA3 - level 3 SA, the projection of future state). Higher scores indicate better SA. SA queries were multiple choice questions that were presented as requests for information from Command. SA scores were averaged across all three missions.

- The SA1 queries were presented along the common route whenever timing would allow, and when presented where the paths diverged were created to be as similar as possible between the two paths (i.e. same query, same density of vehicles and persons, the same type of persons). Each SA1 query had one correct answer and was scored as one point for each correct answer and minus one point for each incorrect answer. Score range for SA1 queries was -18 to +18.
- SA2 queries were presented after the participant had selected a route and answered the SA3 query. SA2 queries were meant to gauge the participants' understanding of the events influencing route selection, and could have multiple correct answers, depending on information presented to the participant at the specific location. Only SA2 queries along the ground truth (correct) route were

scored. Participants received one point for each correct response, and minus one point for each incorrect response. Score range for SA2 queries was -60 to +30.

- SA3 queries were presented immediately after the participant selected a route and gauged the participants perception of how safe their chosen route would be. Each query had one correct response, and score range was -18 to +18.

Perceived workload – After completing three missions, the NASA-TLX was administered to assess the participants' perceived workload. Both global and individual factor workload scores were evaluated. Participants were instructed only to consider the route selection task and their interactions with the RoboLeader agent when completing the NASA-TLX.

Trust – After completing three missions, the Usability and Trust Survey was administered to assess the participants' trust in the agent.

Data Analysis

Data was examined using planned comparisons ($\alpha = .05$). Specifically, ART1 was compared to ART2, ART2 to ART3, and ART1 to ART2+3 (average of ART2 and ART3 scores), unless otherwise noted. Omnibus between-subject ANOVAs/ANCOVAs ($\alpha = .05$) are also reported. Means, SD, and 95% CI are reported for each measure.

Preliminary GPower 3.1.3 analysis indicated that 60 participants, 3 groups, in a between-factors ANOVA, had an estimated power of .83 at a medium-to-large effect size ($f = .35$).

Results

Complacent Behavior Evaluation

> Hypothesis 1: Access to agent reasoning will reduce incorrect acceptances, ART1 > ART2, and increased transparency of agent reasoning will increase incorrect acceptances, ART2 < ART3. When agent reasoning is not available, incorrect acceptances will be greater than when agent reasoning is present, ART1 > ART2+3.

Descriptive statistics for incorrect acceptances and decision times at the locations where the agent recommendation should have been rejected are shown in Table 4.

Table 4. Descriptive statistics for incorrect acceptances and decision times, sorted by agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean
Incorrect Acceptances	ART 1	20	3.25	2.27	0.51	(2.19, 4.31)
	ART 2	20	1.14	1.28	0.29	(0.54, 1.73)
	ART 3	20	2.65	2.32	0.52	(1.56, 3.74)
Overall DT at Reject Locations (sec)	ART 1	20	3.82	1.88	0.42	(2.94, 4.70)
	ART 2	20	2.96	1.44	0.32	(2.29, 3.64)
	ART 3	20	3.41	1.55	0.35	(2.69, 4.14)
DT Correct Rejects (sec)	ART 1	14	7.47	4.29	1.15	(4.99, 9.95)
	ART 2	20	7.49	3.17	0.71	(6.01, 8.98)
	ART 3	18	8.14	3.47	0.82	(6.41, 9.86)
DT Incorrect Accepts (sec)	ART 1	18	8.04	2.86	0.67	(6.62, 9.46)
	ART 2	11	6.09	1.76	0.53	(4.91, 7.28)
	ART 3	14	7.90	3.20	0.86	(6.06, 9.75)

Evaluating incorrect acceptances between ART conditions, there was a violation of the homogeneity of variance assumption. As such, Welch's correction has been reported, and contrast tests do not assume equal variance between conditions. There was a significant effect of ART on incorrect acceptances, $F(2, 34.8) = 7.96$, $p = .001$, $\omega^2 = .14$ (Figure 6), and the data has a significant curvilinear trend, $F(1, 34.8) = 10.80$, $p = .002$, $\omega^2 = .14$. Mean incorrect acceptances were lower in ART2 than in ART1, $t(29.9) = -3.63$, $p = .001$, $r_c = .55$, and ART3, $t(29.5) = 2.55$, $p = .016$, $r_c = .43$. Overall, incorrect acceptances were significantly lower when agent reasoning was provided, $t(31.8) = -2.31$, $p = .028$, $r_c = .38$. The hypothesis was supported, since access to agent reasoning did reduce incorrect acceptances in a low information environment, and increased transparency of agent reasoning began to overwhelm participants, resulting in increased incorrect acceptances.

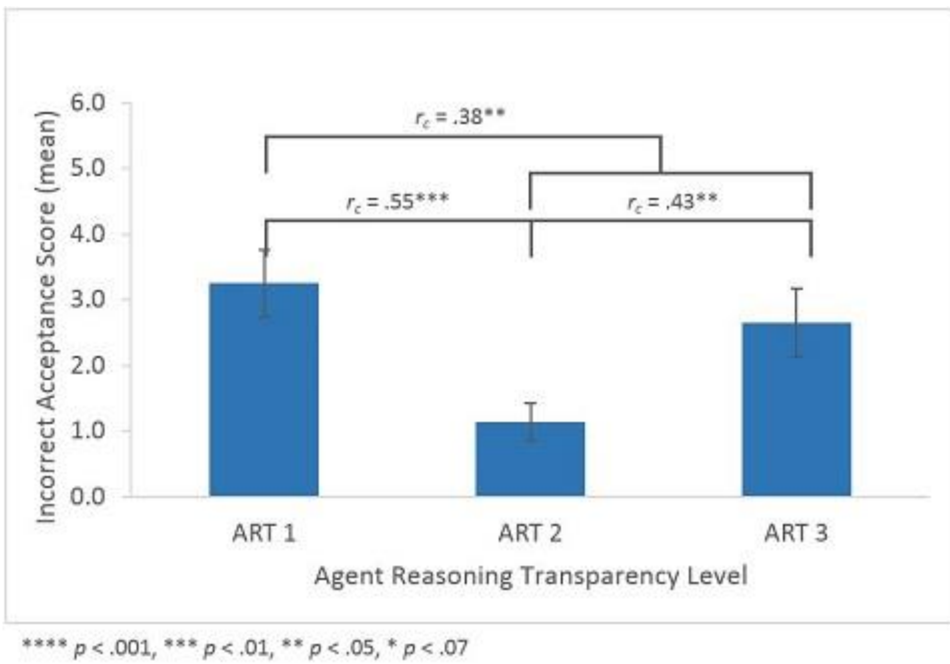
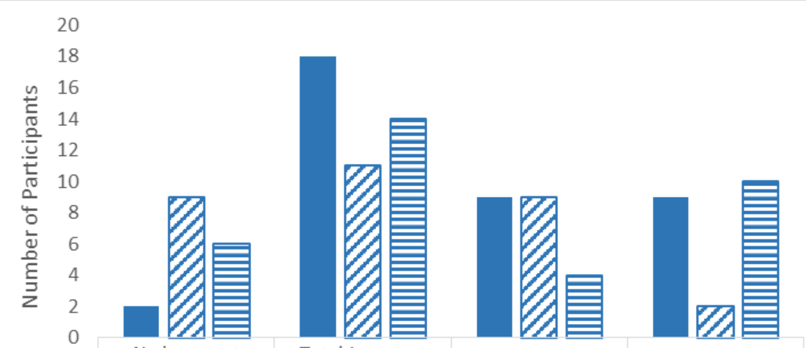


Figure 6. Average incorrect acceptances by agent reasoning transparency (ART) level. Bars denote SE.

Participants' scores were further analyzed by the number of incorrect acceptances per ART level (Figure 7). Chi-square analysis found a significant effect of ART on the number of incorrect acceptances, $X^2(14) = 29.45$, $p = .009$, Cramer's $V = .495$. Across all

ART levels, 17 participants had no incorrect acceptances, 15 of whom were in ARTs 2 and 3, evidence that access to agent reasoning is beneficial in avoiding incorrect acceptances. The range of potential scores for incorrect acceptances was 0 – 6, and the range of participants' scores was 0 – 6. Forty-three participants had at least 1 incorrect acceptance, 42% in ART1, 32% in ART3, and 26% in ART2. The incorrect scores were sorted into groups; < 50% (score 3 or less) or > 50% (score 4 or higher). Participants in ART1 were evenly split between these groups, indicating that in the notification-only condition performance was no better than chance. Also, of the 8 participants who scored 6/6 on incorrect acceptances, 6 were in ART1. However, of the participants who had > 50% incorrect acceptances, these were mostly in ART3. Examining the distribution of scores, it is clear that access to agent reasoning had a beneficial effect on performance. However the increase in incorrect acceptances in ART3 could be an indication that too much access to agent reasoning can have a detrimental effect on performance.



	No incorrect accepts	Total Incorrect Accepts	< 50%	>50%
ART 1	2	18	9	9
ART 2	9	11	9	2
ART 3	6	14	4	10

Figure 7. Distribution of incorrect acceptance scores across agent reasoning transparency (ART) levels.

Complacent behavior could also be indicated by reduced decision time for responses on the route selection task, particularly at those locations where the agent recommendation is incorrect. We hypothesized that decision time would increase as agent reasoning transparency increased, as participants should require additional time to process the extra information. Thus, reduced time could indicate less time spent in deliberation, which could be an indication of complacent behavior. In addition to the overall time to respond, decision times for correct rejections and incorrect acceptances were also examined (Figure 8). There was not a significant effect of ART on overall decision time, $F(2,57) = 0.29$, $p = .747$, $\omega^2 = .02$. Mean decision times were shorter in ART2 than in ART1, $t(57) = -0.49$, $p = .629$, $r_c = .06$, and ART3, $t(57) = 0.76$, $p = .453$, $r_c = .10$. There was not a significant effect of ART on decision time for correct rejections, $F(2,49) = 0.20$, $p = .823$, $\omega^2 = .03$. Mean decision times for correct rejections were essentially the same in ARTs 1 and 2, $t(49) = -0.02$, $p = .985$, $r_c = .00$, and slightly longer in ART3, $t(49) = 0.55$, $p = .583$, $r_c = .08$. There was not a significant main effect of ART on decision time for incorrect acceptances, $F(2,40) = 1.93$, $p = .153$, $\omega^2 = .04$, however there was a significant quadratic trend, $F(2,40) = 3.80$, $p = .058$, $\omega^2 = .06$. Mean decision times for incorrect acceptances were longer in ART1 than in ART2, $t(27.0) = -2.27$, $p = .032$, $r_c = .40$, and shorter in ART2 than in ART3, $t(20.9) = 1.80$, $p = .087$, $r_c = .37$. While decision times remain relatively unchanged across ART levels, decision times for incorrect acceptances drop significantly in ART2, which could be an indication of complacent behavior. Paired t-tests were used to compare differences between decision times for correct and incorrect responses within each ART. The largest difference in decision time was in ART2, $t(10) =$

0.95, p = .363, d = 0.24, while differences in ART3, $t(11)$ = 0.65, p = .527, d = 0.23, and ART1, $t(11)$ = -0.19, p = .851, d = 0.04 were smaller. Although these results did not achieve statistical significance, it is interesting to note that decision times for incorrect responses were lower in ART2, which could indicate that incorrect responses in this condition were a result of more complacent behavior.

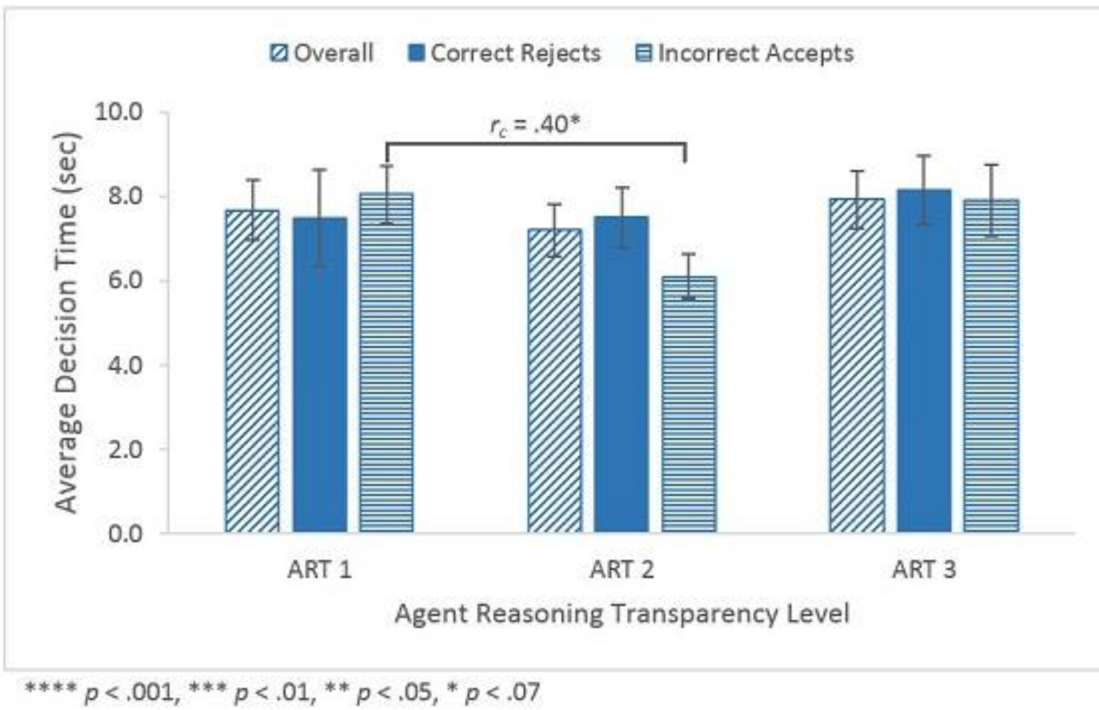


Figure 8. Average decision time in seconds for participant responses at decision points where the agent recommendation was incorrect. Decision times are shown for all responses (overall), correct rejections, and incorrect acceptances, sorted by agent reasoning transparency (ART) level. Bars denote SE.

Route Selection Task Performance

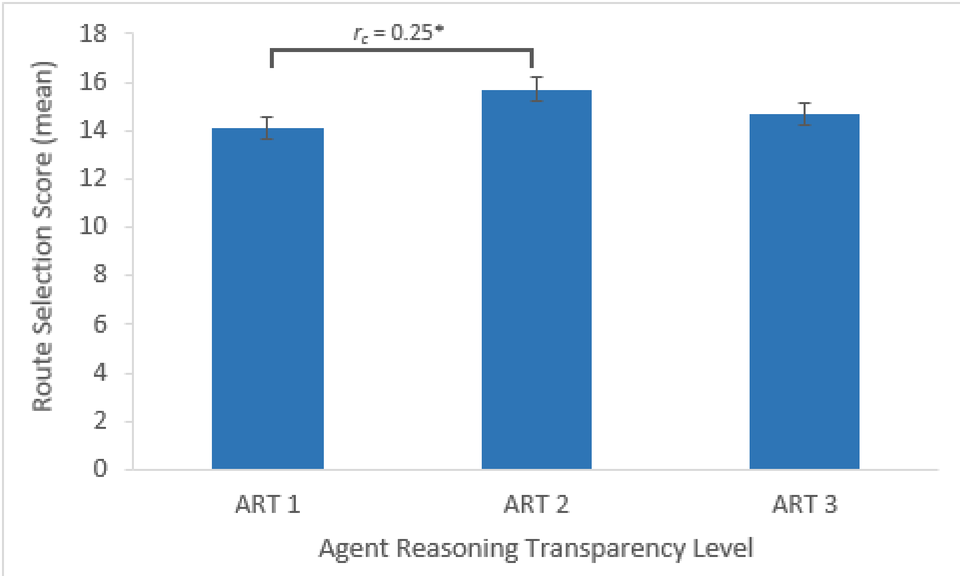
Hypothesis 2: Access to agent reasoning will improve performance (number of correct rejects and accepts) on the route selection task, ART1 < ART2, and increased transparency of agent reasoning will reduce performance on the route selection task, ART2 > ART3. When agent reasoning is not available performance will be lower than when agent reasoning is present, ART1 < ART2+3.

Descriptive statistics for route selection task scores and decision times for all decision points across three missions are shown in Table 5.

Table 5. Descriptive statistics for route selection scores and decision times, sorted by agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean
Route Selection Score	ART 1	20	14.10	2.59	0.58	(12.89, 15.31)
	ART 2	20	15.70	2.23	0.50	(14.66, 16.74)
	ART 3	20	14.70	2.81	0.63	(13.38, 16.02)
Overall Decision Time	ART 1	20	7.64	3.60	0.81	(5.95, 9.32)
	ART 2	20	7.51	3.36	0.75	(5.93, 9.08)
	ART 3	20	8.14	3.62	0.81	(6.45, 9.84)
Decision Time Correct Responses	ART 1	20	7.53	3.52	0.79	(5.88, 9.18)
	ART 2	20	7.42	3.37	0.75	(5.85, 9.00)
	ART 3	20	7.98	3.33	0.74	(6.43, 9.54)
Decision Time Incorrect Responses	ART 1	18	8.02	2.80	0.66	(6.63, 9.42)
	ART 2	17	8.44	4.20	1.02	(6.28, 10.60)
	ART 3	14	9.16	5.20	1.39	(6.16, 12.16)

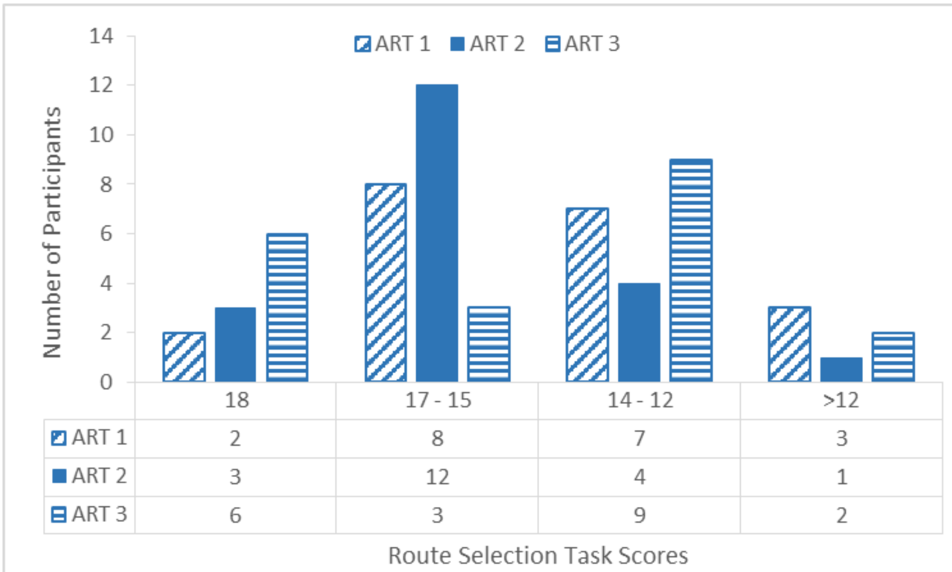
There was no significant effect of ART on the route selection task scores, $F(2,57) = 2.00$, $p = .145$, $\omega^2 = .03$ (Figure 9). Planned comparisons revealed mean performance scores were slightly higher in ART2 than in ART1, $t(57) = 1.98$, $p = .053$, $r_c = .25$. There was no significant difference in performance between ART2 and ART3, $t(57) = -1.24$, $p = .221$, $r_c = .16$. The hypothesis was partially supported, as the medium-large effect size between ARTs 1 and 2 indicates that the addition of agent reasoning did improve route selection. Scores in ART3 were lower than those in ART2. However this difference was not significant, indicating that performance in these two conditions was essentially the same.



**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Figure 9. Average route selection task score by agent reasoning transparency (ART) level. Bars denote SE.

Examining the distribution of scores, the potential range of scores for the route selection task was 0 – 18, and the range of participants' scores was 6 – 18 (Figure 10). Of these, 12 participants scored 18/18, 6 of whom were in ART3. Only 2 participants scored less than 50%; the majority scored 67% or higher. Of these scores, there appeared to be another break point near 80%, so this was used as a natural delineation for sorting the scores into groups (i.e. 17 – 15, 14 – 12, > 12). Participants in ART1 were evenly split between the 17 - 15 and 14 – 12 groups. However, there is an interesting difference between these groups for ARTs 2 & 3, in that ART2 participants makeup 52% of the 17 – 15 group, while ART3 participants make up 45% of the 14 – 12 group. This appears to offer additional support for the hypothesis, as performance in the agent reasoning conditions was better than in the notification-only condition, and performance does appear to be slightly worse in ART3 than in ART2.



	18	17 - 15	14 - 12	>12
ART 1	2	8	7	3
ART 2	3	12	4	1
ART 3	6	3	9	2

Figure 10. Distribution of scores for the route selection task across agent reasoning transparency (ART) levels.

ART had no significant effect on overall decision time (DT), $F(2,57) = 0.18$, $p = .833$, $\omega^2 = .03$. DT in ART1 was slightly longer than in ART2, $t(57) = -0.12$, $p = .908$, $r_c = .02$, and DT in ART2 was shorter than in ART3, $t(57) = 0.57$, $p = .570$, $r_c = .08$. Although this result is contrary to what was expected (DT increasing as ART increased), this could infer additional support for hypothesis 2, as the slight reduction in decision time regardless of the increased amount of information to process could indicate a performance improvement in ART 2 over ART1 when considered jointly with the route selection task performance. There was no significant difference between ARTs 2 and 3 for overall decision time, indicating that the increased access to reasoning had little effect on decision time.

Overall decision times for acceptances were compared to those for rejections [of the agent recommendation] using paired t-tests, and there was no significant difference, $t(54) = -0.79$, $p = .432$, $d = 0.06$ across ART levels. Overall decision times for correct responses were compared to those for incorrect responses using paired t-tests, and were found to be significantly shorter, $t(48) = -2.15$, $p = .037$, $d = 0.17$ across ART levels (Figure 11). Within each ART, this difference was greater in ART 2, $t(16) = -1.91$, $p = .074$, $d = 0.27$, than in either ART3, $t(13) = -1.19$, $p = .256$, $d = 0.18$, or ART1, $t(17) = -0.46$, $p = .651$, $d = 0.06$. Decision times for incorrect responses were evaluated between ARTs, and there was no significant difference between ART1 and ART2, $t(46) = 0.30$, $p = .767$, $d = 0.04$, or ART2 and ART3, $t(46) = 0.49$, $p = .626$, $d = 0.07$. Decision times for correct responses were evaluated between ARTs, and there was no significant difference between ART1 and ART2, $t(57) = -0.10$, $p = .921$, $d = 0.01$, or ART2 and ART3, $t(57) = 0.52$, $p = .607$, $d = 0.07$. While not offering additional support for the hypothesis, the smaller difference in mean decision time for incorrect responses demonstrated in ARTs 1 and 3 could be indicative of some participants' increased complacent behavior in these conditions.

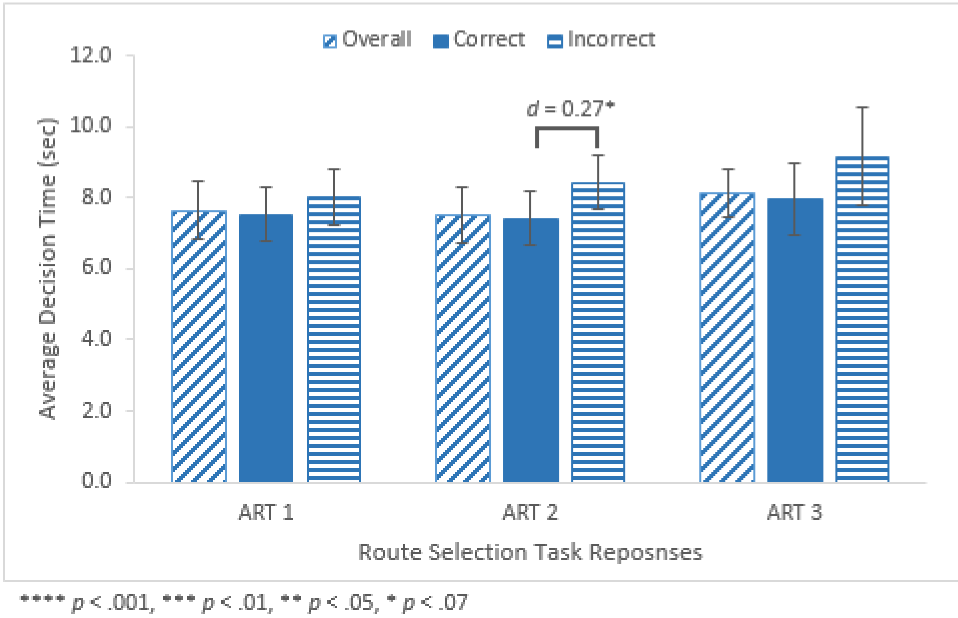


Figure 11. Comparison of average decision times for correct responses and incorrect responses, shown by agent reasoning transparency (ART) level. Bars denote SE.

Operator Trust Evaluation

Hypothesis 3: Access to agent reasoning will increase operator trust in the agent, ART1 < ART2, and increased transparency of agent reasoning will decrease operator trust in the agent, ART2 > ART3.

Descriptive statistics for incorrect rejections and the Usability and Trust Survey scores are shown in Table 6.

Table 6. Descriptive statistics for incorrect rejections and usability and trust survey results, sorted by agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean
Incorrect Rejections	ART 1	20	0.85	1.53	0.34	(0.13, 1.57)
	ART 2	20	1.10	1.33	0.30	(0.48, 1.72)
	ART 3	20	0.75	1.68	0.38	(-0.04, 1.54)
Usability & Trust Survey	ART 1	20	62.75	7.38	1.65	(59.29, 66.21)
	ART 2	20	56.25	9.24	2.07	(51.92, 60.58)
	ART 3	20	62.50	8.27	1.85	(58.63, 66.37)
Usability Responses	ART 1	20	46.75	5.33	1.19	(44.26, 49.24)
	ART 2	20	40.75	6.60	1.48	(37.66, 43.84)
	ART 3	20	45.75	7.03	1.57	(42.46, 49.04)
Trust Responses	ART 1	20	58.55	8.28	1.85	(54.67, 62.43)
	ART 2	20	54.40	10.23	2.29	(49.61, 59.19)
	ART 3	20	61.60	11.72	2.62	(56.12, 67.08)

Examining the distribution of incorrect rejections at those locations where the agent recommendation was correct across ARTs, 33 participants had no incorrect rejections. These were predominately in ARTs 1 and 3, ART2 having half as many perfect scores as the other two conditions (Figure 12). The range for potential scores for incorrect rejections was 0 – 12, and the range of participants' scores was 0 – 6. Twenty-seven (27) participants had at least 1 incorrect rejection, and these scores were sorted into < 50% (score 3 or less) and > 50% (score 4 or higher). Half of the participants in ART2 (10) had only 1 incorrect rejection. Considering perfect scores and 1 incorrect rejection together, it would appear that performance between the ARTs was relatively consistent. However, this may also be evidence of more complacent behavior in ARTs 1 and 3, where the agent recommendation was accepted more often, compared to more engaged, critical behavior in ART2 which resulted in occasional errors in judgment and incorrect responses.

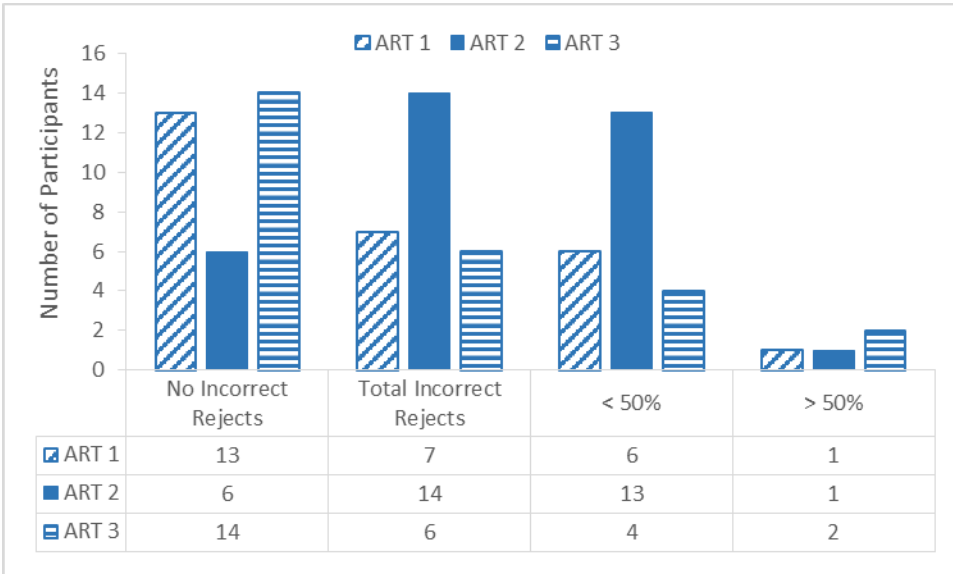


Figure 12. Distribution of scores for incorrect rejections, sorted by agent reasoning transparency (ART) level.

Evaluating incorrect rejections of the agent suggestions, there was no significant effect of ART on incorrect rejections, $F(2,57) = 0.28$, $p = .756$, $\omega^2 = .02$ (Figure 13). Planned comparisons revealed incorrect rejections were slightly higher in ART2 than in ART1, $t(57) = 0.52$, $p = .606$, $r_c = .07$, and ART3, $t(57) = -0.73$, $p = .470$, $r_c = .10$. Although incorrect rejections were higher in ART2 than in either ART1 or 3, which is contrary to predicted results, these findings were not significant.

Figure 13. Average incorrect rejections by agent reasoning transparency (ART) level. Bars denote SE.

Decision time for responses at the locations where the agent recommendation was correct was evaluated as a potential indicator of operator trust. It was hypothesized that decision time would increase as agent reasoning transparency increased, as participants should require additional time to process the extra information. Thus, increased time could indicate more time spent on deliberation, which may infer lower trust (e.g. less complacent behavior). However, reduced decision times for incorrect rejections of the agent recommendation at those locations could be indicative of complacent behavior. There was not a significant effect of ART on overall decision time at the agent correct locations, $F(2, 57) = 0.11$, $p = .896$, $\omega^2 = .03$ (Figure 14). Planned comparisons show that overall decision times in ART2 were similar to those in ART1, $t(57) = 0.09$, $p = .931$, $r_c = .01$, and those in ART3, $t(57) = 0.36$, $p = .723$, $r_c = .05$. Overall decision times for correct accepts were not significantly across ART, $F(2, 57) = 0.31$, $p = .738$, $\omega^2 = .02$. Planned comparisons show that decision times for correct responses in ART2 were similar to those in ART1, $t(57) = 0.03$, $p = .979$, $r_c = .00$, and those in ART3, $t(57) = 0.66$, $p = .510$, $r_c = .09$. Overall decision times for incorrect rejections were not significantly different across ART, $F(2, 57) = 0.09$, $p = .918$, $\omega^2 = .07$. Planned comparisons show that decision times for incorrect rejections in ART2 were similar to those in ART1, $t(24) = 0.40$, $p = .691$, $r_c = .08$, and those in ART3, $t(24) = -0.03$, $p = .975$, $r_c = .01$.

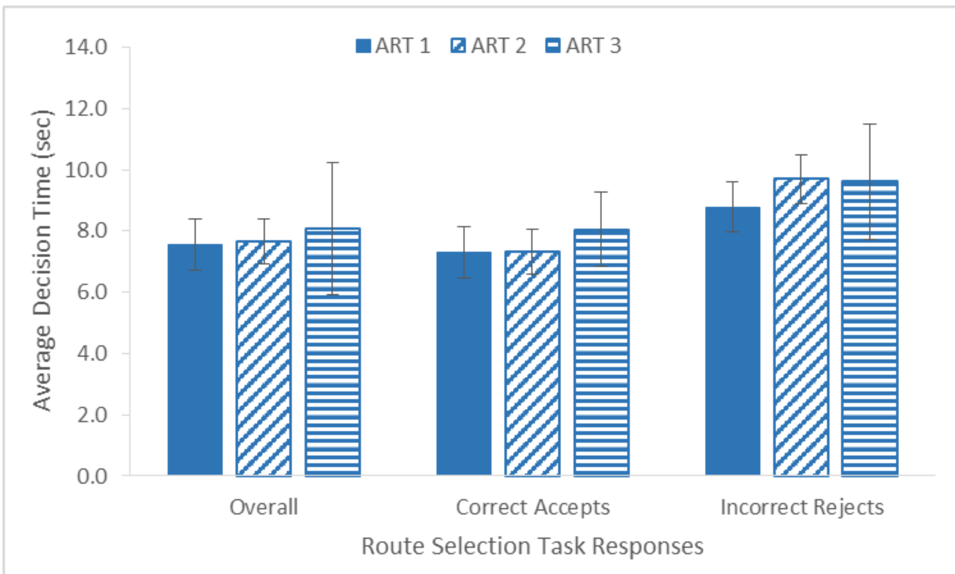


Figure 14. Average Decision Times, in seconds, at the locations where the agent recommendation was correct, sorted by correct/incorrect selections, for each agent reasoning transparency (ART) level. Bars denote SE.

Paired t-tests were used to compare differences between decision times for correct acceptances and incorrect rejections within each ART at those locations where the agent recommendation was correct (Figure 15). Decision times for incorrect rejections were significantly longer than for correct acceptances in ART2, $t(13) = -2.56$, $p = .024$, $d = 0.47$. However, there was no difference between the two in ART1, $t(6) = -0.81$, $p = .448$, $d =$

0.24, or ART3, $t(5) = 0.61$, $p = .572$, $d = 0.13$. This lack of difference in decision times in ARTs 1 and 3 could indicate a more complacent stance towards critiquing the agent recommendation in those conditions, while participants in ART2 appeared to maintain a more engaged, critical stance.

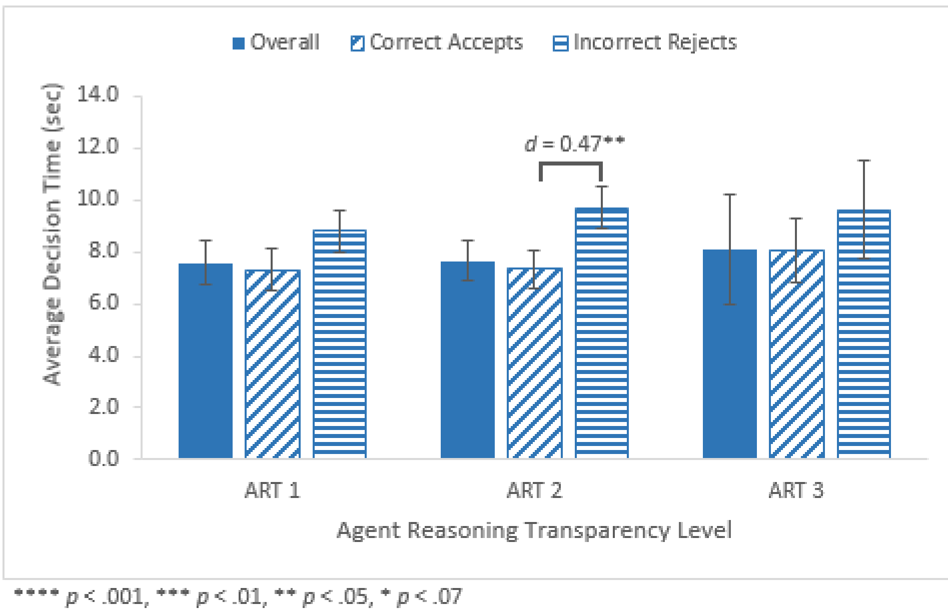


Figure 15. Average Decision Time, in seconds, for correct acceptances and incorrect rejections within each agent reasoning transparency (ART) level. Bars denote SE.

Operator trust was also evaluated using the Usability and Trust Survey. Perceived Attentional Control (PAC) score correlated significantly with ($r = .29$, $p = .013$), and was found to be significant predictor of Usability and Trust Survey scores, $R^2 = .083$, $b = .630$, $t(58) = 2.29$, $p = .025$. Participants who scored higher on PAC, indicating a greater ability to focus their attention, also scored higher on the Usability and Trust survey than their counterparts. PAC scores did not have a consistent relationship with Usability and Trust

Survey scores for each ART group (heterogeneity of regression), as such was not included as a covariate in subsequent analyses.

A between-groups ANOVA was conducted to assess the effect of ART on Usability and Trust Survey scores and found a significant effect, $F(2,57) = 3.00$, $p = .057$, $\omega^2 = .06$ (Figure 16). There was also a significant curvilinear trend to the data, $F(1,57) = 5.76$, $p = .020$, $\omega^2 = .07$. Usability and Trust scores in ART2 were lower than in either ART1, $t(57) = -1.83$, $p = .073$, $r_c = .24$, and ART3, $t(57) = 2.33$, $p = .023$, $r_c = .29$, which is contrary to the hypothesis. These scores indicate that participants trusted the agent more in ARTs 1 and 3 than in ART2. Adding agent reasoning transparency reduced perceived usability and trust, however, increased transparency of agent reasoning appeared to improve perceived usability and trust of the agent.

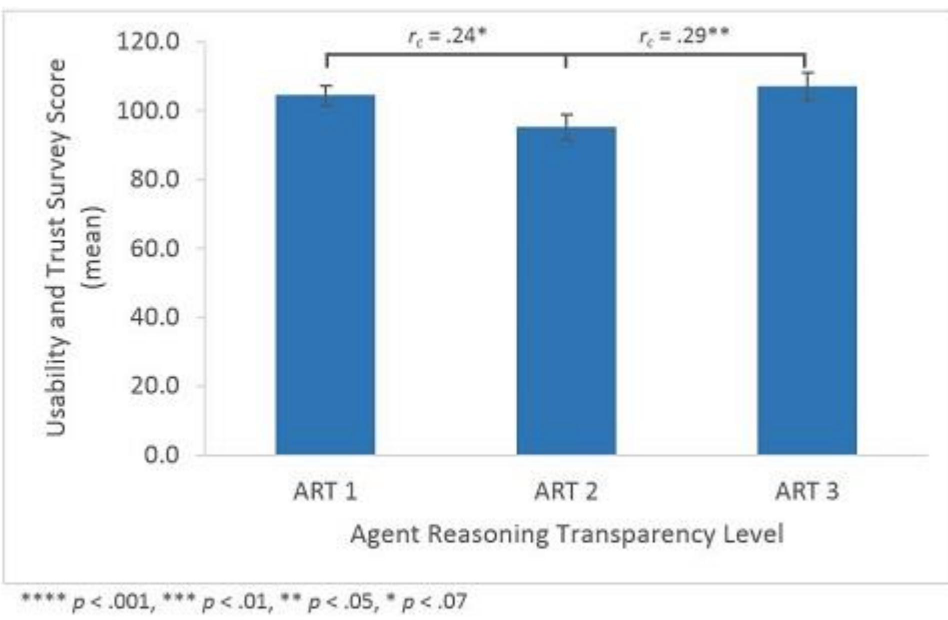


Figure 16. Average usability and trust survey scores by agent reasoning transparency (ART) level. Bars denote SE.

The Usability and Trust survey is a combination of surveys measuring Usability and Trust. These individual surveys were also evaluated separately to assess whether the overall findings were due to mainly operator trust or perceived usability.

Perceived Attentional Control (PAC) scores correlated significantly with (r = .28, p = .03), and were found to be significant predictors of Trust Survey scores, R^2 = .078, b = .384, t(58) = 2.21, p = .031. Participants who scored higher on PAC also scored higher on the Trust survey than their counterparts.

There was not a significant effect of ART on Trust score, F(2,57) = 2.52, p = .089, ω^2 = .05, (Figure 17). There was a significant curvilinear trend to the data, F(1,57) = 4.15, p = .046, ω^2 = .05. Planned comparisons revealed that trust scores in ART2 were slightly lower than in ART1, t(57) = -1.29, p = .202, r_c = .17, and significantly lower than ART3 scores, t(57) = 2.24, p = .029, r_c = .28. These findings do not support the hypothesis, as ART2 had the lowest Trust scores while ART3 had the highest.

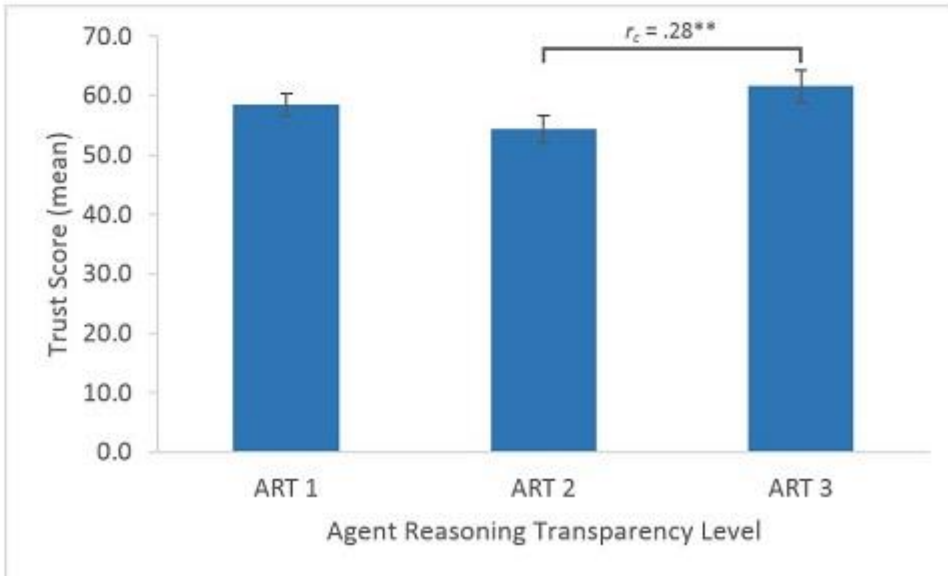


Figure 17. Average trust scores by agent reasoning transparency (ART) level. Bars denote SE.

Perceived Attentional Control (PAC) scores correlated significantly with ($r = .29$, $p = .03$), and were found to be significant predictors of Usability Survey scores, $R^2 = .084$, $b = .260$, $t(58) = 2.31$, $p = .025$. Participants who scored higher on PAC also scored higher on the Usability survey than their counterparts.

There was a significant effect of ART on Usability scores, $F(2,57) = 5.11$, $p = .009$, $\omega^2 = .12$, (Figure 18). There was also a significant curvilinear trend to the data, $F(1,57) = 9.96$, $p = .003$, $\omega^2 = .13$. Planned comparisons show that Usability scores in ART2 were significantly lower than those in either ART1, $t(57) = -2.98$, $p = .004$, $r_c = .37$, or ART3, $t(57) = 2.49$, $p = .049$, $r_c = .31$. Overall, Usability scores were significantly lower when agent reasoning was present than when it was not, $t(57) = -2.01$, $p = .049$, $r_c = .26$. Increased access to agent reasoning appeared to improve perceived usability of the agent.

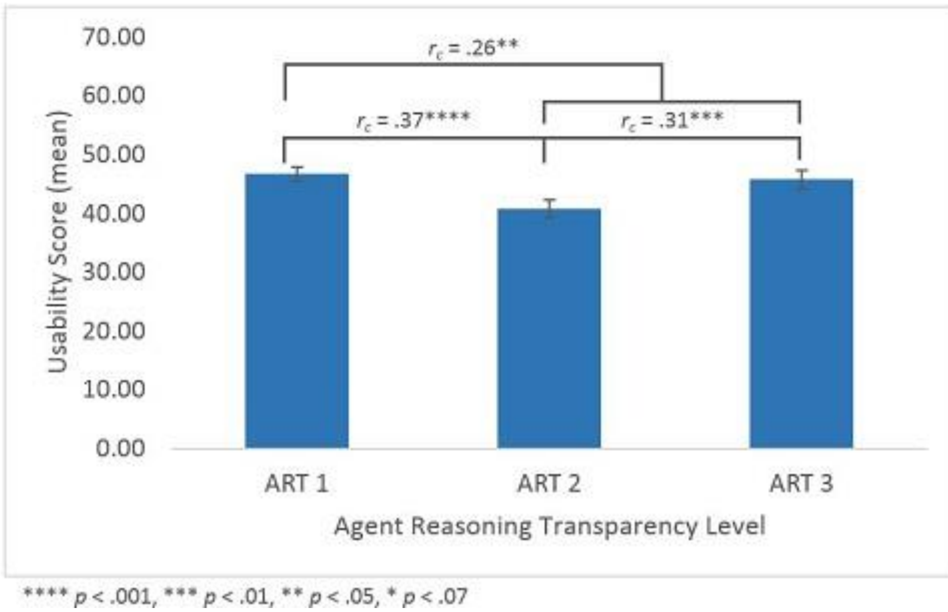


Figure 18. Average usability scores by agent reasoning transparency (ART) level. Bars denote SE.

Workload Evaluation

Hypothesis 4*:* Access to agent reasoning will increase operator workload, ART1 < ART2; and increased transparency of agent reasoning will increase operator workload, ART2 < ART3. When agent reasoning is not available workload will be lower than when agent reasoning is present, ART1 < ART2+3.

Spatial Orientation Test (SOT) scores correlated significantly with (r = .31, p = .015), and were found to be significant predictors of Global NASA-TLX scores, R^2 = .098, b = .570, t(58) = 2.52, p = .015. Participants who scored higher on the SOT, indicating a lesser ability to orient and navigate in their environment, also scored higher on the Global NASA-TLX than their counterparts. SOT scores did not have a consistent relationship with Global NASA-TLX scores for each ART group (heterogeneity of regression), as such was not included as a covariate in subsequent analyses.

A between-groups ANOVA was conducted to assess the effect of ART on Global NASA-TLX scores, and found that ART had no significant effect on participants' global workload, $F(2,57) = 0.68$, $p = .509$, $\omega^2 = .01$ (Figure 19). Planned contrasts revealed there was no overall difference in participant workload when agent reasoning was available compared to the no reasoning condition, (ART1 - ART2+3), $t(57) = -0.48$, $p = .631$, $r_c = .06$. Participants in ART1 ($M = 64.70$, SD = 13.47) reported lower workload than those in ART2 ($M = 65.18$, SD = 12.38), $t(57) = 0.12$, $p = .909$, $r_c = .02$, and workload was higher in ART2 than in ART3 ($M = 60.70$, SD = 14.01), $t(57) = -1.07$, $p = .291$, $r_c = .14$. The non-significant omnibus p-value, along with the small effect sizes, indicate that although workload scores decreased in ART3, there was no significant difference between ARTs.

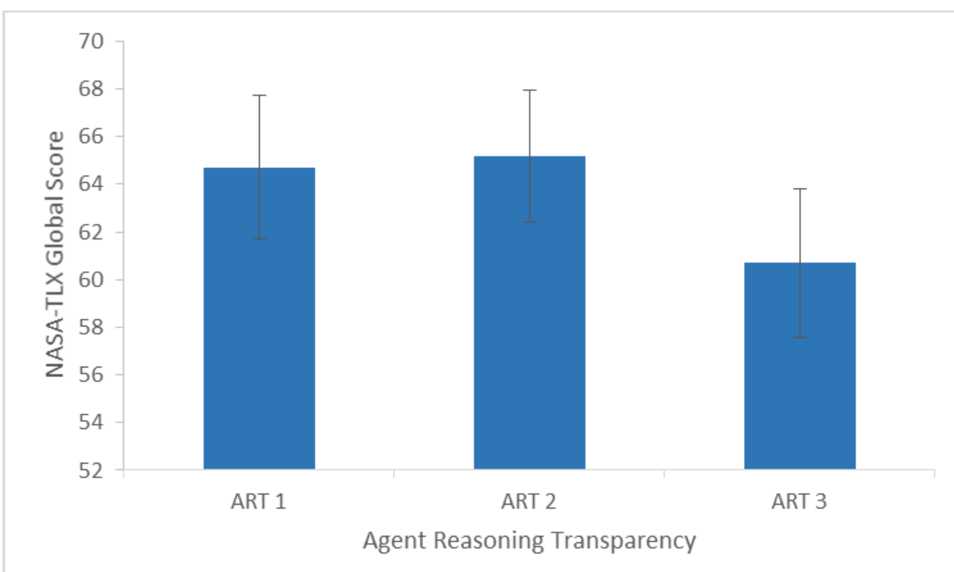


Figure 19. Average global NASA-TLX scores by agent reasoning transparency (ART) level. Bars denote SE.

Cognitive workload was also evaluated using several ocular indices. Descriptive statistics are shown in Table 7. Not all participants had complete eye measurement data, so this N was reduced (N = 12 for each ART). Eye tracking data was evaluated using the same planned comparisons as the subjective workload measure.

Table 7. Descriptive statistics for eye tracking measures by agent reasoning transparency (ART) condition.

		N	Mean	SD	SE	95% C.I. for Mean
Pupil Diameter (mm)	ART 1	12	3.71	0.32	0.09	(3.50, 3.91)
	ART 2	12	3.56	0.32	0.09	(3.36, 3.76)
	ART 3	12	3.46	0.39	0.11	(3.21, 3.70)
Fixation Duration (ms)	ART 1	12	264.54	42.16	12.17	(237.75, 291.33)
	ART 2	12	288.53	42.21	12.18	(261.71, 315.35)
	ART 3	12	265.71	25.23	7.28	(249.68, 281.74)
Fixation Count	ART 1	12	4895.18	513.60	148.26	(4568.85, 5221.51)
	ART 2	12	4809.97	875.08	252.61	(4253.97, 5365.97)
	ART 3	12	5076.82	421.63	121.72	(4808.93, 5344.71)

ART had no significant effect on participants' pupil diameter, $F(2,33) = 1.57$, $p = .224$, $\omega^2 = .03$. There was a marginally significant linear trend, $F(1,33) = 3.11$, $p = .087$, $\omega^2 = .06$, indicating that workload decreased as ART increased. Planned contrasts revealed that there was no difference in participant workload (as measured via pupil diameter) when agent reasoning was available, compared to the no reasoning condition, (ART1 - ART2+3), $t(33) = -1.61$, $p = .116$, $r_c = .27$. Participants in ART1 had larger pupil diameters than those in ART2, $t(33) = -1.03$, $p = .309$, $r_c = .18$, who in turn had larger pupil diameters than those in ART3, $t(33) = -0.73$, $p = .470$, $r_c = .13$.

ART had no significant effect on participants' fixation count, $F(2,33) = 0.55$, $p = .580$, $\omega^2 = .03$. Planned contrasts revealed there was no overall difference in participant workload (as measured via fixation count) when agent reasoning was available compared to the no reasoning condition, (ART1 - ART2+3), $t(33) = 0.22$, $p = .831$, $r_c = .04$. Participants in ART2 had slightly fewer fixations than those in ART1, $t(33) = -0.33$, $p = .744$, $r_c = .06$, and in ART3, $t(33) = 1.03$, $p = .310$, $r_c = .18$. These planned comparisons did not reach statistical significance.

ART had no significant effect on participants' fixation duration, $F(2,33) = 1.57$, $p = .223$, $\omega^2 = .03$ (Figure 20). There was a marginally significant curvilinear trend, $F(1,33) = 3.13$, $p = .086$, $\omega^2 = .06$. Planned contrasts revealed there was no overall difference in participant workload (as measured via fixation duration) when agent reasoning was available compared to the no reasoning condition, (ART1 - ART2+3), $t(33) = 0.95$, $p = .348$, $r_c = .16$. Participants in ART2 had longer fixation durations than those in ART1, $t(33) = 1.57$, $p = .126$, $r_c = .26$, and fixation durations were longer in ART2 than in ART3, $t(33) = -1.50$, $p = .144$, $r_c = .25$. These planned comparisons did not reach statistical significance.

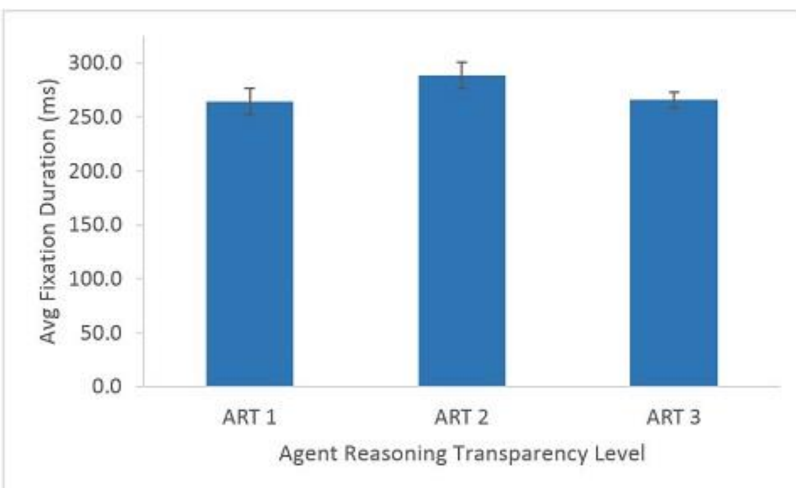


Figure 20. Participant average fixation duration by agent reasoning transparency level. Bars denote SE.

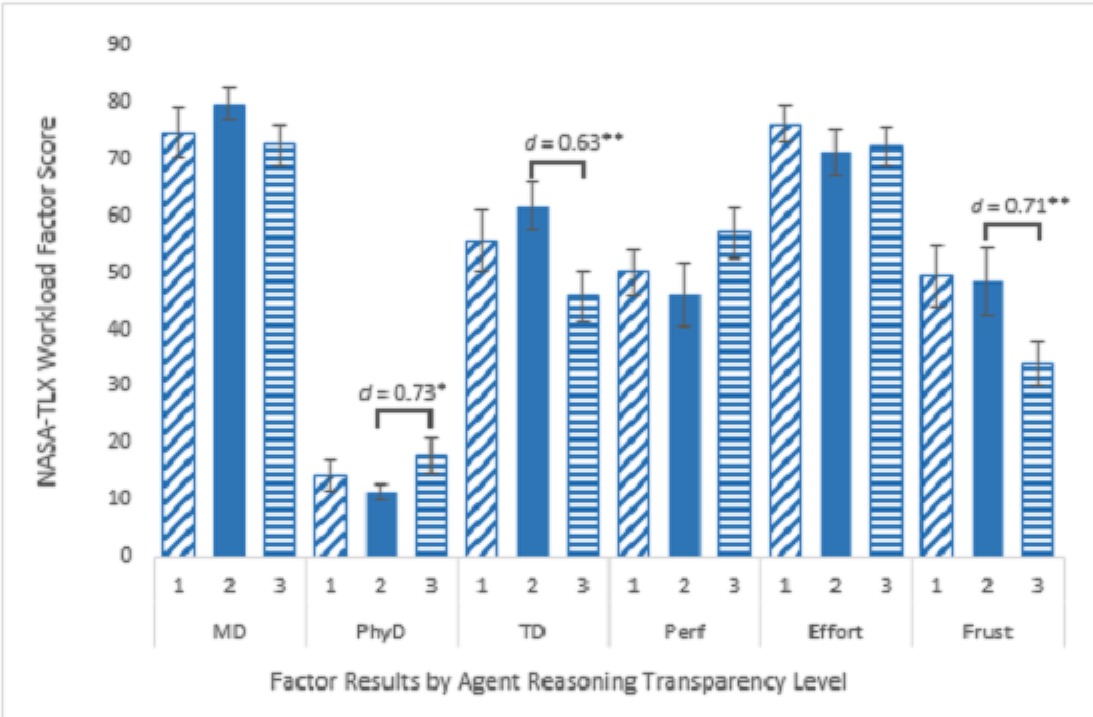
The NASA-TLX Global score is a composite score made up of six factors. Examining these factors separately, correlations between factors were low or nonexistent. An omnibus MANOVA indicated that each factor had no significant difference across ARTs, Wilks' λ = .860, $F(12,104) = 0.68$, p =.770, η_p^2 = .07 (Figure 21). Individual evaluations of each factor across ART were made by one-way ANOVAs using Bonferroni correction, α = .008, see Table 8.

Table 8. Evaluation of NASA-TLX workload factors across agent reasoning transparency (ART) levels.

	Mean (SD)			One-way ANOVA (α = .008)		Planned Comparisons (Cohen's *d*)		
	ART 1	ART 2	ART 3	$F(2,57)$	ω^2	ART 1 - ART 2	ART 2 - ART 3	ART 1 - ART 2+3
MD	74.75 (20.10)	79.75 (13.33)	72.50 (16.34)	0.97	.00	0.25	0.36	0.08
PhyD	14.25 (12.06)	11.25 (6.46)	17.75 (13.91)	1.95	.02	0.36	0.73 *	0.03
TD	55.50 (24.49)	61.75 (19.08)	45.75 (19.49)	2.90 *	.06	0.25	0.63 **	0.10
Perf	50.00 (18.92)	46.25 (25.23)	57.00 (20.16)	1.28	.01	0.15	0.42	0.07
Effort	76.25 (15.29)	71.25 (18.13)	72.25 (15.26)	0.53	.02	0.26	0.05	0.27
Frust	49.25 (24.40)	48.50 (27.00)	34.00 (17.29)	3.49 **	.05	0.03	0.71 **	0.41

**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Mental Demand (MD) was the factor contributing the most to workload, and ART2 elicited greater Mental Demand than ARTs 1 or 3. However, the effect size for the difference between ARTs was small, indicating there is little to no difference in Mental Demand. Physical Demand (PhyD) contributed the least to overall workload. PhyD scores were significantly higher in ART 3 than in ART2.



**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Figure 21. NASA-TLX workload factor average scores by agent reasoning transparency level. Bars denote SE.

Effort decreased when access to agent reasoning was available. However, the effect sizes were small. Temporal Demand (TD) and Frustration (Frust) scores were consistent between ARTs 1 and 2, but dropped off in ART 3, indicating the additional access to agent reasoning may have alleviated some of the pressure on participants in these ARTs. Performance scores are inverted, with lower scores indicating greater satisfaction. Performance scores indicate that participants in ARTs 1 and 2 were similarly satisfied with their performance, but those in ART 3 were less satisfied with their performance. Spatial Orientation Test (SOT) scores correlated significantly with Temporal Demand ($r = .36$, $p = .005$) and Effort ($r = .31$, $p = .015$) scores, but no other NASA-TLX factors. Participants with High SOT scores, which infers less Spatial Orientation ability, reported greater Temporal Demand in both ART2 (*d* = 0.82) and ART3 (*d* = 0.74) than their Low SOT counterparts. High SOT score participants also reported greater Effort in ART1 (*d* = 1.09) and ART3 (*d* = 1.37) than their Low SOT counterparts. However there was little difference in Effort due to SOT in ART2 (*d* = 0.24).

Situation Awareness (SA) Evaluation

Hypothesis 5: Access to agent reasoning will improve SA scores, and increased transparency of agent reasoning will improve SA1 and SA2 scores, but will reduce SA3 scores;

- SA1: ART1 < ART2, ART2 < ART3;
- SA2: ART1 < ART2, ART2 < ART3;
- SA3: ART1 < ART2, ART2 > ART3.

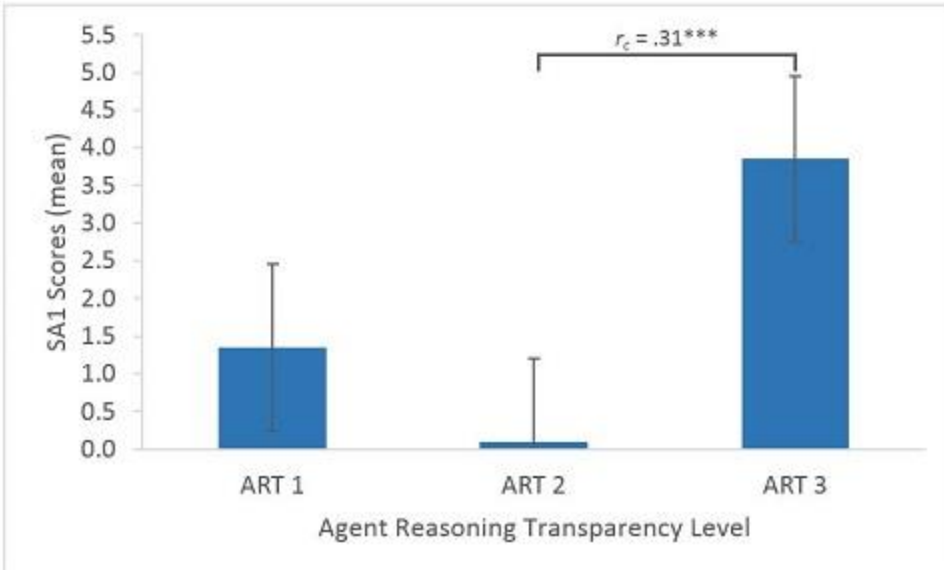
Descriptive statistics for situation awareness scores are shown in Table 9.

Table 9. Descriptive statistics for situation awareness scores by agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean	Min	Max
SA1	ART 1	20	1.35	4.93	1.10	(-0.96, 3.66)	-8	12
	ART 2	20	0.10	5.86	1.31	(-2.64, 2.84)	-10	12
	ART 3	20	3.85	3.65	0.82	(2.14, 5.56)	-5	9
SA2	ART 1	20	11.40	3.89	0.87	(9.58, 13.22)	5	18
	ART 2	20	13.15	3.70	0.83	(11.42, 14.88)	5	18
	ART 3	20	11.20	5.42	1.21	(8.67, 13.73)	1	18
SA3	ART 1	20	1.90	8.56	1.91	(-2.11, 5.91)	-12	14
	ART 2	20	3.85	8.98	2.01	(-0.35, 8.05)	-11	16
	ART 3	20	6.15	8.19	1.83	(2.32, 9.98)	-10	17

Spatial Visualization (SV) scores (r = .31, p = .002) correlated significantly with, and were found to be significant predictors of, SA Level 1 (SA1) scores, R^2 = .130, b = 9.76, $t(58)$ = 2.94, p = .005. Participants who scored higher in SV, indicating a greater ability to manipulate objects mentally in 3D space, also scored higher on SA1 than their counterparts. SV scores did not have a consistent relationship with SA1 scores for each ART group (heterogeneity of regression), as such was not included as a covariate in subsequent analyses.

SA Level 1 (perception of environment) scores indicated a significant effect of ART, $F(2,57)$ = 3.04, p = .056, ω^2 = .06 (Figure 22). Participants in ART2 had lower SA1 scores than those in ART1, $t(57)$ = -0.81, p = .423, r_c = 0.11, and ART3, $t(57)$ = 2.42, p = .019, r_c = .31. The hypothesis was partially supported. SA1 scores were lower in ART2 than in ART1, although the small effect size indicates this difference may not be meaningful. However, SA1 scores were greatest in ART3, supporting the hypothesis that increased transparency of agent reasoning will lead to improved SA1 scores



**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Figure 22. Average SA1 score by agent reasoning transparency (ART) level. Bars denote SE.

Spatial Visualization (SV) scores correlated significantly with ($r = .33$, $p = .006$), and were found to be significant predictors of SA Level 2 (SA2) scores, $R^2 = .106$, $b = 7.71$, $t(58) = 2.62$, $p = .011$. Participants who scored higher in SV, indicating a greater ability to manipulate objects mentally in 3D space, also scored higher on SA2 than their counterparts. SV scores did not have a consistent relationship with SA2 scores for each ART group (heterogeneity of regression), as such was not included as a covariate in subsequent analyses.

SA Level 2 (comprehension) scores indicated no significant effect of ART, $F(2,57) = 0.77$, $p = .469$, $\omega^2 = .01$. SA Level 2 scores were evaluated both regardless of route selection and along the ground truth route only, and no significant difference in results was found. Participants in ART2 had higher SA2 scores than those in ART1, $t(57) = 1.18$, $p = .245$, $r_c = .15$, or ART3, $t(57) = -0.93$, $p = .358$, $r_c = .12$, however these differences were not significant. The hypothesis was not supported, in that access to agent reasoning appeared to have no effect on SA2 scores.

There was unequal variance between groups for SA3 scores, so Welch's statistic was reported. SA Level 3 (projection) scores indicated a marginally significant difference between ARTs, $F(2,36.7) = 2.92$, $p = .067$, $\omega^2 = .04$ (Figure 23). There was also a significant linear trend, $F(1,36.7) = 4.35$, $p = .041$, $\omega^2 = .05$, indicating that SA3 scores increased as ART increased. SA3 was evaluated both regardless of route selection and along the ground truth route only, and no significant difference in results was found. Participants in ART2 had higher SA3 scores than those in ART1, $t(37.9) = 0.44$, $p = .660$, $r_c = .07$, and participants in ART3 had higher SA3 scores than those in ART2, $t(33.7) = 1.68$, $p = .103$, $r_c = .28$. The hypotheses were not supported. Although SA3 scores in ART2 were greater than those in ART1, this difference did not reach significance. SA3 scores in ART3 were predicted to be lower than those in ART2. However they increased as access to agent reasoning increased rather than decreased. While the difference between groups did not reach significance, the significant linear trend indicates that increased access to agent reasoning does help participants' project future status.

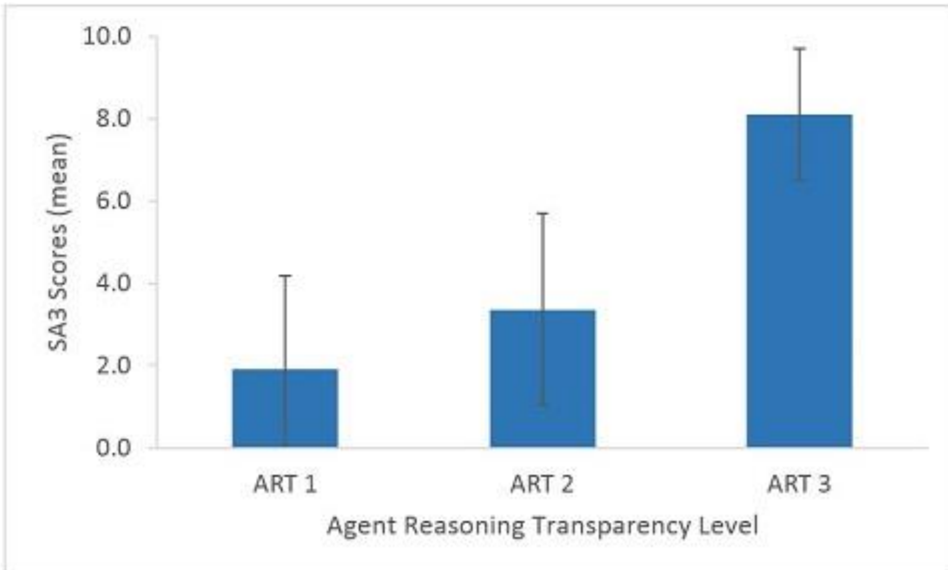


Figure 23. Average SA3 score by agent reasoning transparency (ART) level. Bars denote SE.

Target Detection Task Performance

> Hypothesis 6: Access to agent reasoning will reduce the number of targets detected and the number of FAs, ART1 > ART2, and increased transparency of agent reasoning will again result in fewer targets detected and fewer FAs, ART2 > ART3.

Descriptive statistics for Target Detection measures are shown in Table 10.

Table 10. Descriptive statistics for Target Detection Task measures by agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean	Min	Max
Targets Detected (Count)	ART 1	20	44.45	10.10	2.26	(39.72, 49.18)	30	69
	ART 2	20	45.05	13.64	3.05	(38.66, 51.44)	11	65
	ART 3	20	44.75	10.19	2.28	(39.98, 49.52)	29	65
FAs (Count)	ART 1	20	20.80	6.25	1.40	(17.87, 23.73)	10	33
	ART 2	20	16.35	5.29	1.18	(13.87, 18.83)	7	27
	ART 3	20	17.30	7.53	1.68	(13.78, 20.82)	8	32
d'	ART 1	20	2.20	0.32	0.07	(2.05, 2.35)	1.73	2.94
	ART 2	20	2.31	0.44	0.10	(2.11, 2.52)	1.40	3.19
	ART 3	20	2.29	0.38	0.09	(2.11, 2.46)	1.57	2.94
β	ART 1	20	2.42	0.28	0.06	(2.29, 2.56)	2.00	3.06
	ART 2	20	2.60	0.33	0.07	(2.45, 2.76)	1.90	3.21
	ART 3	20	2.60	0.37	0.08	(2.43, 2.78)	1.91	3.23

Spatial Visualization (SV) scores correlated significantly with (r = .26, p = .022), and were found to be significant predictors of total number of Targets Detected, R^2 = .068, b = 15.71, $t(58)$ = 2.06, p = .044. Participants who scored higher in SV, indicating a greater ability to mentally manipulate objects in 3D space, also detected more targets in their environment than their counterparts. SV scores did not have a consistent relationship with total number of Targets Detected for each ART group (heterogeneity of regression), as such was not included as a covariate in subsequent analyses.

There was no significant effect of ART on the number of targets detected, $F(2,57)$ = 0.01, p = .986, ω^2 = .03. The number of targets detected was slightly greater in ART2 than in ART1, $t(57)$ = 0.17, p = .869, r_c = .02, or ART3, $t(57)$ = -0.08, p = .934, r_c = .01, however these differences were not significant.

Spatial Visualization (SV) scores (r = -.39, p = .001) and Working Memory Capacity (WMC) scores (r = -.31, p = .009) correlated significantly with the total number of False Alarms (FAs) reported. SV scores were found to be significant predictors of FAs, R^2 = .154, b = -14.55, t(57) = -2.80, p = .007, while WMC scores showed to be marginal predictors of number of FAs reported, R^2 = .049, b = -0.16, t(57) = -1.87, p = .067. Participants who scored higher in SV, as well as those who scored higher on WMC measures, also reported fewer FAs than their counterparts. SV scores and Working Memory Capacity (WMC) scores did not have a consistent relationship with total number of False Alarms (FAs) reported for each ART group (heterogeneity of regression), as such was not included as a covariate in subsequent analyses.

There was a marginally significant effect of ART on the number of FAs across ARTs, F(2,57) = 2.66, p = .078, ω^2 = .05 (Figure 24). The number of FAs was lower in ART2 than in ART1, t(57) = -2.19, p = .033, r_c = .28, however there was little-to-no difference in number of reported FAs between ARTs 2 and 3, t(57) = 0.47, p = .642, r_c = .06. Thus the hypothesis was partially supported, as the addition of agent reasoning transparency did result in fewer FAs, however the increased transparency did not further reduce FAs.

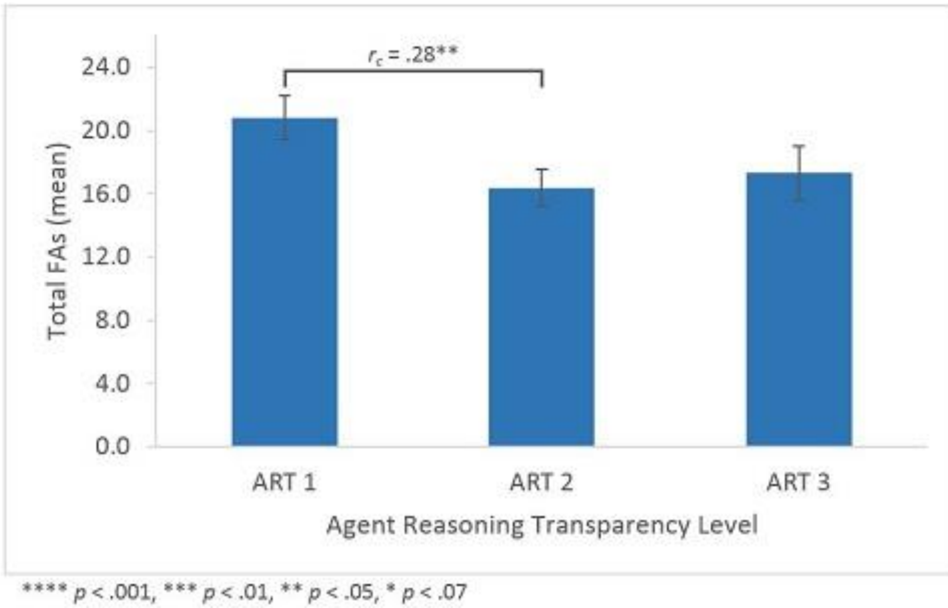


Figure 24. Average number of false alarms (FAs) by agent reasoning transparency (ART) level. Bars denote SE.

Results of the target detection task were also evaluated using SDT to determine if there were differences in sensitivity (*d'*) or selection bias (Beta) between the three ARTs. There was no significant effect of ART on d', $F(2,57) = 0.44$, $p = .647$, $\omega^2 = .02$. Participants were slightly more sensitive to targets in ART2 than in ART1, $t(57) = 0.90$, $p = .374$, $r_c = .12$, or ART3, $t(57) = -0.22$, $p = .831$, $r_c = .03$, however these differences did not achieve statistical significance.

Evaluating Beta across ART, there was no significant effect of ART on Beta scores, $F(2,57) = 1.94$, $p = .153$, $\omega^2 = .03$ (Figure 25). Beta scores were lower in ART1 than in ART2, $t(57) = 1.71$, $p = .094$, $r_c = .22$, and there was no difference in scores between ART2 and ART3, $t(57) = 0.00$, $p = .998$, $r_c = .00$. This indicates that the presence of agent reasoning allowed the participants to use a stricter selection criterion than in the no reasoning condition, but increasing the amount of agent reasoning did not have any further effect on participants' selection criteria. This more lenient selection criterion in ART1 could be the reason there were more FAs reported in ART1 than in either ARTs 2 or 3.

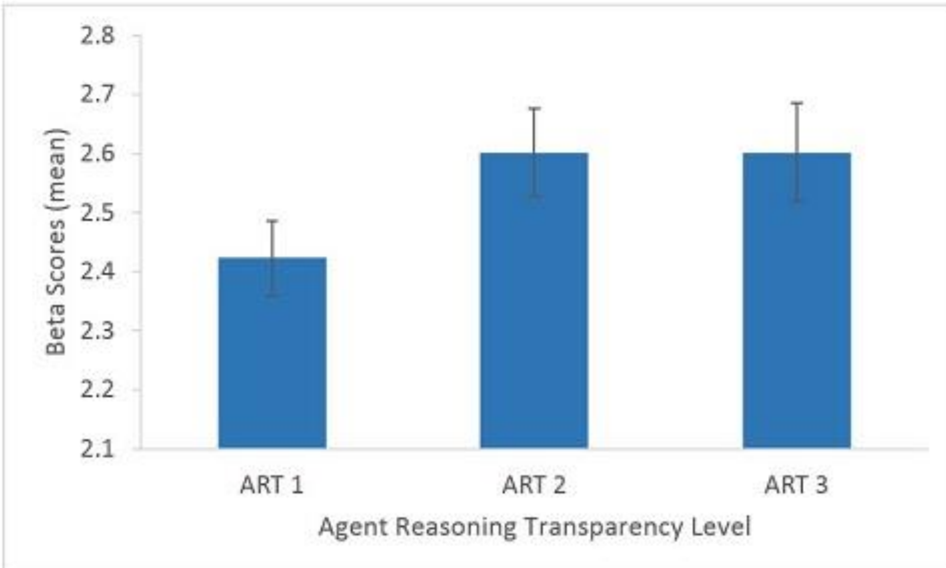


Figure 25. Average Beta scores by agent reasoning transparency (ART) level. Bars denote SE.

Individual Differences Evaluations

Complacency Potential

Complacency Potential (CP) was evaluated via the Complacency Potential Rating Scale (CPRS) scores. The effect of CP on several measures of interest across agent reasoning transparency (ART) level was evaluated via two-way between-groups ANOVAs, $\alpha = .05$. Post hoc t-tests within each ART compared performance differences between high/low group memberships. Descriptive statistics for Complacency Potential, as measured using the Complacency Potential Rating Scale (CPRS), are shown in Tables 11 and 12.

Table 11. Descriptive statistics for Complacency Potential Rating Scale (CPRS) scores by agent reasoning transparency (ART) level.

Group	N	Min	Max	Mdn	Mean	SD	Mdn Split Count	
							Hi	Lo
Overall	60	28	49	39.50	39.90	4.90	30	30
ART 1	20	28	46	38.00	38.50	4.90	8	12
ART 2	20	29	48	41.50	40.90	5.00	10	10
ART 3	20	33	49	41.00	40.30	4.60	12	8

Table 12. Descriptive statistics for High/Low Complacency Potential Rating Scale (CPRS) scores by agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean
ART 1	Low CPRS	12	35.33	3.11	0.90	(33.35, 37.31)
	High CPRS	8	43.25	2.55	0.90	(41.12, 45.38)
ART 2	Low CPRS	10	36.80	3.50	1.11	(34.20, 38.20)
	High CPRS	10	45.10	1.37	0.43	(44.12, 46.08)
ART 3	Low CPRS	8	35.50	1.77	0.63	(34.02, 36.98)
	High CPRS	12	43.50	2.68	0.77	(41.80, 45.20)

Hypothesis 7: High CPRS individuals will have fewer correct rejects on the route planning task than Low CPRS individuals.

A two-way between-groups ANOVA revealed no significant interaction between CPRS and ART on the number of correct rejects in the route planning task, $F(2,54) = 0.39$, $p = .682$, $\eta_p^2 =.01$, nor any significant main effect of CPRS on the number of correct rejects in the route planning task, $F(1,54) = 0.88$, $p = .768$, $\eta_p^2 = .00$.

Hypothesis 8: High CPRS individuals will have higher scores on the usability and trust survey than Low CPRS individuals.

A two-way between-groups ANOVA revealed no significant interaction between CPRS and ART on Usability and Trust Survey scores, $F(2,54) = 0.86$, $p = .429$, $\eta_p^2 = .03$, nor any significant main effect of CPRS on Usability scores, $F(1,54) = 2.25$, $p = .140$, $\eta_p^2 = .04$.

Hypothesis 9: High CPRS individuals will have lower SA scores than Low CPRS individuals.

A two-way between-groups ANOVA revealed no significant interaction between CPRS and ART on SA1 scores, $F(2,54) = 0.23$, $p = .794$, $\eta_p^2 =.01$, nor any significant main effect of CPRS on SA1 scores, $F(2,54) = 0.27$, $p = .608$, $\eta_p^2 = .01$. There was no significant interaction between CPRS and ART on SA2 scores, $F(2,54) = 0.61$, $p = .548$, $\eta_p^2 = .02$, nor any significant main effect of CPRS on SA2 scores, $F(2,54) = 0.24$, $p = .628$, $\eta_p^2 = .00$. There was no significant interaction between CPRS and ART on SA3 scores, $F(2,54) = 1.41$, $p = .254$, $\eta_p^2 =.05$, nor any significant main effect of CPRS on SA3 scores, $F(2,54) = 0.01$, $p = .921$, $\eta_p^2 = .00$.

Spatial Ability (SOT and SV) and Perceived Attentional Control (PAC)

Hypothesis 10: Individual differences, such as SpA and PAC, will have differential effects on the operator's performance on the route selection task and their ability to maintain SA.

The effects of individual difference (ID) factors and agent reasoning transparency (ART) level on route selection performance were evaluated via two-way between-groups ANOVAs, $\alpha = .05$. When Levene's Test of Equality of Error Variance was significant, the evaluation was repeated at $\alpha = .01$. Post hoc t-tests within each ART compared performance differences between high/low group memberships for each ID factor.

Descriptive statistics for Spatial Orientation (SOT), Spatial Visualization (SV), and Perceived Attentional Control (PAC) are shown in Tables 13 and 14.

Table 13. Descriptive statistics for Spatial Orientation (SOT), Spatial Visualization (SV), and Perceived Attentional Control (PAC), by Agent Reasoning Transparency (ART) level.

								Mdn Split Count	
	Group	N	Min	Max	Mdn	Mean	SD	Hi	Lo
SOT	Overall	60	3.97	29.54	12.72	13.59	7.28	30	30
	ART 1	20	5.70	22.00	14.06	13.27	5.20	8	12
	ART 2	20	4.12	29.00	10.10	13.35	7.98	11	9
	ART 3	20	3.97	29.54	11.22	14.15	8.56	11	9
SV	Overall	60	0.19	0.95	0.50	0.53	0.19	35	25
	ART 1	20	0.19	0.93	0.54	0.54	0.19	12	8
	ART 2	20	0.21	0.86	0.54	0.52	0.20	13	7
	ART 3	20	0.21	0.95	0.49	0.52	0.18	10	10
PAC	Overall	60	41.0	74.0	61.00	60.50	7.50	32	28
	ART 1	20	46.0	74.0	65.50	63.00	8.00	13	7
	ART 2	20	47.0	69.0	60.50	60.10	6.00	10	10
	ART 3	20	41.0	74.0	60.00	58.50	8.20	9	11

Table 14. Descriptive statistics for Spatial Orientation (SOT), Spatial Visualization (SV), and Perceived Attentional Control (PAC), by Agent Reasoning Transparency (ART) level, sorted by High/Low group membership.

			N	Mean	SD	SE	95% C.I. for Mean
SOT	ART 1	Low	12	16.88	2.95	0.85	(13.11, 22.00)
		High	8	7.86	1.98	0.70	(5.70, 11.55)
	ART 2	Low	9	20.90	5.28	1.76	(14.64, 29.00)
		High	11	7.16	2.32	0.70	(4.12, 10.43)
	ART 3	Low	9	21.93	6.47	2.16	(12.72, 29.54)
		High	11	7.78	2.56	0.77	(3.97, 12.71)
SV	ART 1	Low	8	0.36	0.09	0.03	(0.19, 0.45)
		High	12	0.66	0.14	0.04	(0.50, 0.93)
	ART 2	Low	7	0.30	0.11	0.04	(0.21, 0.48)
		High	13	0.64	0.12	0.03	(0.50, 0.86)
	ART 3	Low	10	0.39	0.08	0.03	(0.21, 0.48)
		High	10	0.66	0.14	0.04	(0.50, 0.95)
PAC	ART 1	Low	7	53.57	4.24	1.60	(46.0, 60.0)
		High	13	68.08	3.62	1.00	(62.0, 74.0)
	ART 2	Low	10	55.50	4.43	1.40	(47.0, 60.0)
		High	10	64.70	2.95	0.93	(61.0, 69.0)
	ART 3	Low	11	53.18	6.84	2.06	(41.0, 60.0)
		High	9	64.89	3.98	1.33	(61.0, 74.0)

Route Selection Task Evaluation

Spatial Orientation (SOT) was not found to be a significant predictor of performance on the route selection task independent of ART, R^2 = .00, β = -.003, $t(59)$ = -0.02, p = .982. A two-way between-groups ANOVA revealed no significant interaction between SOT and ART on Route Selection scores, $F(2,54)$ = 0.92, p = .406, η_p^2 = .03, nor any significant main effect of SOT on Route Selection scores, $F(1,54)$ = 0.04, p = .848, η_p^2 = .00.

Spatial Visualization (SV) was found to be a significant predictor of performance on the route selection task independent of ART level, $R^2 = .10$, $\beta = .31$, $t(59) = 2.52$, $p = .015$. A two-way between-groups ANOVA, $\alpha = .01$, revealed no significant interaction between SV and ART on Route Selection scores, $F(2,54) = 1.76$, $p = .182$, $\eta_p^2 = .06$, however, there was a significant main effect of SV on Route Selection scores, $F(1,54) = 4.31$, $p = .043$, $\eta_p^2 = .07$ (Figure 26). Post hoc comparisons between high/low SV groups within each ART level show that High SV and Low SV individuals had similar Route Selection scores in ART1, $t(18) = -0.66$, $p = .518$, $d = 0.31$, and ART3, $t(18) = -0.16$, $p = .879$, $d = 0.07$. However in ART2 the High SV individuals had higher Route Selection scores than their Low SV counterparts, $t(18) = -3.08$, $p = .017$, $d = 1.59$.

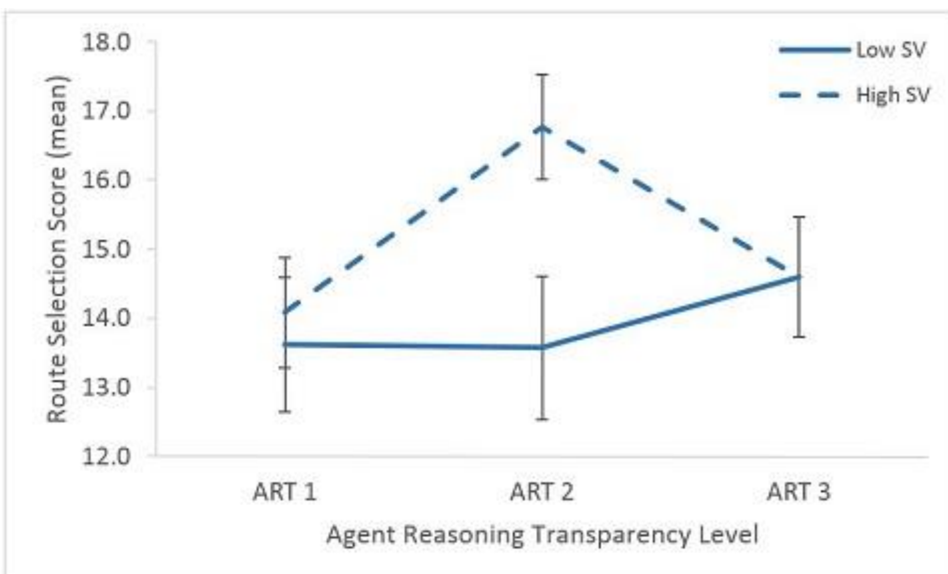


Figure 26. Average route selection scores by High/Low Spatial Visualization (SV) group membership, sorted by agent reasoning transparency (ART) level. Bars denote SE.

A two-way between-groups ANOVA revealed no significant interaction between PAC and ART on Route Selection scores, $F(2,54) = 0.17$, $p = .845$, $\eta_p^2 = .01$, nor any significant main effect of SOT on Route Selection scores, $F(1,54) = 0.32$, $p = .574$, $\eta_p^2 = .01$.

SA1 Evaluation

A two-way between-groups ANOVA revealed no significant interaction between SOT and ART on SA1 scores, $F(2,54) = 0.77$, $p = .469$, $\eta_p^2 = .028$, nor any significant main effect of SOT on SA1 scores, $F(1,54) = 0.43$, $p = .515$, $\eta_p^2 = .008$.

A two-way between-groups ANOVA revealed no significant interaction between SV and ART on SA1 scores, $F(2,54) = 0.34$, $p = .716$, $\eta_p^2 = .01$, however there was a significant main effect of SV on SA1 scores, $F(1,54) = 14.62$, $p < .001$, $\eta_p^2 = .21$ (Figure 27. High SV individuals had higher SA1 scores in all ARTs (ART1, $t(18) = -1.73$, $p = .101$, $d = 0.81$; ART2, $t(18) = -2.39$, $p = .028$, $d = 1.09$; ART3, $t(18) = -2.79$, $p = .012$, $d = 1.25$) than their Low SV counterparts.

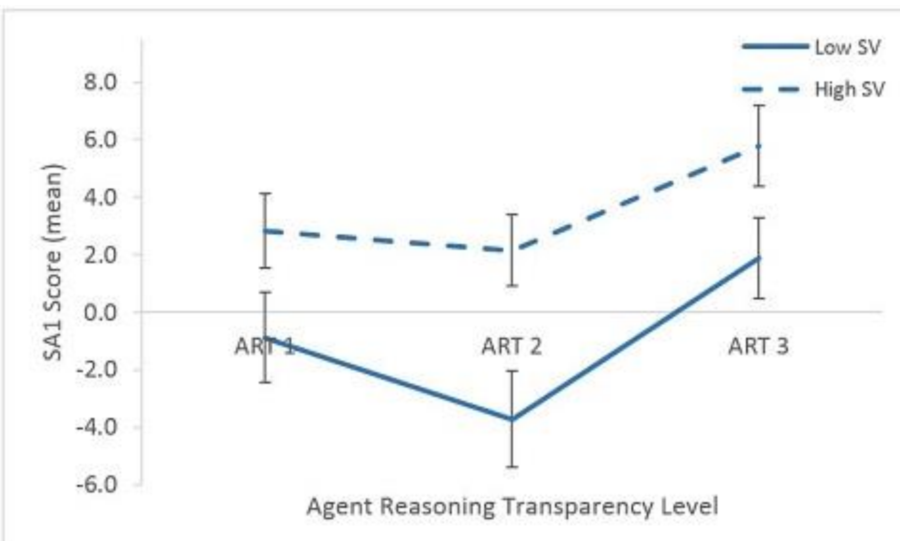


Figure 27. Average SA1 scores by Spatial Visualization (SV) High/Low group membership, sorted by agent reasoning transparency (ART) level. Bars denote SE.

A two-way between-groups ANOVA revealed no significant interaction between PAC and ART on SA1 scores, $F(2,54) = 1.98$, $p = .148$, $\eta_p^2 = .07$, nor any significant main effect of PAC on SA1 scores, $F(1,54) = 2.76$, $p = .102$, $\eta_p^2 = .05$.

SA2 Evaluation

A two-way between-groups ANOVA revealed no significant interaction between SOT and ART on SA2 scores, $F(2,54) = 1.40$, $p = .255$, $\eta_p^2 = .05$, nor any significant main effect of SOT on SA2 scores, $F(1,54) = 0.27$, $p = .603$, $\eta_p^2 = .01$. There was no significant interaction between SV and ART on SA2 scores, $F(2,54) = 0.63$, $p = .534$, $\eta_p^2 = .02$, nor any significant main effect of SV on SA2 scores, $F(1,54) = 3.20$, $p = .079$, $\eta_p^2 = .06$. There was no significant interaction between PAC and ART on SA2 scores, $F(2,54) = 0.06$, $p = .943$, $\eta_p^2 = .00$, nor any significant main effect of PAC on SA2 scores, $F(1,54) = 0.44$, $p = .511$, $\eta_p^2 = .01$.

SA3 Evaluation

A two-way between-groups ANOVA revealed no significant interaction between SOT and ART on SA3 scores, $F(2,54) = 0.51$, $p = .604$, $\eta_p^2 = .02$, nor any significant main effect of SOT on SA3 scores, $F(1,54) = 0.68$, $p = .414$, $\eta_p^2 = .01$.

A two-way between-groups ANOVA revealed no significant interaction between SV and ART on SA3 scores, $F(2,54) = 0.50$, $p = .611$, $\eta_p^2 = .02$, however there was a significant main effect of SV on SA3 scores, $F(1,54) = 6.73$, $p = .012$, $\eta_p^2 = .11$ (Figure 28). High SV individuals had higher SA3 scores in all ARTs (ART1, $t(18) = -1.54$, $p = .142$, $d = 0.69$; ART2, $t(18) = -1.89$, $p = .075$, $d = 0.85$; ART3, $t(18) = -0.93$, $p = .364$, $d = 0.42$).

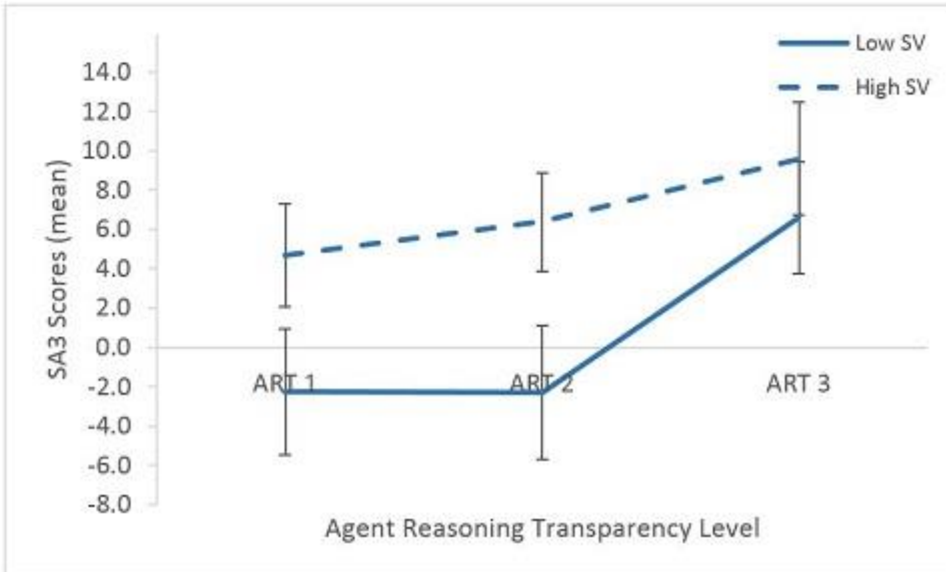


Figure 28. Average SA3 scores by Spatial Visualization (SV) High/Low group membership, sorted by agent reasoning transparency (ART) level. Bars denote SE.

A two-way between-groups ANOVA revealed no significant interaction between PAC and ART on SA3 scores, $F(2,54) = 2.78$, $p = .071$, $\eta_p^2 = .09$, and no significant main effect of PAC on SA3 scores, $F(1,54) = 0.01$, $p = .906$, $\eta_p^2 = .00$.

Working Memory Capacity (WMC)

Hypothesis 11: High WMC individuals will have more correct rejections and higher SA2 and SA3 scores than Low WMC individuals.

The effects of Working Memory Capacity (WMC) and agent reasoning transparency (ART) level were evaluated via two-way between-groups ANOVAs, $\alpha = .05$. Post hoc t-tests within each ART compared performance differences between high/low group memberships. Descriptive statistics for WMC, as measured using the RSPAN test, are shown in Tables 15 and 16.

Table 15. Descriptive statistics for Working Memory Capacity (WMC), by Agent Reasoning Transparency (ART) level.

	Group	N	Min	Max	Mdn	Mean	SD	Mdn Split Count Hi	Mdn Split Count Lo
WMC	Overall	60	5.0	51.0	32.50	31.30	11.10	30	30
	ART 1	20	8.0	51.0	30.50	30.90	10.98	9	11
	ART 2	20	8.0	49.0	36.00	33.85	9.95	13	7
	ART 3	20	5.0	51.0	28.50	29.15	12.39	8	12

Table 16. Descriptive statistics for Working Memory Capacity (WMC), by Agent Reasoning Transparency (ART) level, sorted by High/Low group membership.

			N	Mean	SD	SE	95% C.I. for Mean
WMC	ART 1	Low	11	22.64	6.36	1.92	(18.36, 26.91)
		High	9	41.00	5.22	1.74	(36.99, 45.01)
	ART 2	Low	7	23.29	7.85	2.97	(16.03, 30.54)
		High	13	39.54	5.09	1.41	(36.46, 42.62)
	ART 3	Low	12	20.92	7.59	2.19	(16.10, 25.74)
		High	8	41.50	5.98	2.11	(36.50, 46.50)

Correct Rejections

A two-way between-groups ANOVA revealed no significant interaction between WMC and ART on Correct Rejection scores, $F(2,54) = 0.89$, $p = .418$, $\eta_p^2 = .03$, nor any significant main effect of WMC on Correct Reject scores, $F(1,54) = 0.19$, $p = .664$, $\eta_p^2 = .00$.

SA scores

A two-way between-groups ANOVA revealed no significant interaction between WMC and ART on SA2 scores, $F(2,54) = 1.64$, $p = .203$, $\eta_p^2 = .06$, nor any significant main effect of WMC on SA2 scores, $F(1,54) = 1.51$, $p = .224$, $\eta_p^2 = .03$. There was no significant interaction between WMC and ART on SA3 scores, $F(2,54) = 0.42$, $p = .661$,

$\eta_p^2 = .02$, nor any significant main effect of WMC on SA3 scores, $F(1,54) = 2.36$, $p = .131$, $\eta_p^2 = .04$.

Discussion

The primary goal was to examine how the transparency of an intelligent agent's reasoning in a low information environment affected complacent behavior in a route selection task. Participants supervised a three-vehicle convoy as it traversed a simulated environment and re-routed the convoy when needed with the assistance of an intelligent agent, RoboLeader (RL). Information regarding potential events along the pre-planned route, together with communications from a commander confirming either the presence or absence of activity in the area, were provided to all participants. They did not receive any information about the suggested alternate route. However, they were instructed that the proposed path was at least as safe as their original route. When the convoy approached a potentially unsafe area, the intelligent agent would recommend re-routing the convoy. The agent recommendations were correct 66% of the time. The participant was required to recognize and correctly reject any incorrect suggestions. The secondary goal of this study was to examine how differing levels of agent transparency affected main task and secondary task performance, response time, workload, SA, trust, and system usability, along with implications of individual difference factors such as spatial ability, WMC, PAC, and complacency potential (CP).

Each participant was assigned to a specific level of agent reasoning transparency (ART). The reasoning was provided as to why the agent was making the recommendation and this differed among these levels. ART1 provided no reasoning information; RL notified that a change was recommended without explanation. The type of information the agent supplied varied slightly between ARTs 2 and 3. In ART2 the agent reasoning was a simple statement of fact (e.g. recommend revise convoy route due to Potential IED). In ART3 an additional piece of reasoning information was added, which conveyed when the agent had received the information leading to its recommendation (e.g. recommend revise convoy route due to Potential IED, TOR: 1 [hour]). This additional information did not convey any confidence level or uncertainty but was designed to encourage the operator to evaluate the quality of the information actively rather than simply respond. Therefore, not only was access to agent reasoning examined, but the impact of the type of information the agent supplied was examined as well.

Complacent behavior was examined via primary (route selection) task response, in the form of incorrect acceptances of the agent recommendation, an objective measure of errors of commission (Parasuraman et al., 2000). As predicted, access to agent reasoning reduced these incorrect acceptances and increased access to agent reasoning increased incorrect acceptances. The number of incorrect acceptances was highest when no agent reasoning was available. When the amount of agent reasoning was increased to its highest level, the number of incorrect acceptances increased to nearly the same level as in the no-reasoning condition. This pattern of results indicates that while access to agent reasoning in a decision-supporting agent can counter complacent behavior, too much information resulted in an out-of-the-loop (OOTL) situation and increased complacent behavior. Similar to previous findings (Mercado et. al., 2015) access to agent reasoning did not increase response time. In fact, decision times were reduced in the agent reasoning condition, even though the agent messages in the reasoning conditions were longer than in the no reasoning condition and therefore should require slightly more time to process. Similar studies have suggested that a reduction in accuracy with consistent response times could be attributed to a speed-accuracy trade-off (Wickens et al. 2015). However, the present findings indicated that may not be the case. Initially, there was an increase in accuracy with no accompanying increase in response time (hence no trade-off). What appears to be more likely is that not only does the access to agent reasoning assist the operator in determining the correct course of action, but the type of information the operator receives also influences their behavior.

In all conditions, the operator received all information needed to correctly route the convoy without the agent's suggestion. In the no reasoning condition, the operators were less likely to override the incorrect agent suggestion, demonstrating a clear bias for the agent suggestion. With a moderate amount of information regarding the agent reasoning, the operators were more confident in identifying and overriding erroneous recommendations. In the highest reasoning transparency condition, operators were also given information regarding when the agent had received the information used to generate the recommendation (i.e. its recency). While this information did not imply any confidence or uncertainty pertaining to the agent recommendation, such additional information appeared to encourage more complacent behavior in the operators. This may be due to the increased difficulty in assimilating the additional information, however it is more likely that this type of information was ambiguous, and this ambiguity appeared to encourage the operators to defer to the agent suggestion even when incorrect.

Performance on the route selection task was evaluated via correct rejections and acceptances of the agent suggestion. An increased number of correct acceptances and rejections, as well as reduced response times were all indicative of improved performance. Route selection performance was anticipated as improving with access to agent reasoning and then decline as access to agent reasoning increased. This hypothesis was partially supported. Performance did improve when access to agent reasoning was provided. Increased transparency of agent reasoning did result in a subsequent decline in scores, however the small-medium effect size indicated these results are not strong evidence in support of the latter demand of the hypothesis. Spatial visualization (SV) was predictive of performance on the route selection task. Individuals with High SV scores outperformed their Low SV counterparts on the route selection task in ART2. This demonstrated their advantage in assimilating the agent reasoning information supplied in this condition. However, this advantage was lost when additional reasoning in ART3 was supplied.

Access to agent reasoning did not increase overall operator workload as hypothesized. Workload was evaluated using the NASA-TLX and several ocular indices shown to be informative as to cognitive workload. Global NASA-TLX scores were lower in ART3 than in ARTs 1 and 2, but such changes were not significant. Pupil Diameter (PD), Fixation Count (FC) and Duration (FD) did not differ significantly between the three ARTs. This contradicts the stated hypothesis, and could be evidence that there is no difference in perceived cognitive workload between ARTs. However, examination of the NASA-TLX subscales tells a somewhat different story. The ratings for factors Temporal Demand (i.e. ‘How much time pressure did you feel due to the rate or pace at which the task or tasks elements occurred?') and Frustration ('How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?') were relatively consistent between in ARTs 1 and 2, but dropped off significantly in ART3. Interestingly, 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?') was much higher in ART3 than in ART2. These findings appear to be at odds with one another, however this combination of results may be an indication of increased workload in ART3. As the experiment was delivered via computer simulation, the only sort of physical demand that a participant could encounter would be scanning the OCU for information. Typically, higher physical demand coupled with reduced frustration and temporal demand should result in improved performance. However, the number of incorrect acceptances increased in ART3 from ART2, and participants reported the least satisfaction in their [perceived] Performance (i.e. 'How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?') in ART3. Considered alongside the other findings, the subjective performance rating may be indicative of their awareness of their actual performance in ART3, in that the participant is aware that they are essentially off-loading the route selection task in favor of other tasks. Increased scanning does little to improve performance on the route selection task, however it is key to improved SA1 scores and target detection task performance. While there was no difference in target detection performance between ARTs, SA1 scores in ART3 were much higher than in the other conditions. This observation tends to support the findings of increased complacency in this ART. These findings also indicate that although incorrect acceptances were greatest in ARTs 1 and 3, the reasons behind such may be different. While the incorrect acceptances in ART1 may be due to high workload encouraging complacent behavior, the incorrect acceptances in ART3 may be due to more complex reasons than simply higher workload.

Situation Awareness (SA) scores were hypothesized to improve with access to agent reasoning; this with the exception of SA3 scores in ART3. In this study, SA1 scores evaluated how well the participant maintained a general awareness of their environment, with the idea that increased access to agent reasoning would also give the participant context for events within their environment, thus making certain events and situations more salient. Those who were more successful at this integration would then show improved performance on the route selection task, as well as improved SA2 scores (Hancock and Diaz 2002). SA1 scores did not improve with access to agent reasoning. However with increased agent reasoning transparency SA1 scores improved substantially. This may indicate that additional access to reasoning allowed participants more time to monitor their environment. However, since there was also a reduction in performance on the route selection task, as well as demonstrated complacent behavior in ART3, it is more likely that the improvement in SA1 scores was a result of neglecting duties on other tasks (i.e., an inter-task trade-off). There were no significant difference in SA2 (comprehension) scores between ARTs. However SA3 scores did show a significant upward trend across ARTs. This suggests that, while access to agent reasoning does not improve comprehension, it could incrementally improve an operator's ability to predict future outcomes. In previous studies, increased autonomous assistance did result in improved SA (see Wright et al. 2013). However, the present findings indicate access to agent reasoning does little to improve SA. There were differences in SA scores dependent upon the individual difference factor spatial visualization (SV). High SV individuals had higher SA1 and SA3 scores than their low SV counterparts. This was most likely due to their increased ability to scan their environment (Lathan and Tracey 2002; Chen et al. 2008; Chen et al. 2010).

Access to agent reasoning appeared to have little influence on performance on the target detection task. There were no significant differences in the mean number of targets correctly detected across ART. However, access to agent reasoning did mitigate the number of false alarms reported. Signal Detection Theory measured whether access to agent reasoning had any effect on sensitivity or selection criteria. Sensitivity to targets, assessed as d', appeared to be slightly lower in the no reasoning condition. Selection criteria were also lower in the no reasoning condition. Thus, participants appeared to use a higher selection criterion when targets were more readily identifiable, and then subsequently loosened their selection bias when target sensitivity was lower. This pattern of behavior could explain the greater number of false alarms reported in the no reasoning condition.

Operator trust in the agent was assessed objectively by evaluating incorrect rejections of the agent's suggestions when the agent was correct, along with the associated decision times, and subjectively using the Usability and Trust Survey. The performance measure of operator trust indicated potentially higher trust in the notification-only and the highest agent reasoning transparency conditions, as the number of acceptances (no incorrect rejections) was double that of the moderate agent reasoning transparency condition, with no associated difference in decision time. However, when a moderate amount of agent reasoning was available, there were fewer acceptances, even though the overall score was roughly the same as the other conditions, and decision times in this condition were longer for incorrect rejections than for the correct acceptances. This could be evidence of trustful behavior when agent reasoning transparency was present.

Subjective measures also indicated access to agent reasoning reduced trust and usability evaluations. Increased transparency of agent reasoning resulted in increased trust and usability ratings. However there was no associated overall improvement in performance. Interestingly, operators reported highest trust and usability in the conditions that also had the highest complacency and lowest in the condition that had the highest performance. In the conditions when the agent reasoning was not transparent, and when the agent reasoning was highly transparent, the participant' trust and usability evaluations were highest (albeit for potentially different reasons), even though they knew the agent was not completely reliable. However, in the condition with a moderate amount of agent reasoning transparency, the participants reported lower trust and usability, indicating they were more critical of the agent recommendations in this condition, resulting in reduced complacency and improved performance.

A potential limitation of this work could be the added time information in ART3. While the participants in that agent reasoning condition were instructed that the time reflected when the agent received the information upon which it based its recommendation, they were not instructed how they should use that information in their deliberations. Thus, this information could have appeared ambiguous to the participants, and there could be variability in how they factored this information into their decision based on their personal experience.

Conclusion

The findings of the present study are important for the design of intelligent recommender and decision-aid systems. Keeping the operator engaged and in-the-loop is important for reducing complacency, which could allow lapses in system reliability to go unnoticed. To that end, how agent reasoning transparency affected complacent behavior, as well as task performance and trust, was examined. Access to agent reasoning appears to be an effective deterrent to complacent behavior when the operator has limited information about their task environment. Contrary to the position adopted by Paradis et al. (2005), operators do accept agent recommendations even when they do not know the rationale behind the suggestions. In fact, the absence of agent reasoning appears to encourage complacent behavior. Access to the agent's reasoning appears to allow the operator to calibrate their trust in the system effectively, reducing incorrect acceptances and improving performance. This outcome is similar to findings previously reported by Helldin et al. (2014) and Mercado et al. (2015). However, the addition of information that created ambiguity for the operator again encouraged complacency, resulting in reduced performance and poorer trust calibration. Prior work has shown that irrelevant or ambiguous information can increase workload and encourage complacent behavior (Chen and Barnes, 2014; Westerbeek and Maes, 2013), and these findings align with those. As such, caution should be exercised when considering how transparent to make agent reasoning and what information should be included.

Similar to Mercado et al. (2015), access to agent reasoning did not increase response time or workload. In addition, the presence of agent reasoning appears to have positively affected performance on the secondary target detection task. While the overall number of targets detected did not differ among conditions, the selection criterion appeared to have been higher in the agent reasoning conditions, resulting in fewer reported false alarms. While increased false alarms may be beneficial in some settings, this task environment was non-combat and suburban, and a false alarm meant erroneously identifying a civilian or friendly soldier as a potential enemy combatant. While the route selection task and target detection tasks were not simultaneous, it appears as though the presence of agent reasoning allowed the participants to focus better on the target detection task between route decision locations.

Future Work

This work represents the first of two studies exploring the effect of agent transparency on complacent behavior. In the follow-up study, the amount of information the operator has regarding the task environment will be increased. As a result of this increase, the amount of agent reasoning provided will also be increased to incorporate additional information into agent recommendations. This will allow comparison of differences in operator complacency and performance due to further operator knowledge of their task environment as well as that as a result of greater access to agent reasoning.

CHAPTER 4: EXPERIMENT 2

Study Overview

Experiment 2 investigated how access to the agent's reasoning affected the human operator's decision-making, task performance, situation awareness, and complacent behavior in a multitasking environment when additional, sometimes competing, environmental information is available. It differed from Experiment 1 in two ways: first, the level of environmental information was increased, and second, the degree of agent reasoning transparency, when available, was increased. Environmental information was displayed by icons appearing on the map, with events affecting both the original route and the proposed alternative displayed. Agent reasoning transparency (ART) was manipulated via RoboLeader's detailed notifications, which were expanded from experiment 1 to include each of the icons affecting the area, along with weighing information as to how each event was factored into RL's recommendation.

Hypotheses

Based on the review of the literature, this proposal posits the following hypotheses:

It is hypothesized that access to agent reasoning would reduce complacent behavior, improve task performance, and increase trust in the agent, and increased access to agent reasoning would increase complacent behavior, negatively impact performance, and reduce trust in the agent. Although decision time decreased with the access to agent reasoning in EXP1, the increase in agent transparency in this study was expected to increase decision time (aside from clearly complacent behavior): ART1 < ART2 < ART3. Unlike EXP1, RL's messages were considerably longer in ARTs 2 and 3 than in ART1, as such additional time was expected to be required for reading the messages. Participants were expected to take longer to process the information and reach their decision, resulting in a longer decision time. Shorter response times may indicate less deliberation on the part of the operator before accepting or rejecting the agent recommendation. This could mean either positive complacent behavior or reduced task difficulty.

> Hypothesis 1: Access to agent reasoning will reduce incorrect acceptances, ART1 > ART2, and increased transparency of agent reasoning will increase incorrect acceptances, ART2 < ART3. When agent reasoning is not available, incorrect acceptances will be greater than when agent reasoning is present, ART1 > ART2+3.

> Hypothesis 2: Access to agent reasoning will improve performance (number of correct rejects and accepts) on the route selection task, ART1 < ART2, and increased transparency of agent reasoning will reduce performance on the route selection task, ART2 > ART3. When agent reasoning is not available, performance will be lower than when agent reasoning is present, ART1 < ART2+3.

> Hypothesis 3: Access to agent reasoning will increase operator trust in the agent, ART1 < ART2, and increased transparency of agent reasoning will decrease operator trust in the agent, ART2 > ART3.

It is hypothesized that increasing agent reasoning transparency will, in turn, increase the operators’ workload. In EXP1, increased access to agent reasoning reduced operator perceived workload. However, in this study, as the agent reasoning becomes more transparent the amount of information the operator must process has increased considerably from that presented in EXP1. It is expected that this increased mental demand will be reflected in the workload measures.

> Hypothesis 4: Access to agent reasoning will increase operator workload, ART1 < ART2; and increased transparency of agent reasoning will increase operator workload, ART2 < ART3. When agent reasoning is not available, workload will be lower than when agent reasoning is present, ART1 < ART2+3.

It is hypothesized that agent reasoning transparency will support the operators' situation awareness (SA). Access to the agent reasoning will help the operator better comprehend how objects/events in the task environment affect their mission, thus informing their task of monitoring the environment surrounding the convoy and making them cognizant of potential risks. This understanding will also enable them to make more accurate projections regarding the future safety of their convoy. However, the addition of information that appears ambiguous to the operator will have a detrimental effect on both their ability continuously monitor their environment, as well as their ability to correctly project future status.

> Hypothesis 5: Access to agent reasoning will improve SA scores, and increased transparency of agent reasoning will improve SA2 scores, but will reduce SA1 and SA3 scores;

- SA1: ART1 < ART2, ART2 > ART3;
- SA2: ART1 < ART2, ART2 < ART3;
- SA3: ART1 < ART2, ART2 > ART3.

It is hypothesized that increasing agent reasoning transparency will reduce performance on the target detection task. The increased mental demand on the operator will affect their ability to monitor the environment for threats effectively. The increased amount of environmental information will also affect the operators' selection bias, resulting in increased false alarms.

> Hypothesis 6: Access to agent reasoning will reduce performance on the target detection task (fewer targets detected, higher FAs), ART1 > ART2, and increased transparency of agent reasoning will further reduce performance on the target detection task; ART2 > ART3.

The effects of individual differences in complacency potential, perceived attentional control, spatial ability, and working memory capacity on the operator's task performance, trust, and SA was also investigated. While the results of EXP1 did not always show differences due to ID factors, it is expected that those results occurred because the operators did not experience as heavy a cognitive load as expected. If that is the case, then the increased amount of environmental information and agent reasoning present in EXP2 should increase the cognitive burden, and differences due to ID factors will become apparent.

Hypothesis 7: High CPRS individuals will have fewer correct rejects on the route planning task than Low CPRS individuals.

Hypothesis 8: High CPRS individuals will have higher scores on the usability and trust survey than Low CPRS individuals.

Hypothesis 9: High CPRS individuals will have lower SA scores than Low CPRS individuals.

Hypothesis 10: Individual differences, such as SpA and PAC, will have differential effects on the operator's performance on the route selection task and their ability to maintain SA.

Hypothesis 11: High WMC individuals will have more correct rejects and higher SA2 and SA3 scores than Low WMC individuals.

Overall summaries of expected findings for experiment 2 are shown in Table 17.

Table 17. Anticipated patterns of findings (hypotheses) for Experiment 2. Indicates expected score or performance across agent reasoning transparency conditions (i.e., ART1, ART2, and ART3).

	DV Measure	Add Transparency	Increase Transparency
Route Selection Task (RS)	Correct accepts and rejects	ART 1 < ART 2	ART 2 > ART 3
	Decision Time	ART 1 < ART 2	ART 2 < ART 3
Target Detection Task (TD)	Targets Detected	ART 1 > ART 2	ART 2 > ART 3
	False Alarms	ART 1 < ART 2	ART 2 < ART 3
Complacent Behavior	Incorrect Acceptances	ART 1 > ART 2	ART 2 < ART 3
Situation Awareness Scores	SA1 Queries (Perception)	ART 1 < ART 2	ART 2 > ART 3
	SA2 Queries (Comprehension)	ART 1 < ART 2	ART 2 < ART 3
	SA3 Queries (Projection)	ART 1 < ART 2	ART 2 > ART 3
Workload	Global NASA-TLX	ART 1 < ART 2	ART 2 < ART 3
Trust	Incorrect Rejections	ART 1 > ART 2	ART 2 < ART 3
	Usability and Trust Survey	ART 1 < ART 2	ART 2 > ART 3

Task Environment

There was no change in either the simulation scenario or RoboLeader capabilities from Experiment 1.

Methodology

Experimental Participants

Seventy-three participants (between the ages of 18 and 44) were recruited from the Institute for Simulation and Training's and the Psychology Departments' SONA Systems. Participants received their choice of compensation: either cash payment ($15/hr) or Sona Credit at the rate of 1 credit/hour. Thirteen potential participants were excused or dismissed from the study; eight were dismissed early due to equipment malfunctions, one withdrew during training claiming they did not have time to participate, two fell asleep during their session and were dismissed, one could not pass the training assessments and was dismissed, and one did not pass the color vision screening test and was dismissed. Those who were determined to be ineligible or withdrew from the experiment received payment for the amount of time they participated, with a minimum of one hours' pay. Sixty participants (21 males, 39 females, Min_{age} = 18 years, Max_{age} = 44 years, M_{age} = 21.0 years) successfully completed the experiment and their data was used in the analysis.

Experimental Apparatus

The simulator and eye tracker were the same as in Experiment 1.

Surveys and Tests

All surveys, questionnaires, and tests were the same as in Experiment 1 (EXP1). Descriptive statistics pertaining to Experiment 2 individual differences (ID) measures are listed here. Since the ID measures were dichotomized into High/Low groups similarly to those in EXP1, these groups were also compared between experiments to ensure consistent delineation between high and low group scores. For each ID measure, the high and low groups were found to be distinct from one another, and this difference was consistent between EXPs 1 and 2.

Perceived Attentional Control. High/Low group membership was determined by median split of all participants' scores (Min_{PAC} = 33, Max_{PAC} = 75, Mdn_{PAC} = 58, M_{PAC} = 57.6, SD_{PAC} = 8.16; PAC_{LOW} N = 29, PAC_{HIGH} N = 31).

Cube Comparison Test. High/Low group membership was determined by median split of all participants' scores (Min_{SV} = 0.19, Max_{SV} = 0.88, Mdn_{SV} = 0.50, M_{SV} = 0.52, SD_{SV} = 0.14, SV_{LOW} N = 27, SV_{HIGH} N = 33).

Spatial Orientation Test. High/Low group membership was determined by median split of all participants' scores (Min_{SOT} = 3.96, Max_{SOT} = 50.60, Mdn_{SOT} = 11.19, M_{SOT} = 13.79, SD_{SOT} = 8.48, SOT_{LOW} N = 27, SOT_{HIGH} N = 34).

Complacency Potential Rating Scale. High/Low group membership was determined by median split of all participants' scores (Min_{CPRS} = 25, Max_{CPRS} = 47, Mdn_{CPRS} = 37, M_{CPRS} = 36.8, $CPRS_{LOW}$ N = 28, $CPRS_{HIGH}$ N = 32).

RSPAN. Working memory capacity (WMC) was evaluated by using the participants' total letter set score (sum of all perfectly recalled letter sets), with higher numbers indicating greater working memory capacity, (Min_{RSPAN} = 10.0, Max_{RSPAN} = 54.0, Mdn_{RSPAN} = 31.0, M_{RSPAN} = 31.5, SD_{RSPAN} = 12.1). High/Low group membership was determined by median split of all participants' scores, $RSPAN_{LOW}$ N = 29, $RSPAN_{HIGH}$ N = 31.

Procedure

The procedure and experimental design were the same as in Experiment 1, with the following exception:

When RoboLeader suggested a route revision, in addition to the knowledge of the event potentially affecting their primary route, participants received information regarding potential events that could affect the alternate route. RoboLeader messages in ARTs 2 and 3 included details about events denoted by the map icons for both primary and alternate routes, as well as weighing factors illustrating how RoboLeader used this information in its recommendation. Transcripts of RoboLeader messages for each ART are in Appendix J.

Results

Complacent Behavior Evaluation

> Hypothesis 1: Access to agent reasoning will reduce incorrect acceptances, ART1 > ART2, and increased transparency of agent reasoning will increase incorrect acceptances, ART2 < ART3. When agent reasoning is not available, incorrect acceptances will be greater than when agent reasoning is present, ART1 > ART2+3.

Descriptive statistics for incorrect acceptances and decision times at the locations where the agent recommendation should have been rejected are shown in Table 18.

Table 18. Descriptive statistics for incorrect acceptances and decision times, sorted by agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean
Incorrect Acceptances	ART 1	20	1.00	1.17	0.26	(0.45, 1.55)
	ART 2	20	0.90	0.91	0.20	(0.47, 1.33)
	ART 3	20	1.50	1.64	0.37	(0.73, 2.27)
Overall DT at Reject Locations (sec)	ART 1	20	11.14	3.68	0.82	(9.42, 12.87)
	ART 2	20	11.51	3.35	0.75	(9.94, 13.08)
	ART 3	20	12.30	3.96	0.89	(10.45, 14.16)
DT Correct Rejects (sec)	ART 1	20	10.84	3.45	0.77	(9.23, 12.45)
	ART 2	20	11.25	3.19	0.71	(9.75, 12.74)
	ART 3	20	12.52	4.21	0.94	(10.55, 14.49)
DT Incorrect Accepts (sec)	ART 1	11	12.17	5.76	1.74	(8.30, 16.05)
	ART 2	12	14.37	4.49	1.30	(11.51, 17.22)
	ART 3	12	12.39	4.60	1.33	(9.46, 15.31)

Working Memory Capacity (WMC) score correlated significantly with incorrect acceptances (r = -.28, p = .015), and was found to be a significant predictor of incorrect acceptances, in that participants with lower WMC had more incorrect acceptances than those with greater WMC, R^2 = .079, b = -0.03, t(58) = -2.23, p = .029. WMC scores did not have a consistent relationship with incorrect acceptances for each ART group (heterogeneity of regression), as such was not included as a covariate in subsequent analyses.

There was unequal variance between groups for incorrect acceptances, so Welch's statistic was reported, and contrast tests do not assume equal variance between conditions. A between-groups ANOVA was conducted to assess the effect of ART on incorrect acceptances, and no significant effect was found, F(2,36.23) = 1.04, p = .373, ω^2 = .01 (Figure 29). Planned comparisons revealed the number of incorrect acceptances were lower in ART2 than in ART1, $t(35.9) = -0.30$, $p = .765$, $r_c = .05$, and ART3, $t(29.7) = 1.43$, $p = .163$, $r_c = .25$, however these differences were not significant.

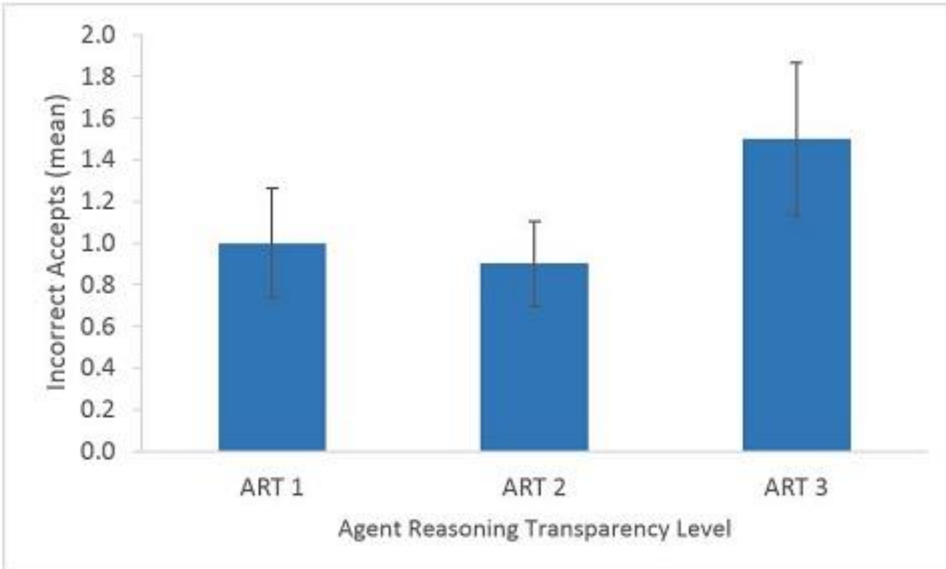
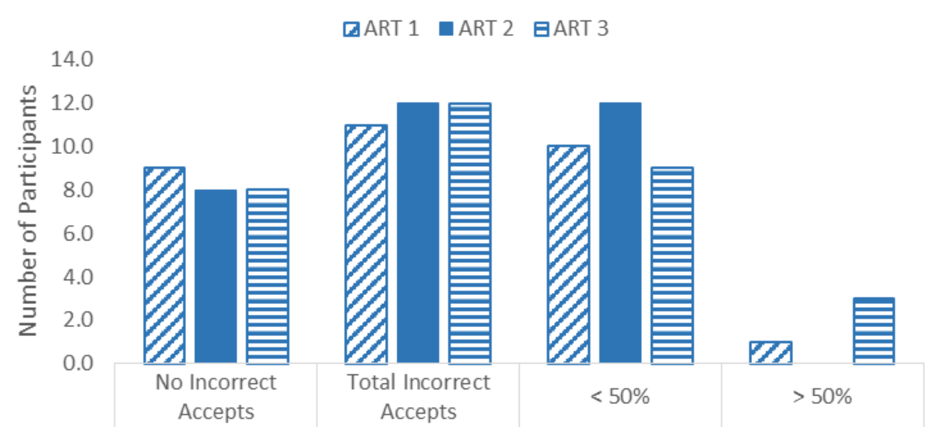


Figure 29. Average number of incorrect acceptances by agent reasoning transparency (ART) level. Bars denote SE.

Participants' scores were further analyzed by the number of incorrect acceptances per ART level (Figure 30). Chi-square analysis found no significant effect of ART on the number of incorrect acceptances, $X^2(10) = 7.36$, $p = .692$, Cramer's $V = .248$. Across all ART levels, 25 participants had no incorrect acceptances, and these were (roughly) equally distributed between ARTs, indicating that the addition of agent reasoning had no more effect on performance than operator knowledge alone. The range of potential scores for incorrect acceptances was 0 – 6, and the range of participants' scores was 0 – 5. Thirty-five participants had at least 1 incorrect acceptance, and these scores were sorted into groups; < 50% (score 3 or less) or > 50% (score 4 or higher). The participants who made incorrect acceptances appeared to be evenly distributed among ARTs. Of these, 31 out of

35 participants scored under 50%. This is evidence that ART had little to no effect on the number of incorrect acceptances. It is interesting to note that no participants in ART2 had more than 3 incorrect acceptances. However, of the participants who had >50% incorrect acceptances, these were mostly in ART3, which could be an indication that too much access to agent reasoning can have a detrimental effect on performance.



	No Incorrect Accepts	Total Incorrect Accepts	< 50%	> 50%
ART 1	9	11	10	1
ART 2	8	12	12	0
ART 3	8	12	9	3

Figure 30. Distribution of number of incorrect acceptances across agent reasoning transparency (ART) level.

As in Experiment 1, decision time for responses at the locations where the agent recommendation was incorrect was evaluated as a potential indicator of complacent behavior. It was hypothesized that decision time would increase as agent reasoning transparency increased, as participants should require additional time to process the extra information, particularly in Experiment 2 as the text conveying agent reasoning in ARTs 2 and 3 was much longer than the notification presented in ART1 (see Appendix J). Thus, reduced time could indicate less time spent on deliberation, which may infer complacent behavior. In addition to the overall time to respond, decision times for correct rejects and incorrect accepts were also examined (Figure 31). There was not a significant effect of ART on overall decision time, $F(2, 57) = 0.52$, $p = .597$, $\omega^2 = .02$ (Figure 32). Overall decision time was slightly shorter in ART1 than in ART2, $t(57) = 0.31$, $p = .755$, $r_c = .04$, and slightly shorter in ART2 than in ART3, $t(57) = 0.68$, $p = .497$, $r_c = .09$. There was not a significant effect of ART on decision time for correct rejections, $F(2,57) = 1.56$, $p = .322$, $\omega^2 = .01$. Mean decision times for correct rejections were slightly shorter in ART1 than in ART2, $t(57) = 0.36$, $p = .724$, $r_c = .05$, and shorter in ART2 than in ART3, $t(57) = 1.10$, $p = .275$, $r_c = .14$. There was not a significant main effect of ART on decision time for incorrect acceptances, $F(2,32) = 0.70$, $p = .504$, $\omega^2 = .02$. Mean decision times for incorrect acceptances were longer in ART2 than in ART1, $t(32) = 1.06$, $p = .297$, $r_c = .18$, and in ART3, $t(32) = -0.98$, $p = .336$, $r_c = .17$. Decision times remained relatively unchanged across ART levels, however in ART2 decision times for incorrect acceptances were longer than decision times for correct rejects. This is evidence that these incorrect responses were most likely due to errors in judgement rather than complacent behavior. Paired t-tests were used to compare differences between decision times for correct and incorrect responses within each ART. The largest difference in decision time was in ART2, $t(11) = -1.57$, $p = .146$, $d = 0.47$, while times in ART1, $t(10) = -1.38$, $p = .198$, $d = 0.34$, and ART3, $t(11) = 0.62$, $p = .551$, $d = 0.12$ were more consistent. Although these results did not achieve statistical significance, it is interesting to note that decision times between correct and incorrect responses are similar in ARTs 1 and 3, while those in ART2 indicate that participants in this condition spent more time in deliberation when their response was incorrect than when it was correct, and the medium effect size indicates this difference is meaningful.

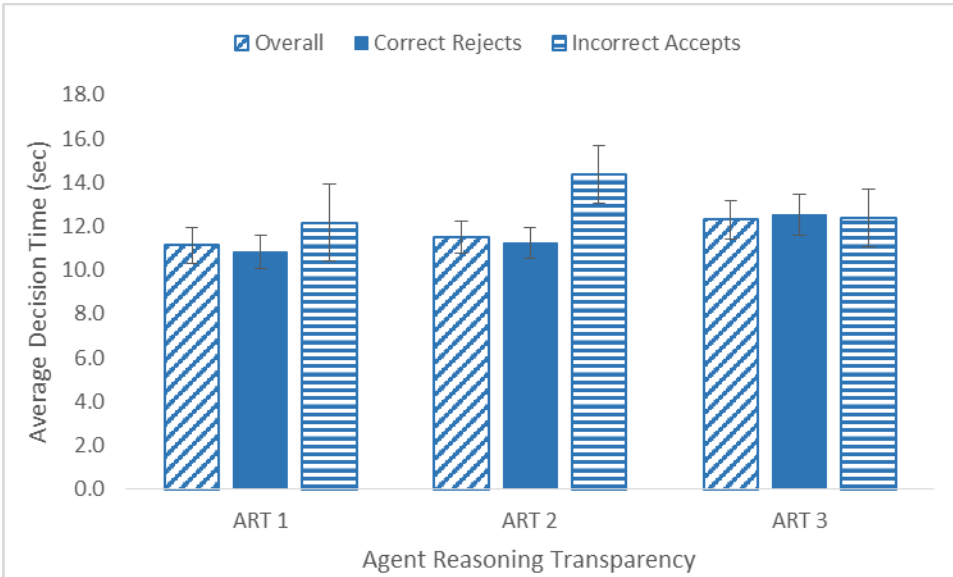


Figure 31. Average decision time in seconds for participant responses at decision points where the agent recommendation was incorrect. Decision times are shown for all responses (overall), correct rejections, and incorrect acceptances, sorted by agent reasoning transparency (ART) level. Bars denote SE.

Route Selection Task Performance

Hypothesis 2: Access to agent reasoning will improve performance (number of correct rejects and accepts) on the route selection task, ART1 < ART2, and increased transparency of agent reasoning will reduce performance on the route selection task, ART2 > ART3. When agent reasoning is not available, performance will be lower than when agent reasoning is present, ART1 < ART2+3.

Descriptive statistics for route selection task scores and decision times for all decision points across three missions are shown in Table 19.

Table 19. Descriptive statistics for route selection scores and decision times, sorted by agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean
Route Selection Score	ART 1	20	13.20	3.46	0.77	(11.58, 14.82)
	ART 2	20	13.30	3.18	0.71	(11.81, 14.79)
	ART 3	20	13.40	3.28	0.73	(11.86, 14.94)
Overall Decision Time (sec)	ART 1	20	10.86	3.04	0.68	(9.44, 12.28)
	ART 2	20	12.53	3.09	0.69	(11.08, 13.97)
	ART 3	20	12.52	4.91	1.10	(10.22, 14.81)
DT Correct Responses (sec)	ART 1	20	10.32	2.79	0.62	(9.02, 11.63)
	ART 2	20	11.95	3.40	0.76	(10.36, 13.54)
	ART 3	20	11.79	3.98	0.89	(9.33, 13.65)
DT Incorrect Responses (sec)	ART 1	20	13.06	5.39	1.21	(10.54, 15.59)
	ART 2	19	15.21	3.05	0.70	(13.74, 16.68)
	ART 3	17	12.65	4.39	1.07	(10.40, 14.91)

Complacency Potential Rating Scale (CPRS) scores (r = -.37, p = .002) and SOT scores (r = -.25, p = .025) correlated significantly with route selection scores, and were found to be significant predictors of route selection scores. Participants who scored higher on the CPRS, indicating a greater potential to demonstrate complacent behavior when interacting with automation, performed worse on the route selection task than their counterparts, R^2 = .138, b = -.276, t(58) = -3.04, p = .004. Participants who scored lower on the SOT demonstrate greater spatial orientation abilities, and also performed better on the route selection task than their counterparts, R^2 = .064, b = -.111, t(58) = -2.00, p = .051. However, neither CPRS scores nor SOT scores had a consistent relationship with route selection scores across ART groups (heterogeneity of regression), and as such were not included as covariates in subsequent analyses.

A between-groups ANOVA was conducted to assess the effect of ART on route selection scores and found no significant effect, $F(2,57) = 0.02$, $p = .982$, $\omega^2 = .03$. Planned comparisons revealed route selection scores were higher in ART2 than in ART1, $t(57) = 0.10$, $p = .924$, $r_c = .01$, and higher in ART3 than in ART2, $t(57) = 0.10$, $p = .924$, $r_c = .01$. The results trended as predicted, however they were not significant.

Examining the distribution of scores, the potential range of scores for the route selection task was 0 – 18, and the range of participants' scores was 7 – 18 (Figure 32). Of these, 4 participants scored 18/18, 3 of whom were in ART3. Only 9 participants scored 50% or less, the majority scored 67% or higher. For comparative purposes, scores were sorted into similar groups as in experiment 1 (i.e. 17 – 15, 14 – 12, > 12). Interestingly, scores in each ART appear to be nearly evenly distributed between the groups. This does not offer support for the hypothesis, as performance in the agent reasoning conditions appears to be no better than in the notification-only condition.

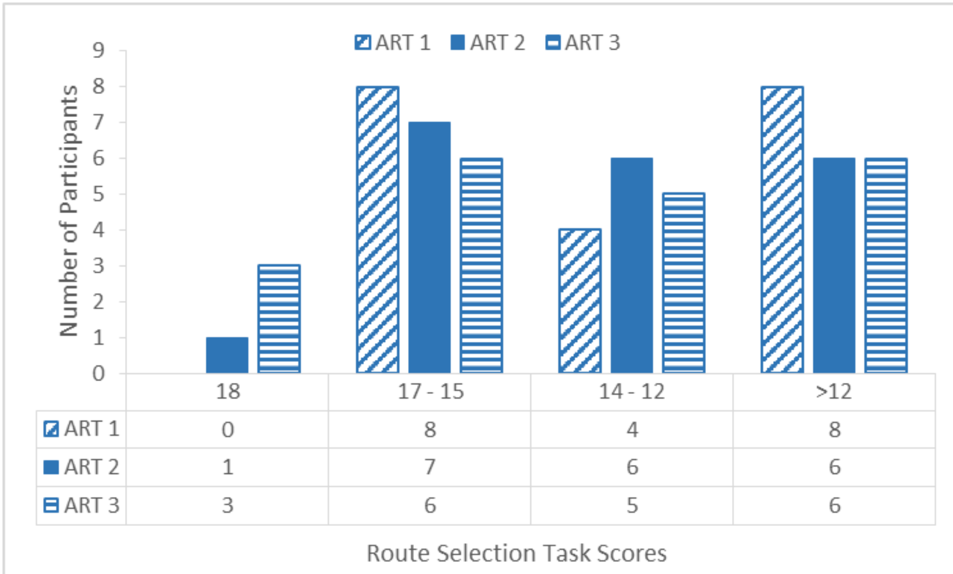
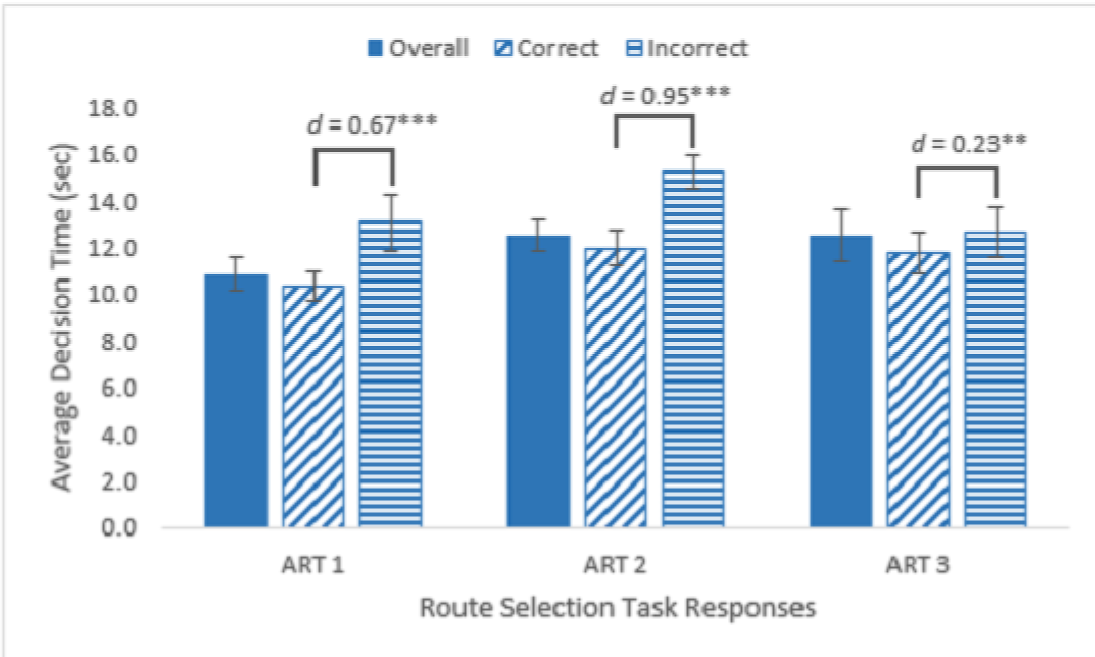


Figure 32. Distribution of scores for the route selection task across agent reasoning transparency (ART) levels.

Overall decision time for route selection responses was evaluated using one-way ANOVA. The homogeneity of variance assumption was violated, as such, Welch's correction has been reported, and contrast tests do not assume equal variance between conditions. There was not a significant effect of ART on elapsed decision time, $F(2, 36.8) = 1.69$, $p = .198$, $\omega^2 = .01$. Planned comparisons revealed decision times were longer in ART2 than in ART1, $t(38.0) = 1.72$, $p = .094$, $r_c = .27$, but not significantly different than in ART3, $t(32.0) = -0.01$, $p = .996$, $r_c = .00$. Overall, decision times were longer in the conditions with agent reasoning than without (ART1 < ART2+3), $t(46.5) = 1.77$, $p = .083$, $r_c = .25$. These results were not significant, but they do follow the same pattern as those for the task performance evaluation.

Overall decision times for acceptances were compared to those for rejections [of the agent recommendation] using paired t-tests, and there was no significant difference,

$t(59) = -1.91$, $p = .061$, $d = 0.17$ across ART levels. Comparing decision times for correct responses to those for incorrect responses using paired t-tests, decision times for correct responses were significantly shorter than those for incorrect responses, $t(55) = -5.20$, $p < .001$, $d = 0.58$ (Figure 33) across ART levels. Within each ART, this difference was greater in ART 2, $t(18) = -3.61$, $p = .002$, $d = 0.95$, than in ART1, $t(19) = -3.21$, $p = .005$, $d = 0.67$, and smallest in ART3, $t(16) = -2.56$, $p = .021$, $d = 0.23$. Decision times for incorrect responses were evaluated between ARTs, and there was no significant difference between ART1 and ART2, $t(30.31) = 1.54$, $p = .134$, $d = 0.56$, or ART1 and ART3, $t(34.35) = -0.25$, $p = .802$, $d = 0.09$, and a marginally significant difference between ART2 and ART3, $t(28.11) = -2.00$, $p = .055$, $d = 0.76$. While not offering additional support for the hypothesis, the smaller difference in mean decision time for incorrect responses demonstrated in ART3 could be indicative of some participants' increased complacent behavior in the highest agent reasoning condition.



**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Figure 33. Comparison of average decision times for correct responses and incorrect responses, shown by agent reasoning transparency (ART) level. Bars denote SE.

Operator Trust Evaluation

Hypothesis 3: Access to agent reasoning will increase operator trust in the agent, ART1 < ART2, and increased transparency of agent reasoning will decrease operator trust in the agent, ART2 > ART3.

Descriptive statistics for incorrect rejections and the Usability and Trust Survey scores are shown in Table 20.

Table 20. Descriptive statistics for incorrect rejections and Usability and Trust Survey results, across agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean
Incorrect Rejections	ART 1	20	3.75	3.49	0.78	(2.12, 5.38)
	ART 2	20	3.80	2.76	0.62	(2.51, 5.09)
	ART 3	20	3.10	3.04	0.68	(1.68, 4.52)
Usability & Trust Survey	ART 1	20	91.30	19.29	4.31	(82.27, 100.33)
	ART 2	20	91.20	15.73	3.52	(83.84, 98.56)
	ART 3	20	93.60	13.03	2.91	(87.50, 99.70)
Usability Responses	ART 1	20	40.35	7.18	1.61	(36.99, 43.71)
	ART 2	20	39.45	6.05	1.35	(36.62, 42.28)
	ART 3	20	41.60	5.70	1.27	(38.93, 44.27)
Trust Responses	ART 1	20	50.95	13.08	2.92	(44.83, 57.07)
	ART 2	20	51.75	11.19	2.50	(46.51, 56.99)
	ART 3	20	52.00	8.61	1.93	(47.97, 56.03)

Scores on the complacency potential rating scale (CPRS) correlated significantly with incorrect rejections (r = .33, p = .005), and CPRS was found to be a significant predictor of incorrect rejections, R^2 = .110, b = 0.23, $t(58)$ = 2.67, p = .010. However, CPRS scores did not have a consistent relationship with incorrect rejections across ART groups (heterogeneity of regression), and as such were not included as covariates in subsequent analyses.

Examining the distribution of incorrect rejections at those locations where the agent recommendation was correct across ARTs, eleven participants had no incorrect rejections, and this number appears to be relatively even across ARTs (Figure 34). The range for potential scores for incorrect rejections was 0 – 12, and the range of participants' scores was 0 – 9. Forty-nine (49) participants had at least 1 incorrect rejection, and these scores were sorted into < 50% (score 5 or less) and > 50% (score 6 or higher). While scores in

ART1 appeared to near the rate for chance, the majority of scores in ARTs 2 and 3 were below 50%, indicating that access to agent reasoning was helpful in reducing incorrect rejections.

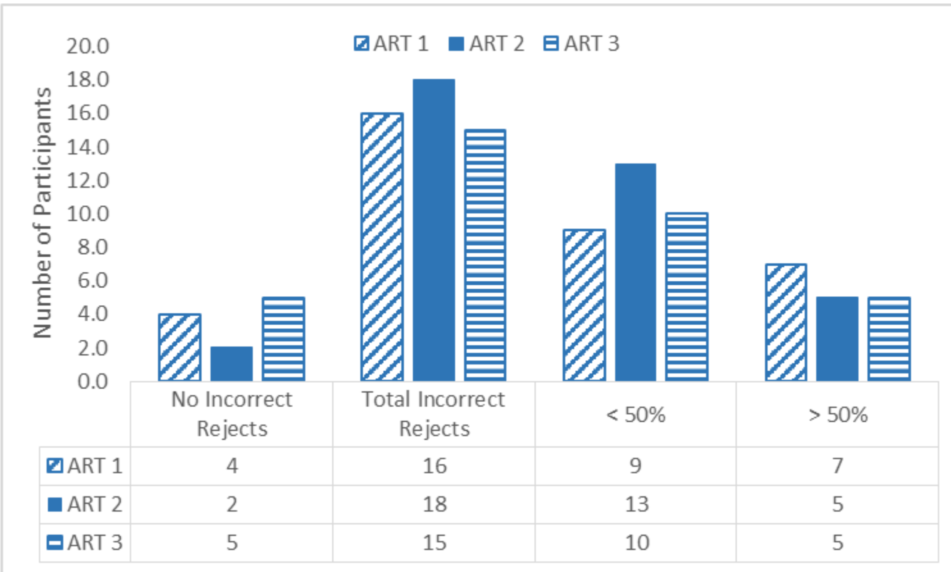


Figure 34. Distribution of scores for incorrect rejections, sorted by agent reasoning transparency (ART) level.

A between-groups ANOVA found no significant effect of ART on incorrect rejections, $F(2,57) = 0.32$, $p = .731$, $\omega^2 = .02$. Planned comparisons revealed incorrect rejections were higher in ART2 than in ART1, $t(57) = 0.51$, $p = .960$, $r_c = .01$, and ART3, $t(57) = -0.71$, $p = .480$, $r_c = .09$, however these differences were not significant.

As in Experiment 1, decision time for responses at the locations where the agent recommendation was correct was evaluated as a potential indicator of operator trust. It was hypothesized that decision time would increase as agent reasoning transparency increased, as participants should require additional time to process the extra information. Thus, increased time could indicate more time spent on deliberation, which may infer lower trust. In addition, decision times for incorrect rejections of the agent recommendation at those locations could be indicative of complacent behavior, i.e. reduced decision times for incorrect responses. There was not a significant effect of ART on overall decision time at the agent correct locations, $F(2, 57) = 2.03$, $p = .141$, $\omega^2 = .03$ (Figure 35). Planned comparisons show that overall decision times in ART2 were longer than those in ART1, $t(57) = 2.00$, $p = .051$, $r_c = .26$, but not significantly longer than those in ART3, $t(57) = -0.77$, $p = .445$, $r_c = .10$. Overall, decision times were longer in the conditions with agent reasoning access than in the notification only condition (ART1 – ART2+3), $t(57) = 1.86$, $p = .068$, $r_c = .24$. Decision times for correct accepts were significantly higher in the agent reasoning conditions than in the notification only condition (ART1 – ART2+3), $t(48.2) = 2.44$, $p = .018$, $r_c = .33$. Decision times for correct responses were shorter ART1 than in ART2, $t(37.4) = 2.48$, $p = .018$, $r_c = .38$, but not significantly different in ART2 than in ART3, $t(34.1) = -0.34$, $p = .736$, $r_c = .06$. Decision times for incorrect responses were longer in ART2 than in ART1, $t(31.0) = 1.45$, $p = .159$, $r_c = .25$, ($d = 0.52$), and significantly longer than in ART3, $t(31.0) = -2.21$, $p = .042$, $r_c = .36$.

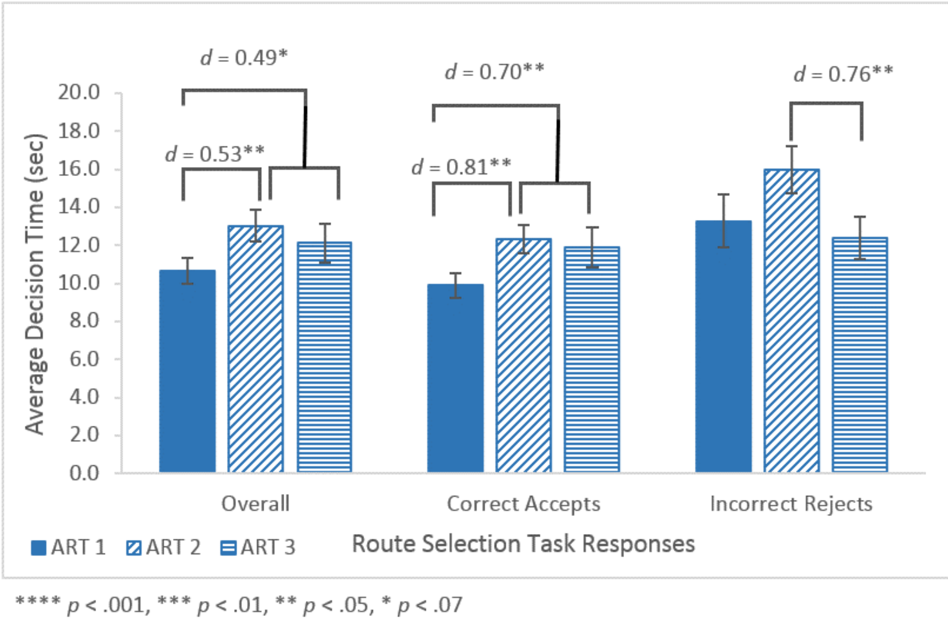


Figure 35. Average Decision Times, in seconds, at the locations where the agent recommendation was correct, sorted by correct/incorrect selections, for each agent reasoning transparency (ART) level. Bars denote SE.

Paired t-tests were used to compare differences between decision times for correct acceptances and incorrect rejections within each ART at those locations where the agent recommendation was correct (Figure 36). Decision times for incorrect rejections were significantly longer than for correct acceptances in ART1, $t(11) = -3.36$, $p = .004$, $d = 0.79$, and ART2, $t(17) = -3.40$, $p = .003$, $d = 0.84$. However, there was no difference between the two in ART3, $t(14) = -0.88$, $p = .395$, $d = 0.21$. While the difference in decision times in ARTs 1 and 2 could indicate difficulty integrating the information, resulting in incorrect choices, the lack of the same difference in ART3 could indicate complacent behavior.

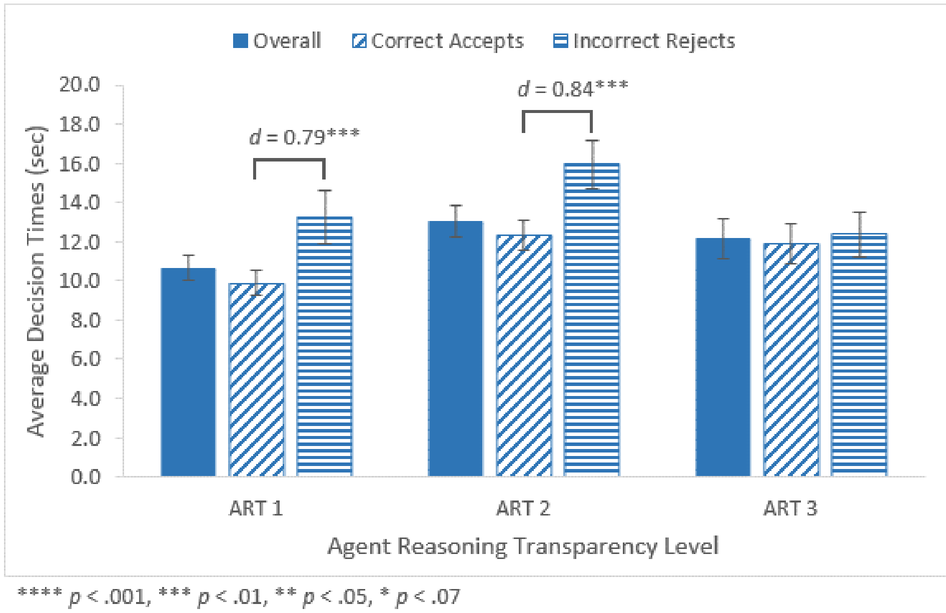


Figure 36. Average Decision Time, in seconds, for correct acceptances and incorrect rejections within each agent reasoning transparency (ART) level. Bars denote SE.

Operator trust was also evaluated using the Usability and Trust Survey. Scores on the complacency potential rating scale, CPRS, correlated significantly with Usability and Trust survey scores (r = -.35, p = .003), and CPRS was found to be a significant predictor of scores on the Usability and Trust survey, R^2 = .120, b = -1.26, t(58) = -2.81, p = .007. Participants who scored higher on the CPRS measure rated the agent as being less usable and trusted than their counterparts. However, CPRS scores did not have a consistent relationship with route selection scores across ART groups (heterogeneity of regression), and as such was not included as a covariate in subsequent analyses.

A one-way ANOVA evaluating overall Usability and Trust scores found no significant effect of ART, F(2,57) = 0.14, p = .870, ω^2 = .03. Planned comparisons revealed scores were higher in ART1 than in ART2, t(57) = -0.19, p = .985, r_c = .00, and higher in

ART3 than in ART2, $t(57) = -.47$, $p = .642$, $r_c = .06$, however these differences were not significant.

The Usability and Trust survey is a combination of surveys measuring Usability and Trust. These individual surveys were also evaluated separately to assess whether the findings were due to mainly operator trust or perceived usability.

A one-way ANOVA evaluating overall Trust scores found no significant effect of ART on Trust scores, $F(2,57) = 0.05$, $p = .952$, $\omega^2 = .03$. Planned comparisons revealed scores were higher in ART2 than in ART1, $t(57) = 0.07$, $p = .944$, $r_c = .01$, and higher in ART3 than in ART2, $t(57) = 0.23$, $p = .821$, $r_c = .03$, however these differences were not significant.

A one-way ANOVA evaluating overall Usability scores found no significant effect of ART, $F(2,57) = 0.58$, $p = .563$, $\omega^2 = .01$. Planned comparisons revealed scores were slightly higher in ART1 than in ART2, $t(57) = -0.45$, $p = .655$, $r_c = .06$, and higher in ART3 than in ART2, $t(57) = 1.07$, $p = .288$, $r_c = .14$, however these differences were not significant.

Workload Evaluation

> Hypothesis 4: Access to agent reasoning will increase operator workload, ART1 < ART2; and increased transparency of agent reasoning will increase operator workload, ART2 < ART3. When agent reasoning is not available, workload will be lower than when agent reasoning is present, ART1 < ART2+3.

ART had no significant effect on participants' global workload, $F(2,57) = 1.14$, p = .327, $\omega^2 = .00$ (Figure 37). Planned contrasts revealed there was no overall difference in participant workload when agent reasoning was available compared to the no reasoning condition, (ART1 - ART2+3), $t(57) = -1.47$, $p = .147$, $r_c = .19$. Participants in ART1 (M = 67.03, SD = 10.87) reported higher workload than those in ART2 (M = 62.80, SD = 13.78), $t(57) = -1.10$, $p = .275$, $r_c = .14$, and workload was higher in ART2 than in ART3 (M = 61.48, SD = 11.58), $t(57) = -0.34$, $p = .733$, $r_c = .05$. The non-significant omnibus p-value, along with the small effect sizes, indicate that although workload scores decreased as ART increased there was no significant difference between ARTs.

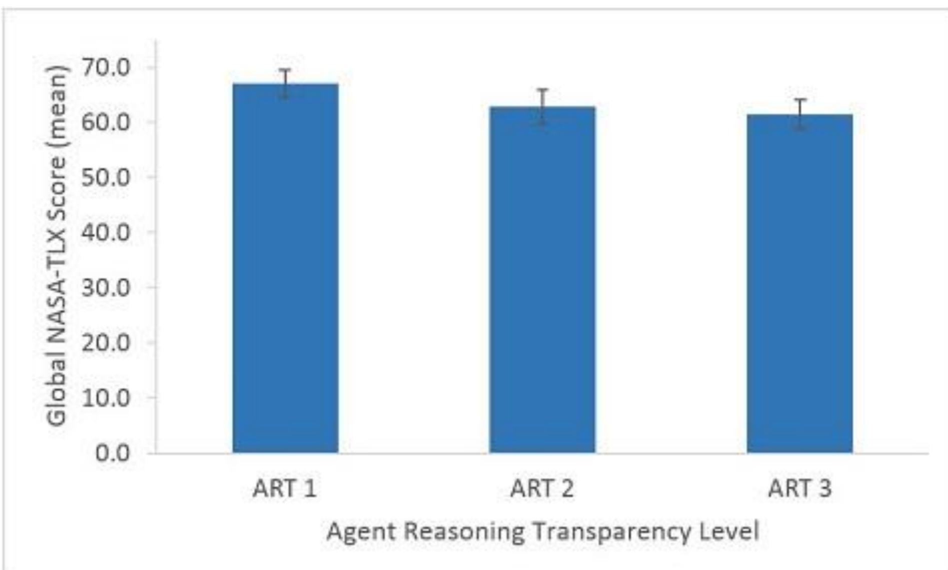


Figure 37. Average global NASA-TLX scores by agent reasoning transparency (ART) level. Bars denote SE.

Cognitive workload was also evaluated using several ocular indices. Descriptive statistics are shown in Table 21. Not all participants had complete eye measurement data, so this N was reduced (ART1 N = 18, ART2 N = 17, ART3 N = 17) and unweighted results reported. Eye tracking data was evaluated using the same planned comparisons as the subjective workload measure.

Table 21. Descriptive statistics for eye tracking measures by agent reasoning transparency (ART) condition.

		N	Mean	SD	SE	95% C.I. for Mean
Pupil Diameter (mm)	ART 1	18	3.77	0.58	0.14	(3.48, 4.06)
	ART 2	17	3.43	0.32	0.08	(3.26, 3.59)
	ART 3	17	3.48	0.36	0.09	(3.29, 3.66)
Fixation Duration (ms)	ART 1	18	4864.48	620.01	146.14	(4556.16, 5172.80)
	ART 2	17	4949.58	701.14	170.05	(4589.09, 5310.07)
	ART 3	17	4995.22	680.51	165.05	(4645.33, 5345.10)
Fixation Count	ART 1	18	279.20	38.57	9.09	(260.01, 298.38)
	ART 2	17	263.89	43.44	10.54	(241.55, 286.22)
	ART 3	17	271.67	32.62	7.91	(254.90, 288.44)

Evaluating average Pupil Diameter (PDia) between ART conditions, there was a violation of the homogeneity of variance assumption. As such, Welch's correction has been reported, and contrast tests do not assume equal variance between conditions. ART did not have a significant effect on participants' pupil diameter, $F(2,31.67) = 2.35$, $p = .112$, $\omega^2 = .07$ (Figure 38), however there was a marginally significant linear trend, $F(1,49) = 3.81$, $p = .057$, $\omega^2 = .05$, indicating that workload decreased as ART increased. Planned contrasts revealed that there was a significant difference in participant workload (as inferred via PDia) when agent reasoning was available, compared to the no reasoning condition, (ART1 - ART2+3), $t(23.1) = -2.12$, $p = .045$, $r_c = .40$. Participants in ART1 had larger pupil diameters than those in ART2, $t(26.5) = -2.18$, $p = .039$, $r_c = .39$. However, there was not significant difference in workload (as inferred via PDia) between ARTs 2 and 3, $t(31.5) = 0.46$, $p = .650$, $r_c = .08$.

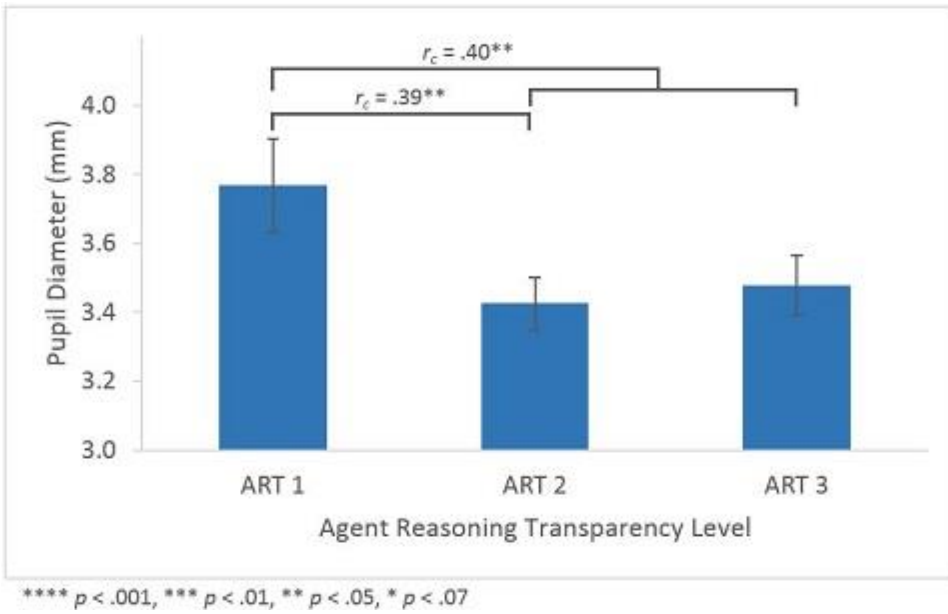


Figure 38. Average participant pupil diameter by agent reasoning transparency (ART) level. Bars denote SE.

ART did not have a significant effect on participants' fixation count, $F(2,49) = 0.17$, $p = .841$, $\omega^2 = .03$. Planned contrasts revealed that there was no significant difference in participant workload (as inferred via fixation count) when agent reasoning was available, compared to the no reasoning condition, (ART1 - ART2+3), $t(49) = 0.56$, $p = .581$, $r_c = .08$. Participants in ART1 had fewer fixations than those in ART2, $t(49) = 0.38$, $p = .708$, $r_c = .05$, who in turn had fewer fixations than those in ART3, $t(49) = 0.20$, $p = .843$, $r_c = .03$. While these results are trending in the hypothesized direction of increasing workload as ART increases, the findings are not significant.

ART did not have a significant effect on participants' fixation duration, $F(2,49) = 0.69$, $p = .505$, $\omega^2 = .01$. Planned contrasts revealed that there was no significant difference in participant workload (as inferred via fixation duration) when agent reasoning was available, compared to the no reasoning condition, (ART1 - ART2+3), $t(49) = -1.02$, $p = .314$, $r_c = .14$. Participants in ART2 had shorter fixations than those in ART1, $t(49) = -1.18$, $p = .245$, $r_c = .17$, and those in ART3, $t(49) = 0.59$, $p = .558$, $r_c = .08$. While these results indicate that the addition of agent reasoning transparency alleviates workload, the results were not significant and the effect sizes were small.

In Experiment 1 the NASA-TLX factors were also examined individually, and so this analysis is repeated for Experiment 2 results. An omnibus MANOVA indicated that there was no significant difference across ARTs for any individual factor, Wilks' $\lambda = .761$, $F(12,104) = 1.27$, $p = .247$, $\eta_p^2 = .13$. Individual evaluations of each factor across ART were made by one-way ANOVA using Bonferroni correction, $\alpha = .008$, see Table 22.

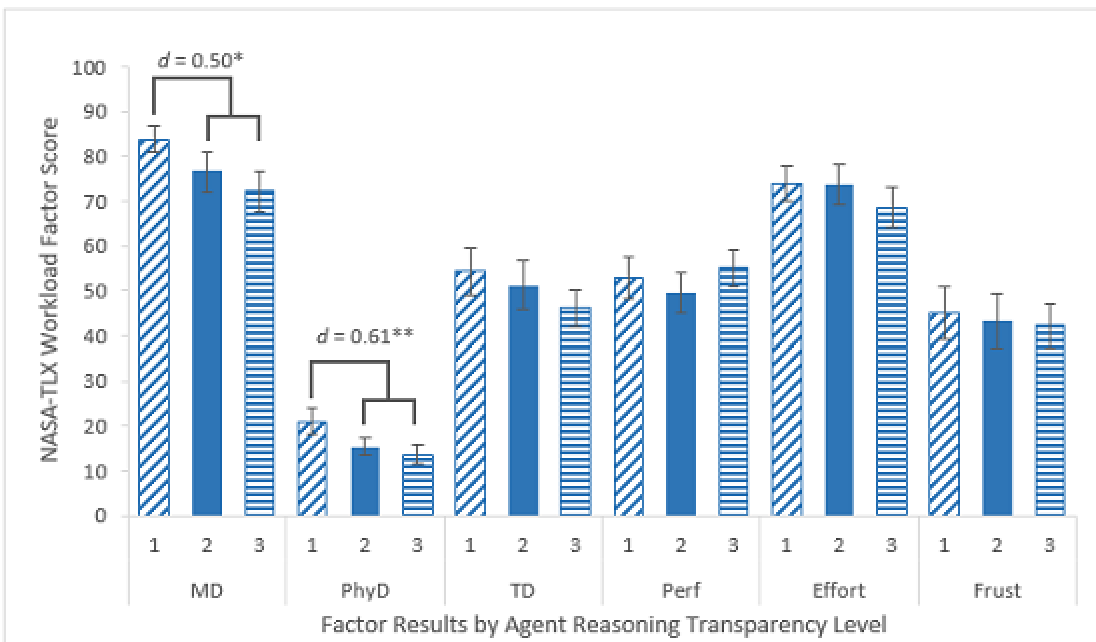
Table 22. Evaluation of NASA-TLX workload factors across agent reasoning transparency (ART) conditions.

	Mean (SD)			One-way ANOVA ($\alpha = .008$)		Planned Comparisons (Cohen's *d*)		
	ART 1	ART 2	ART 3	*F* (2,57)	ω^2	ART 1 - ART 2	ART 2 - ART 3	ART 1 - ART 2+3
MD	83.75 (12.45)	76.50 (20.27)	72.25 (20.10)	2.09	.04	0.34	0.20	0.50 *
PhyD	21.00 (12.94)	15.25 (8.66)	13.50 (9.61)	2.76 *	.06	0.46	0.14	0.61 **
TD	54.25 (23.69)	51.25 (24.00)	46.00 (19.10)	0.70	.01	0.11	0.20	0.24
Perf	52.75 (20.99)	49.50 (19.93)	55.00 (18.06)	0.39	.02	0.14	0.23	0.02
Effort	73.75 (17.08)	73.75 (19.79)	68.50 (19.67)	0.52	.02	0.00	0.23	0.13
Frust	45.00 (25.75)	43.25 (26.77)	42.25 (21.67)	0.06	.03	0.06	0.03	0.09

**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Mental Demand (MD) was the factor contributing the most to workload, and ART1 elicited greater Mental Demand than ARTs 2 or 3 (Figure 39). Although this difference did not reach significance, planned comparisons between ART levels indicate the medium-large effect sizes for the differences between ART1 and the RoboLeader conditions ARTs 2 and 3 were significant. This is evidence that the presence of agent reasoning alleviates mental demand, contradicting the stated hypothesis that workload in ART1 will be lower than in ARTs 2 and 3. Physical Demand (PhyD) contributed the least to overall workload. While the difference between ARTs 1 & 2 had a medium effect size, it did not reach significance (p = .091). However, there was a significant difference between the no reasoning condition (ART1) and the transparent reasoning conditions (ART 2+3).

Unlike Experiment 1, there was no significant difference in factors Temporal Demand (TD) or Effort across ARTs. However, there was an interesting negative correlation between Temporal Demand and the number of hours of sleep the participant reported for the previous night ($r = -.26$, $p = .042$), indicating that those who had less sleep found the task more demanding overall.



**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Figure 39. Average NASA-TLX workload factor scores by agent reasoning transparency (ART) level. Bars denote SE.

Situation Awareness (SA) Evaluation

Hypothesis 5: Access to agent reasoning will improve SA scores; and increased transparency of agent reasoning will improve SA2 scores, but will reduce SA1 and SA3 scores;

- SA1: ART1 < ART2, ART2 > ART3;
- SA2: ART1 < ART2, ART2 < ART3;
- SA3: ART1 < ART2, ART2 > ART3.

Descriptive statistics for Situation Awareness (SA) scores are shown in Table 23.

Table 23. Descriptive statistics for Situation Awareness scores by agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean	Min	Max
SA1	ART 1	20	1.60	4.31	0.96	(-0.42, 3.62)	-6	10
	ART 2	20	2.25	3.84	0.86	(0.45, 4.05)	-6	10
	ART 3	20	1.55	5.43	1.21	(-0.99, 4.09)	-7	10
SA2	ART 1	20	14.80	3.35	0.75	(13.23, 16.37)	9	20
	ART 2	20	13.20	7.15	1.60	(9.85, 16.55)	0	24
	ART 3	20	15.20	6.28	1.40	(12.26, 18.14)	1	25
SA3	ART 1	20	2.90	9.40	2.10	(-1.50, 7.30)	-16	16
	ART 2	20	0.45	8.51	1.90	(-3.53, 4.43)	-18	16
	ART 3	20	2.00	8.78	1.96	(-2.11, 6.11)	-14	18

Working Memory Capacity (WMC) scores correlated significantly with SA1 scores (r = .26, p = .021), and WMC was found to be a significant predictor of SA1 scores, R^2 = .069, b = 0.10, $t(58)$ = 2.07, p = .043. Participants who scored higher on the WMC measure scored higher on SA1 queries than their counterparts. However, WMC scores did not have a consistent relationship with SA1 scores across ART groups (heterogeneity of regression), and as such was not included as a covariate in subsequent analyses.

A one-way ANOVA evaluating SA1 scores found no significant effect of ART, $F(2,57) = 0.15$, $p = .865$, $\omega^2 = .03$. Planned comparisons revealed scores were higher in ART2 than in ART1, $t(57) = 0.50$, $p = .617$, $r_c = .06$, or ART3, $t(57) = -0.48$, $p = .630$, $r_c = .06$, however these differences were not significant.

Working Memory Capacity (WMC) scores ($r = .38$, $p = .001$), Spatial Visualization (SV) scores ($r = .27$, $p = .018$), and Spatial Orientation Test (SOT) scores ($r = -.46$, $p < .001$) correlated significantly with SA2 scores. WMC ($R^2 = .143$, $b = 0.18$, $t(58) = 3.11$, $p = .003$) and SOT ($R^2 = .208$, $b = -0.36$, $t(58) = -3.90$, $p < .001$) were also found to be significant predictors of SA2 scores. Participants who scored higher on the WMC and SV measures, or who performed better on the SOT, scored higher on SA2 queries than their counterparts. However, none of these potential covariates had a consistent relationship with SA2 scores across ART groups (heterogeneity of regression), and as such were not included as covariates in subsequent analyses.

A one-way ANOVA evaluating SA2 scores found no significant effect of ART, $F(2,57) = 0.85$, $p = .434$, $\omega^2 = .01$. Planned comparisons revealed no change in scores between ART1 and ART2, $t(57) = 0.00$, $p = 1.000$, $r_c = .00$, and scores in ART3 were slightly higher than in ART2, $t(57) = 1.13$, $p = .264$, $r_c = .15$, however this difference was not significant.

Complacency Potential Rating Scale (CPRS) scores ($r = -.25$, $p = .026$) and Spatial Orientation Test (SOT) scores ($r = -.27$, $p = .018$) correlated significantly with SA3 scores.

Participants who scored lower on the CPRS, indicating a lower potential for complacent behavior, as well as those who performed better on the SOT, scored higher on SA3 queries than their counterparts. However, neither of these potential covariates had a consistent relationship with SA3 scores across ART groups (heterogeneity of regression), and as such were not included as covariates in subsequent analyses.

A one-way ANOVA evaluating SA3 scores found no significant effect of ART, $F(2,57) = 0.39$, $p = .680$, $\omega^2 = .02$. Planned comparisons revealed SA3 scores in ART1 were higher than those in ART2, $t(57) = -0.87$, $p = .388$, $r_c = .11$, and scores in ART2 were lower than in ART3, $t(57) = 0.55$, $p = .584$, $r_c = .09$. These results were contrary to the stated hypothesis, in that SA3 scores were lowest in ART2, however these results were not significant.

Target Detection Task Performance

> Hypothesis 6: Access to agent reasoning will reduce performance on the target detection task (fewer targets detected, higher FAs), ART1 > ART2, and increased transparency of agent reasoning will further reduce performance on the target detection task; ART2 > ART3.

Descriptive statistics for Target Detection measures are shown in Table 24.

Table 24. Descriptive statistics for Target Detection Task measures by agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean	Min	Max
Targets Detected (Count)	ART 1	20	45.25	10.96	2.45	(40.12, 50.38)	24	59
	ART 2	20	47.65	10.74	2.40	(42.62, 52.68)	30	73
	ART 3	20	40.30	13.27	2.97	(34.09, 46.51)	18	61
FAs (Count)	ART 1	20	16.30	6.18	1.38	(13.41, 19.19)	4	28
	ART 2	20	16.65	4.97	1.11	(14.33, 18.97)	11	26
	ART 3	20	15.90	6.12	1.37	(13.04, 18.76)	6	26
d'	ART 1	20	2.30	0.40	0.09	(2.11, 2.49)	1.62	2.95
	ART 2	20	2.38	0.35	0.08	(2.21, 2.54)	1.81	3.32
	ART 3	20	2.19	0.44	0.10	(1.99, 2.39)	1.49	2.88
β	ART 1	20	2.64	0.34	0.08	(2.48, 2.80)	2.17	3.24
	ART 2	20	2.59	0.28	0.06	(2.46, 2.72)	1.88	2.96
	ART 3	20	2.65	0.39	0.09	(2.47, 2.83)	2.14	3.51

Spatial Visualization (SV) scores correlated significantly with (r = .38, p = .001), and were found to be significant predictors of total number of Targets Detected, R^2 = .143, b = 32.15, $t(58)$ = 3.12, p = .003. Participants who scored higher in SV, indicating a greater ability to mentally manipulate objects in 3D space, also detected more targets in their environment than their counterparts. SV scores did not have a consistent relationship with total number of Targets Detected for each ART group (heterogeneity of regression), as such was not included as a covariate in subsequent analyses.

There was no significant effect of ART on the number of targets detected, $F(2,57)$ = 2.05, p = .138, ω^2 = 0.03 (Figure 40). The number of targets detected was slightly greater in ART2 than in ART1, $t(57)$ = 0.65, p = .520, r_c = .09, and significantly higher in ART2 than in ART3, $t(57)$ = -1.98, p = .052, r_c = .25. While access to agent reasoning did not appear to improve performance on the target detection task, increasing the amount of agent reasoning did reseult in a decline in performance, indicating the participants may have become overwhelmed.

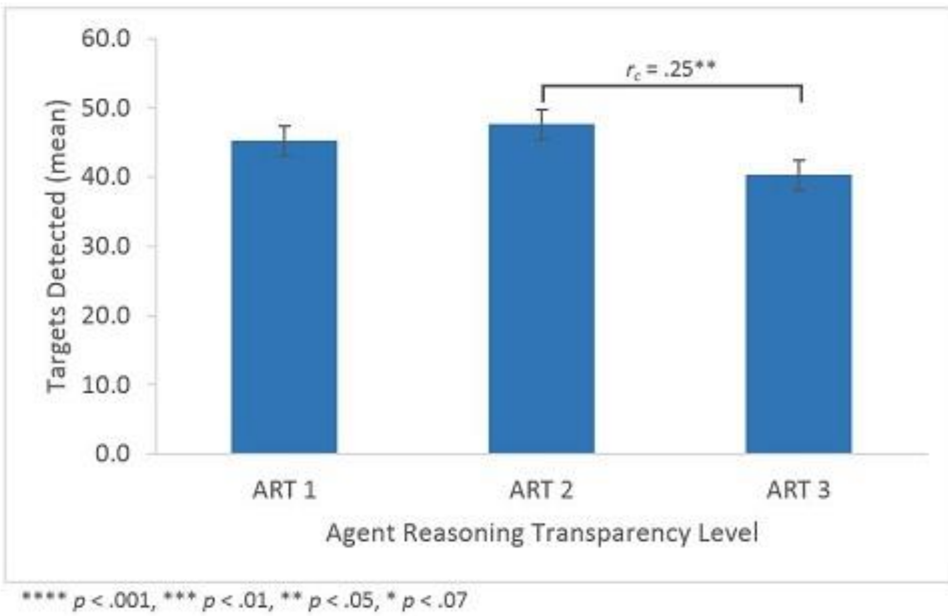


Figure 40. Average number of targets detected by agent reasoning transparency (ART) level. Bars denote SE.

There was no significant effect of ART on the number of FAs across ARTs, $F(2,57) = 0.08$, $p = .919$, $\omega^2 = 0.03$. The number of FAs was higher in ART2 than in ART1, $t(57) = 0.19$, $p = .849$, $r_c = 0.03$, and ART3, $t(57) = -0.41$, $p = .683$, $r_c = 0.05$, however these differences were not significant.

Results of the target detection task were also evaluated using SDT to determine if there were differences in sensitivity (*d'*) or selection bias (Beta) between the three ARTs. There was no significant effect of ART on *d*', $F(2,57) = 1.10$, $p = .341$, $\omega^2 = .00$. Participants were slightly more sensitive to targets in ART2 than in ART1, $t(57) = 0.59$, $p = .555$, $r_c = .08$, or ART3, $t(57) = -1.47$, $p = .147$, $r_c = .19$, however these differences did not achieve statistical significance. Evaluating Beta across ART, there was no significant effect of ART on Beta scores, $F(2,57) = 0.17$, $p = .843$, $\omega^2 = .03$. Beta scores were slightly lower in ART2 than in ART1, $t(57) = -0.48$, $p = .636$, $r_c = .06$, and ART3, $t(57) = 0.53$, $p = .596$, $r_c = .07$, however these differences were not significant. In an information-rich environment, agent reasoning transparency appears to have no effect on sensitivity to targets or target selection criterion.

Individual Differences Evaluations

Complacency Potential

Complacency Potential (CP) was evaluated via the Complacency Potential Rating Scale (CPRS) scores. The effect of CP on several measures of interest across agent reasoning transparency (ART) level were evaluated via two-way between-groups ANOVAs, $\alpha = .05$. Post hoc t-tests within each ART compared performance differences between high/low group memberships. Descriptive statistics for Complacency Potential (CP), as measured using the Complacency Potential Rating Scale (CPRS), are shown in Tables 25 and 26.

Table 25. Descriptive statistics for Complacency Potential Rating Scale (CPRS) scores by agent reasoning transparency (ART) level.

							Mdn Split Count	
Group	N	Min	Max	Mdn	Mean	SD	Hi	Lo
Overall	60	25	47	37.00	36.83	4.38	32	28
ART 1	20	25	41	35.00	35.00	4.21	8	12
ART 2	20	32	47	40.00	39.05	3.53	15	5
ART 3	20	31	47	35.50	36.45	4.54	9	11

Table 26. Descriptive statistics for High/Low Complacency Potential Rating Scale (CPRS) scores by agent reasoning transparency (ART) level.

		N	Mean	SD	SE	95% C.I. for Mean
ART 1	Low CPRS	12	32.42	3.34	0.96	(30.29, 34.54)
	High CPRS	8	38.88	1.36	0.48	(37.74, 40.01)
ART 2	Low CPRS	5	34.80	1.79	0.80	(32.58, 37.02)
	High CPRS	15	40.47	2.72	0.70	(38.96, 41.97)
ART 3	Low CPRS	11	33.18	1.54	0.46	(32.15, 34.21)
	High CPRS	9	40.44	3.64	1.21	(37.64, 43.25)

Hypothesis 7: High CPRS individuals will have fewer correct rejects on the route planning task than Low CPRS individuals.

A two-way between-groups ANOVA revealed no significant interaction between CPRS and ART on the number of correct rejects in the route planning task, $F(2,54) = 2.04$, $p = .140$, $\eta_p^2 = .07$, however there was a significant main effect of CPRS on the number of correct rejects across ART, $F(1,54) = 7.51$, $p = .008$, $\eta_p^2 = .12$ (Figure 41). Post hoc comparisons between high/low CPRS groups within each ART level show that High CPRS and Low CPRS individuals had similar Route Selection scores in ART1, $t(18) = 0.20$, $p = .842$, $d = 0.31$, however low CPRS participants had more correct rejects in ART2, $t(18) = 2.17$, $p = .044$, $d = 1.37$, and ART3, $t(18) = 2.69$, $p = .015$, $d = 1.20$. When agent reasoning was not available there was no difference in correct rejects between High and Low CPRS persons. However, when agent reasoning was available participants with Low Complacency Potential (CP) had more correct rejects than those with High CP, and this difference became greater as agent reasoning transparency increased.

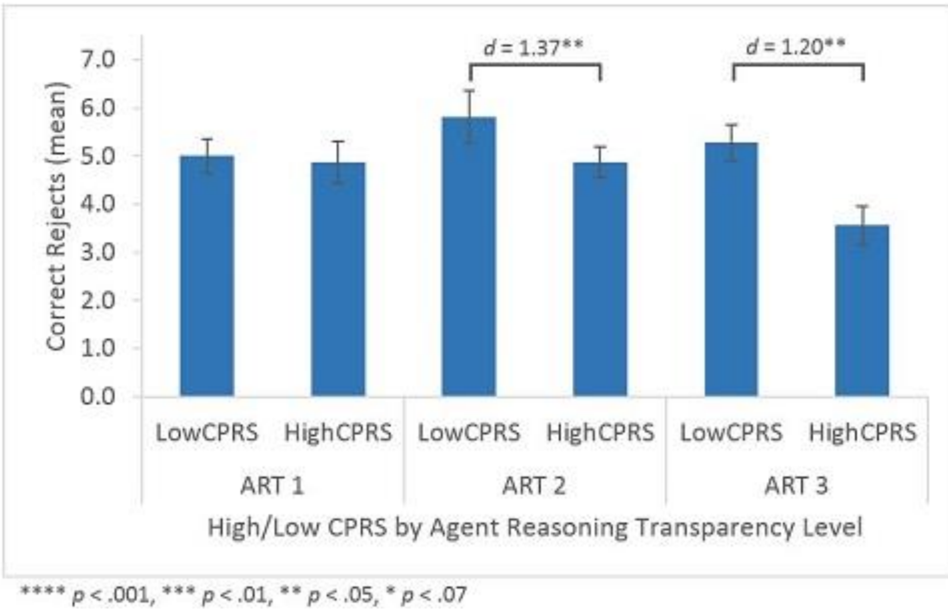


Figure 41. Average number of correct rejects by High/Low complacency potential rating scale (CPRS) score group, sorted by agent reasoning transparency (ART) level. Bars denote SE.

Hypothesis 8: High CPRS score individuals will have higher scores on the usability and trust survey than Low CPRS score individuals.

A two-way between-groups ANOVA revealed no significant interaction between CPRS score and ART on Usability and Trust Survey scores, $F(2,54) = 0.11$, $p = .895$, $\eta_p^2 = .00$, nor any significant main effect of CP on Usability scores, $F(1,54) = 1.79$, $p = .187$, $\eta_p^2 = .03$.

Hypothesis 9: High CPRS score individuals will have lower SA scores than Low CPRS score individuals.

A two-way between-groups ANOVA revealed no significant interaction between CPRS scores and ART on SA1 scores, $F(2,54) = 0.22$, $p = .800$, $\eta_p^2 = .01$, however there was a significant main effect of CP on SA1 scores across ART, $F(1,54) = 4.12$, $p = .047$, $\eta_p^2 = .12$ (Figure 42). Post hoc comparisons between high/low CPRS score groups within each ART level show that Low CP individuals had higher SA1 scores in each ART (ART1, $t(18) = 0.93$, $p = .365$, $d = 0.42$; ART2, $t(18) = 1.05$, $p = .310$, $d = 0.72$; ART3, $t(18) = 1.54$, $p = .142$, $d = 0.69$) than their High CP counterparts. Thus, in a high information environment, Low CP individuals better monitored their environment than High CP individuals.

Figure 42. Average level 1 situation awareness (SA1) scores by High/Low complacency potential rating scale (CPRS) group, sorted by agent reasoning transparency (ART) level. Bars denote SE.

A two-way between-groups ANOVA revealed no significant interaction between CPRS and ART on SA2 scores, $F(2,54) = 0.46$, $p = .636$, $\eta_p^2 = .02$, nor any significant main effect of CPRS on SA2 scores across ART, $F(1,54) = 3.23$, $p = .078$, $\eta_p^2 = .06$. A two-way between-groups ANOVA revealed no significant interaction between CPRS and

ART on SA3 scores, $F(2,54) = 0.78$, $p = .465$, $\eta_p^2 = .03$, nor any significant main effect of CPRS on SA3 scores across ART, $F(1,54) = 1.80$, $p = .185$, $\eta_p^2 = .03$.

Spatial Ability (SOT and SV) and Perceived Attentional Control (PAC)

> Hypothesis 10: Individual differences, such as SpA and PAC, will have differential effects on the operator's performance on the route selection task and their ability to maintain SA.

The effects of individual difference (ID) factors and agent reasoning transparency (ART) level on route selection performance were evaluated via two-way between-groups ANOVAs, $\alpha = .05$. When Levene's Test of Equality of Error Variance was significant, the evaluation was repeated at $\alpha = .01$. Post hoc t-tests within each ART compared performance differences between high/low group memberships for each ID factor. Spatial Orientation (SOT) is reverse-scored, so lower test scores imply greater spatial ability (High SOT group), while Spatial Visualization (SV), and Perceived Attentional Control (PAC) are scored normally (higher test scores implies greater ability). Descriptive statistics for SOT, SV, and PAC are shown in Tables 27 and 28.

Table 27. Descriptive statistics for Spatial Orientation (SOT), Spatial Visualization (SV), and Perceived Attentional Control (PAC), by Agent Reasoning Transparency (ART) level.

								Mdn Split Count	
	Group	N	Min	Max	Mdn	Mean	SD	Hi	Lo
SOT	Overall	60	3.96	33.01	11.19	13.39	7.40	30	30
	ART 1	20	4.58	27.00	9.26	12.75	7.08	12	8
	ART 2	20	4.52	33.01	13.74	14.71	8.14	8	12
	ART 3	20	3.96	27.81	10.23	12.71	7.15	10	10
SV	Overall	60	0.19	0.88	0.50	0.52	0.14	30	30
	ART 1	20	0.36	0.76	0.54	0.52	0.11	12	8
	ART 2	20	0.36	0.88	0.51	0.53	0.13	13	7
	ART 3	20	0.19	0.83	0.48	0.50	0.17	8	12
PAC	Overall	60	33	75	58.00	57.55	8.23	31	29
	ART 1	20	33	74	57.50	56.35	8.87	10	10
	ART 2	20	41	75	60.50	60.05	7.67	13	7
	ART 3	20	41	70	57.00	56.25	7.93	8	12

Table 28. Descriptive statistics for Spatial Orientation (SOT), Spatial Visualization (SV), and Perceived Attentional Control (PAC), by Agent Reasoning Transparency (ART) level, sorted by High/Low group membership.

			N	Mean	SD	SE	95% C.I. for Mean
SOT	ART 1	Low	8	20.03	5.50	1.94	(15.44, 24.63)
		High	12	7.90	1.78	0.51	(6.77, 9.03)
	ART 2	Low	12	19.59	6.82	1.97	(15.25, 23.92)
		High	8	7.40	2.14	0.76	(5.60, 9.19)
	ART 3	Low	10	18.67	5.18	1.64	(14.96, 22.37)
		High	10	6.75	1.54	0.49	(5.65, 7.86)
SV	ART 1	Low	8	0.41	0.05	0.02	(0.37, 0.45)
		High	12	0.59	0.08	0.02	(0.54, 0.64)
	ART 2	Low	7	0.40	0.04	0.01	(0.37, 0.44)
		High	13	0.60	0.11	0.03	(0.54, 0.67)
	ART 3	Low	12	0.38	0.11	0.03	(0.31, 0.45)
		High	8	0.67	0.09	0.03	(0.59, 0.75)
PAC	ART 1	Low	10	50.10	7.42	2.34	(44.80, 55.41)
		High	10	62.60	4.93	1.56	(59.08, 66.12)
	ART 2	Low	7	52.29	5.50	2.08	(47.20, 57.37)
		High	13	64.23	4.90	1.36	(61.27, 67.19)
	ART 3	Low	12	51.25	5.56	1.61	(47.72, 54.78)
		High	8	63.75	3.85	1.36	(60.54, 66.97)

Route Selection Task Evaluation

A two-way between-groups ANOVA revealed no significant interaction between SOT and ART on Route Selection scores, $F(2,54) = 0.20$, $p = .981$, $\eta_p^2 = .01$, however there was a significant main effect of SOT on Route Selection scores, $F(1,54) = 4.40$, $p = .041$, $\eta_p^2 = .08$ (Figure 43). Post hoc comparisons between high/low SOT groups within each ART level show that Low SOT individuals (those who performed better on the SOT) had higher Route Selection scores in each ART (ART1, $t(18) = -1.29$, $p = .214$, $d = 0.61$;

ART2, $t(18) = -1.10$, $p = .287$, $d = 0.50$; ART3, $t(18) = -1.24$, $p = .230$, $d = 0.56$). Although these post hoc analyses did not reach statistical analysis, they had medium effect sizes.

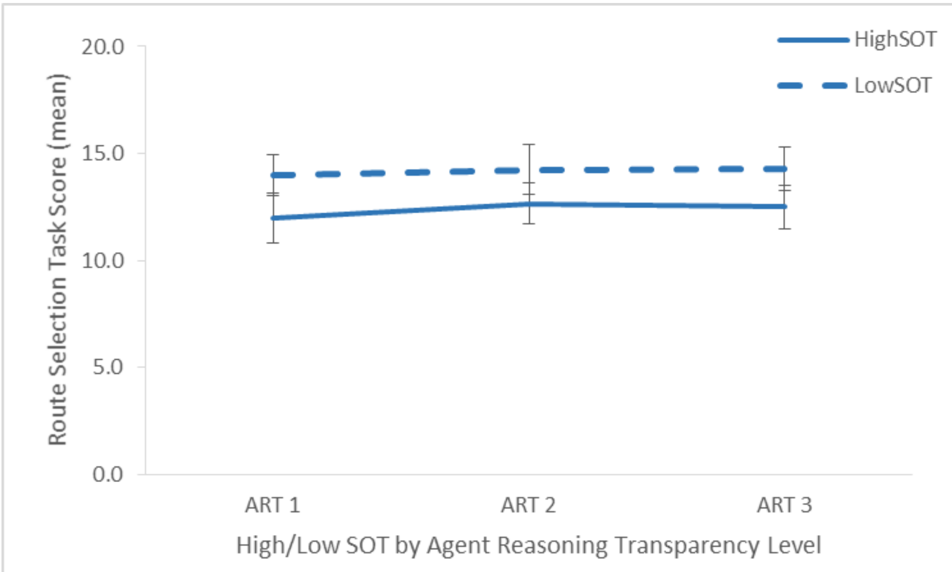


Figure 43. Average route selection scores by High/Low Spatial Orientation Test (SOT) group membership, across agent reasoning transparency level. Bars denote SE.

A two-way between-groups ANOVA revealed no significant interaction between SV and ART on Route Selection scores, $F(2,54) = 0.04$, $p = .965$, $\eta_p^2 = .00$, nor any significant main effect of SV on Route Selection scores, $F(1,54) = 0.08$, $p = .782$, $\eta_p^2 = .00$.

A two-way between-groups ANOVA revealed no significant interaction between PAC and ART on Route Selection scores, $F(2,54) = 0.14$, $p = .869$, $\eta_p^2 = .01$, however there was a significant main effect of PAC on Route Selection scores, $F(1,54) = 3.98$, $p = .051$, $\eta_p^2 = .07$ (Figure 44). Post hoc comparisons between high/low PAC groups within each ART level show that High PAC individuals had higher Route Selection scores in each ART (ART1, $t(18) = -1.18$, $p = .255$, $d = 0.53$; ART2, $t(18) = -0.74$, $p = .467$, $d = 0.34$;

ART3, $t(18) = -1.56$, $p = .137$, $d = 0.69$). Although these post hoc analyses did not reach statistical analysis, they had medium effect sizes.

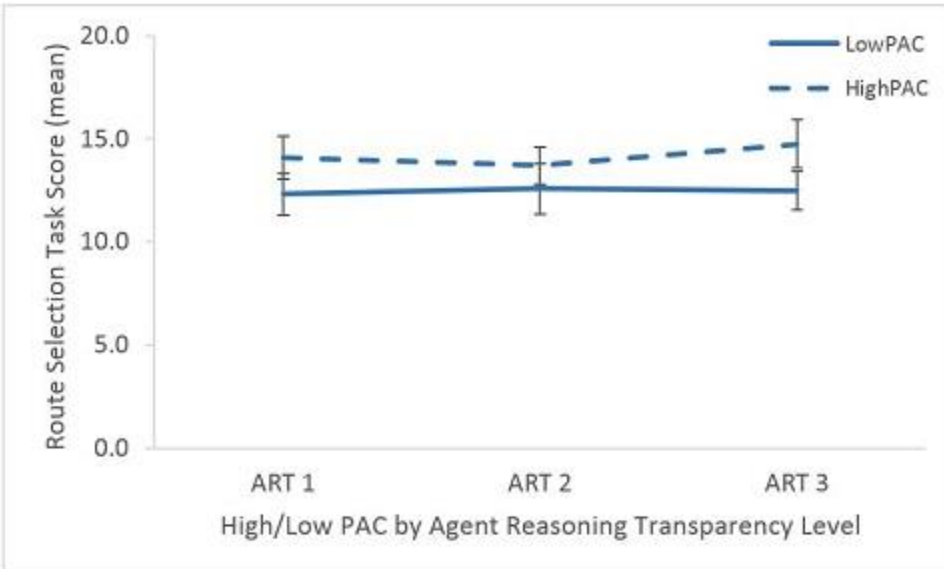


Figure 44. Average route selection scores by High/Low Perceived Attentional Control (PAC) group membership, across agent reasoning transparency level. Bars denote SE.

<u>SA1 Evaluation</u>

A two-way between-groups ANOVA revealed no significant interaction between SOT and ART on SA1 scores, $F(2,54) = 1.80$, $p = .175$, $\eta_p^2 = .06$, nor any significant main effect of SOT on SA1 scores, $F(1,54) = 3.30$, $p = .075$, $\eta_p^2 = .06$. A two-way between-groups ANOVA revealed no significant interaction between SV and ART on SA1 scores, $F(2,54) = 0.35$, $p = .709$, $\eta_p^2 = .01$, nor any significant main effect of SV on SA1 scores, $F(1,54) = 2.63$, $p = .111$, $\eta_p^2 = .05$. A two-way between-groups ANOVA revealed no significant interaction between PAC and ART on SA1 scores, $F(2,54) = 0.52$, $p = .598$, η_p^2

= .02, nor any significant main effect of PAC on SA1 scores, $F(1,54) = 0.73$, $p = .399$, $\eta_p^2 = .01$.

SA2 Evaluation

A two-way between-groups ANOVA revealed no significant interaction between SOT and ART on SA2 scores, $F(2,54) = 2.42$, $p = .099$, $\eta_p^2 = .08$, however there is a significant main effect of SOT on SA2 scores, $F(1,54) = 16.98$, $p < .001$, $\eta_p^2 = .24$ (Figure 45). Post hoc comparisons between high/low SOT groups within each ART level show that High SOT and Low SO individuals had similar SA2 scores in ART1, $t(18) = -0.87$, $p = .398$, $d = 0.39$, however high SO participants had higher SA2 scores in ART2, $t(18) = -2.78$, $p = .012$, $d = 1.29$, and ART3, $t(18) = -3.09$, $p = .006$, $d = 1.42$. When agent reasoning was not available there was no significant difference in SA2 scores between High and Low SO persons. However, when agent reasoning was available participants who performed better on the SOT also had higher SA2 scores than their counterparts.

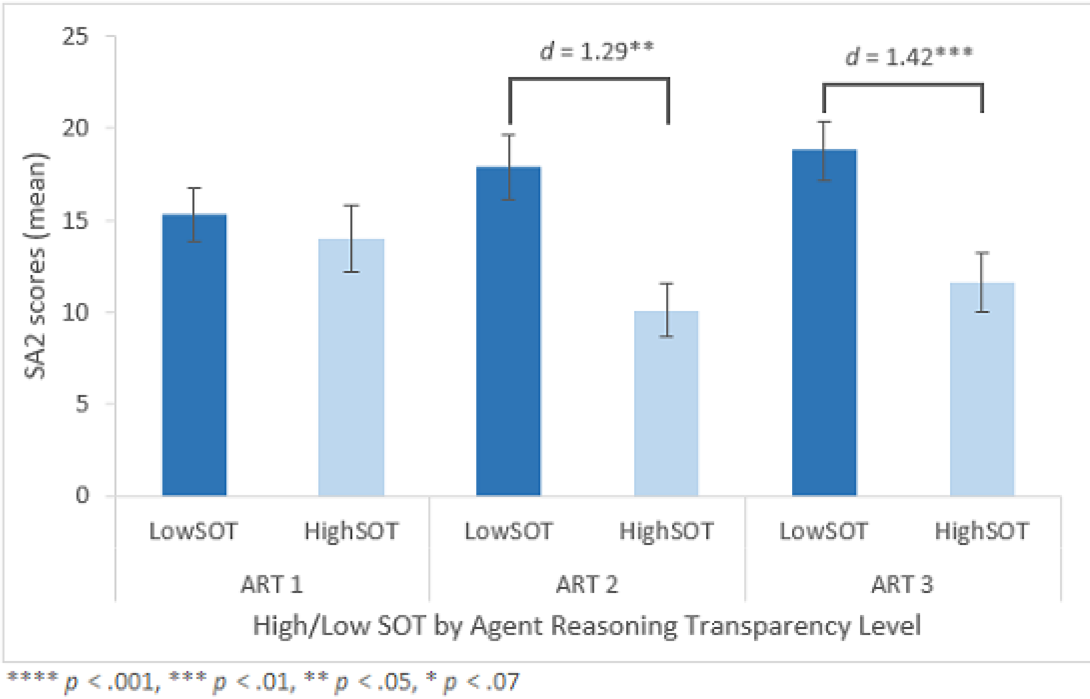


Figure 45. Average SA2 scores by Spatial Orientation Test (SOT) High/Low group membership, sorted by agent reasoning transparency level. Bars denote SE.

A two-way between-groups ANOVA revealed no significant interaction between SV and ART on SA2 scores, $F(2,54) = 0.70$, $p = .501$, $\eta_p^2 = .03$, nor any significant main effect of SV on SA2 scores, $F(1,54) = 3.57$, $p = .064$, $\eta_p^2 = .06$. A two-way between-groups ANOVA revealed no significant interaction between PAC and ART on SA2 scores, $F(2,54) = 0.21$, $p = .809$, $\eta_p^2 = .01$, nor any significant main effect of PAC on SA2 scores, $F(1,54) = 0.08$, $p = .775$, $\eta_p^2 = .00$.

SA3 Evaluation

A two-way between-groups ANOVA revealed no significant interaction between SOT and ART on SA3 scores, $F(2,54) = 0.62$, $p = .541$, $\eta_p^2 = .02$, nor any significant main effect of SOT on SA3 scores, $F(1,54) = 3.34$, $p = .073$, $\eta_p^2 = .06$. A two-way between-groups ANOVA revealed no significant interaction between SV and ART on SA3 scores,

$F(2,54) = 0.63$, $p = .537$, $\eta_p^2 = .02$, nor any significant main effect of SV on SA3 scores, $F(1,54) = 0.65$, $p = .423$, $\eta_p^2 = .01$. A two-way between-groups ANOVA revealed no significant interaction between PAC and ART on SA3 scores, $F(2,54) = 0.42$, $p = .661$, $\eta_p^2 = .02$, nor any significant main effect of PAC on SA3 scores, $F(1,54) = 1.27$, $p = .265$, $\eta_p^2 = .02$.

Working Memory Capacity (WMC)

Hypothesis 11: High WMC individuals will have more correct rejects and higher SA2 and SA3 scores than Low WMC individuals.

The effects of Working Memory Capacity (WMC) and agent reasoning transparency (ART) level were evaluated via two-way between-groups ANOVAs, $\alpha = .05$. Post hoc t-tests within each ART compared performance differences between high/low group memberships. Descriptive statistics for WMC, as measured using the RSPAN test, are shown in Tables 29 and 30.

Table 29. Descriptive statistics for Working Memory Capacity (WMC), by Agent Reasoning Transparency (ART) level.

								Mdn Split Count	
	Group	N	Min	Max	Mdn	Mean	SD	Hi	Lo
WMC	Overall	60	10	54	31.00	31.47	12.06	31	29
	ART 1	20	17	54	31.00	33.15	11.86	11	9
	ART 2	20	11	54	32.50	31.10	13.75	11	9
	ART 3	20	10	54	28.00	30.15	11.17	9	11

Table 30. Descriptive statistics for Working Memory Capacity (WMC), by Agent Reasoning Transparency (ART) level, sorted by High/Low group membership.

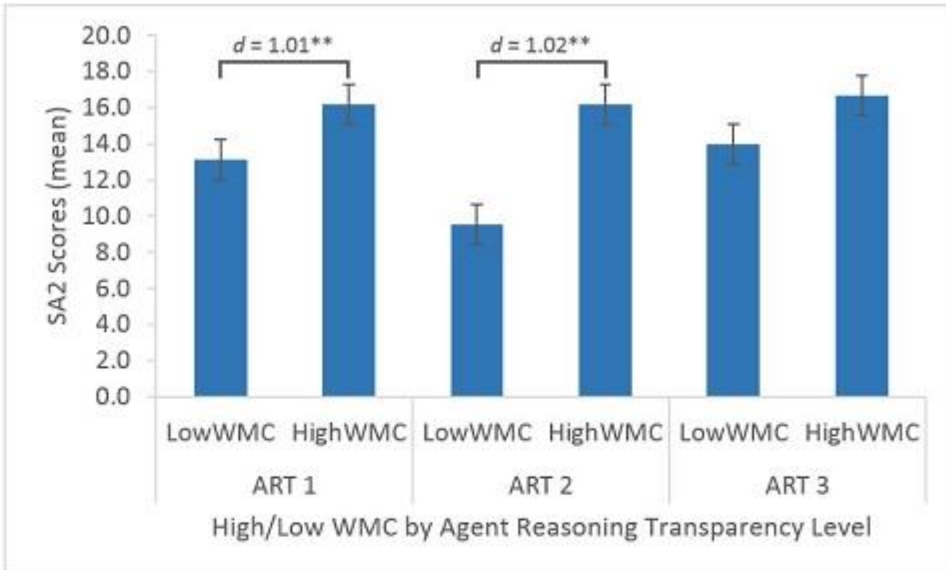
			N	Mean	SD	SE	95% C.I. for Mean
WMC	ART 1	Low	9	22.11	3.55	1.18	(19.38, 24.84)
		High	11	42.18	7.59	2.29	(37.08, 47.28)
	ART 2	Low	9	18.00	4.61	1.54	(14.46, 21.54)
		High	11	41.82	7.83	2.36	(36.56, 47.08)
	ART 3	Low	11	22.09	5.65	1.70	(18.30, 25.88)
		High	9	40.00	7.62	2.54	(34.15, 45.85)

Correct Rejects

A two-way between-groups ANOVA revealed no significant interaction between WMC and ART on Correct Rejection scores, $F(2,54) = 1.61$, $p = .210$, $\eta_p^2 = .06$, nor any significant main effect of WMC on Correct Reject scores, $F(1,54) = 0.61$, $p = .439$, $\eta_p^2 = .01$.

SA Scores

A two-way between-groups ANOVA revealed no significant interaction between WMC and ART on SA2 scores, $F(2,54) = 0.78$, $p = .465$, $\eta_p^2 = .03$, however there was a significant main effect of WMC on SA2 scores across ARTs, $F(1,54) = 8.33$, $p = .006$, $\eta_p^2 = .13$ (Figure 46). High WMC participants had higher SA2 scores in all ART conditions (ART1, $t(18) = -2.25$, $p = .037$, $d = 1.01$; ART2, $t(18) = -2.28$, $p = .035$, $d = 1.02$; ART3, $t(18) = -1.94$, $p = .359$, $d = 0.44$) than their Low WMC counterparts. Performance of the High WMC group was consistent between ARTs, while the Low WMC participants' SA2 scores varied. This difference was greatest in ART2, where access to agent reasoning resulted in Low WMC participants to have lower SA2 scores than in the no reasoning condition, and smallest in ART3, where increased access to agent reasoning appears to have helped Low WMC participants SA2 scores increase to almost that of their High WMC counterparts.



**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Figure 46. Average SA2 scores by Working Memory Capacity (WMC) High/Low group membership, sorted by agent reasoning transparency level (ART). Bars denote SE.

There was no significant interaction between WMC and ART on SA3 scores, $F(2,54) = 1.17$, $p = .317$, $\eta_p^2 = .04$, nor any significant main effect of WMC on SA3 scores, $F(1,54) = 0.45$, $p = .503$, $\eta_p^2 = .01$.

Discussion

The primary goal of this study was to examine how the transparency of an intelligent agent's reasoning in a high information environment affected complacent behavior in a route selection task. Participants supervised a three-vehicle convoy as it traversed a simulated environment and re-routed the convoy when needed with the assistance of an intelligent agent, RoboLeader (RL). Information regarding potential events along the pre-planned route, together with communications from a commander confirming either the presence or absence of activity in the area, were provided to all participants. They received information about both their current route and the agent-recommended alternative route. When the convoy approached a potentially unsafe area, the intelligent agent would recommend re-routing the convoy. The agent recommendations were correct 66% of the time. The participant was required to recognize and correctly reject any incorrect suggestions. The secondary goal of this study was to examine how differing levels of agent transparency affected main task and secondary task performance, response time, workload, SA, trust, and system usability, along with implications of individual difference factors such as spatial ability, WMC, PAC, and complacency potential.

Each participant was assigned to a specific level of agent reasoning transparency (ART). The reasoning was provided as to why the agent was making the recommendation and this differed among these levels. ART1 provided no reasoning information; RL notified that a change was recommended without explanation. The type of information the agent supplied varied slightly between ARTs 2 and 3. In ART2 the agent reasoning were simple statements of fact corresponding to the information icons that appeared on the map, along with reasoning as to how the agent factored each piece of information into its final recommendation (e.g. recommend revise convoy route: Potential IED (H), Potential Sniper (M), Dense Fog (L)). In ART3 an additional piece of information was added, which conveyed when the agent had received the information leading to its recommendation (e.g. recommend revise convoy route: Potential IED (H), TOR: 1 [hour], Potential Sniper (M), TOR: 2, Dense Fog (L), TOR: 4). This additional information did not convey any confidence level or uncertainty but was designed to encourage the operator to actively evaluate the quality of the information rather than simply respond. Therefore, not only was access to agent reasoning examined, but the impact of the type of information the agent supplied was reviewed as well.

Complacent behavior was investigated via primary (route selection) task response at those decision points where the agent recommendation was incorrect, in the form of incorrect acceptances of the agent recommendation, an objective measure of errors of commission (Parasuraman et al., 2000). Access to agent reasoning was predicted to reduce the number of incorrect acceptances while an increase in agent reasoning transparency was expected to increase incorrect acceptances. The trend in the data appeared to support this prediction, even though the findings were not significant. While there was a slight decrease in the mean score for incorrect acceptances when agent reasoning transparency was added, the highest mean score for incorrect acceptances was in ART3, when agent reasoning transparency was highest. Response times for incorrect acceptances were longer than those for correct rejections in the agent reasoning transparency condition, indicating that these incorrect acceptances could be the result of errors in judgment rather than an indication of complacent behavior. However, in the condition with the highest amount of agent reasoning transparency, not only are there more incorrect acceptances of the agent suggestion, but the decision times for these responses is no different from those for correct rejections. Considered together, this may be an indication that the combination of high information and increased access to agent reasoning could overwork the operator, resulting in an out-of-the-loop (OOTL) situation. Differences due to individual differences offer support for this notion, as individuals with higher working memory capacity had fewer incorrect acceptances overall, demonstrating an ability to process more information more effectively than their counterparts. Additionally, individuals who scored low on complacency potential (CP) had fewer incorrect acceptances in the agent reasoning transparency conditions. There was no difference in performance between High and Low CP individuals in the information-only condition. However, when agent reasoning was transparent, Low CP individuals had more correct rejections than the High CP individuals, and when agent reasoning transparency was increased the difference in performance became more pronounced. The better performance of low CP individuals could be an indication of either their willingness to engage with the agent rather than defer, or of their calibrated trust in the ability of the intelligent agent (Parasuraman & Manzey, 2010).

As in experiment 1, the operator received all information needed to route the convoy correctly without the agent's suggestion. While the addition of agent reasoning did result in fewer incorrect acceptances than in the no reasoning condition, the difference was not significant. However, the small reduction in the number of incorrect acceptances considered with the increased response times does provide evidence that the addition of agent reasoning transparency is effective at keeping the operator engaged in the task, even if the performance gains are small. In the highest reasoning transparency condition, operators were also given information that could have seemed ambiguous, and as a result, the number of incorrect acceptances increased while the response times were unchanged from those for correct responses. Thus, the addition of information that isn't clear how it should be used created a situation that encouraged the operator to defer to the agent suggestion.

Performance on the route selection task was evaluated via correct rejections and acceptances of the agent suggestion. An increased number of correct acceptances and rejections, as well as reduced decision times, were all indicative of improved performance. Route selection performance was anticipated to improve with access to agent reasoning and then decline as access to agent reasoning increased. This hypothesis was not supported. Performance was unchanged in the agent reasoning transparency conditions compared to the information-only condition. Decision times (overall and correct responses) were slightly longer in the agent reasoning transparency conditions compared to the information only condition, which is to be expected due to the additional processing required for the agent reasoning transparency. However, decision times for incorrect responses did not follow this trend, with mean decision time in the most transparent agent reasoning condition being shortest of all conditions. This shortening of deliberation time could indicate complacent behavior is occurring in this condition.

Complacency Potential (CP), as evaluated using the Complacency Potential Rating Scale (CPRS), and Spatial Orientation (SO) test scores were found to be predictive of performance on the route selection task, in that individuals with low CP and those with high SO ability were found to score higher on the route selection task overall. There were also performance differences due to Perceived Attentional Control (PAC); individuals with higher PAC had better performance on the route selection task in all ART conditions. When considered together, these findings offer support for the notion that automation bias is, at least to some degree, an issue stemming from attention resource issues (Parasuraman & Manzey, 2010).

Operator trust in the agent was assessed objectively by evaluating incorrect rejections of the agent's suggestions, and subjectively using the Usability and Trust Survey. As in experiment 1, the objective measure of operator trust indicated no difference in trust due to agent reasoning transparency. However, unlike experiment 1, the subjective measures also indicated no difference in trust or perceived usability due to agent reasoning transparency. Complacency Potential (CP), as evaluated using the Complacency Potential Rating Scale (CPRS), was found to be predictive of operator trust, as evaluated via incorrect rejections and scores on the usability and trust survey. Individuals with low CP were found to have fewer incorrect rejections of the agent recommendation overall and rated the agent as more trustworthy and usable than their high CP counterparts. However, there was no difference in incorrect rejections, trust, or usability evaluations across agent reasoning transparency conditions between high and low CP individuals, which indicates these findings were not affected by the presence (or lack thereof) of agent reasoning transparency.

Operator workload was expected to increase as agent reasoning transparency was increased. However, this hypothesis was not supported. Workload was evaluated using the NASA-TLX and several ocular indices that have been shown to be informative as to cognitive workload. Global NASA-TLX scores decreased as ART increased, but such changes were not significant. Pupil Diameter (PD) also decreased as ART increased, indicating that overall cognitive workload decreased as ART increased. Operator PD was larger in the information-only condition compared to the agent reasoning transparency conditions, indicating that the presence of agent reasoning transparency reduced cognitive workload. This finding contradicts our stated hypothesis. Fixation Count (FC) and Duration (FD) did not differ significantly between the three ART levels, indicating no difference in cognitive workload.

Similar to Global scores, Mental Demand and Physical Demand were greater in ART1 than in ARTs 2 or 3, suggesting that the access to agent reasoning reduced cognitive workload. The ratings for NASA-TLX Temporal Demand and Effort were higher in ART1 than in either ART2 or 3, albeit not significantly different, which would support the Mental Demand ratings. Interestingly, participants also reported higher satisfaction in their Performance in ART2 than in ART3. Although participants reported greater mental demand in ART2 than in ART3, they also stayed more engaged in the task, as indicated by their increased decision times for incorrect responses, resulting in higher performance ratings. Alternatively, the addition of the recency information in ART3 created an overwork condition for the operator which encouraged complacent behavior. The combination of decreased satisfaction in their performance and reduced decision times for incorrect responses in ART3 could be indicative of an OOTL situation.

Situation Awareness (SA) scores were hypothesized to improve with access to agent reasoning; this with the exception of SA1 and SA3 scores in ART3. In this study, SA1 scores evaluated how well the participant maintained a general awareness of their environment. The additional context gained by access to agent reasoning would make certain events and situations more salient, which in turn would lead to improved performance on the route selection task (Hancock & Diaz, 2002). However, increased access to agent transparency was expected to overwhelm the participant, leading to a decline in SA1 and SA3 scores. The hypotheses were not supported, SA scores did not improve with access to agent reasoning, nor did they vary across ART levels. In a high information environment, access to agent reasoning does not appear to affect operator situation awareness. These results offer limited support for experiment 1 findings, in which access to agent reasoning does little to improve SA.

While there were no differences in situation awareness due to the effect of agent reasoning access, there were notable distinctions in SA scores for several individual difference factors. Low complacency potential (CP) individuals overall had higher SA1 scores than their high CP counterparts in all ART levels, which could be due to reduced trust in the agent encouraging them to monitor their surroundings more carefully (Pop, Shrewsbury, & Durso, 2015), in effect, supervising the agent. High working memory capacity (WMC) individuals had higher SA2 scores across all ART levels than their low WMC counterparts, demonstrating their improved ability to assimilate the information from various sources into a coherent understanding (Wickens & Hollands, 2000). Low WMC individuals' SA2 scores were lowest in ART2, which could indicate that the access to agent reasoning overtasked them, and the scoring their lowest SA2 scores in ART2. High spatial orientation (SO) individuals had higher SA2 scores when agent reasoning transparency was available than their low SO counterparts. While both groups had similar SA2 scores in the absence of agent reasoning, when access to agent reasoning became available the High SO individuals' SA2 scores improved while the Low SO individuals'

SA2 scores decreased. Gugerty and Brooks (2004) found that high SO individuals were better able to overlook slight disparities in reference-frame alignments. This ability could explain why high SO individuals appear to have increased skill when combining information from several sources (one of which being a map of the area) into a comprehensive understanding of the environment surrounding the convoy's route.

Access to agent reasoning appeared to have little influence on performance on the target detection task. The number of targets detected in ART3 was significantly lower than the other two conditions, indicating that increased agent reasoning transparency interfered with this task. However, access to agent reasoning had no effect on the number of false alarms reported. Signal Detection Theory was used to evaluate whether access to agent reasoning had any effect on sensitivity or selection criteria. There was no significant difference in either sensitivity to targets, assessed as d', or selection criteria, assessed as Beta, across ART levels. In an information-rich environment, agent reasoning transparency appears to have no effect on sensitivity to targets or target selection criterion.

As in experiment 1, a potential limitation of this work could be the added time information in ART3. Participants in that agent reasoning condition were instructed that the time reflected when the agent received the information upon which it based its recommendation, however they were not instructed how they should use that information in their deliberations. Thus, this information could have appeared ambiguous to the participants, and there could be variability in how they factored this information into their decision based upon their personal experience.

Conclusion

The findings of the present study are important for the design of intelligent recommender and decision-aid systems. Keeping the operator engaged and in-the-loop is important for reducing complacency, which could allow lapses in system reliability to go unnoticed. To that end, we examined how agent reasoning transparency affected complacent behavior, as well as task performance, workload, and trust when the operator had complete information about their task environment.

Access to agent reasoning was found to have little effect on complacent behavior when the operator has complete information about the task environment. However, the addition of information that created ambiguity for the operator appeared to encourage complacency, as indicated by reduced performance and shorter decision times. Agent reasoning transparency did not increase overall workload, which agrees with previous studies (see Mercado et al., 2015), and operators reported higher satisfaction with their performance and reduced mental demand. Contrary to findings previously reported by Helldin et al. (2014) and Mercado et al. (2015), access to agent reasoning did not improve operators' secondary task performance, situation awareness, or operator trust. However, this access did not have a negative effect until transparency increased to such a level as to include ambiguous information, thus encouraging complacency. As such, these findings would suggest that when the operator has complete information regarding their task environment, access to agent reasoning may be beneficial, but not dramatically so. However, agent reasoning transparency that includes ambiguous information does have negative effects, as such, the amount of transparency and the type of information conveyed to the operator should be carefully considered.

CHAPTER 5: COMPARISON OF EXPERIMENTS 1 & 2

Objective

Results from Experiments 1 and 2 were compared to evaluate how differences in the level of information available to the operator interacted with access to the agents' reasoning and uncertainty information. In ART1, the only difference between experiment 1 (EXP1) and experiment 2 (EXP2) was the amount of information the participant received via the map icons. In ARTs 2 and 3, agent reasoning transparency was similar between the two experiments in that participants were shown the agent reasoning equating to each map icon; there were simply more icons in EXP2 to explain. However, in EXP2 participants were also told how the agent factored each piece of information into its recommendation via the weighing factor, thus there was a slight increase in agent reasoning transparency in ARTs 2 and 3 compared to EXP1.

Hypotheses:

Based on the review of the literature, this proposal posits the following hypotheses:

It is hypothesized that complacent behavior in the high information environment (EXP2) will be lower than in the low information environment (EXP1) in the absence of agent reasoning (ART1). The additional information should help the participant successfully maneuver their environment more safely. The presence of agent reasoning (ART2) will assist the operator in understanding the additional environmental information, resulting in reduced incorrect acceptances in the high information environment (EXP2) from the low information environment (EXP1). However, the increase in agent reasoning transparency (ART3) will overload the operator, and as a result, incorrect acceptances will be greater in the high information environment (EXP2) than in the low information environment (EXP1).

Hypothesis 1: Incorrect acceptances will be lower in EXP2 than in EXP1 in ART1 (EXP1 > EXP2), as the additional environmental information will reduce the operator's dependency on the agent's recommendations. In ART2, incorrect acceptances will be lower in EXP2 than in EXP1 due to the presence of agent reasoning (EXP1 > EXP2). In ART3, incorrect acceptances will be higher in EXP2 than in EXP1 (EXP1 < EXP2) due to overloading the operator with information.

Hypothesis 2: Performance (number of correct rejects and accepts) on the route selection task in Experiment 2, compared to Experiment 1, will be:

- Lower in ART1, due to increased environmental information without access to agent reasoning (EXP1 > EXP2).
- Greater in ART2, due to access to agent reasoning, (EXP1 < EXP2), and
- Lower in ART3, due to information overload as a result of the increase in transparency of the agent reasoning which included ambiguous information (EXP1 > EXP2).
- In all conditions, time to decide on the route selection task will be higher in Experiment 2 than Experiment 1 (EXP1 < EXP2).

Hypothesis 3: Operator trust in the agent will be greater in EXP2 than in EXP1 for ARTs 1 and 2 (EXP1 < EXP2). However, operator trust will be lower in EXP2 than in EXP1 for ART3 (EXP1 > EXP2).

Hypothesis 4: Operator perceived workload will be greater in Experiment 2 than in Experiment 1 for all ARTs, (EXP1 < EXP2). Inferred measures of workload (i.e., pupil diameter, fixation count, and fixation duration) will also show increased workload.

Hypothesis 5: The increased environmental information will result in lower SA scores in EXP2 than in EXP1 in ARTs 1 and 3 (EXP1 > EXP2) for SA 1 and SA 3 measures. SA 2 scores will be higher in EXP2 than in EXP1 in ARTs 1 and 2, however, will be lower in ART3.

- SA1: ARTs 1, 2 & 3: EXP1 > EXP2
- SA2: ARTs 1 & 2: EXP1 < EXP2; ART3: EXP1 > EXP2.
- SA3: ARTs 1, 2 & 3: EXP1 > EXP2

Hypothesis 6: Performance in the Target Detection Task, in both Targets Detected and False Alarms, will be worse in EXP2 than in EXP1 in all ARTs, due to information overload.

- Number of Targets Detected: EXP1 > EXP2
- False Alarms: EXP1 < EXP2.

Table 31. The anticipated pattern of findings (hypotheses) for comparison of Experiment 2 results to Experiment 1 results.

	DV Measure	ART 1	ART 2	ART 3
Route Selection Task (RS)	Correct accepts and rejects	LI > HI	LI < HI	LI > HI
	Decision Time	LI < HI	LI < HI	LI < HI
Target Detection Task (TD)	Targets Detected	LI > HI	LI > HI	LI > HI
	False Alarms	LI < HI	LI < HI	LI < HI
Complacent Behavior	Incorrect Acceptances	LI > HI	LI > HI	LI < HI
Situation Awareness Scores	SA Level 1 Queries (Perception)	LI > HI	LI > HI	LI > HI
	SA Level 2 Queries (Comprehension)	LI < HI	LI < HI	LI > HI
	SA Level 3 Queries (Projection)	LI > HI	LI > HI	LI > HI
Workload	Global NASA-TLX	LI < HI	LI < HI	LI < HI
Trust	Incorrect Rejections	LI < HI	LI < HI	LI < HI
	Usability and Trust Survey	LI < HI	LI < HI	LI < HI

LI - Low Information (EXP1), HI - High Information (EXP2)

Results

Data was examined using independent samples t-tests ($\alpha = .05$) within each agent reasoning transparency (ART) level between Experiment 1 (EXP1) and Experiment 2 (EXP2). Equal variances between groups were not assumed. Specifically, ART1 was compared to ART1, ART2 to ART2, and ART3 to ART3 for each measure of interest. Means, SD, SE, and 95% CI are reported for each measure.

Complacent Behavior Evaluation

Hypothesis 1: Incorrect acceptances will be lower in EXP2 than in EXP1 in ART1 (EXP1 > EXP2), as the additional environmental information will reduce the operator’s dependency on the agent’s recommendations. In ART2, incorrect acceptances will be lower in EXP2 than in EXP1 due to the presence of agent reasoning (EXP1 > EXP2). In ART3, incorrect acceptances will be higher in EXP2 than in EXP1 (EXP1 < EXP2) due to overloading the operator with information.

Descriptive statistics for incorrect acceptances and EXP1 – EXP2 t-test results are shown in Table 32.

Table 32. Descriptive statistics for incorrect acceptances, sorted by experiment (EXP) for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	3.25	2.27	0.51	(2.19, 4.31)	27.6	4.03	< .001	1.35
	EXP 2	20	0.98	1.11	0.25	(0.46, 1.49)				
ART 2	EXP 1	20	1.15	1.31	0.29	(0.54, 1.76)	33.9	0.70	.488	0.23
	EXP 2	20	0.90	0.91	0.20	(0.47, 1.33)				
ART 3	EXP 1	20	2.65	2.32	0.52	(1.56, 3.74)	34.2	1.81	.079	0.58
	EXP 2	20	1.50	1.64	0.37	(0.73, 2.27)				

Evaluating incorrect acceptances between experiments, it is evident that, overall, more incorrect acceptances occurred in EXP1 than EXP2 (Figure 47). There was a significant correlation between EXP and the number of incorrect acceptances, regardless of ART, r = -.26, p = .013. In ART1, which had no agent reasoning available for the operator, there were fewer incorrect acceptances in EXP2 than EXP1. This supports the hypothesis and is strong evidence that operator knowledge of the task environment can reduce complacent behavior even in the absence of agent reasoning. As predicted, incorrect acceptances were also lower in EXP2 than in EXP1 in ART2. However, this result was not statistically significant. It was expected that the increased agent reasoning transparency in ART3 would overwhelm the operator in EXP2, resulting in higher incorrect acceptances. However, this was not the case. Although EXP2 mean scores in ART3 were greater than those in ARTs 1 or 2, indicating that the increased transparency was not without its cost, scores were significantly lower than in EXP1. Overall, these findings are evidence of the importance of information in addition to agent reasoning transparency for reducing the complacent behavior.

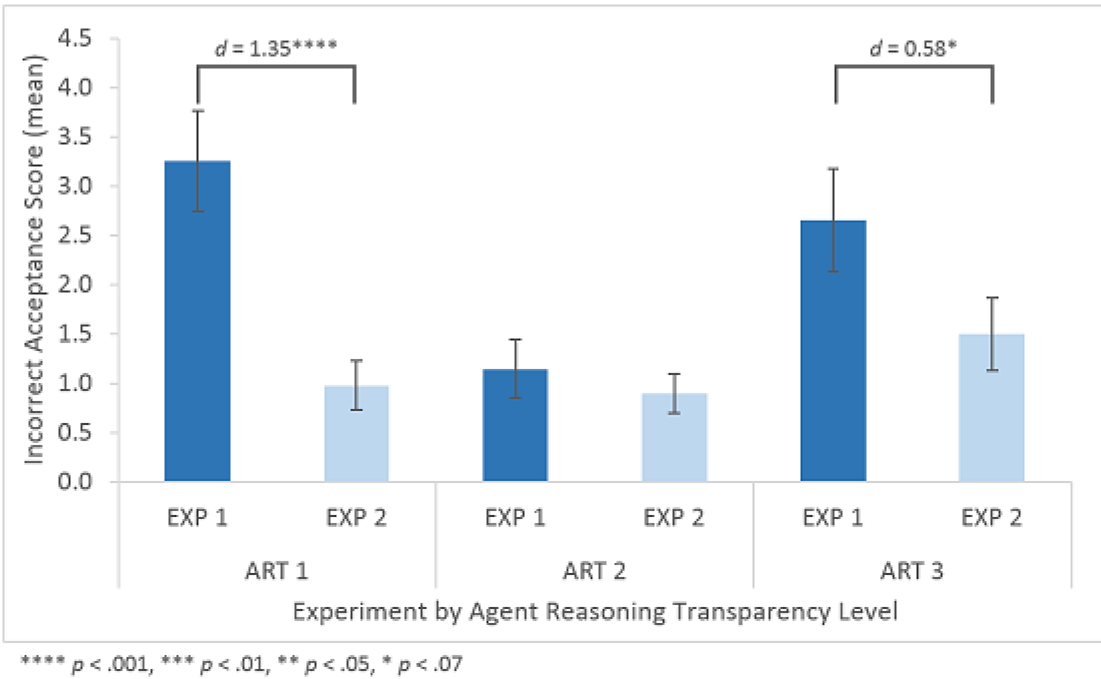


Figure 47. Average incorrect acceptances by experiment for each agent reasoning transparency (ART) level. Bars denote SE.

Participants' scores were further analyzed by comparing the number of participants who had no incorrect acceptances, by ART level, between EXP1 and EXP2 (Figure 48). Chi-square analysis found a significant difference in ART1 in the number of participants with no incorrect acceptances, $X^2(6) = 15.26$, $p = .018$, Cramer's $V = .618$, but no difference in ART2, $X^2(5) = 3.35$, $p = .646$, Cramer's $V = .290$, or ART3, $X^2(6) = 8.23$, $p = .222$, Cramer's $V = .454$. In ART1, the increased information in EXP2 appeared to improve the participants' ability to discern when the agent was incorrect compared to EXP1. However, the addition of agent reasoning in ARTs 2 and 3 appeared to improve EXP1 participants' ability to discern when the agent was incorrect to the same degree as in EXP2. When participants did incorrectly accept the agent's recommendation, more participants made incorrect acceptances in EXP1 (n = 43) than in EXP2 (n = 35), across all ARTs. Of these, 89% of participants in EXP2 scored less than 50% on incorrect acceptances, compared to 51% of those in EXP1.

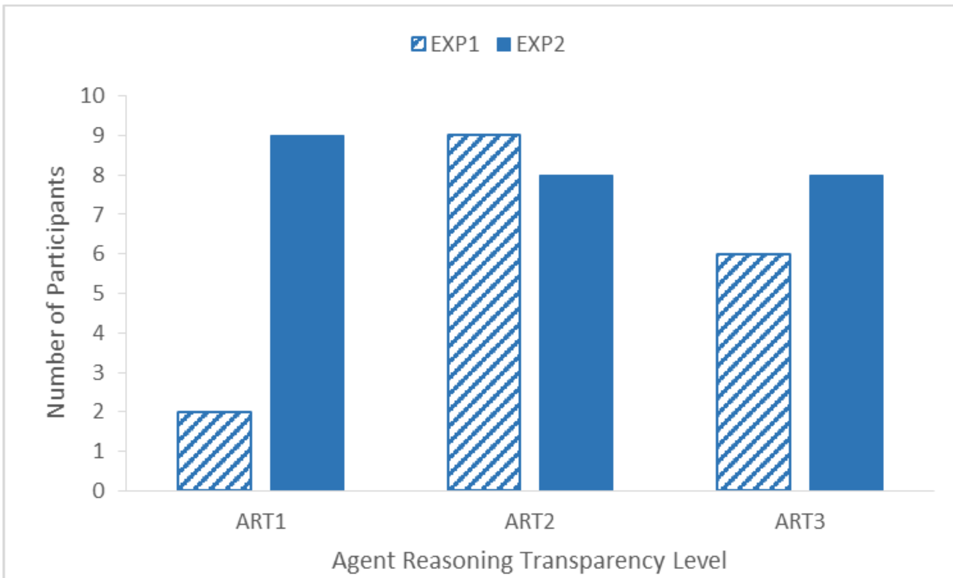


Figure 48. Between experiment comparisons of the number of participants who had no incorrect acceptances in each agent reasoning transparency (ART) level.

Decision time for responses on the route selection task at those locations where the agent recommendation was incorrect was evaluated. It was hypothesized that decision time would increase as agent reasoning transparency increased, and decision times in EXP2 would be longer than those in EXP1 as participants should require additional time to process the extra information. Thus, reduced time could indicate less time spent in deliberation, which could be an indication of complacent behavior. Descriptive statistics for Decision Times and EXP1 – EXP2 t-test results are shown in Table 33.

Table 33. Descriptive statistics for average decision time at those locations where the agent recommendation is incorrect, sorted by experiment (EXP) for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	7.63	3.10	0.69	(6.18, 9.08)	36.9	-3.27	.002	1.04
	EXP 2	20	11.14	3.68	0.82	(9.42, 12.87)				
ART 2	EXP 1	20	7.20	2.77	0.62	(5.91, 8.50)	36.7	-4.43	< .001	1.41
	EXP 2	20	11.51	3.35	0.75	(9.94, 13.08)				
ART 3	EXP 1	20	7.89	3.01	0.67	(6.48, 9.30)	35.5	-3.97	< .001	1.27
	EXP 2	20	12.30	3.96	0.89	(10.45, 14.16)				

Evaluating decision times at those locations where the agent recommendation was incorrect between experiments, it is evident that participants took longer deliberating in EXP2 than EXP1 (Figure 49) across all ARTs, which supports the hypothesis. This difference was smallest in ART1 (ΔM = 3.52), and larger when agent reasoning transparency was present (ART2, ΔM = 4.31; ART3, ΔM = 4.42). Participants took longer to reach their decisions in EXP2 than in EXP1, most likely due to the increased environmental information and increased agent reasoning. It is interesting that in ART3, when agent reasoning transparency was at its highest, decision time was the roughly the same as in ART2. In order to understand this lack of difference, decision times were also evaluated by correct/incorrect responses, see Table 34.

Figure 49. Average decision time in seconds for participant responses at decision points where the agent recommendation was incorrect, sorted by experiment, for each agent reasoning transparency (ART) level. Bars denote SE.

Table 34. Descriptive statistics for decision times (in seconds) for participant responses at decision points where the agent recommendation was incorrect. Decision times are sorted by correct rejections, incorrect acceptances, and experiment (EXP) for each agent reasoning transparency (ART) level, and include t-test results for between-experiment comparisons.

			N	Mean	SD	SE	*df*	*t*	*p*	Cohen's *d*
Correct Rejections	ART 1	EXP 1	14	8.96	8.69	2.32	32.0	-0.98	.337	0.34
		EXP 2	20	11.15	4.25	0.95				
	ART 2	EXP 1	20	7.49	3.17	0.71	38.0	-3.73	.001	1.18
		EXP 2	20	11.25	3.19	0.71				
	ART 3	EXP 1	18	8.14	3.47	0.82	36.0	-3.36	.002	1.12
		EXP 2	20	12.94	5.09	1.14				
Incorrect Acceptances	ART 1	EXP 1	18	8.72	4.88	1.15	27.0	-1.73	.096	0.65
		EXP 2	11	12.17	5.76	1.74				
	ART 2	EXP 1	11	6.09	1.76	0.53	14.6	-5.91	< .001	2.65
		EXP 2	12	14.37	4.49	1.30				
	ART 3	EXP 1	14	8.94	5.27	1.41	24.0	-2.01	.056	0.82
		EXP 2	12	15.70	11.23	3.24				

Response times for both correct rejections and incorrect acceptances were significantly longer in EXP2 than EXP1 in all ARTs. However, the differences in response times between EXP1 and EXP2 were greater for the incorrect responses than the associated correct responses in each ART (Figure 50). There was no significant difference in response times between experiments for the notification-only condition, indicating that the increase in information alone did not result in an associated increase in decision time, regardless of correct or incorrect status. Considered along with the reduced number of incorrect acceptances in EXP2, this could be evidence that information alone appears to be effective at mitigating complacent behavior. For correct rejections, differences in response time for the agent reasoning conditions were similar but longer than the response time for the notification-only condition. Response times for incorrect acceptances were considerably longer than those for correct rejections in the same ARTs, which could be evidence that the incorrect responses were due to difficulty integrating all of the available information. In ART3 the difference in response time for incorrect acceptances is considerably longer than that for correct rejections, and not significantly different between the two experiments. This is mainly due to the increased variability of response times in EXP2 in this ART level. The increased variability could indicate that while some participants erred due to difficulty in assimilating the information, others were exhibiting complacent behavior.

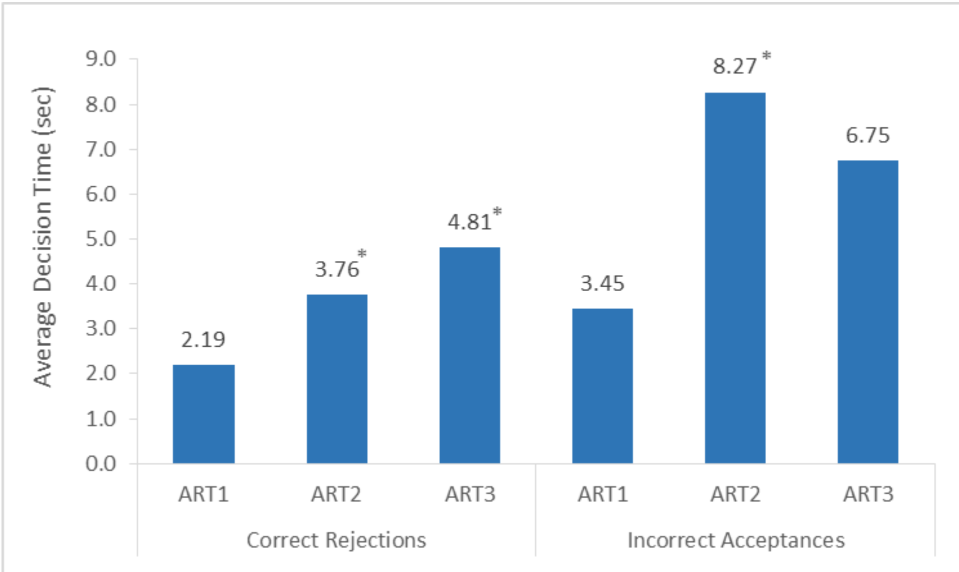


Figure 50. Differences in mean decision times (EXP2-EXP1) for average decision times (in seconds) for correct rejections and incorrect acceptances, sorted by agent reasoning transparency (ART) level. An asterisk (*) denotes that the difference between experiments was significant.

Route Selection Task Performance

Hypothesis 2: Performance (number of correct rejects and accepts) on the route selection task in Experiment 2, compared to Experiment 1, will be:

- Lower in ART1, due to increased environmental information without access to agent reasoning (EXP1 > EXP2).
- Greater in ART2, due to access to agent reasoning, (EXP1 < EXP2), and
- Lower in ART3, due to information overload as a result of the increase in transparency of the agent reasoning which included ambiguous information (EXP1 > EXP2).
- In all conditions, time to decide on the route selection task will be higher in Experiment 2 than Experiment 1 (EXP1 < EXP2).

Descriptive statistics for Route Selection Task scores and EXP1 – EXP2 t-test results are shown in Table 35.

Table 35. Descriptive statistics for route selection task scores, sorted by experiment (EXP) for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	14.10	2.59	0.58	(12.89, 15.31)	35.2	0.93	.358	0.30
	EXP 2	20	13.20	3.46	0.77	(11.58, 14.82)				
ART 2	EXP 1	20	15.90	1.80	0.40	(15.06, 16.74)	30.1	3.18	.003	1.04
	EXP 2	20	13.30	3.18	0.71	(11.81, 14.79)				
ART 3	EXP 1	20	14.70	2.81	0.63	(13.38, 16.02)	37.1	1.35	.187	0.43
	EXP 2	20	13.40	3.28	0.73	(11.86, 14.94)				

Evaluating Route Selection scores between experiments, it is evident that, overall, scores were higher in EXP1 than in EXP2 (Figure 51), although this difference was only significant in ART2. In ART1, which had no agent reasoning available for the operator, and ART3, which had the greatest access to agent reasoning, route selection scores were essentially the same between the two experiments. Increasing the amount of information available to the operator did not improve overall performance on the primary task as predicted, nor did performance improve when agent reasoning transparency was at its highest level. This is evidence that too much access to agent reasoning can have a similar effect on performance as too little. Results in ART2 are contrary to the predicted direction, where performance in EXP2 was expected to be greater than in EXP1. Instead, route selection scores were significantly higher in EXP1 than in EXP2. These results indicate that the combination of high environment information and access to agent reasoning can have a detrimental effect on task performance.

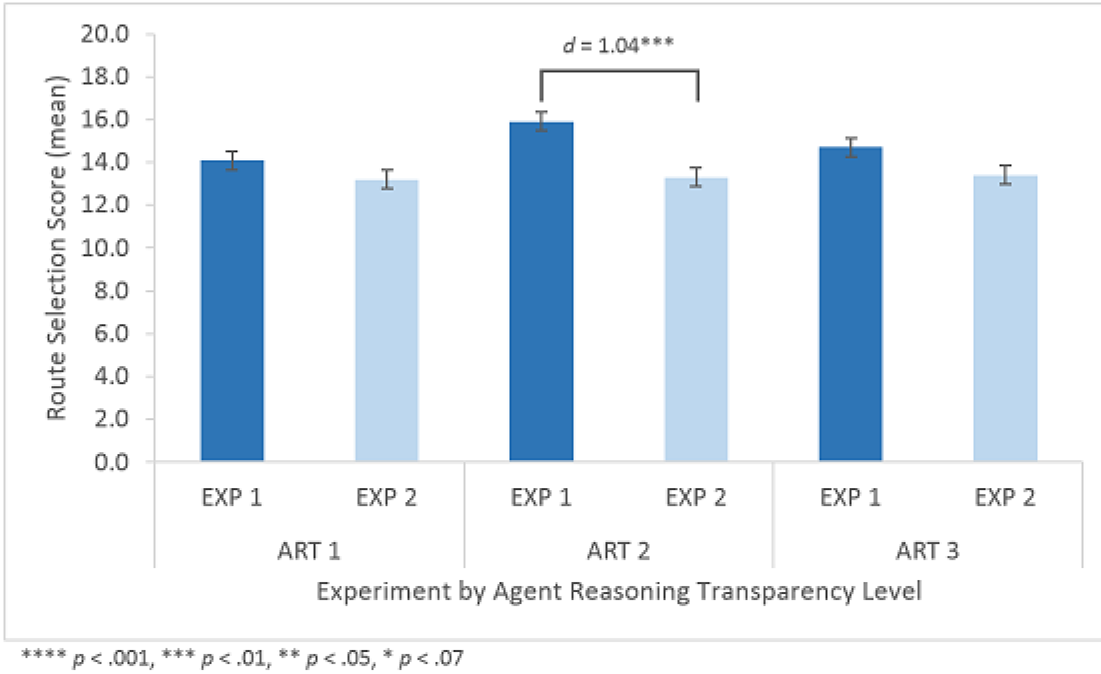


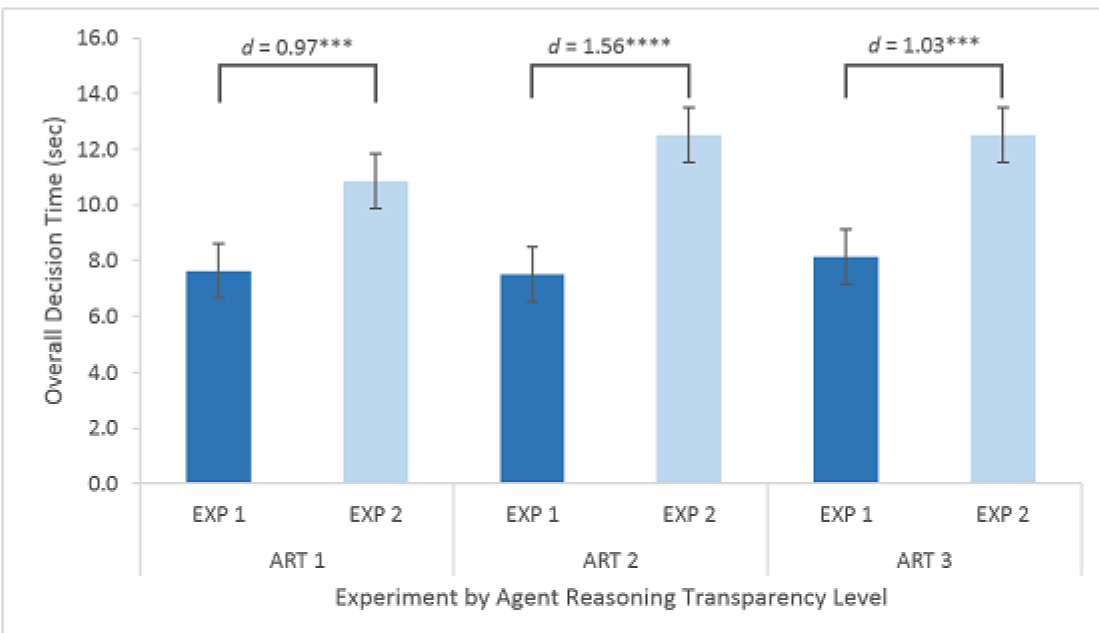
Figure 51. Average route selection task score by experiment for each agent reasoning transparency (ART) level. Bars denote SE.

Operator performance was also evaluated via response time on the route selection task. Descriptive statistics for Overall Decision Times and EXP1 – EXP2 t-test results are shown in Table 36.

Table 36. Descriptive statistics for overall decision times (in seconds) for the route selection task, sorted by experiment (EXP) for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	7.64	3.60	0.81	(5.95, 9.32)	37.0	-3.06	.004	0.97
	EXP 2	20	10.86	3.04	0.68	(9.44, 12.82)				
ART 2	EXP 1	20	7.51	3.36	0.75	(5.93, 9.08)	37.7	-4.92	< .001	1.56
	EXP 2	20	12.53	3.09	0.69	(11.08, 13.97)				
ART 3	EXP 1	20	8.14	3.62	0.81	(6.46, 9.84)	34.9	-3.21	.003	1.03
	EXP 2	20	12.52	4.91	1.10	(10.22, 14.81)				

Overall decision time on the route selection task was hypothesized to be longer in EXP2 than in EXP1, and the findings support the hypothesis. Comparing decision times between experiments, it is evident that times were significantly longer in EXP2 than in EXP1 (Figure 52). This difference was smallest in ART1 (ΔM = 3.22), and larger when agent reasoning transparency was present (ART2, ΔM = 5.02; ART3, ΔM = 4.38). Participants took longer to reach their decisions in EXP2 than in EXP1, most likely due to the increased environmental information and increased agent reasoning. It is interesting that in ART3 when agent reasoning transparency was at its highest, decision time was the same as in ART2. In order to understand this lack of difference, decision times were also evaluated by correct/incorrect responses, see Table 37.



**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Figure 52. Average overall decision times (in seconds) by experiment for each agent reasoning transparency (ART) level. Bars denote SE.

Table 37. Descriptive statistics for decision times (in seconds) for the route selection task, sorted by correct and incorrect responses, and experiment (EXP) for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

			N	Mean	SD	SE	*df*	*t*	*p*	Cohen's *d*
Correct Responses	ART 1	EXP 1	20	7.52	3.50	0.78	38.0	-2.80	.008	0.89
		EXP 2	20	10.32	2.79	0.62				
	ART 2	EXP 1	20	7.42	3.37	0.75	38.0	-4.23	< .001	1.34
		EXP 2	20	11.95	3.40	0.76				
	ART 3	EXP 1	20	7.98	3.33	0.74	38.0	-3.42	.002	1.04
		EXP 2	20	12.10	4.60	1.03				
Incorrect Responses	ART 1	EXP 1	18	8.85	5.38	1.27	36.0	-2.40	.022	0.78
		EXP 2	20	13.06	5.39	1.21				
	ART 2	EXP 1	17	8.44	4.20	1.02	34.0	-4.67	< .001	1.57
		EXP 2	19	15.58	4.89	1.12				
	ART 3	EXP 1	14	9.16	5.20	1.39	29.0	-2.16	.039	0.82
		EXP 2	17	14.77	8.46	2.05				

Response times for both correct and incorrect responses were significantly longer in EXP2 than EXP1 in all ARTs. However, the differences in response times between EXP1 and EXP2 were greater for the incorrect responses than the associated correct responses in each ART (Figure 53). For correct responses, the difference in response time for the agent reasoning conditions was similar but longer than the response time for the notification-only condition. Response times for incorrect responses were longer than those for correct responses in the same ARTs, which could be evidence that the incorrect responses were due to difficulty integrating all of the available information. The reduced route selection score along with the increased decision times in ART2 supports this notion. However, if this were the case, the difference in response times for incorrect responses in ART3 would be at least as long as that in ART2; instead, it is shorter, and there is no difference in route selection task scores between experiments in ART3. This reduction in response time may be an indication of some participants' exhibiting complacent behavior in the highest ART.

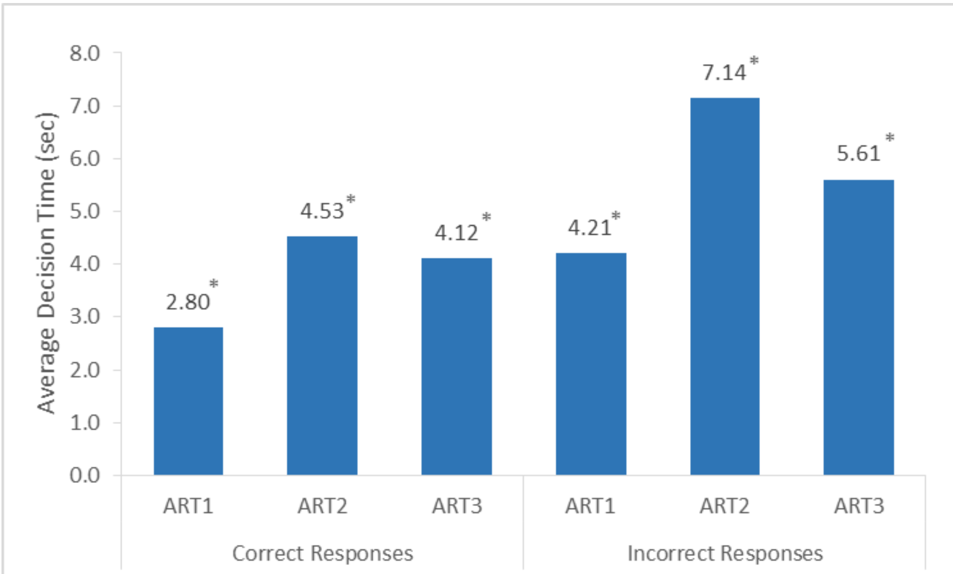


Figure 53. Differences in mean decision times (EXP2-EXP1) for average decision times (in seconds) for correct and incorrect responses, sorted by agent reasoning transparency (ART) level. An asterisk (*) denotes that the difference between experiments was significant.

Operator Trust Evaluation

Hypothesis 3: Operator trust in the agent will be greater in EXP2 than in EXP1 for ARTs 1 and 2 (EXP1 < EXP2). However, operator trust will be lower in EXP2 than in EXP1 for ART3 (EXP1 > EXP2).

Descriptive statistics for incorrect rejections and EXP1 – EXP2 t-test results are shown in Table 38.

Table 38. Descriptive statistics for incorrect rejections, sorted by experiment (EXP) for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	0.75	1.14	0.26	(0.19, 1.26)	23.0	-3.68	< .001	1.31
	EXP 2	20	3.75	3.49	0.78	(2.12, 5.39)				
ART 2	EXP 1	20	0.93	0.77	0.17	(0.57, 1.28)	21.9	-4.48	< .001	1.63
	EXP 2	20	3.80	2.76	0.62	(2.51, 5.09)				
ART 3	EXP 1	20	0.34	0.54	0.12	(0.08, 0.59)	20.2	-4.00	< .001	1.54
	EXP 2	20	3.10	3.04	0.68	(1.68, 4.52)				

Incorrect rejections of the agent recommendation at those locations where the agent recommendation was correct were evaluated as indicative of operator trust. There were significantly more incorrect rejections in EXP2 than in EXP1 in all ARTs (Figure 54). Incorrect rejections in ARTs 1 and 2 were expected to be lower in EXP2 than in EXP1, as such these findings are contrary to the stated hypothesis. Incorrect rejections in ART3 were expected to be higher in EXP2 than in EXP1, due to the combination of the high information environment and increased access to agent reasoning transparency, and this was supported. Across all ARTs, more participants had no incorrect rejections in EXP1 (33 out of 60) than in EXP2 (11 out of 60). The increased number of incorrect rejections in EXP2 is most likely due to the increase in task environment information, which was consistent across ARTs.

**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

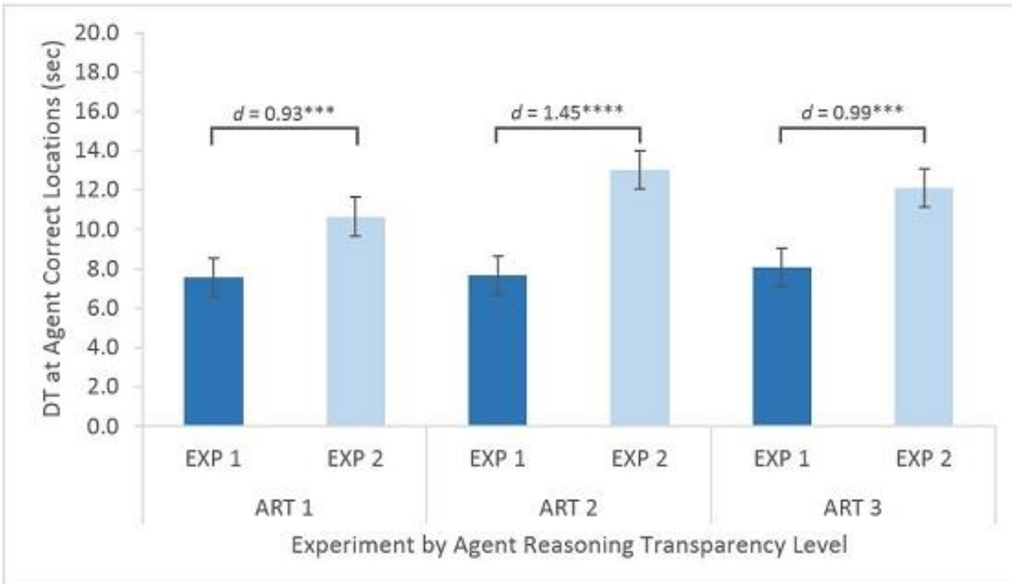
Figure 54. The average number of incorrect rejections of agent recommendations, by experiment, for each agent reasoning transparency (ART) level. Bars denote SE.

Decision time on the route selection task for the locations where the agent recommendation was correct was also compared between experiments. It was hypothesized that decision time would increase as agent reasoning transparency increased, and decision times in EXP2 would be longer than those in EXP1 as participants should require additional time to process the extra information. Descriptive statistics for decision times and EXP1 – EXP2 t-test results are shown in Table 39.

Table 39. Descriptive statistics for average decision time at those locations where the agent recommendation is correct, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	7.55	3.77	0.84	(5.79, 9.32)	35.8	-2.91	.006	0.93
	EXP 2	20	10.65	2.92	0.65	(9.29, 12.02)				
ART 2	EXP 1	20	7.66	3.75	0.84	(5.90, 9.41)	38.0	-4.59	< .001	1.45
	EXP 2	20	13.03	3.67	0.82	(11.32, 14.75)				
ART 3	EXP 1	20	8.07	3.60	0.80	(6.39, 9.76)	36.1	-3.12	.004	0.99
	EXP 2	20	12.12	4.54	1.02	(9.99, 14.24)				

Evaluating decision times at those locations where the agent recommendation was correct between experiments, it is evident that participants took longer deliberating in EXP2 than EXP1 (Figure 55) across all ARTs, which supports the hypothesis. This difference was smallest in ART1 (ΔM = 3.10), and larger when agent reasoning transparency was present (ART2, ΔM = 5.38; ART3, ΔM = 4.04). Participants took longer to reach their decisions in EXP2 than in EXP1, most likely due to the increased environmental information. Decision times were also evaluated by correct/incorrect responses, see Table 40.



**** p < .001, *** p < .01, ** p < .05, * p < .07

Figure 55. Average decision times (in seconds) for operator responses at decision locations where the agent recommendation was correct, sorted by experiment, for each agent reasoning transparency (ART) level. Bars denote SE.

Table 40. Descriptive statistics for decision times (in seconds) for participant responses at decision points where the agent recommendation was correct. Decision times are sorted by correct acceptances, incorrect rejections, and experiment (EXP) for each agent reasoning transparency (ART) level, and include t-test results for between-experiment comparisons.

			N	Mean	SD	SE	*df*	*t*	*p*	Cohen's *d*
Correct Acceptances	ART 1	EXP 1	20	8.21	5.82	1.30	38.0	-1.15	.256	0.38
		EXP 2	20	9.89	2.91	0.65				
	ART 2	EXP 1	20	7.53	3.75	0.84	38.0	-3.79	.001	1.20
		EXP 2	20	12.35	4.28	0.96				
	ART 3	EXP 1	20	8.04	3.59	0.80	38.0	-2.89	.006	0.93
		EXP 2	20	12.10	5.14	1.15				
Incorrect Rejections	ART 1	EXP 1	7	10.79	9.82	3.71	21.0	-0.77	.448	0.32
		EXP 2	16	13.26	5.57	1.39				
	ART 2	EXP 1	14	9.69	4.57	1.22	30.0	-3.54	.001	1.28
		EXP 2	18	15.95	5.24	1.24				
	ART 3	EXP 1	6	9.62	4.59	1.88	19.0	-2.21	.242	0.64
		EXP 2	15	13.20	6.62	1.71				

Response times for both correct acceptances and incorrect rejections were longer in EXP2 than EXP1 in all ARTs (Figure 56). There was no significant difference in response times between experiments for the notification-only condition (ART1), indicating that the increase in information alone did not result in an associated increase in decision time, regardless of correct or incorrect response status. Decision times in ART2 were significantly longer in EXP2 than in EXP1 regardless of correct or incorrect response status. This could indicate either more distrustful behavior, the participant's level of engagement with the agent, or difficulty integrating the information. However, it is likely that the large increase in decision time for EXP2 for incorrect rejections is an indication of difficulty integrating the available information.

In ART3, decision times for incorrect rejections were shorter than those for correct acceptances. This difference was significant for correct acceptances. However there was no significant difference in decision times for incorrect rejections even though there were considerably more incorrect rejections in EXP2 than in EXP1. This could be an indication that the incorrect rejections in ART3 were due to an overwork situation, rather than difficulty integrating information, i.e. complacent behavior or overtrust.

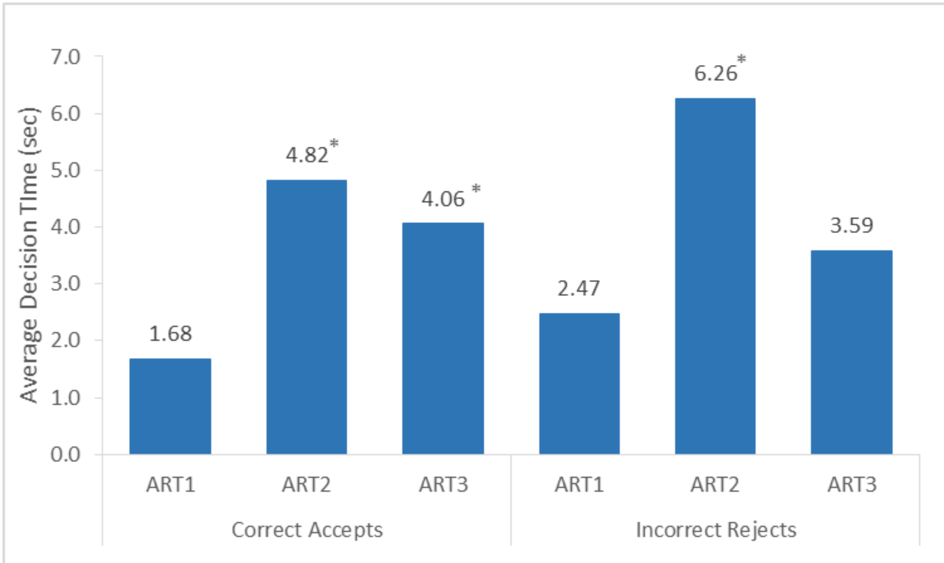


Figure 56. Differences in mean decision times (EXP2-EXP1) for average decision times (in seconds) for correct acceptances and incorrect rejections, sorted by agent reasoning transparency (ART) level. An asterisk (*) denotes that the difference between experiments was significant.

Usability and Trust survey results were also compared between experiments. Descriptive statistics for Usability and Trust Survey scores and EXP1 – EXP2 t-test results are shown in Table 41.

Table 41. Descriptive statistics for Usability and Trust Survey score, sorted by experiment (EXP) for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	104.40	12.91	2.89	(98.36, 110.44)	33.2	2.52	.017	0.81
	EXP 2	20	91.30	19.29	4.31	(82.27, 100.33)				
ART 2	EXP 1	20	95.15	16.94	3.79	(87.22, 103.08)	37.8	0.76	.449	0.24
	EXP 2	20	91.20	15.73	3.52	(83.84, 98.56)				
ART 3	EXP 1	20	106.95	17.79	3.98	(98.63, 115.27)	34.8	2.71	.010	0.87
	EXP 2	20	93.60	13.03	2.91	(87.50, 99.70)				

Independent samples t-tests were used to compare overall Usability and Trust scores between experiments (Figure 57). Usability and Trust survey scores were higher in

EXP1 than in EXP2 across all ART levels, although this difference was not significant in ART2.

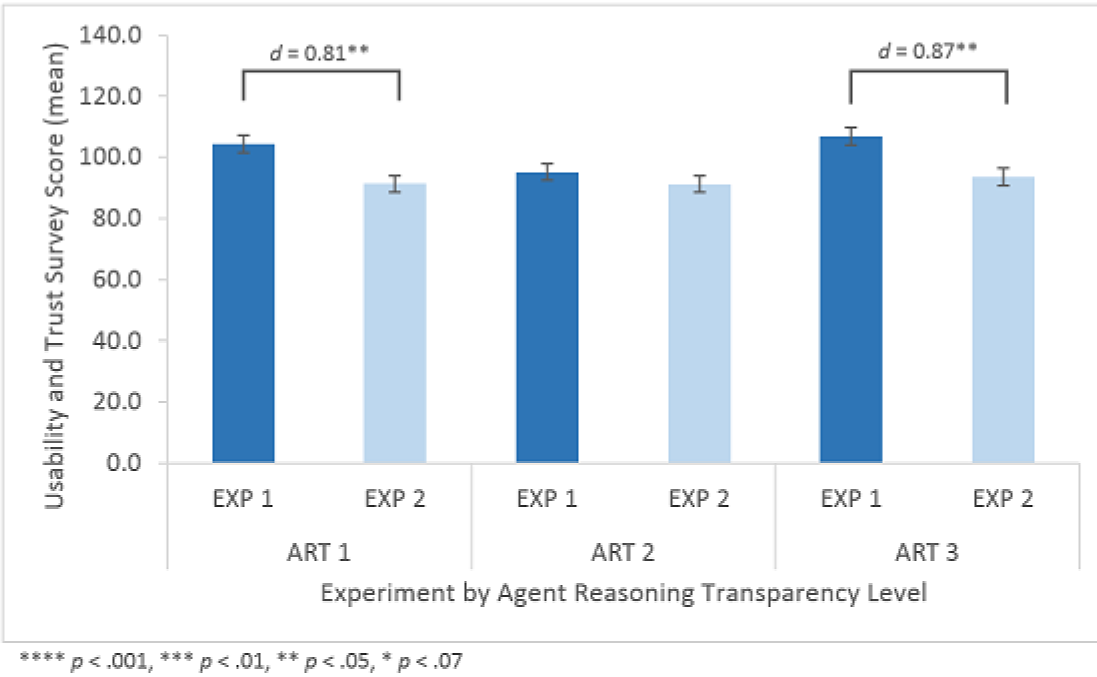


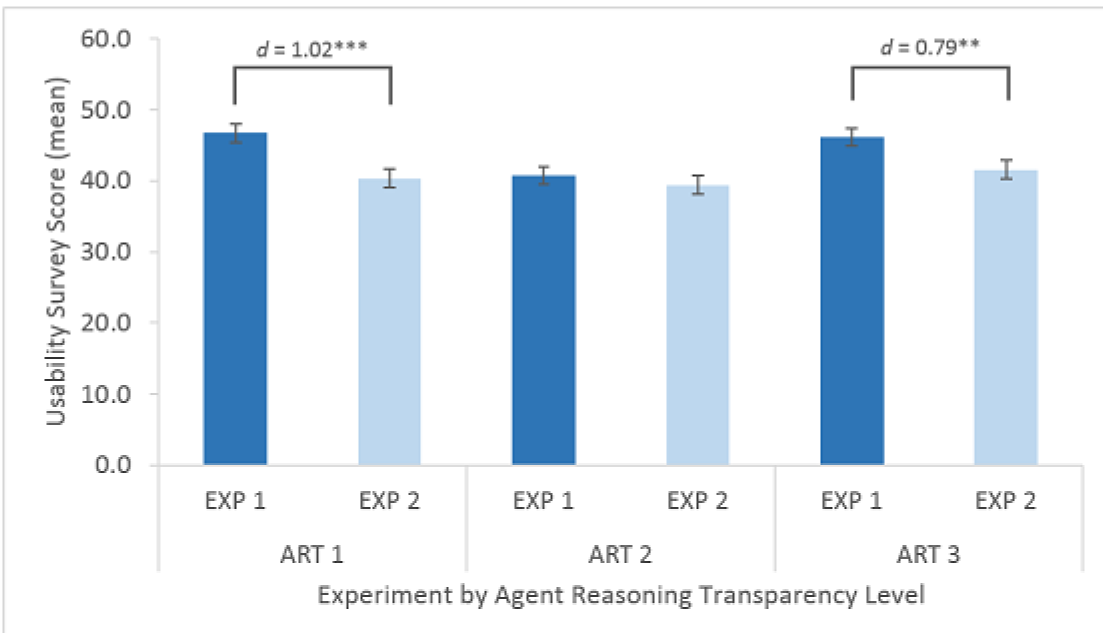
Figure 57. Average usability and trust survey score, by experiment, for each agent reasoning transparency (ART) level. Bars denote SE.

Usability survey results were compared between experiments. Descriptive statistics for Usability Survey scores and EXP1 – EXP2 t-test results are shown in Table 42.

Table 42. Descriptive statistics for Usability Survey score, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	46.75	5.33	1.19	(44.26, 49.24)	35.1	3.20	.003	1.02
	EXP 2	20	40.35	7.18	1.61	(36.99, 43.71)				
ART 2	EXP 1	20	40.75	6.60	1.48	(37.66, 43.84)	37.7	0.65	.520	0.21
	EXP 2	20	39.45	6.05	1.35	(36.62, 42.28)				
ART 3	EXP 1	20	46.20	5.90	1.32	(43.44, 48.96)	38.0	2.51	.017	0.79
	EXP 2	20	41.60	5.70	1.27	(38.93, 44.27)				

Examining the Usability scores separately from the Trust survey scores, there is a significant difference in perceived usability between the two experiments. Usability scores were higher for EXP1 than EXP2 in ARTs 1 and 3 (Figure 58). This indicates that the extra information provided in EXP2 affected the operator perception of agent usability in these ARTs. However, this appears to have been mitigated in ART2, where there was no significant difference in evaluation between the two experiments.



**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Figure 58. Average usability survey scores, by experiment, for each agent reasoning transparency (ART) level. Bars denote SE.

Trust survey results were compared between experiments. Descriptive statistics for Trust Survey scores and EXP1 – EXP2 t-test results are shown in Table 43.

Table 43. Descriptive statistics for Trust Survey score, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	58.55	8.28	1.85	(54.67, 62.43)	32.1	2.20	.035	0.71
	EXP 2	20	50.95	13.08	2.92	(44.83, 57.07)				
ART 2	EXP 1	20	54.40	10.23	2.29	(49.61, 59.19)	37.7	0.78	.439	0.25
	EXP 2	20	51.75	11.19	2.50	(46.51, 56.99)				
ART 3	EXP 1	20	61.60	11.72	2.62	(56.12, 67.08)	34.9	2.95	.006	0.94
	EXP 2	20	52.00	8.61	1.93	(47.97, 56.03)				

Examining the Trust scores separately from the Usability survey scores, there is a significant difference in operator subjective trust between the two experiments. Trust scores were higher for EXP1 than EXP2 in all ART levels (Figure 59), and this difference was significant in ARTs 1 and 3. This indicates that the extra information provided in EXP2 reduced operator trust in the agent. However, the access to agent reasoning in ART2 also reduced operator trust in EXP1, where there was no significant difference in trust survey scores between the two experiments.

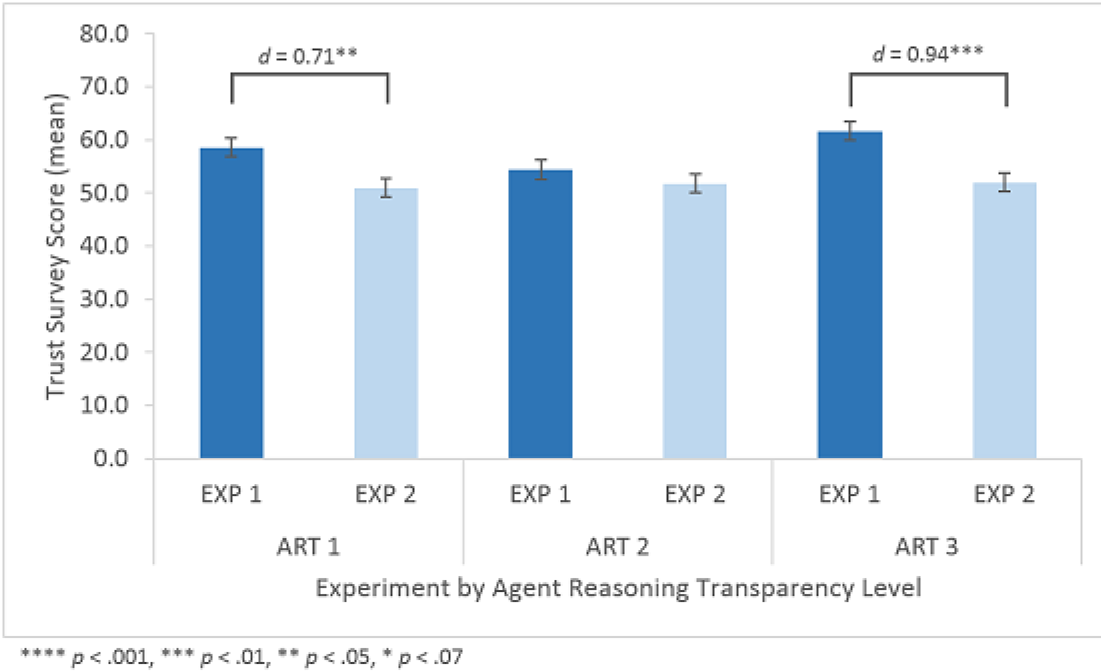


Figure 59. Average trust survey scores, by experiment, for each agent reasoning transparency (ART) level. Bars denote SE.

Workload Evaluation

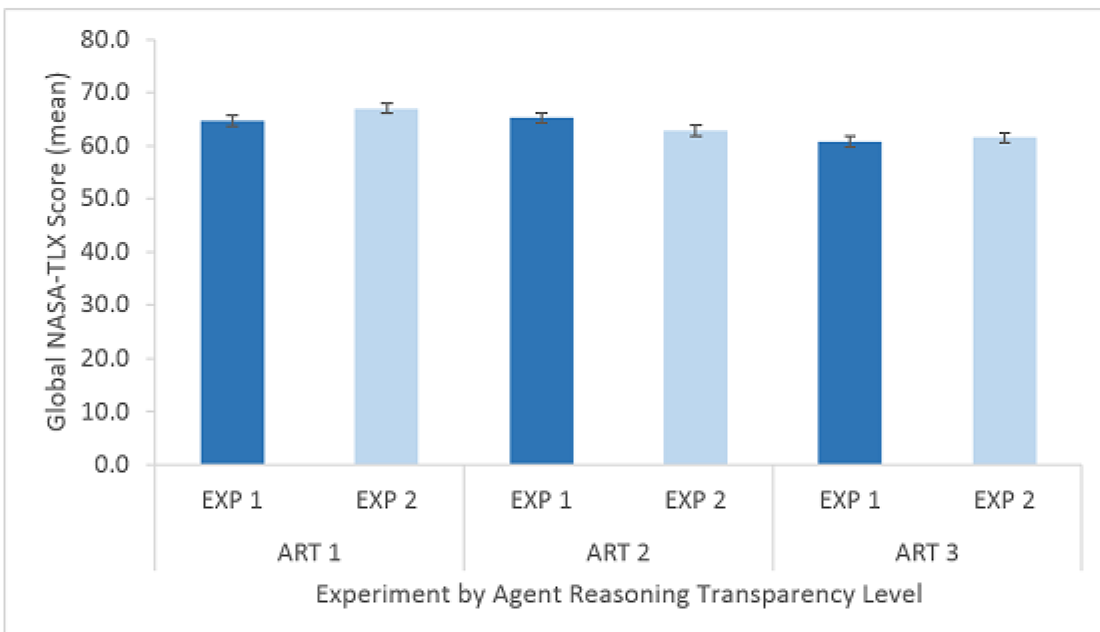
Hypothesis 4: Operator perceived workload will be greater in Experiment 2 than in Experiment 1 for all ARTs, (EXP1 < EXP2). Objective measures of workload (i.e., pupil diameter, fixation count, and fixation duration) will also show increased workload.

Operator perceived workload was evaluated using the NASA-TLX workload survey, and results were compared between experiments. Descriptive statistics for Global NASA-TLX scores and EXP1 – EXP2 t-test results are shown in Table 44.

Table 44. Descriptive statistics for Global NASA-TLX scores, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

	N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
EXP 1	20	64.70	13.47	3.01	(58.40, 70.01)	36.4	-0.60	.550	0.19
EXP 2	20	67.03	10.87	2.43	(61.95, 72.12)				
EXP 1	20	65.19	12.38	2.77	(59.39, 70.98)	37.6	0.58	.569	0.18
EXP 2	20	62.80	13.89	3.08	(56.35, 69.25)				
EXP 1	20	60.70	14.01	3.13	(54.15, 67.26)	36.7	-0.19	.848	0.06
EXP 2	20	61.48	11.58	2.59	(56.06, 66.90)				

Using independent samples t-test to compare findings, no significant difference in global NASA-TLX scores was found between experiments (Figure 60).



**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Figure 60. Average global NASA-TLX score, by experiment, for each agent reasoning transparency (ART) level. Bars denote SE.

Cognitive workload was also evaluated using several ocular indices, and results were compared between experiments. Descriptive statistics for Pupil Diameter, Fixation

Count, and Fixation Duration and EXP1 – EXP2 t-test results are shown in Tables 45, 46, and 47, respectively.

Table 45. Descriptive statistics for Pupil Diameter, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	19	3.74	0.31	0.07	(3.58, 3.94)	25.7	-0.20	.844	0.07
	EXP 2	18	3.77	0.58	0.14	(3.48, 4.06)				
ART 2	EXP 1	20	3.62	0.35	0.08	(3.46, 3.78)	34.8	1.79	.082	0.59
	EXP 2	17	3.43	0.32	0.08	(3.26, 3.59)				
ART 3	EXP 1	19	3.51	0.40	0.09	(3.31, 3.70)	34.0	0.23	.820	0.08
	EXP 2	17	3.48	0.36	0.09	(3.29, 3.66)				

Table 46. Descriptive statistics for Fixation Count, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	19	4830.81	689.30	158.14	(4498.58, 5163.04)	34.9	-0.16	.877	0.05
	EXP 2	18	4864.48	620.01	146.14	(4556.16, 5172.80)				
ART 2	EXP 1	20	5109.85	819.94	183.34	(4726.10, 5493.59)	35.0	0.64	.526	0.21
	EXP 2	17	4949.58	701.14	170.05	(4589.09, 5310.07)				
ART 3	EXP 1	19	4897.41	667.18	153.06	(4575.84, 5218.98)	33.4	-0.43	.667	0.15
	EXP 2	17	4995.22	680.51	165.05	(4645.33, 5345.10)				

Table 47. Descriptive statistics for Fixation Duration, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	19	260.82	40.24	9.23	(241.43, 280.22)	35.0	-1.42	.165	0.47
	EXP 2	18	279.20	38.57	9.09	(260.01, 298.38)				
ART 2	EXP 1	20	276.59	37.11	8.30	(259.23, 293.96)	31.7	0.95	.351	0.32
	EXP 2	17	263.89	43.44	10.54	(241.55, 286.22)				
ART 3	EXP 1	19	267.18	38.98	8.94	(248.39, 285.97)	33.9	-0.38	.709	0.13
	EXP 2	17	271.67	32.62	7.91	(254.90, 288.44)				

Using independent samples t-test to compare findings, no significant difference in workload between experiments was found for any agent reasoning transparency level, as evaluated using eye measure metrics.

Situation Awareness (SA) Evaluation

Hypothesis 5: The increased environmental information will result in lower SA scores in EXP2 than in EXP1 in ARTs 1 and 3 (EXP1 > EXP2) for SA1 and SA3 measures. SA2 scores will be higher in EXP2 than in EXP1 in ARTs 1 and 2, however, will be lower in ART3.

- SA1: ARTs 1, 2 & 3: EXP1 > EXP2
- SA2: ARTs 1 & 2: EXP1 < EXP2; ART3: EXP1 > EXP2.
- SA3: ARTs 1, 2 & 3: EXP1 > EXP2

Descriptive statistics for SA1 scores and EXP1 – EXP2 t-test results are shown in Table 48.

Table 48. Descriptive statistics for SA1 scores, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	1.35	4.93	1.10	(-0.96, 3.66)	37.3	-0.17	.865	0.05
	EXP 2	20	1.60	4.31	0.96	(-0.42, 3.62)				
ART 2	EXP 1	20	0.10	5.86	1.31	(-2.64, 2.84)	32.8	-1.37	.179	0.44
	EXP 2	20	2.25	3.84	0.86	(0.45, 4.05)				
ART 3	EXP 1	20	3.85	3.65	0.82	(2.14, 5.56)	33.2	1.57	.125	0.51
	EXP 2	20	1.55	5.43	1.22	(-0.99, 4.09)				

SA1 scores were expected to be lower in EXP2 than in EXP1 in all ART levels. When comparing results from EXP1 to EXP2, it is evident that SA1 scores varied widely between experiments and ART levels, however, there were no significant differences between EXP2 and EXP1 at any ART level. The hypothesis was not supported.

Descriptive statistics for SA2 scores and EXP1 – EXP2 t-test results are shown in Table 49.

Table 49. Descriptive statistics for SA2 scores, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	10.90	4.51	1.01	(8.79, 13.01)	35.1	-3.11	.004	0.99
	EXP 2	20	14.80	3.35	0.75	(13.23, 16.37)				
ART 2	EXP 1	20	12.55	3.76	0.84	(10.79, 14.31)	28.8	-0.36	.722	0.12
	EXP 2	20	13.20	7.15	1.60	(9.85, 16.55)				
ART 3	EXP 1	20	11.25	4.96	1.11	(8.93, 13.57)	36.1	-2.21	.034	0.70
	EXP 2	20	15.20	6.28	1.40	(12.26, 18.14)				

SA2 scores were expected to be lower in EXP1 than in EXP2 in ART levels 1 and 2, but higher in EXP1 than EXP2 in ART3. Comparing results from EXP1 to EXP2, it is evident that SA2 scores were higher in EXP2 than in EXP1 for all ART levels, although this difference was not significant in ART2 (Figure 61). Thus, the hypothesis was partially supported. The additional environmental information in EXP2 did improve SA2 scores in ART1, compared to EXP1, which supported the hypothesis. In ART3, the high information environment and the increased access to agent transparency were expected to overload the operator, resulting in lower SA2 scores in EXP2 than in EXP1. However, this was not the case. Participants in EXP2 had higher SA2 scores than their EXP1 counterparts, contrary to the stated hypothesis.

**** *p* < .001, *** *p* < .01, ** *p* < .05, * *p* < .07

Figure 61. Average SA2 scores, by experiment, for each agent reasoning transparency (ART) level. Bars denote SE.

SA3 scores were compared between experiments. Descriptive statistics for SA3 scores and EXP1 – EXP2 t-test results are shown in Table 50.

Table 50. Descriptive statistics for SA3 scores, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	1.90	10.22	2.29	(-2.88, 6.68)	37.7	-0.32	.749	0.10
	EXP 2	20	2.90	9.40	2.10	(-1.50, 7.30)				
ART 2	EXP 1	20	3.35	10.43	2.33	(-1.53, 8.23)	36.5	-0.96	.342	0.31
	EXP 2	20	0.45	8.51	1.90	(-3.53, 4.43)				
ART 3	EXP 1	20	8.10	7.18	1.61	(4.74, 11.46)	36.6	2.41	.021	0.76
	EXP 2	20	2.00	8.78	1.96	(-2.11, 6.11)				

SA3 scores were expected to be lower in EXP2 than in EXP1 in all ART levels. Comparing results from EXP1 to EXP2, SA3 scores were significantly higher in EXP1 than in EXP2 for ART3, but not significantly different in ARTs 1 and 2 (Figure 62). Thus, the hypothesis was partially supported. In ART3, the high information environment and the increased access to agent transparency were expected to overload the operator, resulting in lower SA3 scores in EXP2 than in EXP1.

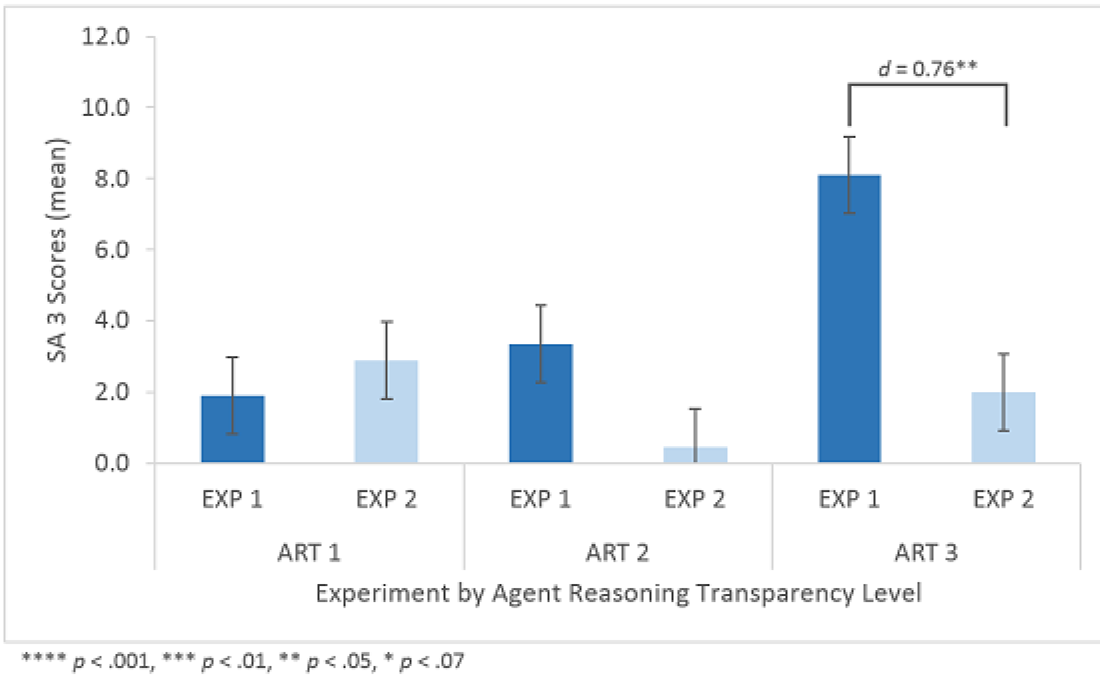


Figure 62. Average SA3 score, by experiment, for each agent reasoning transparency (ART) level. Bars denote SE.

Target Detection Task Performance

Hypothesis 6: Performance in the Target Detection Task, in both Targets Detected and False Alarms, will be worse in EXP2 than in EXP1 in all ARTs, due to information overload.

- Number of Targets Detected: EXP1 > EXP2.
- False Alarms: EXP1 < EXP2.

Descriptive statistics for Target Detection task scores and EXP1 – EXP2 t-test results are shown in Table 51.

Table 51. Descriptive statistics for Target Detection scores, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	44.45	10.10	2.26	(39.72, 49.18)	37.8	-0.24	.812	0.08
	EXP 2	20	45.25	10.96	2.45	(40.12, 50.38)				
ART 2	EXP 1	20	45.05	13.64	3.05	(38.66, 51.44)	36.0	-0.67	.507	0.21
	EXP 2	20	47.65	10.74	2.40	(42.62, 52.68)				
ART 3	EXP 1	20	44.75	10.19	2.28	(39.98, 49.52)	35.6	1.19	.242	0.38
	EXP 2	20	40.30	13.28	2.97	(34.09, 46.51)				

Target Detection task scores were expected to be lower in EXP2 than in EXP1 in all ART levels. Comparing results from EXP1 to EXP2, Target detection scores were not significantly different in any ART level. Thus, the hypothesis was not supported.

Descriptive statistics for the number of reported False Alarms and EXP1 – EXP2 t-test results are shown in Table 52.

Table 52. Descriptive statistics for False Alarms (count), sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	20.80	6.25	1.40	(17.87, 23.73)	38.0	2.29	.028	0.72
	EXP 2	20	16.30	6.18	1.38	(13.41, 19.19)				
ART 2	EXP 1	20	16.35	5.29	1.18	(13.87, 18.83)	37.8	-0.19	.854	0.06
	EXP 2	20	16.65	4.97	1.11	(14.33, 18.97)				
ART 3	EXP 1	20	15.25	3.89	0.87	(13.43, 17.07)	32.2	-0.40	.691	0.13
	EXP 2	20	15.90	6.12	1.37	(13.04, 18.76)				

Reported False Alarms were expected to be lower in EXP1 than in EXP2 in all ART levels. When comparing results from EXP1 to EXP2, there are significantly more

False Alarms reported in EXP1 than in EXP2 in ART1, but no significant difference in ARTs 2 and 3 (Figure 63). Thus, the hypothesis was partially supported.

**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Figure 63. Average reported False Alarms, by experiment, for each agent reasoning transparency (ART) level. Bars denote SE.

In each experiment, results of the target detection task were also evaluated using SDT to determine if there were differences in sensitivity (*d'*) or selection bias (Beta) between the three ARTs. These comparisons follow. Descriptive statistics and EXP1 – EXP2 t-test results for sensitivity (*d'*) are shown in Table 53.

Table 53. Descriptive statistics for *d*' scores, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	2.20	0.32	0.07	(2.05, 2.35)	36.4	-0.85	.400	0.27
	EXP 2	20	2.30	0.40	0.09	(2.11, 2.49)				
ART 2	EXP 1	20	2.31	0.43	0.10	(2.11, 2.52)	36.6	-0.49	.626	0.16
	EXP 2	20	2.38	0.35	0.08	(2.21, 2.54)				
ART 3	EXP 1	20	2.29	0.38	0.09	(2.11, 2.46)	37.3	0.73	.467	0.23
	EXP 2	20	2.19	0.44	0.10	(1.99, 2.39)				

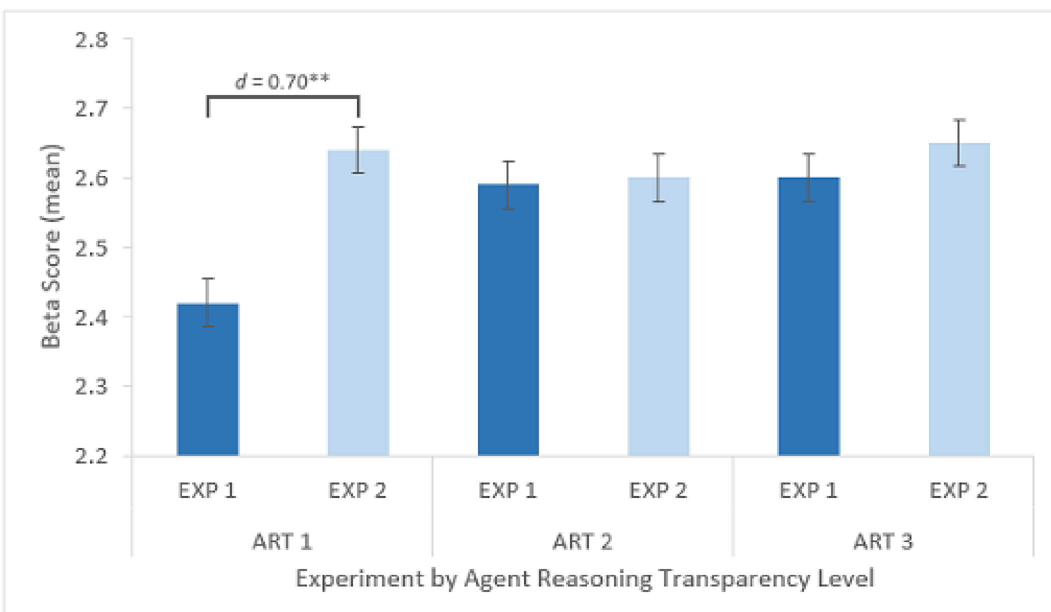
Target Detection task scores were expected to be lower in EXP2 than in EXP1 in all ART levels, so it would be expected that sensitivity to target presence would be higher in EXP1 compared to EXP2. Comparing results from EXP1 to EXP2, mean *d'* scores for EXP2 were higher than those in EXP1 in ARTs 1 and 2, which was contrary to the expected results. However, these results were not significant. The mean *d'* scores in ART3 were higher in EXP1 than in EXP2, which was in the expected direction. However, this finding was not significant. Thus, the hypothesis was not supported.

Descriptive statistics and EXP1 – EXP2 t-test results for selection bias (Beta) are shown in Table 54.

Table 54. Descriptive statistics for Beta scores, sorted by experiment (EXP), for each agent reasoning transparency (ART) level, and t-test results for between-experiment comparisons.

		N	Mean	SD	SE	95% C.I. for Mean	*df*	*t*	*p*	Cohen's *d*
ART 1	EXP 1	20	2.42	0.28	0.06	(2.29, 2.56)	36.8	-2.22	.033	0.70
	EXP 2	20	2.64	0.34	0.08	(2.48, 2.80)				
ART 2	EXP 1	20	2.59	0.35	0.08	(2.43, 2.76)	34.0	-0.11	.912	0.04
	EXP 2	20	2.60	0.25	0.06	(2.49, 2.72)				
ART 3	EXP 1	20	2.60	0.37	0.08	(2.43, 2.78)	37.9	-0.39	.701	0.12
	EXP 2	20	2.65	0.39	0.09	(2.47, 2.83)				

The number of reported False Alarms were expected to be lower in EXP1 than in EXP2 in all ART levels, so it would be expected that selection bias (Beta) would be stricter (higher Beta scores) in EXP1 compared to EXP2. Comparing results from EXP1 to EXP2, mean Beta scores for EXP2 were significantly higher than those in EXP1 in ART1. However, there was no significant difference in Beta scores between the two experiments in ARTs 2 and 3 (Figure 64). The lower Beta scores for EXP1 for ART1 indicate a looser selection criterion was used in this setting, agreeing with the finding that there were more reported False Alarms in this condition. This is evidence that the additional environmental information supplied in EXP2 added support for this task, most likely by removing ambiguity for the operator, thus freeing their attention from the route selection task so that it could be directed to the target detection task. However, the hypothesis was not supported.



**** $p < .001$, *** $p < .01$, ** $p < .05$, * $p < .07$

Figure 64. Average Beta scores, by experiment, for each agent reasoning transparency (ART) level. Bars denote SE.

Discussion

The primary goal of this study was to examine how differing levels of information regarding the task environment and agent reasoning transparency affected complacent behavior in a route selection task. In two experiments, participants supervised a three-vehicle convoy as it traversed a simulated environment and re-routed the convoy when needed with the assistance of an intelligent agent, RoboLeader (RL). Participants received communications from a commander confirming either the presence or absence of activity in the area. They also received information regarding potential events along their route via icons that appeared on a map displaying the convoy route and surrounding area. Participants in Experiment 1 (low information setting) received information about their current route only; they did not receive any information about the suggested alternate route. However, they were instructed that the proposed path was at least as safe as their original route. Participants in Experiment 2 (high information setting) received information about both their current route and the agent recommended alternative route. When the convoy approached a potentially unsafe area, the intelligent agent would recommend re-routing the convoy. The agent recommendations were correct 66% of the time. The participant was required to recognize and correctly reject any incorrect suggestions. The secondary goal of this study was to examine how differing levels of information affected main task and secondary task performance, response time, workload, SA, trust, and system usability.

Complacent behavior was quantified as incorrect acceptances of agent suggestion (Parasuraman et al., 2000), and evaluated via primary (route selection) task response at those decision points where the agent recommendation was incorrect. Increased environmental information was predicted to reduce the number of incorrect acceptances, except when the agent reasoning included information that may be ambiguous for the operator. This prediction was partially supported, as the number of incorrect acceptances were lower in all ARTs in EXP2 than in EXP1. However, the participants in the high information setting (in all ART conditions) may have been more inclined to reject the agent suggestion overall, as the information manipulation gave them more reasons to reject rather than accept (Shafir, 1993). As such, the low number of incorrect acceptances in EXP2 is not particularly informative on its own.

In ART2, participants in EXP1 reduced their incorrect acceptances to nearly the same as those in EXP2. Considering that the number of incorrect acceptances for EXP2 were the same in all ARTs, this result serves to underscore how effective the addition of agent reasoning transparency was in EXP1 in mitigating complacent behavior. There were also interesting differences in the amount of time it took participants to reach their decisions. Even though there was more information available in EXP2 than in EXP1, participants in EXP2 did not take any more time to respond (whether correctly or incorrectly) to the agent suggestion in ART1 than those in EXP1, which may suggest that the additional route information also encouraged more complacent behavior in the absence of agent reasoning. Decision times were significantly longer in ART2 in EXP2 than those in EXP1, particularly for incorrect acceptances, which were nearly twice as long as their decision times for correct rejections. This could indicate difficulty integrating the information, or more likely, difficulty deciding to accept (albeit incorrectly) the agent suggestion in the face of the additional inducement to reject.

Participants in ART3 in EXP2 also had significantly longer decision times for correct rejections than their EXP1 counterparts. However there was no significant difference in their decision times for incorrect acceptances. Considering the results from the other ARTs, it would be reasonable to deduce that this lack of difference in decision times could indicate an overwork situation which encouraged more complacent behavior.

Overall performance on the route selection task was predicted to be worse in the high information setting, except in ART2, when performance in the high information setting would be improved. These predictions were not supported; there was no difference in route selection scores in ARTs 1 or 3 between the two experiments and route selection task scores were lower in ART2 for EXP2 than for EXP1. As previously discussed, these results are most likely due to the added inducement to reject present in EXP2. While decision times were longer in EXP2 than in EXP1 for route selection choices, these findings were anticipated and did not indicate any supervisory control issues.

Operator trust of the agent was expected to be greater in EXP2 than in EXP1, except when access to agent reasoning was at its highest (ART3). Incorrect rejections of the agent recommendation when the agent was correct, along with the associated decision times, were assessed as objective indicators of operator trust. There were significantly more incorrect rejections in EXP2 than in EXP1 in all ARTs. The increased number of incorrect rejections in EXP2 is most likely due to the increase in task environment information, which most likely encouraged participants to reject the agent suggestion. Participants took longer deliberating in EXP2 than EXP1 in all ARTs. The difference in decision times between experiments for ART1 was not significant, which could indicate that the increase in information alone did not result in any associated increase in decision time. In ART2 decision times were significantly longer in EXP2 than in EXP1, and this difference was twice as long for incorrect rejections as for correct acceptances. Considering this, it is most likely that this increase is an indication of difficulty integrating the available information rather than a reflection of the operators trust in the agent. In ART3, the difference in decision times between experiments was significant for correct acceptances. However there was no significant difference in decision times for incorrect rejections even though there were considerably more incorrect rejections in EXP2 than in EXP1. This could be an indication that the incorrect rejections in ART3 were due to an overwork situation, rather than difficulty integrating information, i.e. complacent behavior or overtrust. Taken as a whole, the objective assessments of operator trust indicate no discernable distrust of the agent. However there could be indications of overtrust when agent reasoning transparency was at its highest.

The Usability and Trust survey, the subjective measure of operator trust, indicates that in two conditions, ART1 - when no agent reasoning was available, and ART3 – when agent reasoning transparency was greatest, operators reported higher trust and greater usability in EXP1 than in EXP2. However, in ART2 – when agent reasoning transparency was available but contained no information that would be considered ambiguous or subjective, there was no difference in operator trust of reported usability. Therefore, the hypothesis was only partially supported. In the high information setting, operators appeared to question the agent suggestions more and reported lower trust and usability than in the low information setting. These findings agree with previous research that found when operators question the agent's accuracy and rationale they will demonstrate reduced trust and reliance on the agent (Linegang et al., 2006, Lyons & Havig, 2014).

Operator workload was expected to be greater in the high information setting than in the low information setting. However, this hypothesis was not supported. Workload was evaluated using the NASA-TLX and several ocular indices that have been shown to be informative as to cognitive workload. Similar to findings by Mercado et al. (2016), there were no significant differences in global NASA-TLX scores or eye behavior metrics due to information level.

Situation Awareness (SA) scores were hypothesized to be lower in the high information setting than the low information setting, with the exception of SA2 scores in ART2. There was no difference in SA1 scores between experiments. Contrary to the predicted outcome, SA2 scores were higher in the high information setting when agent reasoning transparency was not available, and again when agent reasoning transparency was at its highest. However, there was no difference in SA2 scores between experiments in ART2. There was no difference in SA3 scores between the two experiments, except in the highest agent reasoning transparency condition, where scores in the low information setting were much higher than those in the high information setting. These findings provide partial support for the hypothesis. Operator comprehension (SA2) benefitted from the increased level of information in EXP2 when agent reasoning transparency was not available, and again when it was ambiguous.

Performance on the secondary task, target detection, was not different between the two experiments. However, false alarms were greater in the low information setting than in the high information setting when agent reasoning transparency was not available. Higher Beta scores indicate that participants were using a looser selection criterion in ART1 in the low information setting than in the high, indicating that having more information about their task environment allowed them to be more discerning when conducting the target detection task.

There were several limitations to this comparative analysis. First, the agent reasoning transparency in EXP2 was arguably greater than that in EXP1, as it contained the weight factors that were not present in EXP1. Therefore, within-condition comparisons contained analysis that attempted to tease apart the effects that resulted from the increase in agent reasoning transparency from those that resulted from the increase in environmental information. A second limitation would be the study paradigm itself. At each decision point, the participant is not choosing which path to take so much as they are deciding whether to reject the agent suggestion. In EXP1, where there is no other information available about the agent's recommended route, there is no strong reason to reject the route. However, in EXP2 when the participants receive information about the alternative route they receive two pieces of information, as compared to the one piece of information they have about their original route. According to decision theory, this additional information would make it more likely that the participant would reject the agent suggestion (Shafir, 1993). Thus the comparison of the effect of information level between the two experiments is not equitable. A third limitation is a difference in information between EXP1 and EXP2. In EXP1, the participant is given one piece of information about their main path, and no information about the alternative route. In EXP2 the participant is given one piece of information about the main path and two pieces of information about the alternative route. Hence, the comparison is not of the effects of an increase in information as much as it is comparing the difference between no information and some information. While these limitations do not negate the findings of the comparative analysis, understanding how they would potentially affect the outcome of this comparison warrants caution in the interpretation of the comparison and generalizing the findings to larger populations.

Conclusion

Understanding the interaction between the amount of information available to the operator and the transparency of agent reasoning is important to designers of intelligent recommender and decision-aid systems. To that end, we examined how the amount of information pertaining to the task environment the operator had, and the increase in agent reasoning transparency, affected complacent behavior, as well as task performance, workload, and trust.

The amount of information the operator had regarding the task environment had a profound effect their proper use of the agent. Increased environmental information resulted in more rejections of the agent recommendation regardless of the transparency of agent reasoning. The way in which the information was presented appeared to create a situation wherein operators were encouraged to reject the agent recommendation. Even so, the addition of agent reasoning transparency appeared to be effective at countering this bias by keeping the operator engaged.

Objective evidence indicated probable complacent behavior in the high information setting when agent reasoning was either not transparent or so transparent as to become ambiguous. However, operators reported lower trust and usability for the agent than when environmental information was limited. This suggests dissonance between operator performance and operator perception of the agent.

Situation awareness (SA2) scores were also higher in the high information environment when agent reasoning was either not transparent or so transparent as to become ambiguous, compared to the low information environment. However, when a moderate amount of agent reasoning was available to the operator, the amount of information available to the operator had no effect on the operator's complacent behavior, subjective trust, or SA. These findings indicate that some negative outcomes resulting from the incongruous transparency of agent reasoning may be mitigated by increasing the information the operator has regarding the task environment.

APPENDIX A:

INFORMED CONSENT FORM

Informed Consent Form

Army Research Laboratory, Human Research & Engineering Directorate
Orlando, FL 32826

Title of Project: Transparency of automation reasoning and its effect on automation-induced complacency.

Project Number: 14-043

Sponsor: Army Research Laboratory

Principal Investigator

Name:	Julia Wright
Division:	Human Factors Integration Division
Branch:	Information Systems Branch
Phone Number:	(407) 208-3348 (DSN 970)
Email:	Julia.l.wright8.civ@mail.mil

You are being asked to join a research study. This consent form explains the research study and your part in it. Please read this form carefully before you decide to take part. You can take as much time as you need. Please ask questions at any time about anything you do not understand. You are a volunteer. If you join the study, you can change your mind later. You can decide not to take part now or you can quit at any time later on.

Location of Research:

University of Central Florida Institute for Simulation and Technology, 3100 Technology Pkwy (Partnership II building), Orlando, FL 32826.

Purpose of the Study:

The purpose of this study is to determine how understanding the reasoning behind an autonomous agents' suggestions affects decision-making and performance. You will play the role of vehicle commander of a manned ground vehicle (MGV), guiding your convoy through an urban environment. In addition to the MGV, you will have an unmanned ground vehicle (UGV) and an unmanned aerial system (UAV) under your control. While supervising the robots, you will also try to maintain awareness of the surroundings of your own vehicle.

Procedures to be followed:

First, you will fill out a demographics questionnaire and complete a complete a working memory capacity test (RSPAN) and a brief color vision evaluation. The score on the RSPAN and color vision tests will determine your eligibility to continue with the experiment. After completing the RSPAN, you will complete some surveys that will assess your attentional control, trait trust in automation, and complacency potential. After these surveys, you will complete two tests which measure your spatial ability. After these tests, you will receive training on the experimental tasks. Your task will be to supervise a convoy of these three vehicles (your own MGV, the UAV, and the UGV) as it moves along a predetermined route from point A to point B. If route revisions are required, the autonomous agent will automatically suggest a new route, however you will have access to the information that the agent has and will need to agree or disagree with the proposed route changes. The autonomous agent will not always recommend the best route. There will be three experimental scenarios. You will learn how to differentiate between insurgents and civilians, and what to do once you detect targets.

The preliminary session (questionnaires and tests) and training will last about 1.5 hours, which will be followed by the experimental session, which will consist of three scenarios and will last about 1.5 hrs. In the experimental scenarios, you will supervise a convoy as it travels through an urban environment. You will try to find targets that are in your immediate environment as well. After completing three scenarios, you will assess your workload by completing a workload questionnaire developed by NASA (NASA-TLX) and complete the usability and trust survey. There will be a 2-minute break between scenarios. You can take longer breaks if necessary. During the experimental session, we will measure your eye movement (where you look at on the screen) using eye tracking equipment. A camera will be used to measure your eye movement; however, only aggregate eye movement data from all the participants will be reported in reports and presentations on the experiment. Your individual data will not be made public. There will not be any video recording of your eyes and face. A calibration process will take place prior to the training session and each scenario.

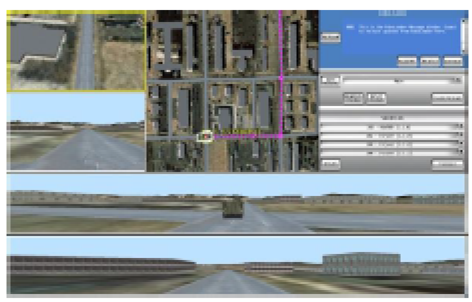


Figure 1. RoboLeader Operator Control Unit.

Discomforts and Risks:

There is minimal risk associated with using simulators such as the one used in this study that is no greater than normal use of a computer.

Benefits:

There are no personal benefits for you for taking part in this study. The results of this study might help us understand how access to agent reasoning affects human performance when interacting with multiple semi-autonomous robots for reconnaissance missions in a multi-tasking environment.

Compensation for Participation:

You will receive your choice of compensation: either payment ($15/hr) or Sona Credit at the rate of 1 credit/hour for taking part in this experiment. You will receive at least 1 hour payment for participating. You must take all compensation in the same method, and will not be allowed to change compensation method once payment has been delivered. You cannot be paid if you are a member of the military, a civilian employee of the U.S. Government, or a family member of an employee of the Human Research & Engineering Directorate.

You will be paid cash by the UCF-IST Prodigy lab payment clerk. You will be given instructions how to receive payment upon completion of the study.

Duration: It will take about 3.5 hours for you to take part in this study.

Confidentiality:

Your participation in this research is confidential. The data will be stored and secured in a locked file cabinet in the Principal Investigator's office. Data with no identifying information (i.e., your name will not be associated with your data) will be transferred to a password-protected computer for data analysis. After the data is put in the computer file, the paper copies of the data will be shredded. This consent form will be sent to the Army Research Laboratory's Institution Review Board, where it will be retained in a secure location for a minimum of three years.

In the event of a publication or presentation resulting from the research, no personally identifiable information will be shared. Publication of the results of this study in a journal, technical report, or presentation at a meeting will not reveal personally identifiable information. The research staff will protect your data from disclosure to people not connected to this study. However, complete confidentiality cannot be guaranteed because officials of the U.S. Army Human Research Protections Office and the Army Research Laboratory's Institutional Review Board are permitted by law to inspect the records obtained in this study to insure compliance with laws and regulations covering experiments using human subjects.

Participation terminated by the investigator:

If you are unable to demonstrate sufficient ability in task performance at the end of your training, participation will be terminated by the investigator.

Consequences of withdrawal:

You may end your participation in the study at any time and there will be no penalty for withdrawing from the study. If in the rare event you ask to stop the study because you do not feel well, you will be asked to remain at the site until you feel better. You will be paid $15.00 an hour for the amount of time you participated in the study, with a minimum of one hour paid.

Contact Information for Additional Questions:

You have the right to obtain answers to any questions you might have about this research both while you take part in the study and after you leave the research site. Please contact anyone listed at the top of the first page of this consent form for more information about this study. You may also contact the Institution Review Board, at (410) 278-5928 with questions, complaints, or concerns about this research or if you feel this study has harmed you. They can also answer questions about your rights as a research participant. You may also call this number if you cannot reach the research team or wish to talk to someone else.

Voluntary Participation:

Your decision to be in this research is voluntary. You can stop at any time. You do not have to answer any questions you do not want to answer. Refusal to take part in or withdrawing from this study will involve no penalty or loss of benefits you would receive by staying in it.

Military personnel cannot be punished under the Uniform Code of Military Justice for choosing not to take part in or withdrawing from this study, and cannot receive administrative sanctions for choosing not to participate.

Civilian employees of the U.S. Government or contractors cannot receive administrative sanctions for choosing not to participate in or withdrawing from this study.

You must be 18 years of age or older to take part in this research study. If you agree to take part in this research study based on the information outlined above, please sign your name and the date below.

You will be given a copy of this consent form for your records.

_______________________________ _______________
Participant Signature Date

_______________________________ _______________
Person Obtaining Consent Date

APPENDIX B:

DEMOGRAPHICS QUESTIONNAIRE

Demographics Questionnaire

Date: ________________ Participant ID: _________

1. General Information

a. Age: _____ Gender: _____ Handedness (L/R): ______

b. How long ago did you have an eye exam? Within the last (Circle one):
6 months 1 year 2 years 4+ years never

c. Do you have any of the following (Circle all that apply):
Astigmatism Near-sightedness Far-sightedness

Other (explain):________________

d. Do you have corrected vision (Circle one)? No Glasses Contact Lenses
If so, are you wearing them today? Yes No

e. Are you in your good/ comfortable state of health physically? YES NO
If NO, please briefly explain:

f. How many hours of sleep did you get last night? ______ hours

2. Military Experience

a. Do you have prior military service? YES NO
If Yes, how long __________

3. Educational Data

a. What is your highest level of education received? Select one.

____ GED ____ Bachelor's Degree
____ High School ____ M.S/M.A
____ Some College ____ Ph.D.
____ Associates or Technical Degree

b. What subject is your degree/education in (for example, Engineering)? ________________

4. Computer Experience

a. How long have you been using a computer?
__Less than 1 year ___ 1-3 years ___4-6 years ___7-10 years ___10+ years

b. How often do you play computer/video games? (Circle one)
Daily 3-4X/ Week Weekly Monthly Once or twice a year Never

c. Enter the names of the games you play most frequently:

____

d. How often do you operate a radio-controlled vehicle (car, boat, or plane)?
Daily 3-4X/ Week Weekly Monthly Once or twice a year Never

e. How often do you use graphics/drawing features in software packages?
Daily 3-4X/ Week Weekly Monthly Once or twice a year Never

APPENDIX C:

ATTENTIONAL CONTROL SURVEY

Attentional Control Survey Participant #_______ Date _________

For each of the following questions, circle the response that best describes you.

It is very hard for me to concentrate on a difficult task when there are noises around.
Almost never, Sometimes, Often, Always

When I need to concentrate and solve a problem, I have trouble focusing my attention.
Almost never, Sometimes, Often, Always

When I am working hard on something, I still get distracted by events around me.
Almost never, Sometimes, Often, Always

My concentration is good even if there is music in the room around me.
Almost never, Sometimes, Often, Always

When concentrating, I can focus my attention so that I become unaware of what's going on in the room around me.
Almost never, Sometimes, Often, Always

When I am reading or studying, I am easily distracted if there are people talking in the same room.
Almost never, Sometimes, Often, Always

When trying to focus my attention on something, I have difficulty blocking out distracting thoughts.
Almost never, Sometimes, Often, Always

I have a hard time concentrating when I'm excited about something.
Almost never, Sometimes, Often, Always

When concentrating, I ignore feelings of hunger or thirst. Almost never, Sometimes, Often, Always

I can quickly switch from one task to another. Almost never, Sometimes, Often, Always

It takes me a while to get really involved in a new task. Almost never, Sometimes, Often, Always

It is difficult for me to coordinate my attention between the listening and writing required when taking notes during lectures. Almost never, Sometimes, Often, Always

I can become interested in a new topic very quickly when I need to.
Almost never, Sometimes, Often, Always

It is easy for me to read or write while I'm also talking on the phone.
Almost never, Sometimes, Often, Always

I have trouble carrying on two conversations at once. Almost never, Sometimes, Often, Always

I have a hard time coming up with new ideas quickly. Almost never, Sometimes, Often, Always

After being interrupted or distracted, I can easily shift my attention back to what I was doing before.
Almost never, Sometimes, Often, Always

When a distracting thought comes to mind, it is easy for me to shift my attention away from it.
Almost never, Sometimes, Often, Always

It is easy for me to alternate between two different tasks. Almost never, Sometimes, Often, Always

It is hard for me to break from one way of thinking about something and look at it from another point of view.
Almost never, Sometimes, Often, Always

APPENDIX D:

CUBE COMPARISON TEST

Cube Comparisons Test   Participant # _______   Date _______

CUBE COMPARISONS TEST -- S-2 (Rev.)

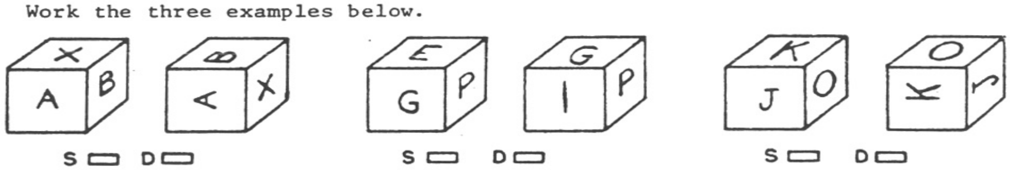
Wooden blocks such as children play with are often cubical with a different letter, number, or symbol on each of the six faces (top, bottom, four sides). Each problem in this test consists of drawings of pairs of cubes or blocks of this kind. Remember, there is a different design, number, or letter on each face of a given cube or block. Compare the two cubes in each pair below.

The first pair is marked D because they must be drawings of <u>different</u> cubes. If the left cube is turned so that the A is upright and facing you, the N would be to the left of the A and hidden, not to the right of the A as is shown on the right hand member of the pair. Thus, the drawings must be of different cubes.

The second pair is marked S because they could be drawings of the same cube. That is, if the A is turned on its side the X becomes hidden, the B is now on top, and the C (which was hidden) now appears. Thus the two drawings could be of the same cube.

<u>Note</u>: No letters, numbers, or symbols appear on more than one face of a given cube. Except for that, <u>any</u> letter, number or symbol can be on the hidden faces of a cube.

Work the three examples below.



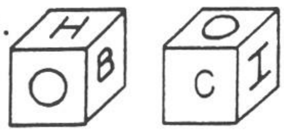
The first pair immediately above should be marked D because the X cannot be at the peak of the A on the left hand drawing and at the base of the A on the right hand drawing. The second pair is "different" because P has its side next to G on the left hand cube but its top next to G on the right hand cube. The blocks in the third pair are the same, the J and K are just turned on their side, moving the O to the top.

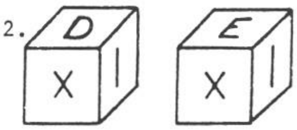
Your score on this test will be the number marked correctly minus the number marked incorrectly. Therefore, it will <u>not</u> be to your advantage to guess unless you have some idea which choice is correct. Work as quickly as you can without sacrificing accuracy.

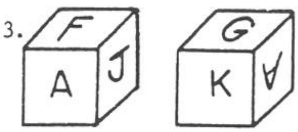
You will have <u>3 minutes</u> for each of the two parts of this test. Each part has one page. When you have finished Part 1, STOP.

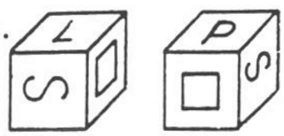
DO NOT TURN THE PAGE UNTIL YOU ARE ASKED TO DO SO.

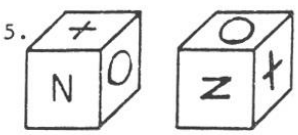
Part 1 (3 minutes)

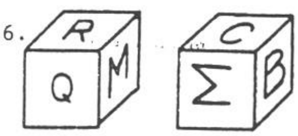
1. 
S ☐ D ☐

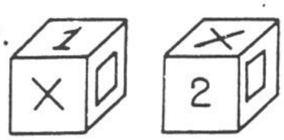
2. 
S ☐ D ☐

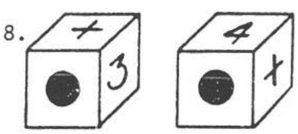
3. 
S ☐ D ☐

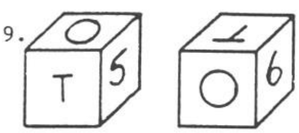
4. 
S ☐ D ☐

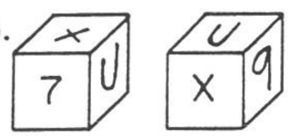
5. 
S ☐ D ☐

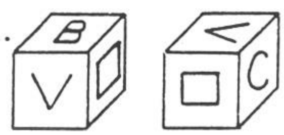
6. 
S ☐ D ☐

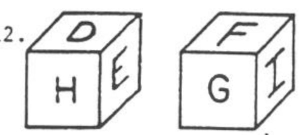
7. 
S ☐ D ☐

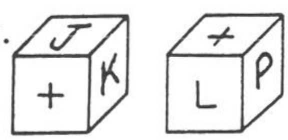
8. 
S ☐ D ☐

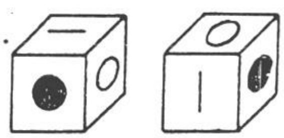
9. 
S ☐ D ☐

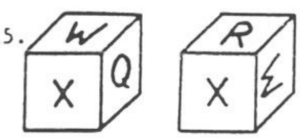
10. 
S ☐ D ☐

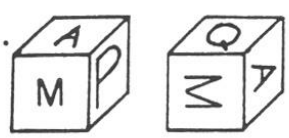
11. 
S ☐ D ☐

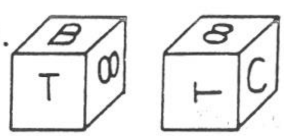
12. 
S ☐ D ☐

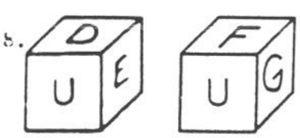
13. 
S ☐ D ☐

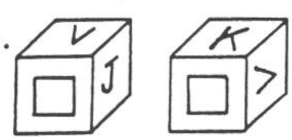
14. 
S ☐ D ☐

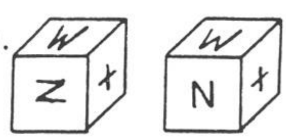
15. 
S ☐ D ☐

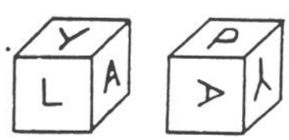
16. 
S ☐ D ☐

17. 
S ☐ D ☐

18. 
S ☐ D ☐

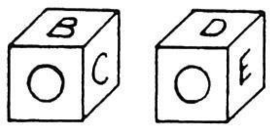
19. 
S ☐ D ☐

20. 
S ☐ D ☐

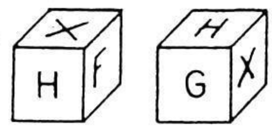
21. 
S ☐ D ☐

DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO. STOP.

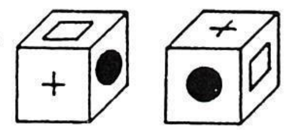
Part 2 (3 minutes)

22. 

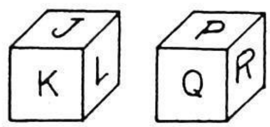
S ☐ D ☐

23. 

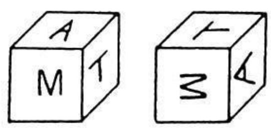
S ☐ D ☐

24. 

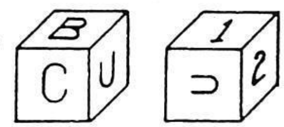
S ☐ D ☐

25. 

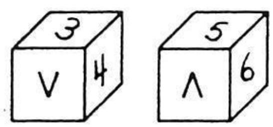
S ☐ D ☐

26. 

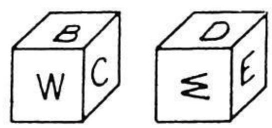
S ☐ D ☐

27. 

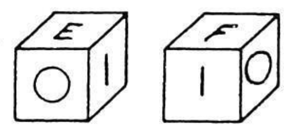
S ☐ D ☐

28. 

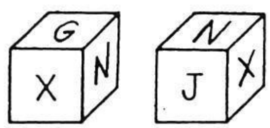
S ☐ D ☐

29. 

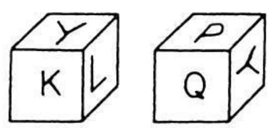
S ☐ D ☐

30. 

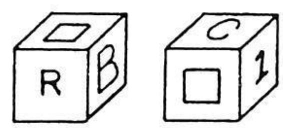
S ☐ D ☐

31. 

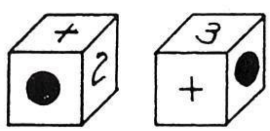
S ☐ D ☐

32. 

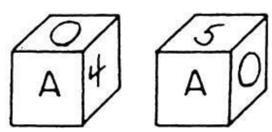
S ☐ D ☐

33. 

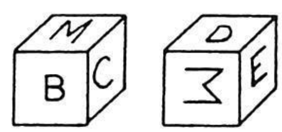
S ☐ D ☐

34. 

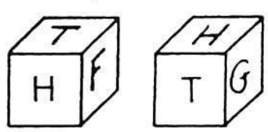
S ☐ D ☐

35. 

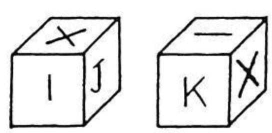
S ☐ D ☐

36. 

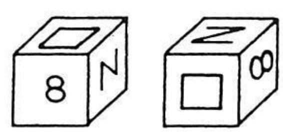
S ☐ D ☐

37. 

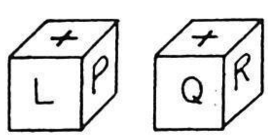
S ☐ D ☐

38. 

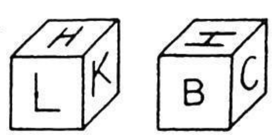
S ☐ D ☐

39. 

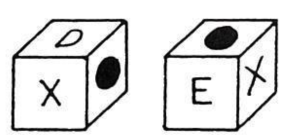
S ☐ D ☐

40. 

S ☐ D ☐

41. 

S ☐ D ☐

42. 

S ☐ D ☐

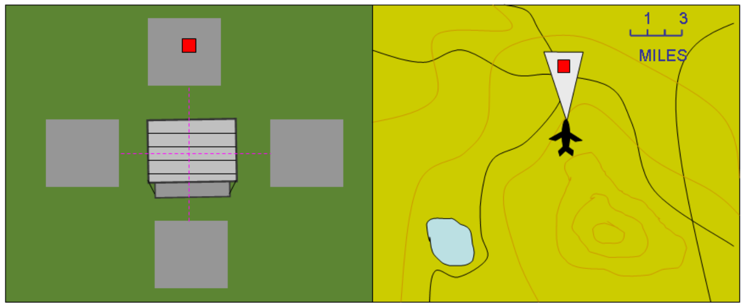
DO NOT GO BACK TO PART 1 AND

DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO. STOP.

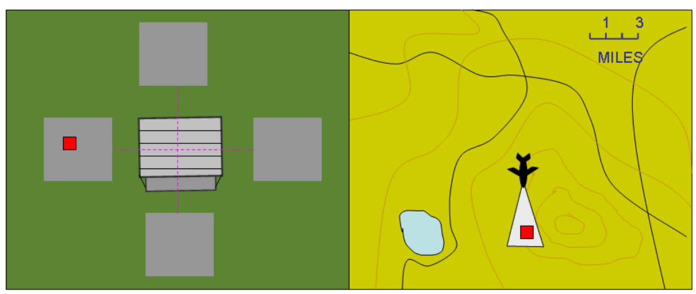
APPENDIX E:

SPATIAL ORIENTATION TEST

The Spatial Orientation Test, modeled after the cardinal direction test developed by Gugerty and his colleagues (Gugerty & Brooks, 2004), is a computerized test consisting of a brief training segment and 32 test questions. The program automatically captures both accuracy and response time. Participants are shown the following image:



The right side image is a map showing a plane flying. The left side of the display is the pilot's view (from the cockpit of the plane) of several parking lots surrounding a building. The participants' task is to use the right side of the display to learn in which direction the plane is flying. They then use this information to identify which parking lot (north, south, east, or west) in the left side image has the dot. In the example shown above, the plane is heading north, and so the dot appears in the north parking lot. In the example shown below, the plane is heading south, and so the dot appears in the east parking lot.



Participants are shown 32 of these images in succession; each time the direction the plane is flying and the location of the dot are randomized. Participants answer by clicking on one of four buttons (North, South, East, or West). This test is self-paced; the participant may take as long as they wish to answer, and when they answer one question the next question automatically appears. No questions can be skipped, and the order of images is randomized among participants.

APPENDIX F:

COMPLACENCY POTENTIAL RATING SCALE

Complacency Potential Rating Scale

Participant # __________ 2

Read each statement carefully and circle the one response that you feel most accurately describes your views and experiences. THERE ARE NO RIGHT OR WRONG ANSWERS. Please answer honestly and do not skip any questions.

SA	A	U	D	SD
Strongly agree	Agree	Undecided	Disagree	Strongly disagree

1. Manually sorting through emails is more reliable than computer-aided searches for finding emails in my inbox.	SA A U D SD
2. If I need to have a tumor in my body removed, I would choose to undergo computer-aided surgery using laser technology because computerized surgery is more reliable and safer than manual surgery.	SA A U D SD
3. People save time by using automatic teller machines (ATMs) rather than a bank teller in making transactions.	SA A U D SD
4. I do not trust automated devices such as ATMs and computerized pay stations for parking lots.	SA A U D SD
5. People who work frequently with automated devices have lower job satisfaction because they feel less involved in their job than those who work manually.	SA A U D SD
6. I feel safer depositing my money at an ATM than with a human teller.	SA A U D SD
7. I have to pay an important bill. To ensure that the bill is paid with the correct amount and on time, I would use the automatic bill pay facility on my online banking rather than pay the bill manually.	SA A U D SD
8. People whose jobs require them to work with automated systems are lonelier than people who do not work with such devices.	SA A U D SD
9. Automated systems used in modern aircraft, such as the automatic landing system, have made air journey safer.	SA A U D SD
10. ATMs provide safeguard against the inappropriate use of an individual's bank account by dishonest people.	SA A U D SD
11. Automated devices used in aviation and banking have made work easier for both employees and customers.	SA A U D SD
12. I often use automated devices.	SA A U D SD
13. People who work with automated devices have greater job satisfaction because they feel more involved than those who work manually.	SA A U D SD
14. Automated devices in medicine save time and money in the diagnosis and treatment of disease.	SA A U D SD
15. Even though the automatic cruise control in my car is set at a speed below the speed limit, I worry when I pass a police radar speed-trap in case the automatic control is not working properly.	SA A U D SD
16. Bank transactions have become safer with the introduction of computer technology for the direct deposit of checks.	SA A U D SD
17. I would rather purchase an item using a computer than have to deal with a sales representative on the phone because my order is more likely to be correct using the computer.	SA A U D SD
18. Work has become more difficult with the increase of automation in aviation and banking.	SA A U D SD
19. I do not like to use ATMs because I feel that they are sometimes unreliable.	SA A U D SD
20. I think that automated devices used in medicine, such as CAT-scans and ultrasound, provide very reliable medical diagnosis.	SA A U D SD

APPENDIX G:

NASA-TLX WORKLOAD ASSESSMENT

NASA TLX Questionnaire

NASA TLX Workload Assessment

Instructions: Ratings Scales

We are interested in the "workload" you experienced during this scenario. Workload is something experienced individually by each person. One way to find out about workload is to ask people to describe what they experienced. Workload may be caused by many different factors and we would like you to evaluate them individually. The set of six workload rating factors was developed for you to use in evaluating your experiences during different tasks. Please read them. If you have a question about any of the scales in the table, please ask about it. It is extremely important that they be clear to you.

Definitions

Title	Endpoints	Descriptions
MENTAL DEMAND	Low / High	How much mental and perceptual activity was required (that is, thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low / High	How much physical activity was required (that is, pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low / High	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?
PERFORMANCE	Poor / Good	How successful do you think you were in accomplishing the goals of the task? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low / High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low / High	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

We want you to evaluate workload. Rate the workload on each factor on a scale. Each scale has two end descriptions, and 20 slots (hash marks) between the end descriptions. Place an "x" in the slot (between the hash marks) that you feel most accurately reflects your workload.

After you have finished the entire series, we will be able to use the pattern of your choices to create a weighted combination of ratings into a summary workload score.

We ask you to evaluate your workload for this scenario. This includes all the duties involved in your job (e.g., detecting targets and using display).

Participant ID: ________________

TLX Workload Scale

Please rate your workload by putting a mark on each of the six scales at the point which matches your experience.

Mental Demand

Low — High

Physical Demand

Low — High

Temporal Demand

Low — High

Performance

Good — Poor

Effort

Low — High

Frustration

Low — High

APPENDIX H:

RSPAN TEST

Participants will be administered a computerized version of the RSPAN task (Daneman & Carpenter, 1980; Unsworth, Heitz, Schrock, & Engle, 2005) in order to evaluate their working memory capacity, as well as remove participants with potential reading comprehension issues.

RSPAN Instructions for automated presentation

The RSPAN task is broken down into two sections. First, participants receive practice and second, the participants perform the actual task. The practice sessions are further broken down into three sections.

The first practice is simple letter span. They see letters appear on the screen one at a time, and then must recall these letters in the same order they saw them. In all experimental conditions, letters remain on-screen for 800 ms. Recall consists of filling in boxes with the appropriate letters. Entering a letter or space in a box should advance the cursor to the next box. At the final box, hitting the spacebar will advance to the next slide. After each recall slide, the computer provides feedback about the number of letters correctly recalled.

Next, participants practice the sentence portion of the task. Participants first see a sentence (e.g., "Andy was stopped by the policeman because he crossed the yellow heaven"). Once the participant has read the sentence, they are required to answer YES or NO (did the sentence make sense). After each sentence sense-verification, participants are given feedback. The reading practice serves to familiarize participants with the sentence portion of the task as well as calculate how long it takes a given person to solve the sentence problems. Thus, it attempts to account for individual differences in the time it takes to solve reading problems. After the reading practice, the program calculates the individual's mean time required to solve the problems. This time (plus 2.5 standard deviations) is then used as a time limit for the reading portion of the experimental session.

The final practice session has participants perform both the letter recall and reading portions together, just as they will do in the task block. As with traditional RSPAN, participants first see the sentence and after verifying that it makes sense or not, they see the letter be recalled. If participants take more time to verify the sentence than their average time plus 2.5 SD, then the program automatically moves on. This serves to prevent participants from rehearsing the letters when they should be verifying the sense of the sentences. After the participant completes all of the practice sessions, the program moves them on to the real trials.

The task trials consist of 3 trials of each set-size, with the set sizes ranging from 3 - 6. This makes for a total of 54 letters and 54 sentence problems. Subjects are instructed to keep their reading accuracy at or above 80% at all times. During recall, a percentage in red is presented in the upper right-hand corner. Subjects are instructed to keep a careful watch on the percentage in order to keep it above 80%. Subjects get feedback at the end of each trial. Subjects that do not finish with a reading accuracy score of 80% or better will be excused from continuing with the study.

RSPAN Timing

Sentence verification screen: Min=none, Max=Mean of practice trials +2.5 SD.

Letter presentation: 800 ms.

Recall screen: Min=none, Max= 2 min (Have a next button or something they can click to continue faster)

READY screen: 3 seconds (No keys active, cannot skip this screen)

RSPAN Scoring

Scores should report five values at the conclusion of the task: Absolute score, RSPAN score, total number correct, sentence errors, and speed errors.

The absolute score combines sentence verification with letter recall. In order to be eligible to earn a point, the participant must first correctly answer yes or no, identifying the statement as correct or not. Then the corresponding letter to the statement must be correctly entered in the correct blank for that set. Example: Q1 (F) was incorrectly identified No, Q2 (M), Q3 (P), and Q4 (B) were correctly identified Yes. The participant entered the letters F, M, B, & P. The participant scored 1 point. Although four correct letters were entered, Q1 was answered incorrectly, and Q3 & Q4 letters were entered in the wrong blanks.

RSPAN score is the sum of all perfectly recalled sets. So, for example, if an individual recalled correctly 2 letters in a set size of 2, 3 letters in a set size of 3, and 3 letters in a set size of 4, their RSPAN score would be 5 (2 + 3 + 0).

Total number correct is the total number of letters recalled in the correct position. For example, if an individual recalled correctly 2 letters in a set size of 2, 3 letters in a set size of 3, and 3 letters in a set size of 4, their RSPAN score would be 8 (2 + 3 + 3).

Sentence errors are reported as accuracy errors where the subject verified the sense of the sentence incorrectly.

Speed errors are where the subject ran out of time in attempting to verify a given sentence.

APPENDIX I:

USABILITY AND TRUST SURVEY

Usability and Trust Survey

1. I made use of RoboLeader's recommendations.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

2. I sometimes felt 'lost' using the RoboLeader display.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

3. I do not feel the RoboLeader display was helpful in the task.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

4. I relied heavily on the RoboLeader for the task.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

5. Threats were visible on the screen(s) long enough to accurately detect them.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

6. The RoboLeader display was confusing.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

7. The RoboLeader display was annoying.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

8. The RoboLeader display improved my performance on the task.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

9. The RoboLeader display can be deceptive.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

10. The RoboLeader display sometimes behaves in an unpredictable manner.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

11. I am often suspicious of the RoboLeader system's intent, action, or outputs.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

12. I am sometimes unsure of the RoboLeader system.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

13. The RoboLeader system may have harmful effects on the task.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

14. I am confident in the RoboLeader system.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

15. The RoboLeader system can provide security.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

16. The RoboLeader system has integrity.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

17. The RoboLeader system is dependable.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

18. The RoboLeader system is consistent.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

19. I can trust the RoboLeader system.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

20. I am familiar with the RoboLeader display.

Strongly DISAGREE	1	2	3	4	5	6	7	Strongly AGREE

APPENDIX J:

ROBOLEADER NOTIFICATIONS

RoboLeader communications vary across experiment and level of reasoning. Experiment 1 is the low information study, and will only have one icon present on the map that affects each route decision. Experiment 2 is the high information study and has several icons present on the map that will need to be considered when accepting or rejecting RL's suggestion.

Experiment 1 Mission 1 RoboLeader Notifications:

Experiment 1 ART1:
All Areas:
Change to convoy path recommended

Experiment 1 ART2:
Area 1 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 2 -
Change to convoy path recommended
Activity in area: Gunfire/Sniper
Recommended action: Reroute to avoid Gunfire/Sniper

Area 3 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 4 -
Change to convoy path recommended
Activity in area: Potential IED
Recommended action: Reroute to avoid Potential IED

Area 5 -
Change to convoy path recommended
Activity in area: Comm Dead Zone
Recommended action: Reroute to avoid Comm Dead Zone

Area 6 -
Change to convoy path recommended
Activity in area: Dense Fog
Recommended action: Reroute to avoid Dense Fog

Experiment 1 ART3:
Area 1 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock TOR: 1
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 2 -
Change to convoy path recommended
Activity in area: Gunfire/Sniper TOR: 1
Recommended action: Reroute to avoid Gunfire/Sniper

Area 3 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock TOR: 1
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 4 -
Change to convoy path recommended
Activity in area: Potential IED TOR: 1
Recommended action: Reroute to avoid Potential IED

Area 5 -
Change to convoy path recommended
Activity in area: Comm Dead Zone TOR: 1
Recommended action: Reroute to avoid Comm Dead Zone

Area 6 -
Change to convoy path recommended
Activity in area: Dense Fog TOR: 1
Recommended action: Reroute to avoid Dense Fog

Experiment 2 Mission 1 RoboLeader Notifications

Experiment 2 ART1:
All Areas
Change to convoy path recommended

Experiment 2 ART2:
Area 1 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock (H)
Comm Dead Zone (M)
Dense Fog (L)
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 2 -
Change to convoy path recommended
Activity in area: Gunfire/Sniper (H)
Congested Area/Roadblock (M)
Comm Dead Zone (L)
Recommended action: Reroute to avoid Gunfire/Sniper

Area 3 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock (H)
Potential IED (L)
Comm Dead Zone (M)
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 4 -
Change to convoy path recommended
Activity in area: Potential IED (H)
Gunfire/Sniper (M)
Dense Fog (L)
Recommended action: Reroute to avoid Potential IED

Area 5 -
Change to convoy path recommended
Activity in area: Comm Dead Zone (H)
Potential IED (M)
Dense Fog (L)
Recommended action: Reroute to avoid Comm Dead Zone

Area 6 -
Change to convoy path recommended
Activity in area: Dense Fog (H)
Gunfire/Sniper (M)
Congested Area/Roadblock (L)
Recommended action: Reroute to avoid Dense Fog

Experiment 2 ART3:
Area 1 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock (H) TOR: 1
Comm Dead Zone (M) TOR: 4
Dense Fog (L) TOR: 2
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 2 -
Change to convoy path recommended
Activity in area: Gunfire/Sniper (H) TOR: 1
Congested Area/Roadblock (M) TOR: 2
Comm Dead Zone (L) TOR: 4
Recommended action: Reroute to avoid Gunfire/Sniper

Area 3 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock (H) TOR: 1
Potential IED (L) TOR: 3
Comm Dead Zone (M) TOR: 6
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 4 -
Change to convoy path recommended
Activity in area: Potential IED (H) TOR: 1
Gunfire/Sniper (M) TOR: 4
Dense Fog (L) TOR: 6
Recommended action: Reroute to avoid Potential IED

Area 5 -
Change to convoy path recommended
Activity in area: Comm Dead Zone (H) TOR: 1
Potential IED (M) TOR: 4
Dense Fog (L) TOR: 6
Recommended action: Reroute to avoid Comm Dead Zone

Area 6 -
Change to convoy path recommended
Activity in area: Dense Fog (H) TOR: 1
Gunfire/Sniper (M) TOR: 6
Congested Area/Roadblock (L) TOR: 3
Recommended action: Reroute to avoid Dense Fog

Experiment 1 Mission 2 RoboLeader Notifications:

Experiment 1 ART1:
All Areas:
Change to convoy path recommended

Experiment 1 ART2:

Area 1 -
Change to convoy path recommended
Activity in area: Dense Fog
Recommended action: Reroute to avoid Dense Fog

Area 2 -
Change to convoy path recommended
Activity in area: Comm Dead Zone
Recommended action: Reroute to avoid Comm Dead Zone

Area 3 -
Change to convoy path recommended
Activity in area: Potential IED
Recommended action: Reroute to avoid Explosion

Area 4 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 5 -
Change to convoy path recommended
Activity in area: Gunfire/Sniper
Recommended action: Reroute to avoid Gunfire/Sniper

Area 6 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock
Recommended action: Reroute to avoid Potential IED

Experiment 1 ART3:
Area 1 -
Change to convoy path recommended
Activity in area: Dense Fog TOR: 1
Recommended action: Reroute to avoid Dense Fog

Area 2 -
Change to convoy path recommended
Activity in area: Comm Dead Zone TOR: 1
Recommended action: Reroute to avoid Comm Dead Zone

Area 3 -
Change to convoy path recommended
Activity in area: Potential IED TOR: 1
Recommended action: Reroute to avoid Explosion

Area 4 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock TOR: 1
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 5 -
Change to convoy path recommended
Activity in area: Gunfire/Sniper TOR: 2
Recommended action: Reroute to avoid Gunfire/Sniper

Area 6 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock TOR: 1
Recommended action: Reroute to avoid Potential IED

Experiment 2 Mission 2 RoboLeader Notifications:

Experiment 2 ART1:
All Areas
Change to convoy path recommended

Experiment 2 ART2:
Area 1 -
Change to convoy path recommended
Activity in area: Dense Fog (H)
Gunfire/Sniper (M)
Congested Area/Roadblock (L)
Recommended action: Reroute to avoid Dense Fog

Area 2 -
Change to convoy path recommended
Activity in area: Comm Dead Zone (H)
Dense Fog (M)
Potential IED (L)
Recommended action: Reroute to avoid Comm Dead Zone

Area 3 -
Change to convoy path recommended
Activity in area: Potential IED (H)
Gunfire/Sniper (M)
Congested Area/Roadblock (M)
Recommended action: Reroute to avoid Explosion

Area 4 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock (H)
Comm Dead Zone (M)
Dense Fog (L)
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 5 -
Change to convoy path recommended
Activity in area: Gunfire/Sniper (H)
Comm Dead Zone (M)
Congested Area/Roadblock (L)
Recommended action: Reroute to avoid Gunfire/Sniper

Area 6 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock (H)
Potential IED (L)
Comm Dead Zone (M)
Recommended action: Reroute to avoid Potential IED

Experiment 2 ART3:
Area 1 -
Change to convoy path recommended
Activity in area: Dense Fog (H) TOR: 1
Gunfire/Sniper (M) TOR: 4
Congested Area/Roadblock (L) TOR: 2
Recommended action: Reroute to avoid Dense Fog

Area 2 -
Change to convoy path recommended
Activity in area: Comm Dead Zone (H) TOR: 1
Dense Fog (M) TOR: 2
Potential IED (L) TOR: 4
Recommended action: Reroute to avoid Comm Dead Zone

Area 3 -
Change to convoy path recommended
Activity in area: Potential IED (H) TOR: 1
Gunfire/Sniper (M) TOR: 3
Congested Area/Roadblock (M) TOR: 2
Recommended action: Reroute to avoid Explosion

Area 4 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock (H) TOR: 1
Comm Dead Zone (M) TOR: 2
Dense Fog (L) TOR: 4
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 5 -
Change to convoy path recommended
Activity in area: Gunfire/Sniper (H) TOR: 2
Comm Dead Zone (M) TOR: 1
Congested Area/Roadblock (L) TOR: 6
Recommended action: Reroute to avoid Gunfire/Sniper

Area 6 -
Change to convoy path recommended

Activity in area:	Congested Area/Roadblock (H) TOR: 1
	Potential IED (L) TOR: 6
	Comm Dead Zone (M) TOR: 3

Recommended action: Reroute to avoid Potential IED

Experiment 1 Mission 3 RoboLeader Notifications:

Experiment 1 ART1:
All Areas:
Change to convoy path recommended

Experiment 1 ART2:
Area 1 -
Change to convoy path recommended
Activity in area: Potential IED
Recommended action: Reroute to avoid Potential IED

Area 2 -
Change to convoy path recommended
Activity in area: Dense Fog
Recommended action: Reroute to avoid Dense Fog

Area 3 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 4 -
Change to convoy path recommended
Activity in area: Comm Dead Zone
Recommended action: Reroute to avoid Comm Dead Zone

Area 5 -
Change to convoy path recommended
Activity in area: Gunfire/Sniper
Recommended action: Reroute to avoid Gunfire/Sniper

Area 6 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock
Recommended action: Reroute to avoid Congested Area/Roadblock

Experiment 1 ART3:
Area 1 -
Change to convoy path recommended
Activity in area: Potential IED TOR: 1
Recommended action: Reroute to avoid Potential IED

Area 2 -
Change to convoy path recommended
Activity in area: Dense Fog TOR: 2
Recommended action: Reroute to avoid Dense Fog

Area 3 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock TOR: 1
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 4 -
Change to convoy path recommended
Activity in area: Comm Dead Zone TOR: 3
Recommended action: Reroute to avoid Comm Dead Zone

Area 5 -
Change to convoy path recommended
Activity in area: Gunfire/Sniper TOR: 1
Recommended action: Reroute to avoid Gunfire/Sniper

Area 6 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock TOR: 1
Recommended action: Reroute to avoid Congested Area/Roadblock

Experiment 2 Mission 3 RoboLeader Notifications:

Experiment 2 ART1:
All Areas
Change to convoy path recommended

Experiment 2 ART2:
Area 1 -
Change to convoy path recommended
Activity in area: Potential IED (H)
Congested Area/Roadblock (M)
Gunfire/Sniper (L)
Recommended action: Reroute to avoid Potential IED

Area 2 -
Change to convoy path recommended

Activity in area:	Dense Fog (H)
	Comm Dead Zone (M)
	Potential IED (L)

Recommended action: Reroute to avoid Dense Fog

Area 3 -
Change to convoy path recommended

Activity in area:	Congested Area/Roadblock (H)
	Gunfire/Sniper (L)
	Dense Fog (M)

Recommended action: Reroute to avoid Congested Area/Roadblock

Area 4 -
Change to convoy path recommended

Activity in area:	Comm Dead Zone (H)
	Gunfire/Sniper (M)
	Dense Fog (L)

Recommended action: Reroute to avoid Comm Dead Zone

Area 5 -
Change to convoy path recommended

Activity in area:	Gunfire/Sniper (H)
	Comm Dead Zone (M)
	Congested Area/Roadblock (L)

Recommended action: Reroute to avoid Gunfire/Sniper

Area 6 -
Change to convoy path recommended

Activity in area:	Congested Area/Roadblock (H)
	Potential IED (M)
	Comm Dead Zone (L)

Recommended action: Reroute to avoid Congested Area/Roadblock

Experiment 2 ART3:
Area 1 -
Change to convoy path recommended
Activity in area: Potential IED (H) TOR: 1
Congested Area/Roadblock (M) TOR: 4
Gunfire/Sniper (L) TOR: 2
Recommended action: Reroute to avoid Potential IED

Area 2 -
Change to convoy path recommended
Activity in area: Dense Fog (H) TOR: 2
Comm Dead Zone (M) TOR: 1
Potential IED (L) TOR: 3
Recommended action: Reroute to avoid Dense Fog

Area 3 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock (H) TOR: 1
Gunfire/Sniper (L) TOR: 3
Dense Fog (M) TOR: 4
Recommended action: Reroute to avoid Congested Area/Roadblock

Area 4 -
Change to convoy path recommended
Activity in area: Comm Dead Zone (H) TOR: 3
Gunfire/Sniper (M) TOR: 1
Dense Fog (L) TOR: 2
Recommended action: Reroute to avoid Comm Dead Zone

Area 5 -
Change to convoy path recommended
Activity in area: Gunfire/Sniper (H) TOR: 1
Comm Dead Zone (M) TOR: 2
Congested Area/Roadblock (L) TOR: 4
Recommended action: Reroute to avoid Gunfire/Sniper

Area 6 -
Change to convoy path recommended
Activity in area: Congested Area/Roadblock (H) TOR: 1
Potential IED (M) TOR: 2
Comm Dead Zone (L) TOR: 3
Recommended action: Reroute to avoid Congested Area/Roadblock

APPENDIX K:

SITUATION AWARENESS PROBES

Situation awareness probes were common across all ART levels and both experiments.

Level 1 – What is happening?

SA1 queries gauged how well the participant perceived information about the experimental environment.

Mission 1

1. How many Dump trucks have you passed?
   Answer: B. 2
   A. 1
   B. 2
   C. 3
   D. 4
   E. None

2. What vehicle was positioned between the two walls?
   Answer: E. Tank
   A. Personnel Carrier
   B. Pickup Truck
   C. Fuel Truck
   D. Dump Truck
   E. Tank

3. What vehicle/object of interest did you just pass?
   Answer: B. Garbage Truck
   A. Personnel Carrier
   B. Garbage Truck
   C. Fuel Truck
   D. Dump Truck
   E. Propane Tank

4. You have just passed a person standing behind the wall. Identify them.
   Answer: A. Male Civilian
   A. Male Civilian
   B. Female Civilian
   C. US Military
   D. Armed Civilian
   E. None

5. Who was standing next to the Dump truck you just passed?
   Answer: D. 1 Male & 1 Female Civilian
   A. 1 Male Civilian
   B. 1 Female Civilian
   C. 2 Male Civilians
   D. 1 Male & 1 Female Civilian
   E. None

6. What object/vehicle of interest was next to the Garbage Truck you just passed?
   Answer: C. Fuel Truck
   A. Personnel Carrier
   B. Garbage Truck
   C. Fuel Truck
   D. Dump Truck
   E. Propane Tank

Mission 2

1. Who was standing next to the Dump truck you just passed?
   Answer: C. 2 Male Civilians
   A. 1 Male Civilian
   B. 1 Female Civilian
   C. 2 Male Civilians
   D. 1 Male & 1 Female Civilian
   E. None

2. How many U.S. Military were standing by the Garbage truck?
   Answer: C. 3
   A. 1
   B. 2
   C. 3
   D. 4
   E. None

3. What vehicle/object of interest did you just pass?
   Answer: C. Fuel Truck
   A. Personnel Carrier
   B. Garbage Truck
   C. Fuel Truck
   D. Dump Truck
   E. Propane Tank

4. How many destroyed vehicles were near the Dump truck?
   Answer: A. 1
   A. 1
   B. 2
   C. 3
   D. 4
   E. None

5. What vehicle/object of interest was near the Propane Tank that you just passed?
   Answer: C. Fuel Truck
   A. Personnel Carrier
   B. Garbage Truck
   C. Fuel Truck
   D. Dump Truck
   E. Propane Tank

6. What was behind the wall that you just passed?
   Answer: B. Propane Tank
   A. Pickup Truck
   B. Propane Tank
   C. Fuel Truck
   D. Tank
   E. Dump Truck

Mission 3

1. How many Propane Tanks have you passed?
   Answer: B. 2
   A. 1
   B. 2
   C. 3
   D. 4
   E. None

2. Who was standing next to the Dump truck you just passed?
   Answer: D. 3 Male Civilians
   A. 1 Male Civilian
   B. 1 Female Civilian
   C. 2 Male Civilians
   D. 3 Male Civilians
   E. None

3. Since your last route selection, how many Dump Trucks have you passed?
   Answer: B. 2
   A. 1
   B. 2
   C. 3
   D. 4
   E. None

4. How many U.S. Military were standing by the Personnel Carrier?
   Answer: D. 4
   A. 1
   B. 2
   C. 3
   D. 4
   E. None

5. What was behind the wall that you just passed?
   Answer: D. Dump Truck
   A. Personnel Carrier
   B. Garbage Truck
   C. Fuel Truck
   D. Dump Truck
   E. Propane Tank

6. Who was standing next to the Personnel Carrier you just passed?
   Answer: C. 2 Male Civilians
   A. 1 Male Civilian
   B. 1 Female Civilian
   C. 2 Male Civilians
   D. 2 Female Civilians
   E. None

Level 2 – Why is it happening?

SA2 Queries evaluated how well the participant was integrating information from multiple sources in their decision-making. These questions appeared shortly after the SA3 queries. Each mission contained 6 SA2 queries.

Mission 1

Area 1
Both Paths:
Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Potential IED
E. Route Clear

Area 2
Both Paths:
Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Gunfire/Sniper
D. Avoid Potential IED
E. Main path Clear (or Proposed path clear)

Area 3
Both Paths:
Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Potential IED
E. Main path Clear (or Proposed path clear)

Area 4
Both Paths:
Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Gunfire/Sniper
C. Avoid Dense Fog
D. Avoid Potential IED
E. Main path Clear (or Proposed path clear)

Area 5

Both Paths:

Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Potential IED
E. Main path Clear (or Proposed path clear)

Area 6

Both Paths:

Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Gunfire/Sniper
C. Avoid Dense Fog
D. Avoid Potential IED
E. Main path Clear (or Proposed path clear)

Mission 2

Area 1

Both Paths:

Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Gunfire/Sniper
E. Main Path Clear (or Proposed Path Clear)

Area 2

Both Paths:

Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Potential IED
E. Main Path Clear (or Proposed Path Clear)

Area 3

Both Paths:

Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Potential IED

D. Avoid Gunfire/Sniper
E. Main Path Clear (or Proposed Path Clear)

Area 4
Both Paths:
Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Gunfire/Sniper
E. Main Path Clear (or Proposed Path Clear)

Area 5
Both Paths:
Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Gunfire/Sniper
E. Main Path Clear (or Proposed Path Clear)

Area 6
Both Paths:
Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Potential IED
E. Main Path Clear (or Proposed Path Clear)

Mission 3

Area 1
Both Paths:
Bravo unit - Why are you on your current route? (Select all that apply)

F. Avoid Congested Area/Roadblock
G. Avoid Potential IED
H. Avoid Dense Fog
I. Avoid Gunfire/Sniper
J. Main Path Clear (or Proposed Path Clear)

Area 2
Both Paths:

Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Potential IED
E. Main Path Clear (or Proposed Path Clear)

Area 3

Both Paths:

Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Gunfire/Sniper
E. Main Path Clear (or Proposed Path Clear)

Area 4

Both Paths:

Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Gunfire/Sniper
E. Main Path Clear (or Proposed Path Clear)

Area 5

Both Paths:

Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Gunfire/Sniper
E. Main Path Clear (or Proposed Path Clear)

Area 6

Both Paths:

Bravo unit - Why are you on your current route? (Select all that apply)

A. Avoid Congested Area/Roadblock
B. Avoid Comm Dead Zone
C. Avoid Dense Fog
D. Avoid Potential IED
E. Main Path Clear (or Proposed Path Clear)

Level 3 – What will happen?

SA3 Queries evaluated how well the participant could predict the consequences of their chosen action. This question was asked immediately after passing every decision point, regardless of route selection. There were 6 SA3 queries in each mission.

Bravo unit -

Please evaluate how safe your current route will be.

A – Completely Safe	C – Relatively Unsafe
B – Relatively Safe	D – Completely Unsafe

APPENDIX L:

COMMAND COMMUNICATIONS

Messages from command were the same in both experiments and across ARTs, but unique for each Mission. Messages were both informative notifications (i.e., All Units), and 'chatter' that was included to create noise and keep the rate of incoming messages near one message (from all sources) every 30 seconds. The following lists are in order of appearance.

Mission 1 Script/Messages

Area 1 MP (Main Path)

1. Echo Unit: Report Status
2. All Units: Accident Reported in Sector D9
3. SA1 Prompt 1
4. **RB prompt**
5. SA3 Prompt 1
6. Alpha Unit Report Status
7. SA2 prompt 1

Area 1 PP (Proposed Path)

8. SA3 Prompt
9. Alpha Unit: Report status
10. Victor Unit: Rally at Checkpoint
11. All Units: Gunman subdued, All Clear Sector
12. All Units: Dense Fog Reported - Sector C7
13. Charlie Unit: Return to Base
14. SA 2 prompt

Area 2 MP

15. SA1 prompt
16. All Units: Gunfire reported Sector C9
17. **RB prompt**
18. SA3 prompt
19. All Units: Communications Down Sector D5
20. Echo Unit: Report Status
21. All Units: Sector C7 – Fog Cleared
22. Alpha Unit: Report Status
23. SA2 prompt

Area 2 PP

24. SA3 prompt
25. All Units: Communications Down Sector D5
26. SA2 prompt

Area 3 MP

27. SA1 prompt
28. All Units: Crowd dispersed All Clear Sector C9
29. RB prompt
30. SA3 Prompt
31. All Units: Suspicious Activity: Large Congregation Sector B4
32. SA2 Prompt

Area 3 PP

33. SA3 Prompt
34. All Units: Suspicious Activity: Large Congregation Sector B4
35. Charlie Unit: Recon Sector B4 and Report
36. Lima Unit: Rendezvous Sierra Unit Sector E11
37. All Units: All Clear Sector E9
38. SA2 Prompt

Area 4 MP

39. SA1 Prompt
40. Echo Unit: Report Status
41. All Units: Explosion reported Sector B7
42. SA1 Prompt
43. RB Prompt
44. SA3 Prompt
45. All Units: Communications restored Sector B6
46. SA2 Prompt

Area 4 PP

47. SA3 Prompt
48. Alpha Unit: Report Status
49. Delta Unit: Recon Sector D4 and Report
50. Zulu Unit: Report Status
51. Echo Unit: Return to Base
52. All Units: Communications restored Sector B6
53. SA2 Prompt

Area 5 MP

54. RB Prompt
55. SA3 Prompt
56. All Units, Dense Fog Reported Sector C6
57. SA2 Prompt

Area 5 PP

58. SA3 Prompt
59. Sierra Unit: Report to Rally Point
60. Lima Unit: Report Status

61. All Units, Dense Fog Reported Sector C6
62. SA2 Prompt

Area 6 MP

63. RB Prompt
64. SA3 Prompt
65. All Units: Accident/Road Blocked Sector C9
66. SA1 Prompt
67. All Units: IED Cleared All Clear Sector D7
68. SA2 Prompt

Area 6 PP

69. SA3 Prompt
70. Charlie Unit: Report Status
71. All Units: Accident/Road Blocked Sector C9
72. SA1 Prompt
73. Lima Unit: Return to Base
74. All Units: IED Cleared All Clear Sector D7
75. SA2 Prompt

Mission 2 Script/Messages

Area 1 MP

1. Alpha Unit: Report status
2. Victor Unit: Rally at Checkpoint
3. All Units: Dense Fog Reported Sector D2
4. SA1 Prompt
5. **RB Prompt**
6. SA3 Prompt
7. All Units: Communications Down Sector E5
8. Charlie Unit: Return to Base
9. Echo Unit: Report Status
10. Victor Unit: Rally at Checkpoint
11. All Units: Road Clear Sector C6
12. SA2 prompt

Area 1 PP

13. SA3 Prompt
14. All Units: Communications Down Sector E5
15. Charlie Unit: Return to Base
16. Echo Unit: Report Status
17. Victor Unit: Rally at Checkpoint
18. All Units: Road Clear Sector C6

19. SA 2 prompt

Area 2 MP

20. Echo Unit: Report Status
21. All Units: Fog Clear Sector C7
22. SA1 prompt
23. Alpha Unit: Report Status
24. All Units: Gunfire reported Sector B4
25. Delta Unit: Return to Base
26. **RB prompt**
27. SA3 prompt
28. Zulu Unit: Proceed to Rally Point
29. SA2 prompt

Area 2 PP

30. SA3 prompt
31. Zulu Unit: Proceed to Rally Point
32. Charlie Unit: Recon Sector B4 and Report
33. Lima Unit: Rendezvous Sierra Unit Sector E11
34. All Units: All Clear Sector E9
35. SA2 prompt

Area 3 MP

36. Delta Unit: Recon Sector D4 and Report
37. Zulu Unit: Report Status
38. SA1 prompt
39. Echo Unit: Return to Base
40. All Units: Explosion Reported Sector C5
41. Sierra Unit: Report Status
42. **RB prompt**
43. SA3 Prompt
44. All Units: Suspicious Activity: Large Congregation Sector D4
45. Echo Unit: Report Status
46. All Units: Sector C7 – Fog Cleared
47. Alpha Unit: Report Status
48. Victor Unit: Rally at Checkpoint
49. All Units: Accident/Road Blocked Sector E6
50. Tango Unit: Recon Sector D6 and Report
51. SA2 Prompt
52. **RB Prompt (for Area 4)**

Area 3 PP

53. SA3 Prompt

54. All Units: Suspicious Activity: Large Congregation Sector D4
55. Echo Unit: Report Status
56. All Units: Sector C7 – Fog Cleared
57. Alpha Unit: Report Status
58. Victor Unit: Rally at Checkpoint
59. All Units: Accident/Road Blocked Sector E6
60. Tango Unit: Recon Sector D6 and Report
61. SA2 Prompt
62. **RB Prompt (for Area 4)**

Area 4 MP

63. SA3 prompt
64. Sierra Unit: Report to Rally Point
65. All Units: All Clear Sector B11
66. All Units: High Wind Reported Sector D3
67. Victor Unit: Report Status
68. Lima Unit: Return to Base
69. All Units: Communications restored Sector B6
70. SA2 Prompt

Area 4 PP

71. SA3 Prompt
72. Sierra Unit: Report to Rally Point
73. Victor Unit: Report Status
74. All Units: Communications restored Sector B6
75. SA2 Prompt

Area 5 MP

76. Echo Unit: Report Status
77. SA1 Prompt
78. All Units: Gunfire Reported Sector E3
79. Lima Unit: Rendezvous Sierra Unit Sector C2
80. Tango Unit: Report Status
81. **RB Prompt**
82. SA3 Prompt
83. Sierra Unit: Report Status
84. Charlie Unit: Report Status
85. All Units: Dense Fog Sector B6
86. SA2 Prompt

Area 5 PP

87. SA3 Prompt
88. Sierra Unit: Report Status
89. Charlie Unit: Report Status

90. All Units: Dense Fog Sector B6
91. SA2 Prompt

Area 6 MP
92. SA1 Prompt
93. All Units: Congestion Cleared Sector D4
94. RB Prompt
95. SA3 Prompt
96. All Units: All Clear Sector D7
97. Charlie Unit: Return to Base
98. Lima Unit: Report to Rally Point
99. SA2 Prompt

Area 6 PP
100. SA3 Prompt
101. All Units: All Clear Sector D7
102. Charlie Unit: Return to Base
103. Lima Unit: Report to Rally Point
104. SA2 Prompt

Mission 3 Script/Messages

Area 1 MP
1. Alpha Unit: Report status
2. All Units: Explosion Reported in Sector C10
3. RB prompt
4. SA3 Prompt
5. Alpha Unit Report Status
6. SA2 prompt

Area 1 PP
7. SA3 Prompt
8. Alpha Unit: Report status
9. All Units: Dense Fog Reported Sector C7
10. Victor Unit: Rally at Checkpoint
11. All Units: Road Clear Sector C6
12. Charlie Unit: Return to Base
13. SA 2 prompt

Area 2 MP
14. SA1 prompt
15. All Units: Wind Calm All Clear Sector D9
16. RB prompt

17. SA3 prompt
18. Lima Unit: Rendezvous Sierra Unit Sector E11
19. Echo Unit: Report Status
20. All Units: Fog Cleared All Clear Sector C7
21. Alpha Unit: Report Status
22. SA2 prompt

Area 2 PP

23. SA3 prompt
24. Lima Unit: Rendezvous Sierra Unit Sector E11
25. Echo Unit: Report Status
26. All Units: Fog Cleared All Clear Sector C7
27. Alpha Unit: Report Status
28. SA2 prompt

Area 3 MP

29. SA1 prompt
30. Charlie Unit: Return to Base
31. Lima Unit: Report to Rally Point
32. All Units: Accident Reported in Sector D9
33. SA1 Prompt
34. **RB prompt (Area 3)**
35. SA3 Prompt
36. Delta Unit: Recon Sector D4 and Report
37. SA1 Prompt
38. All Units: Communications restored Sector C6
39. SA2 Prompt
40. **RB prompt (Area 4)**

Area 3 PP

41. SA3 Prompt
42. Delta Unit: Recon Sector D4 and Report
43. SA1 Prompt
44. All Units: Communications restored Sector C6
45. SA2 Prompt
46. **RB prompt (Area 4)**

Area 4 MP

47. SA3 Prompt
48. Sierra Unit: Report to Rally Point
49. Lima Unit: Report Status
50. All Units: Gunfire Reported Sector D5
51. SA2 Prompt
52. **RB prompt (Area 5)**

Area 4 PP

53. SA3 Prompt
54. All Units: IED Cleared All Clear Sector D7
55. Tango Unit: Recon Sector D6 and Report
56. Sierra Unit: Report to Rally Point
57. Lima Unit: Report Status
58. All Units: Gunfire Reported Sector D5
59. SA2 Prompt
60. **RB prompt (Area 5)**

Area 5 MP

61. SA3 Prompt
62. All Units, Dense Fog Reported Sector C6
63. SA2 Prompt

Area 5 PP

64. SA3 Prompt
65. Delta Unit: Recon Sector D4 and Report
66. Zulu Unit: Report Status
67. Echo Unit: Return to Base
68. All Units, Dense Fog Reported Sector C6
69. SA2 Prompt

Area 6 MP

70. Zulu Unit: Proceed to Rally Point
71. SA1 Prompt
72. Lima Unit: Rendezvous Sierra Unit Sector E11
73. All Units: Accident Reported in Sector E6
74. **RB Prompt**
75. SA3 Prompt
76. Charlie Unit: Recon Sector B4 and Report
77. All Units: IED Cleared All Clear Sector E10
78. SA2 Prompt

Area 6 PP

79. SA3 Prompt
80. Charlie Unit: Recon Sector B4 and Report
81. All Units: Accident/Road Blocked Sector C9
82. Lima Unit: Return to Base
83. All Units: IED Cleared All Clear Sector E10
84. SA2 Prompt
85. Echo Unit: Report Status
86. SA1 prompt
87. All Units: Road Cleared Sector C9

APPENDIX M:

UCF IRB APPROVAL OF HUMAN RESEARCH

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Daniel J. Barber**

Date: **February 20, 2015**

Dear Researcher:

On 2/20/2015, the IRB approved the following human participant research until 02/19/2016 inclusive:

Type of Review:	UCF Initial Review Submission Form Expedited Review Category #4 and 7
Project Title:	Transparency of automation reasoning and its effect on automation-induced complacency
Investigator:	Daniel J Barber
IRB Number:	SBE-15-10979
Funding Agency:	Army Research Laboratory(ARL)
Grant Title:	Human System Interfaces for Emerging Simulation Applications
Research ID:	1056451

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 02/19/2016, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a signed and dated copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 02/20/2015 10:26:18 AM EST

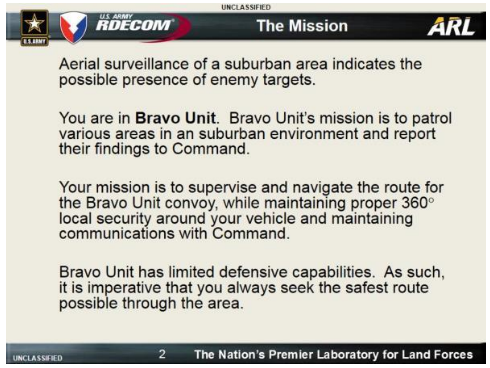
IRB manager

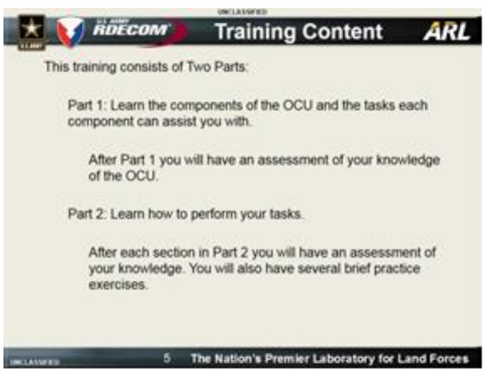
APPENDIX N:

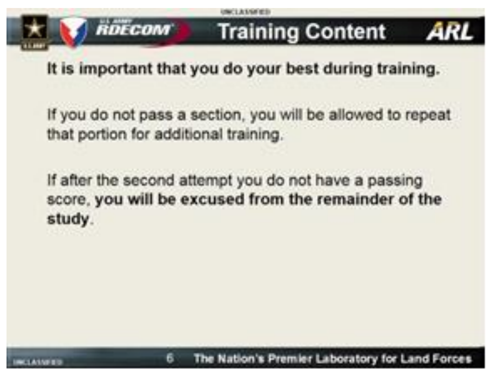
TRAINING MATERIALS

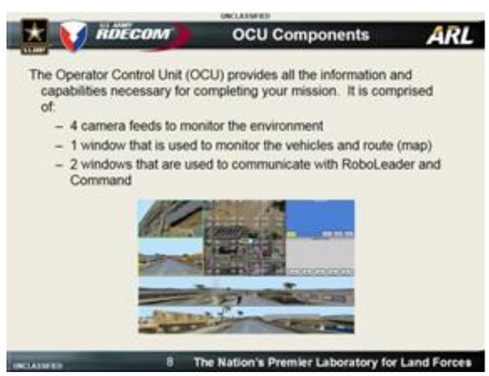
Experiment 1 Training Slides

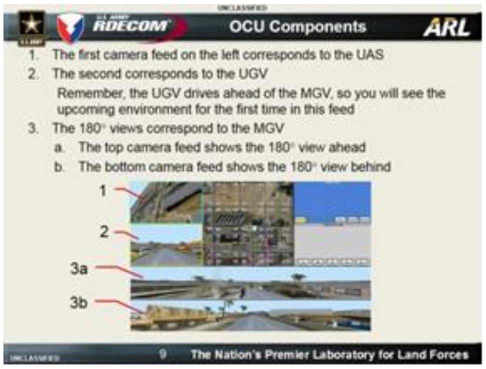
Slides are common across ARTs unless otherwise noted.

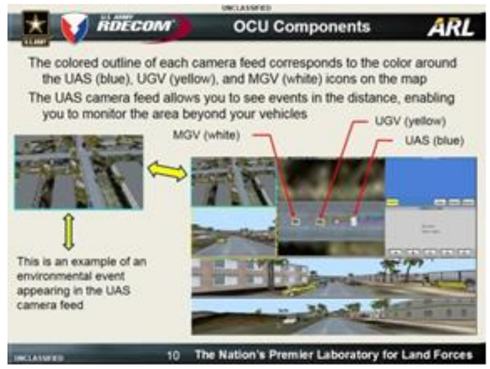




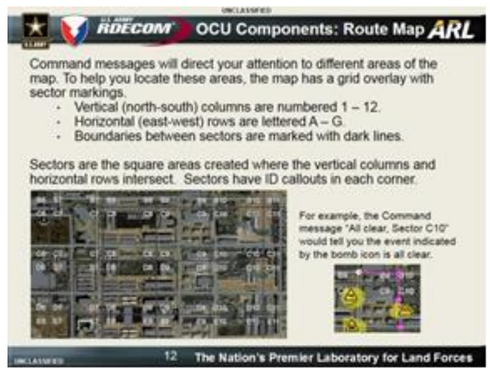


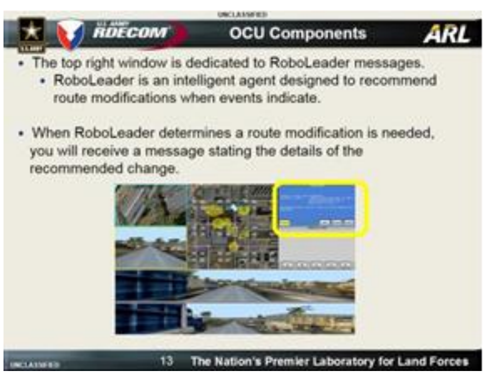


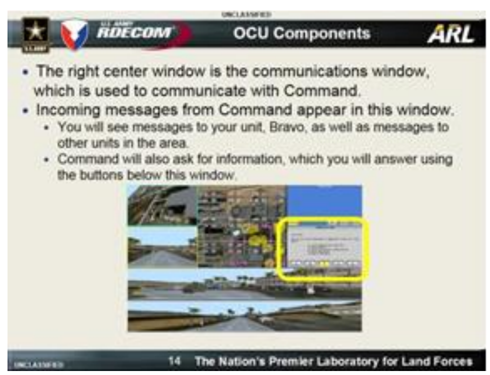

UNCLASSIFIED
U.S. ARMY RDECOM
OCU Components
ARL
1. The first camera feed on the left corresponds to the UAS
2. The second corresponds to the UGV
Remember, the UGV drives ahead of the MGV, so you will see the upcoming environment for the first time in this feed
3. The 180° views correspond to the MGV
a. The top camera feed shows the 180° view ahead
b. The bottom camera feed shows the 180° view behind
1
2
3a
3b
UNCLASSIFIED
9
The Nation's Premier Laboratory for Land Forces

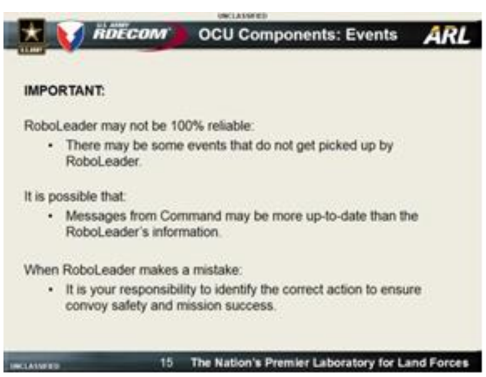

UNCLASSIFIED
U.S. ARMY RDECOM
OCU Components
ARL
The colored outline of each camera feed corresponds to the color around the UAS (blue), UGV (yellow), and MGV (white) icons on the map
The UAS camera feed allows you to see events in the distance, enabling you to monitor the area beyond your vehicles
UGV (yellow)
MGV (white)
UAS (blue)
This is an example of an environmental event appearing in the UAS camera feed
UNCLASSIFIED
10
The Nation's Premier Laboratory for Land Forces

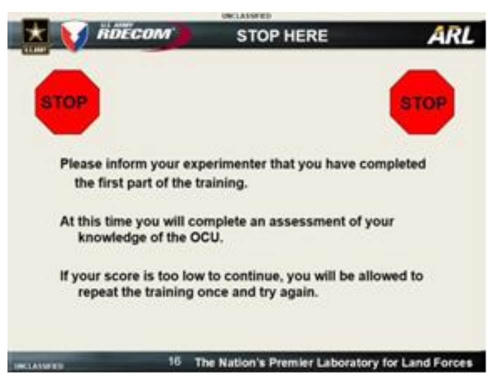

UNCLASSIFIED
U.S. ARMY RDECOM
OCU Components: Route Map
ARL
The upper center window is the route map
The route map is an aerial view of the operational area
• The map allows you to maintain awareness of where your vehicles are along the route, and allows you to plan new routes when necessary.
UNCLASSIFIED
11
The Nation's Premier Laboratory for Land Forces

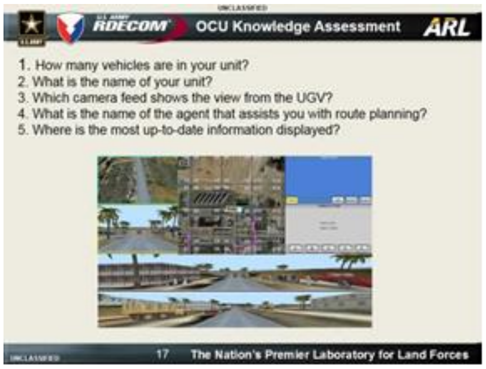

UNCLASSIFIED
U.S. ARMY RDECOM
OCU Components: Route Map
ARL
Command messages will direct your attention to different areas of the map. To help you locate these areas, the map has a grid overlay with sector markings.
• Vertical (north-south) columns are numbered 1 – 12.
• Horizontal (east-west) rows are lettered A – G.
• Boundaries between sectors are marked with dark lines.
Sectors are the square areas created where the vertical columns and horizontal rows intersect. Sectors have ID callouts in each corner.
For example, the Command message "All clear, Sector C10" would tell you the event indicated by the bomb icon is all clear.
UNCLASSIFIED
12
The Nation's Premier Laboratory for Land Forces

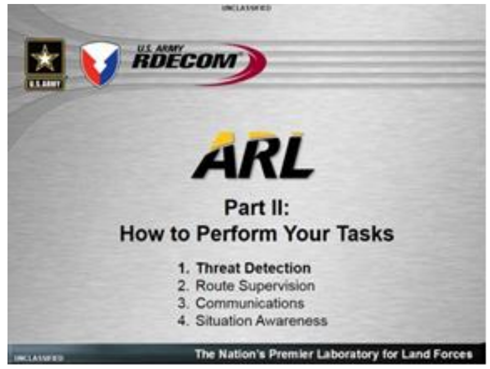

UNCLASSIFIED
U.S. ARMY RDECOM
OCU Components
ARL
• The top right window is dedicated to RoboLeader messages.
• RoboLeader is an intelligent agent designed to recommend route modifications when events indicate.
• When RoboLeader determines a route modification is needed, you will receive a message stating the details of the recommended change.
UNCLASSIFIED
13
The Nation's Premier Laboratory for Land Forces

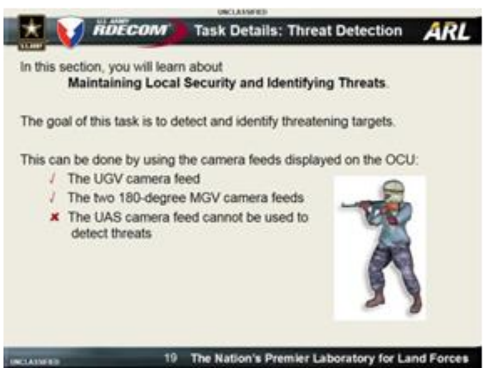

UNCLASSIFIED
U.S. ARMY RDECOM
OCU Components
ARL
• The right center window is the communications window, which is used to communicate with Command.
• Incoming messages from Command appear in this window.
• You will see messages to your unit, Bravo, as well as messages to other units in the area.
• Command will also ask for information, which you will answer using the buttons below this window.
UNCLASSIFIED
14
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
OCU Components: Events
ARL
IMPORTANT:
RoboLeader may not be 100% reliable:
• There may be some events that do not get picked up by RoboLeader.
It is possible that:
• Messages from Command may be more up-to-date than the RoboLeader's information.
When RoboLeader makes a mistake:
• It is your responsibility to identify the correct action to ensure convoy safety and mission success.
UNCLASSIFIED
15
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed the first part of the training.
At this time you will complete an assessment of your knowledge of the OCU.
If your score is too low to continue, you will be allowed to repeat the training once and try again.
UNCLASSIFIED
16
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
OCU Knowledge Assessment
ARL
1. How many vehicles are in your unit?
2. What is the name of your unit?
3. Which camera feed shows the view from the UGV?
4. What is the name of the agent that assists you with route planning?
5. Where is the most up-to-date information displayed?
UNCLASSIFIED
17
The Nation's Premier Laboratory for Land Forces

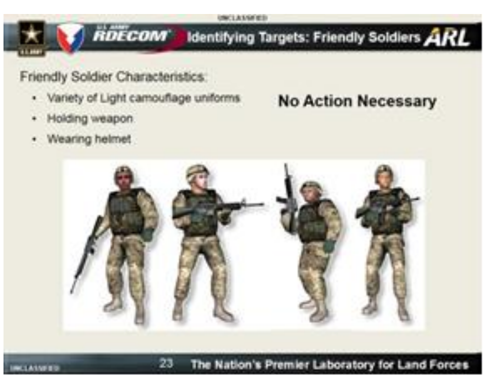

UNCLASSIFIED
U.S. ARMY RDECOM
ARL
Part II:
How to Perform Your Tasks
1. Threat Detection
2. Route Supervision
3. Communications
4. Situation Awareness
UNCLASSIFIED
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
Task Details: Threat Detection
ARL
In this section, you will learn about
Maintaining Local Security and Identifying Threats.
The goal of this task is to detect and identify threatening targets.
This can be done by using the camera feeds displayed on the OCU:
✓ The UGV camera feed
✓ The two 180-degree MGV camera feeds
✗ The UAS camera feed cannot be used to detect threats
UNCLASSIFIED
19
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
Task Details: Threat Detection
ARL
Your task is to scan the UGV and MGV camera feeds in search of threats.
• When you find a threat, click directly on it.
• It is helpful to aim for the center of the body.
Every time you click inside a camera feed window, you will hear a camera shutter sound.
You are asked to identify every threat you see in the environment.
• Only identify each threat ONE time!
• If you identify a threat using the UGV camera feed, do not click on it again in the MGV camera feeds.
Threat!
Insurgent
UNCLASSIFIED
20
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
Task Details: Threat Detection
ARL
Keep in mind that some threats cannot be seen in the UGV camera feed
– Insurgents may be hiding behind trucks or other objects
Some threats can ONLY be seen in the back 180° MGV camera feed
Make sure to consistently scan the camera feeds!
UNCLASSIFIED
21
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
Task Details: Threat Detection
ARL
You will encounter three types of people:
• Friendly Soldiers
• Friendly Civilians
• Armed Civilians (Insurgents)
You must identify and report all armed civilians.
The following slides will show representations of each type
UNCLASSIFIED
22
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
Identifying Targets: Friendly Soldiers
ARL
Friendly Soldier Characteristics:
• Variety of Light camouflage uniforms
• Holding weapon
• Wearing helmet
No Action Necessary
UNCLASSIFIED
23
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
Identifying Targets: Friendly Soldiers
ARL
No Action Necessary
UNCLASSIFIED
24
The Nation's Premier Laboratory for Land Forces


RDECOM
Identifying Targets: Friendly Soldiers
ARL
No Action Necessary
25
The Nation's Premier Laboratory for Land Forces


RDECOM
Identifying Targets: Friendly Civilians
ARL
Friendly Civilian Characteristics:
• Civilian clothing
• No weapon in hand
No Action Necessary
26
The Nation's Premier Laboratory for Land Forces


RDECOM
Identifying Targets: Friendly Civilians
ARL
No Action Necessary
27
The Nation's Premier Laboratory for Land Forces

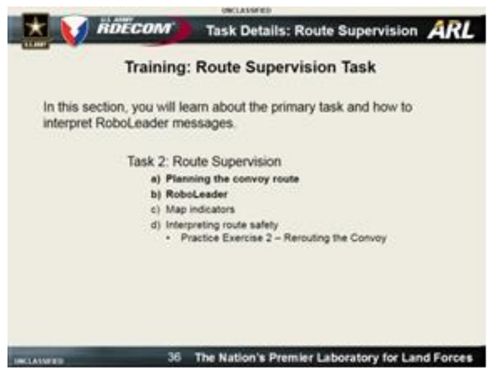

RDECOM
Identifying Targets: Friendly Civilians
ARL
Friendly Civilian Characteristics:
• Civilian clothing
• No weapon in hand
No Action Necessary
28
The Nation's Premier Laboratory for Land Forces

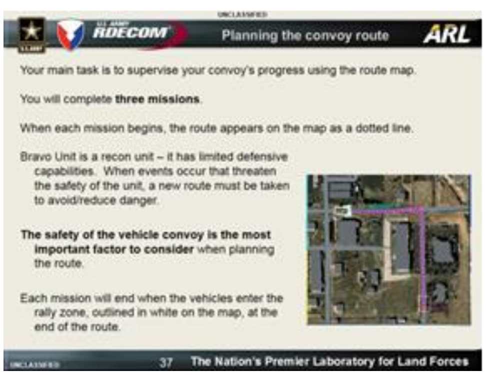

RDECOM
Identifying Targets: Armed Civilians
ARL
Armed Civilian (Insurgent) Characteristics:
• Holding weapon
• Casual clothing
• Most will have masked face
Must Identify the Threat!
29
The Nation's Premier Laboratory for Land Forces


RDECOM
Identifying Targets: Armed Civilians
ARL
Must Identify the Threat!
30
The Nation's Premier Laboratory for Land Forces


RDECOM
Identifying Targets: Armed Civilians
ARL
Must Identify the Threat!
31
The Nation's Premier Laboratory for Land Forces

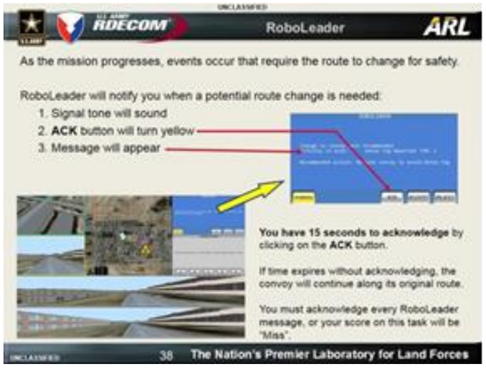

RDECOM
Identifying Targets: Armed Civilians
ARL
Must Identify the Threat!
32
The Nation's Premier Laboratory for Land Forces

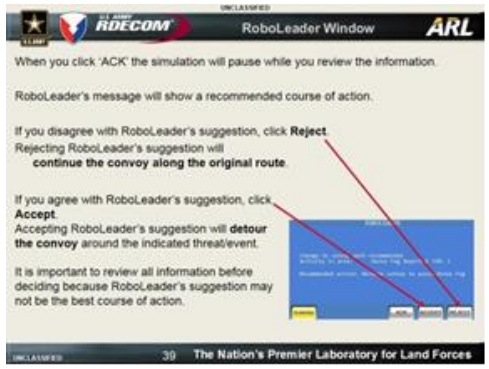


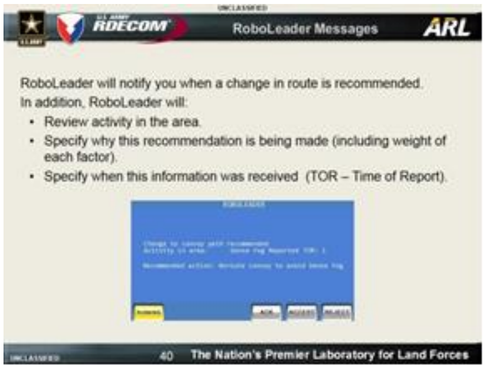


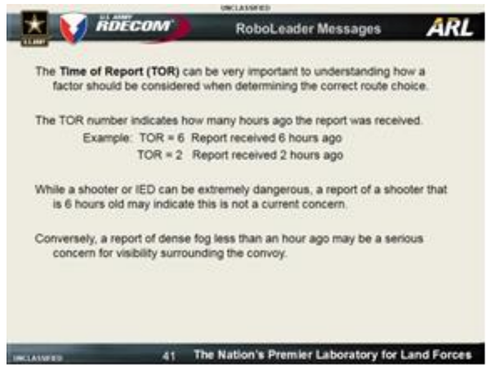
The following slides in the section "Route Supervision," parts a and b, vary according to Agent Reasoning Transparency (ART) level.

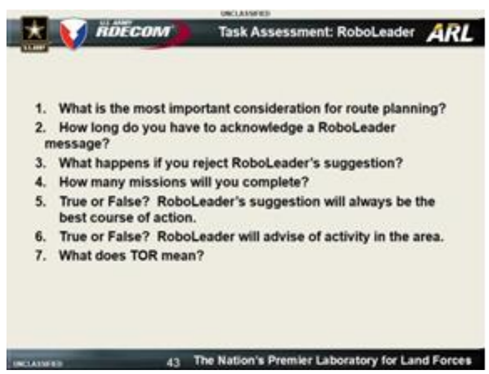
Route Supervision training slides, ART 3



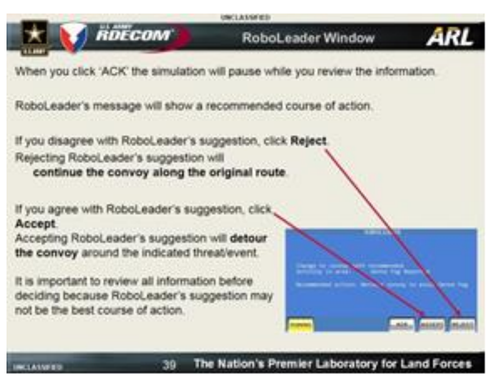


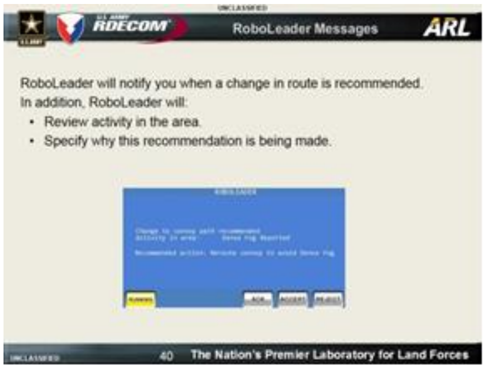


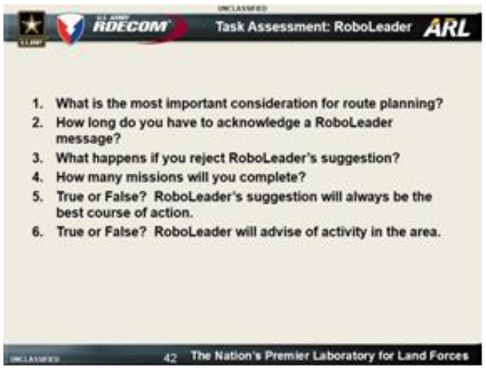




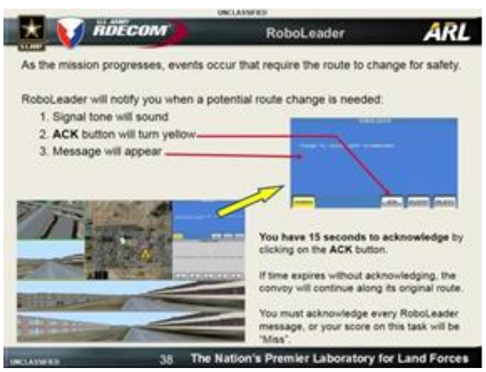
Route Supervision training slides, ART 2

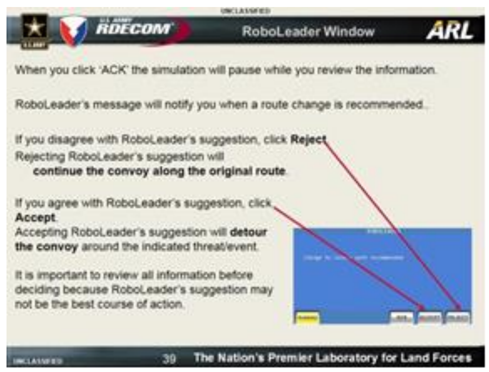


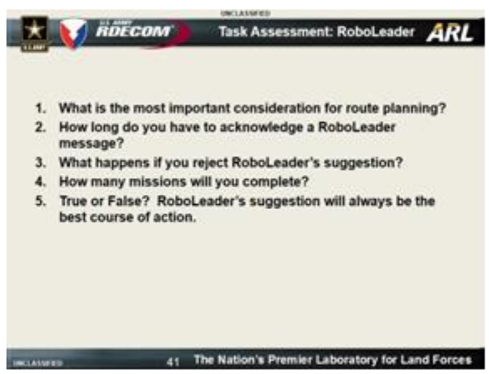




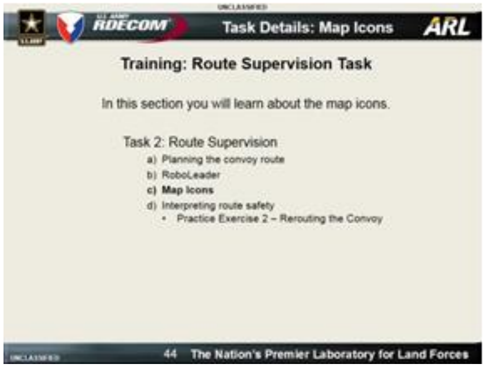
Route Supervision training slides, ART 1

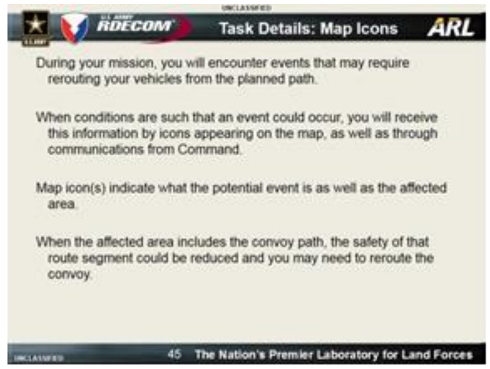


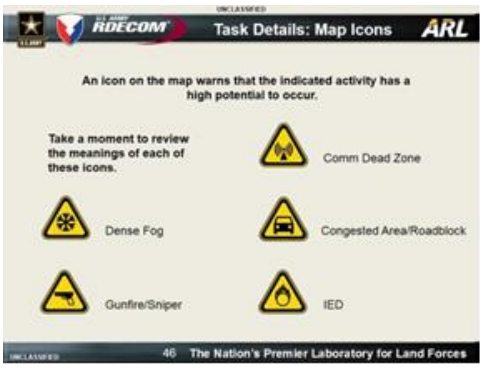


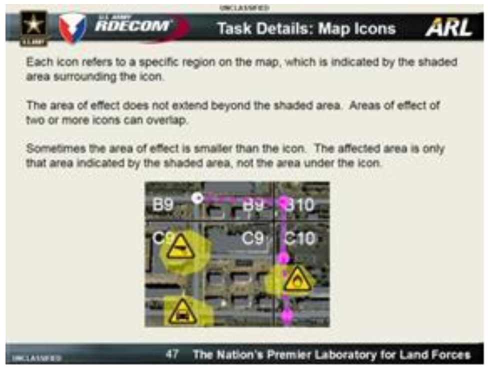


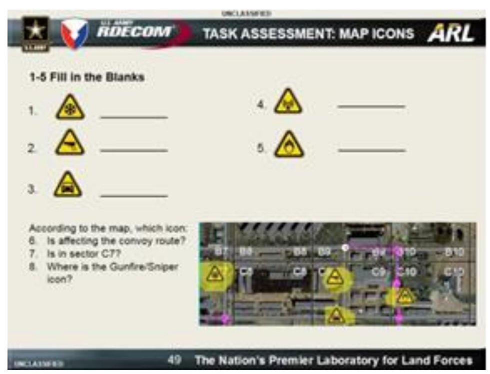
The following slides are common to all ART levels.

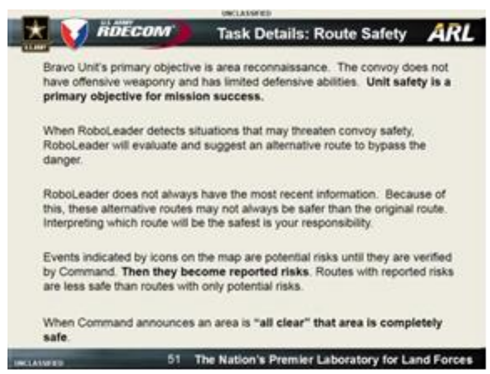


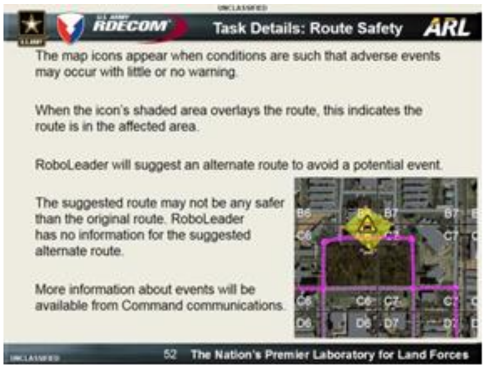


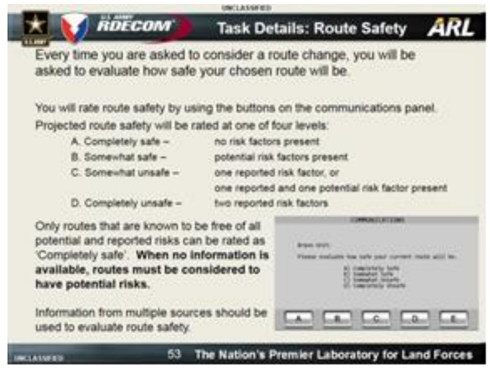




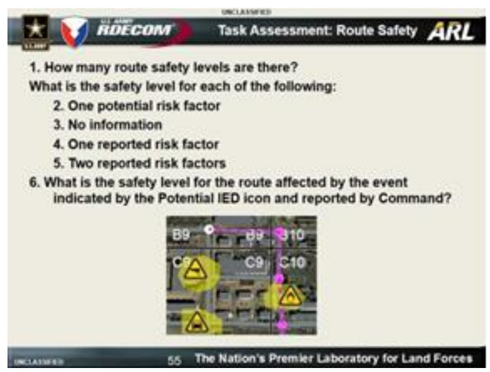




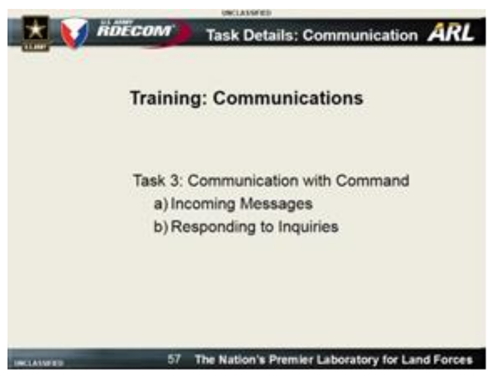

UNCLASSIFIED
RDECOM
Task Details: Route Safety
ARL
The map icons appear when conditions are such that adverse events may occur with little or no warning.
When the icon's shaded area overlays the route, this indicates the route is in the affected area.
RoboLeader will suggest an alternate route to avoid a potential event.
The suggested route may not be any safer than the original route. RoboLeader has no information for the suggested alternate route.
More information about events will be available from Command communications.
UNCLASSIFIED
52 The Nation's Premier Laboratory for Land Forces

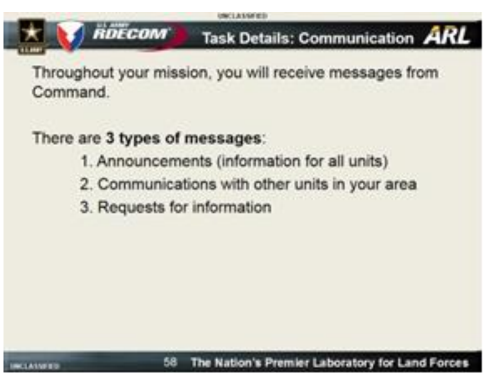

UNCLASSIFIED
RDECOM
Task Details: Route Safety
ARL
Every time you are asked to consider a route change, you will be asked to evaluate how safe your chosen route will be.
You will rate route safety by using the buttons on the communications panel.
Projected route safety will be rated at one of four levels:
A. Completely safe – no risk factors present
B. Somewhat safe – potential risk factors present
C. Somewhat unsafe – one reported risk factor, or one reported and one potential risk factor present
D. Completely unsafe – two reported risk factors
Only routes that are known to be free of all potential and reported risks can be rated as 'Completely safe'. When no information is available, routes must be considered to have potential risks.
Information from multiple sources should be used to evaluate route safety.
UNCLASSIFIED
53 The Nation's Premier Laboratory for Land Forces

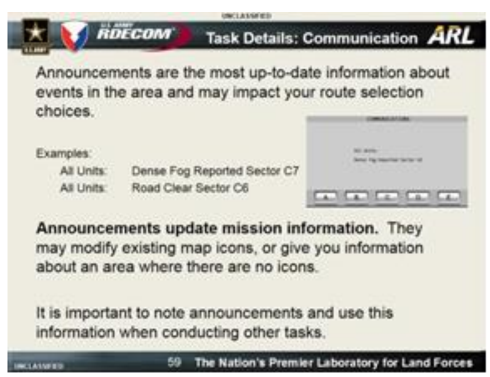

UNCLASSIFIED
RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed this part of the training.
At this time you will complete an assessment of your knowledge of assessing and reporting route safety.
If your score is too low to continue, you will be allowed to repeat the training once and try again.
UNCLASSIFIED
54 The Nation's Premier Laboratory for Land Forces

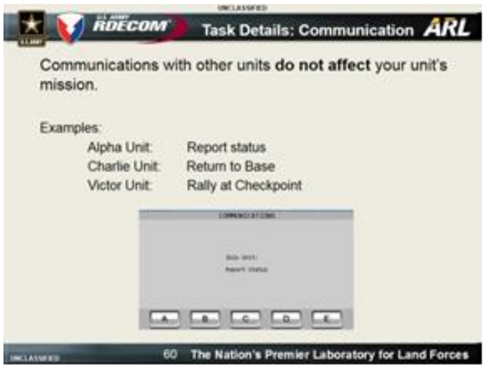

UNCLASSIFIED
RDECOM
Task Assessment: Route Safety
ARL
1. How many route safety levels are there?
What is the safety level for each of the following:
2. One potential risk factor
3. No information
4. One reported risk factor
5. Two reported risk factors
6. What is the safety level for the route affected by the event indicated by the Potential IED icon and reported by Command?
UNCLASSIFIED
55 The Nation's Premier Laboratory for Land Forces

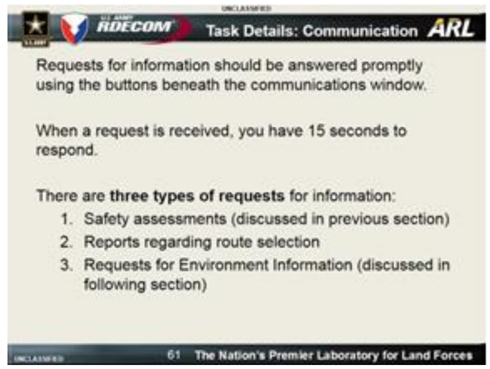

UNCLASSIFIED
RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed this part of the training.
At this time you will practice the route supervision tasks.
When you complete this practice mission, you will return to these training slides.
UNCLASSIFIED
56 The Nation's Premier Laboratory for Land Forces

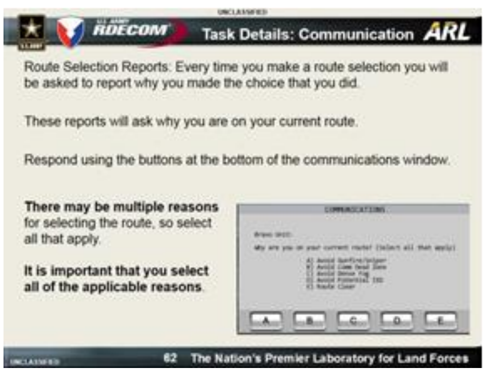

UNCLASSIFIED
RDECOM
Task Details: Communication
ARL
Training: Communications
Task 3: Communication with Command
a) Incoming Messages
b) Responding to Inquiries
UNCLASSIFIED
57 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Task Details: Communication
ARL
Throughout your mission, you will receive messages from Command.
There are 3 types of messages:
1. Announcements (information for all units)
2. Communications with other units in your area
3. Requests for information
UNCLASSIFIED
58 The Nation's Premier Laboratory for Land Forces

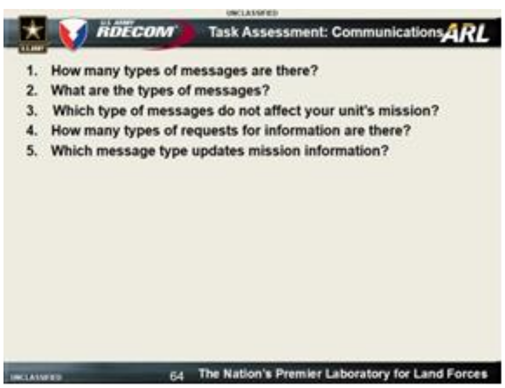

UNCLASSIFIED
RDECOM
Task Details: Communication
ARL
Announcements are the most up-to-date information about events in the area and may impact your route selection choices.
Examples:
All Units: Dense Fog Reported Sector C7
All Units: Road Clear Sector C6
Announcements update mission information. They may modify existing map icons, or give you information about an area where there are no icons.
It is important to note announcements and use this information when conducting other tasks.
UNCLASSIFIED
59 The Nation's Premier Laboratory for Land Forces


RDECOM
Task Details: Communication
ARL
Communications with other units do not affect your unit's mission.
Examples:
Alpha Unit: Report status
Charlie Unit: Return to Base
Victor Unit: Rally at Checkpoint
A
B
C
D
E
60 The Nation's Premier Laboratory for Land Forces

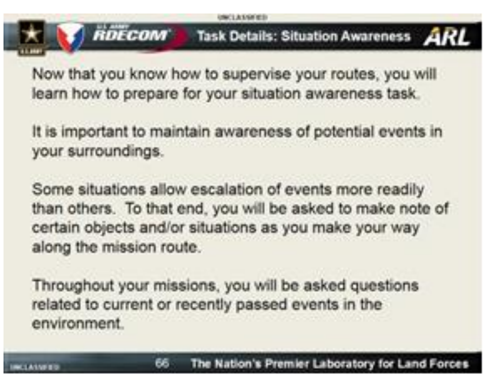

RDECOM
Task Details: Communication
ARL
Requests for information should be answered promptly using the buttons beneath the communications window.
When a request is received, you have 15 seconds to respond.
There are three types of requests for information:
1. Safety assessments (discussed in previous section)
2. Reports regarding route selection
3. Requests for Environment Information (discussed in following section)
61 The Nation's Premier Laboratory for Land Forces


RDECOM
Task Details: Communication
ARL
Route Selection Reports: Every time you make a route selection you will be asked to report why you made the choice that you did.
These reports will ask why you are on your current route.
Respond using the buttons at the bottom of the communications window.
There may be multiple reasons for selecting the route, so select all that apply.
It is important that you select all of the applicable reasons.
A
B
C
D
E
62 The Nation's Premier Laboratory for Land Forces

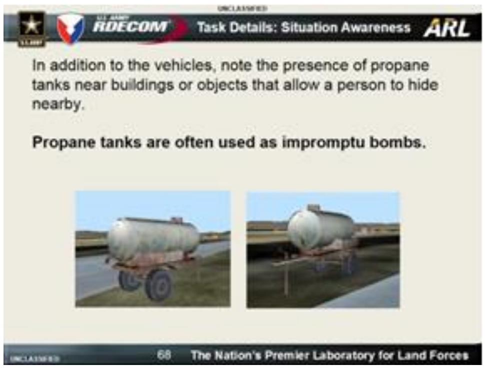

RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed this part of the training.
At this time you will complete an assessment of your knowledge of communications.
If your score is too low to continue, you will be allowed to repeat the training once and try again.
63 The Nation's Premier Laboratory for Land Forces

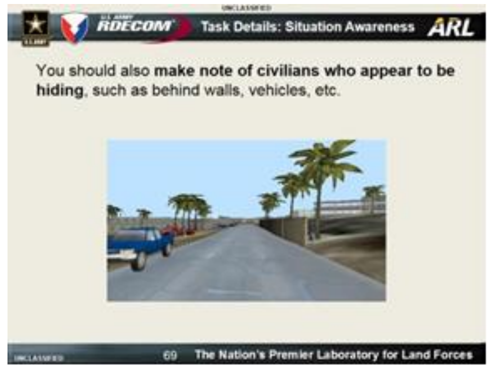

RDECOM
Task Assessment: Communications
ARL
1. How many types of messages are there?
2. What are the types of messages?
3. Which type of messages do not affect your unit's mission?
4. How many types of requests for information are there?
5. Which message type updates mission information?
64 The Nation's Premier Laboratory for Land Forces

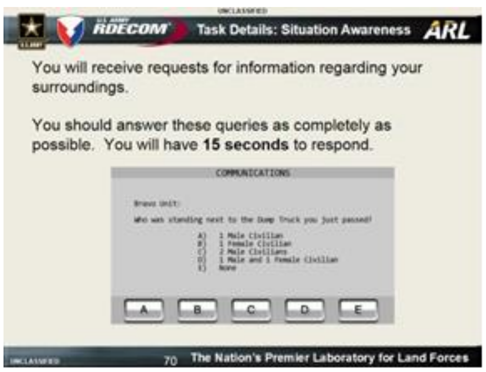

RDECOM
Task Details: Situation Awareness
ARL
Training: Situation Awareness
Task 4: Situation Awareness
65 The Nation's Premier Laboratory for Land Forces


RDECOM
Task Details: Situation Awareness
ARL
Now that you know how to supervise your routes, you will learn how to prepare for your situation awareness task.
It is important to maintain awareness of potential events in your surroundings.
Some situations allow escalation of events more readily than others. To that end, you will be asked to make note of certain objects and/or situations as you make your way along the mission route.
Throughout your missions, you will be asked questions related to current or recently passed events in the environment.
66 The Nation's Premier Laboratory for Land Forces

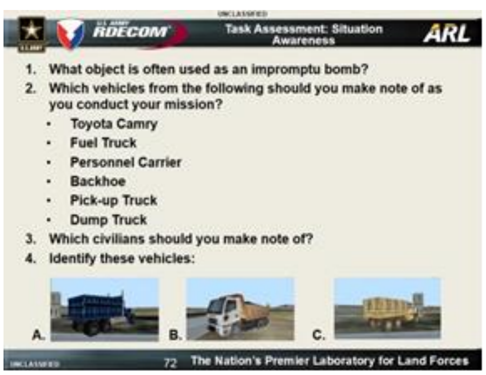

RDECOM
Task Details: Situation Awareness
ARL
Certain vehicles are used for enemy activity more than others. Make note of these vehicles and if people, particularly civilians, are hanging around them:
Personnel Carrier
Blue Garbage Truck
Fuel Truck
Dump Truck
67 The Nation's Premier Laboratory for Land Forces


U.S. ARMY RDECOM
Task Details: Situation Awareness
ARL
In addition to the vehicles, note the presence of propane tanks near buildings or objects that allow a person to hide nearby.
Propane tanks are often used as impromptu bombs.
68
The Nation's Premier Laboratory for Land Forces


U.S. ARMY RDECOM
Task Details: Situation Awareness
ARL
You should also make note of civilians who appear to be hiding, such as behind walls, vehicles, etc.
69
The Nation's Premier Laboratory for Land Forces

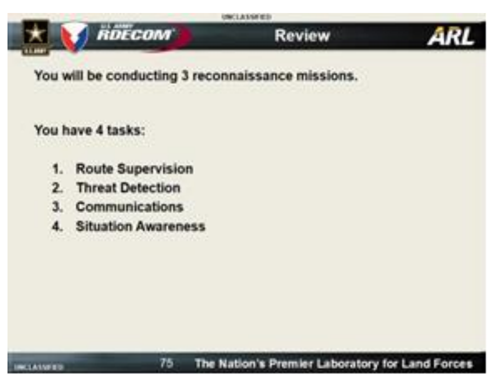

U.S. ARMY RDECOM
Task Details: Situation Awareness
ARL
You will receive requests for information regarding your surroundings.
You should answer these queries as completely as possible. You will have 15 seconds to respond.
COMMUNICATIONS
A
B
C
D
E
70
The Nation's Premier Laboratory for Land Forces

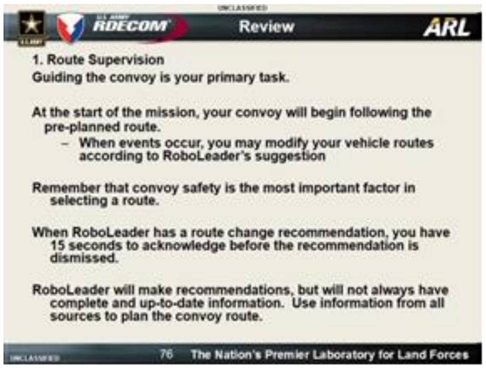

U.S. ARMY RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed this part of the training.
At this time you will complete an assessment of your knowledge of the situation awareness task.
If your score is too low to continue, you will be allowed to repeat the training once and try again.
71
The Nation's Premier Laboratory for Land Forces

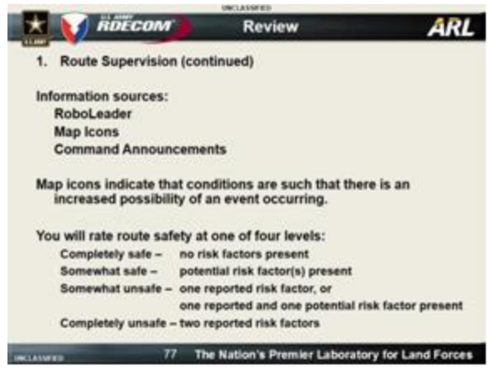

U.S. ARMY RDECOM
Task Assessment: Situation Awareness
ARL
1. What object is often used as an impromptu bomb?
2. Which vehicles from the following should you make note of as you conduct your mission?
• Toyota Camry
• Fuel Truck
• Personnel Carrier
• Backhoe
• Pick-up Truck
• Dump Truck
3. Which civilians should you make note of?
4. Identify these vehicles:
A.
B.
C.
72
The Nation's Premier Laboratory for Land Forces

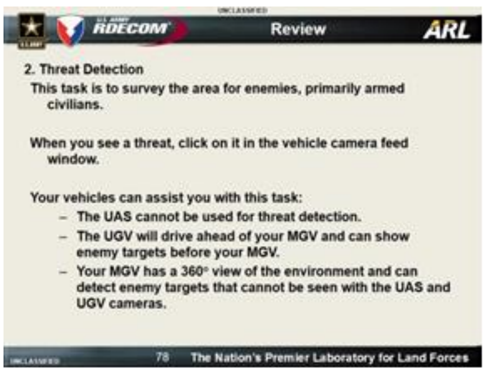

U.S. ARMY RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed this part of the training.
At this time you will practice the communication and situation awareness tasks.
When you complete this practice mission, you will return to these training slides.
73
The Nation's Premier Laboratory for Land Forces

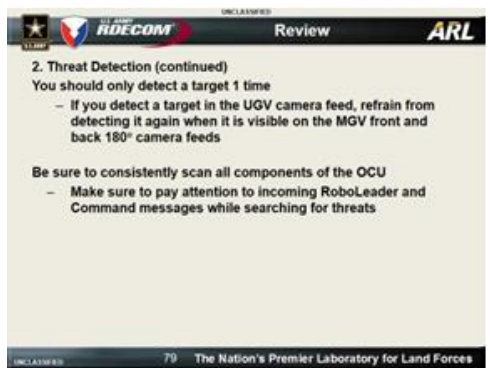

U.S. ARMY RDECOM
ARL
Review and Helpful Reminders
74
The Nation's Premier Laboratory for Land Forces

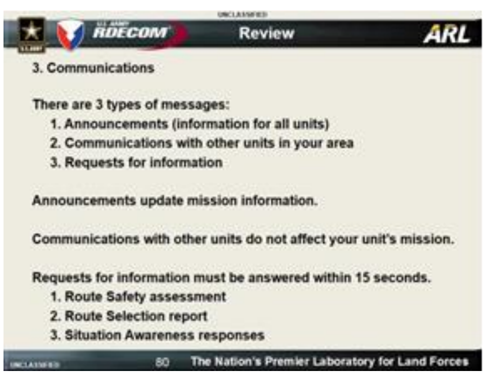

U.S. ARMY RDECOM
Review
ARL
You will be conducting 3 reconnaissance missions.
You have 4 tasks:
1. Route Supervision
2. Threat Detection
3. Communications
4. Situation Awareness
75
The Nation's Premier Laboratory for Land Forces

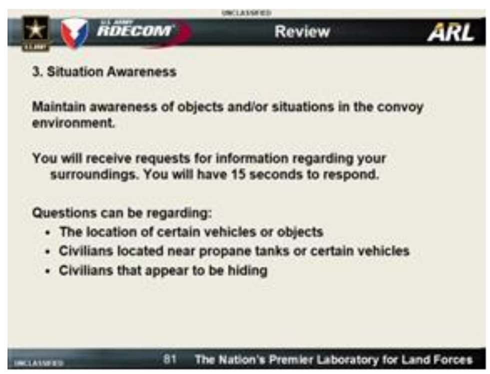

UNCLASSIFIED
U.S. ARMY RDECOM
Review
ARL
1. Route Supervision
Guiding the convoy is your primary task.
At the start of the mission, your convoy will begin following the pre-planned route.
– When events occur, you may modify your vehicle routes according to RoboLeader's suggestion
Remember that convoy safety is the most important factor in selecting a route.
When RoboLeader has a route change recommendation, you have 15 seconds to acknowledge before the recommendation is dismissed.
RoboLeader will make recommendations, but will not always have complete and up-to-date information. Use information from all sources to plan the convoy route.
UNCLASSIFIED
76 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
Review
ARL
1. Route Supervision (continued)
Information sources:
RoboLeader
Map Icons
Command Announcements
Map icons indicate that conditions are such that there is an increased possibility of an event occurring.
You will rate route safety at one of four levels:
Completely safe – no risk factors present
Somewhat safe – potential risk factor(s) present
Somewhat unsafe – one reported risk factor, or one reported and one potential risk factor present
Completely unsafe – two reported risk factors
UNCLASSIFIED
77 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
Review
ARL
2. Threat Detection
This task is to survey the area for enemies, primarily armed civilians.
When you see a threat, click on it in the vehicle camera feed window.
Your vehicles can assist you with this task:
– The UAS cannot be used for threat detection.
– The UGV will drive ahead of your MGV and can show enemy targets before your MGV.
– Your MGV has a 360° view of the environment and can detect enemy targets that cannot be seen with the UAS and UGV cameras.
UNCLASSIFIED
78 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
Review
ARL
2. Threat Detection (continued)
You should only detect a target 1 time
– If you detect a target in the UGV camera feed, refrain from detecting it again when it is visible on the MGV front and back 180° camera feeds
Be sure to consistently scan all components of the OCU
– Make sure to pay attention to incoming RoboLeader and Command messages while searching for threats
UNCLASSIFIED
79 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY RDECOM
Review
ARL
3. Communications
There are 3 types of messages:
1. Announcements (information for all units)
2. Communications with other units in your area
3. Requests for information
Announcements update mission information.
Communications with other units do not affect your unit's mission.
Requests for information must be answered within 15 seconds.
1. Route Safety assessment
2. Route Selection report
3. Situation Awareness responses
UNCLASSIFIED
80 The Nation's Premier Laboratory for Land Forces

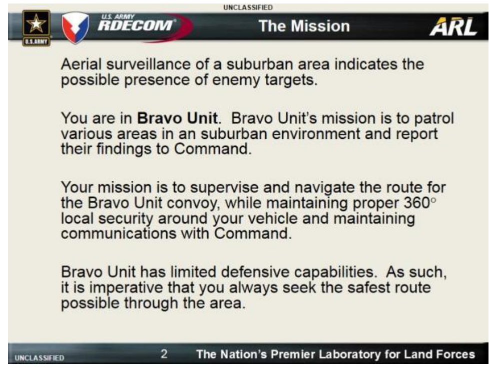

UNCLASSIFIED
U.S. ARMY RDECOM
Review
ARL
3. Situation Awareness
Maintain awareness of objects and/or situations in the convoy environment.
You will receive requests for information regarding your surroundings. You will have 15 seconds to respond.
Questions can be regarding:
• The location of certain vehicles or objects
• Civilians located near propane tanks or certain vehicles
• Civilians that appear to be hiding
UNCLASSIFIED
81 The Nation's Premier Laboratory for Land Forces

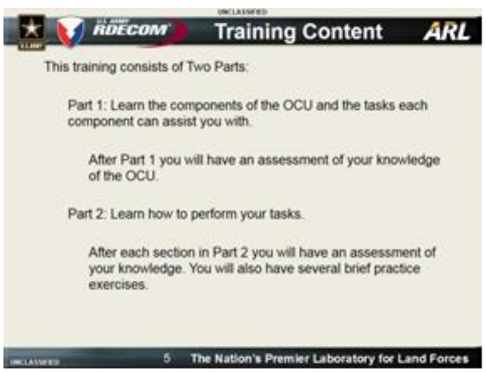

UNCLASSIFIED
U.S. ARMY RDECOM
Review
ARL
For all Missions -
• Mission is complete when the vehicles arrive at the rally zone
• The mission will end automatically
UNCLASSIFIED
82 The Nation's Premier Laboratory for Land Forces

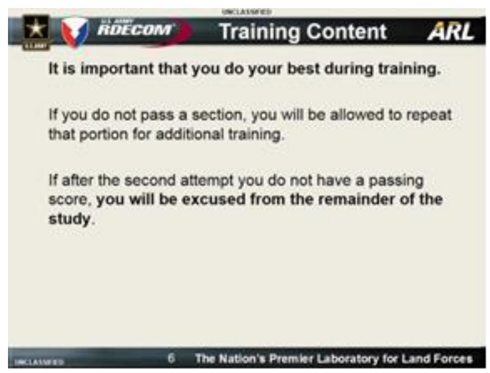

UNCLASSIFIED
U.S. ARMY RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed this part of the training.
At this point, you will perform one full practice scenario with all of the task components
• Route Supervision
• Threat Detection
• Communications
• Situation Awareness
When you have completed this practice mission, you have a short break, and then will begin your first mission
UNCLASSIFIED
83 The Nation's Premier Laboratory for Land Forces

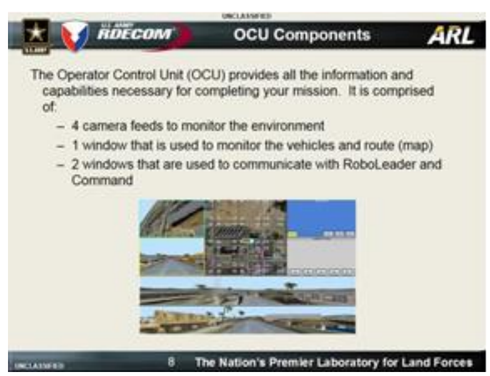
Experiment 2 Training Slides

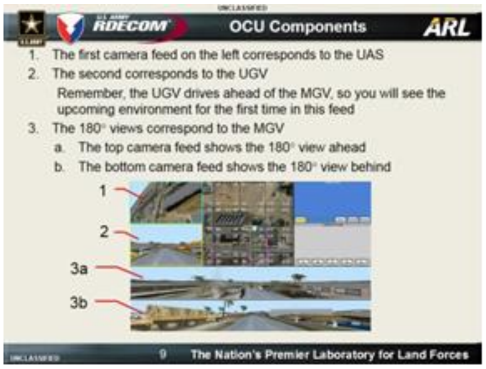
Slides are common across ARTs unless otherwise noted.

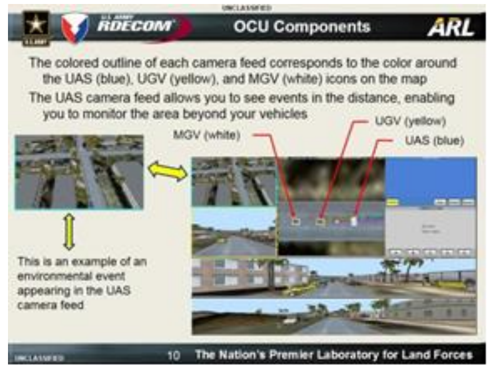




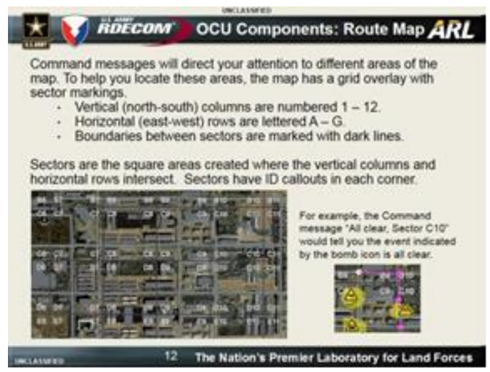


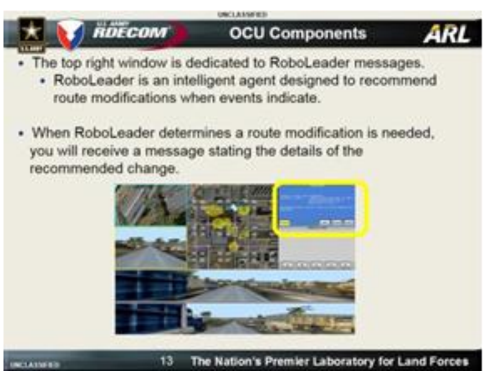


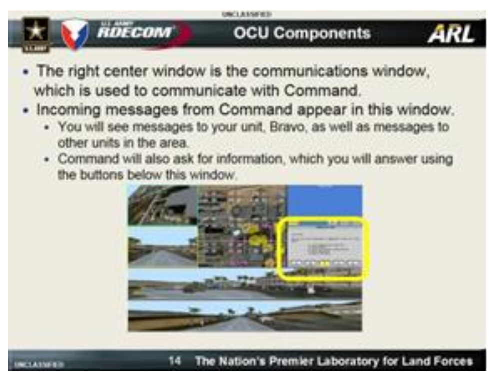

RDECOM
OCU Components
ARL
1. The first camera feed on the left corresponds to the UAS
2. The second corresponds to the UGV
Remember, the UGV drives ahead of the MGV, so you will see the upcoming environment for the first time in this feed
3. The 180° views correspond to the MGV
a. The top camera feed shows the 180° view ahead
b. The bottom camera feed shows the 180° view behind
1
2
3a
3b
9 The Nation's Premier Laboratory for Land Forces

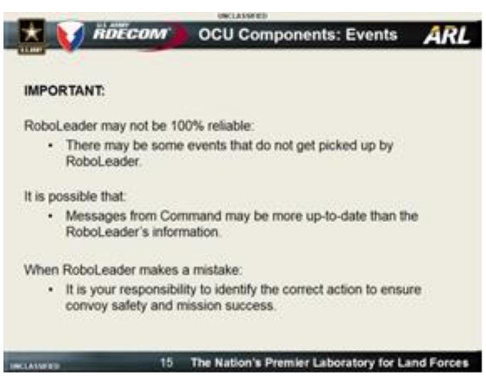

RDECOM
OCU Components
ARL
The colored outline of each camera feed corresponds to the color around the UAS (blue), UGV (yellow), and MGV (white) icons on the map
The UAS camera feed allows you to see events in the distance, enabling you to monitor the area beyond your vehicles
UGV (yellow)
MGV (white)
UAS (blue)
This is an example of an environmental event appearing in the UAS camera feed
10 The Nation's Premier Laboratory for Land Forces

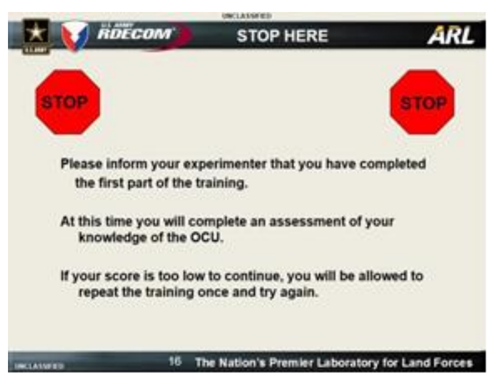

RDECOM
OCU Components: Route Map
ARL
The upper center window is the route map
The route map is an aerial view of the operational area
• The map allows you to maintain awareness of where your vehicles are along the route, and allows you to plan new routes when necessary.
11 The Nation's Premier Laboratory for Land Forces

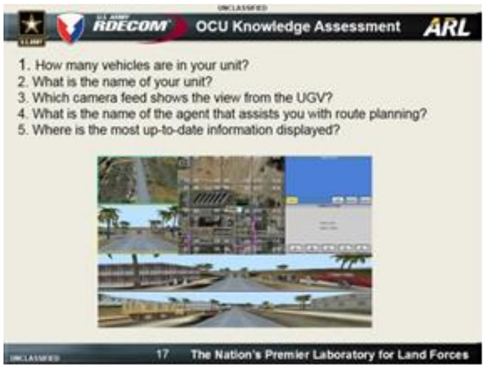

RDECOM
OCU Components: Route Map
ARL
Command messages will direct your attention to different areas of the map. To help you locate these areas, the map has a grid overlay with sector markings.
• Vertical (north-south) columns are numbered 1 – 12.
• Horizontal (east-west) rows are lettered A – G.
• Boundaries between sectors are marked with dark lines.
Sectors are the square areas created where the vertical columns and horizontal rows intersect. Sectors have ID callouts in each corner.
For example, the Command message "All clear, Sector C10" would tell you the event indicated by the bomb icon is all clear.
12 The Nation's Premier Laboratory for Land Forces

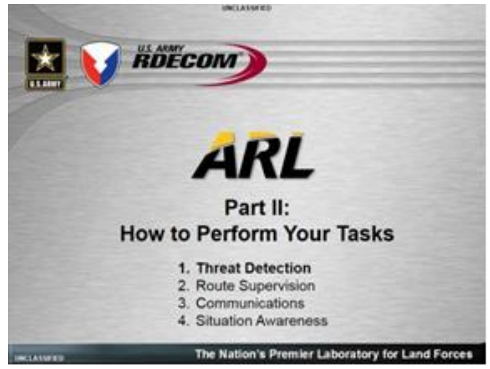

RDECOM
OCU Components
ARL
• The top right window is dedicated to RoboLeader messages.
• RoboLeader is an intelligent agent designed to recommend route modifications when events indicate.
• When RoboLeader determines a route modification is needed, you will receive a message stating the details of the recommended change.
13 The Nation's Premier Laboratory for Land Forces

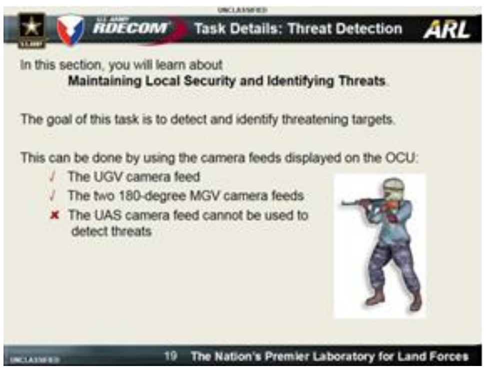

RDECOM
OCU Components
ARL
• The right center window is the communications window, which is used to communicate with Command.
• Incoming messages from Command appear in this window.
• You will see messages to your unit, Bravo, as well as messages to other units in the area.
• Command will also ask for information, which you will answer using the buttons below this window.
14 The Nation's Premier Laboratory for Land Forces


RDECOM
OCU Components: Events
ARL
IMPORTANT:
RoboLeader may not be 100% reliable:
• There may be some events that do not get picked up by RoboLeader.
It is possible that:
• Messages from Command may be more up-to-date than the RoboLeader's information.
When RoboLeader makes a mistake:
• It is your responsibility to identify the correct action to ensure convoy safety and mission success.
15 The Nation's Premier Laboratory for Land Forces

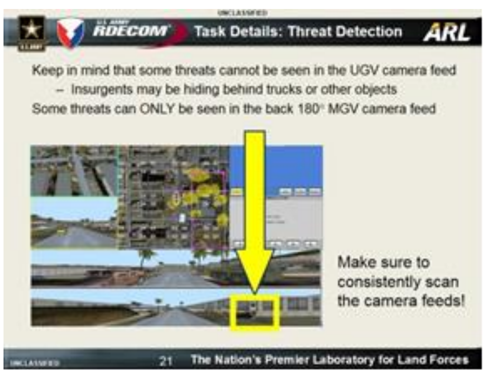

RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed the first part of the training.
At this time you will complete an assessment of your knowledge of the OCU.
If your score is too low to continue, you will be allowed to repeat the training once and try again.
16 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
OCU Knowledge Assessment
ARL
1. How many vehicles are in your unit?
2. What is the name of your unit?
3. Which camera feed shows the view from the UGV?
4. What is the name of the agent that assists you with route planning?
5. Where is the most up-to-date information displayed?
UNCLASSIFIED
17
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
U.S. ARMY
RDECOM
ARL
Part II:
How to Perform Your Tasks
1. Threat Detection
2. Route Supervision
3. Communications
4. Situation Awareness
UNCLASSIFIED
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Task Details: Threat Detection
ARL
In this section, you will learn about
Maintaining Local Security and Identifying Threats.
The goal of this task is to detect and identify threatening targets.
This can be done by using the camera feeds displayed on the OCU:
√ The UGV camera feed
√ The two 180-degree MGV camera feeds
✘ The UAS camera feed cannot be used to detect threats
UNCLASSIFIED
19
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Task Details: Threat Detection
ARL
Your task is to scan the UGV and MGV camera feeds in search of threats.
• When you find a threat, click directly on it.
• It is helpful to aim for the center of the body.
Every time you click inside a camera feed window, you will hear a camera shutter sound.
You are asked to identify every threat you see in the environment.
• Only identify each threat ONE time!
• If you identify a threat using the UGV camera feed, do not click on it again in the MGV camera feeds.
Threat!
Insurgent
UNCLASSIFIED
20
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Task Details: Threat Detection
ARL
Keep in mind that some threats cannot be seen in the UGV camera feed
– Insurgents may be hiding behind trucks or other objects
Some threats can ONLY be seen in the back 180° MGV camera feed
Make sure to consistently scan the camera feeds!
UNCLASSIFIED
21
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Task Details: Threat Detection
ARL
You will encounter three types of people:
• Friendly Soldiers
• Friendly Civilians
• Armed Civilians (Insurgents)
You must identify and report all armed civilians.
The following slides will show representations of each type
UNCLASSIFIED
22
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Identifying Targets: Friendly Soldiers
ARL
Friendly Soldier Characteristics:
• Variety of Light camouflage uniforms
• Holding weapon
• Wearing helmet
No Action Necessary
UNCLASSIFIED
23
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Identifying Targets: Friendly Soldiers
ARL
No Action Necessary
UNCLASSIFIED
24
The Nation's Premier Laboratory for Land Forces


RDECOM
Identifying Targets: Friendly Soldiers
ARL
No Action Necessary
25
The Nation's Premier Laboratory for Land Forces


RDECOM
Identifying Targets: Friendly Civilians
ARL
Friendly Civilian Characteristics:
• Civilian clothing
• No weapon in hand
No Action Necessary
26
The Nation's Premier Laboratory for Land Forces


RDECOM
Identifying Targets: Friendly Civilians
ARL
No Action Necessary
27
The Nation's Premier Laboratory for Land Forces

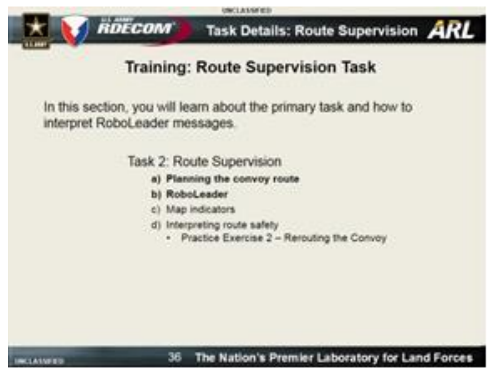

RDECOM
Identifying Targets: Friendly Civilians
ARL
Friendly Civilian Characteristics:
• Civilian clothing
• No weapon in hand
No Action Necessary
28
The Nation's Premier Laboratory for Land Forces

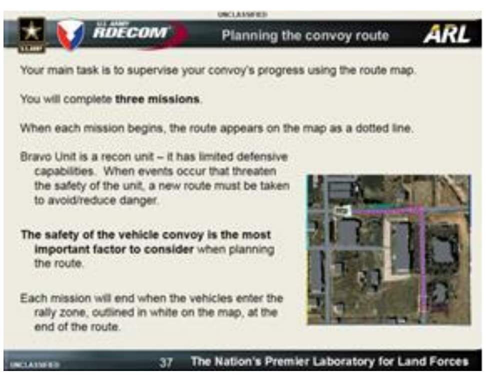

RDECOM
Identifying Targets: Armed Civilians
ARL
Armed Civilian (Insurgent) Characteristics:
• Holding weapon
• Casual clothing
• Most will have masked face
Must Identify the Threat!
29
The Nation's Premier Laboratory for Land Forces


RDECOM
Identifying Targets: Armed Civilians
ARL
Must Identify the Threat!
30
The Nation's Premier Laboratory for Land Forces


RDECOM
Identifying Targets: Armed Civilians
ARL
Must Identify the Threat!
31
The Nation's Premier Laboratory for Land Forces

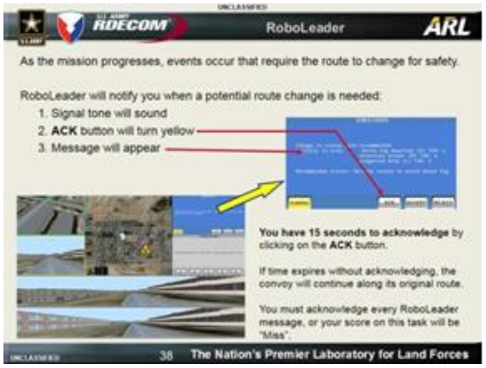

RDECOM
Identifying Targets: Armed Civilians
ARL
Must Identify the Threat!
32
The Nation's Premier Laboratory for Land Forces

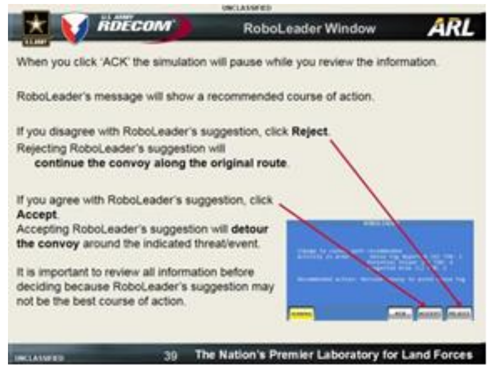


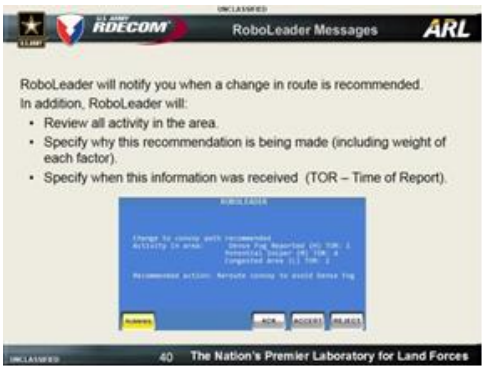


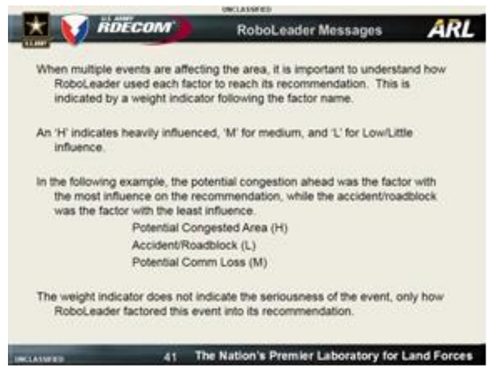
The following slides in the section "Route Supervision," parts a and b, vary according to Agent Reasoning Transparency (ART) level.

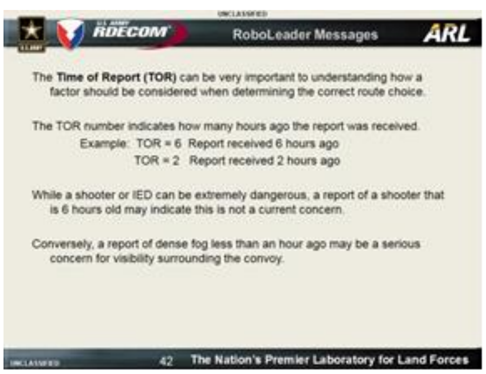
Route Supervision training slides, ART 3

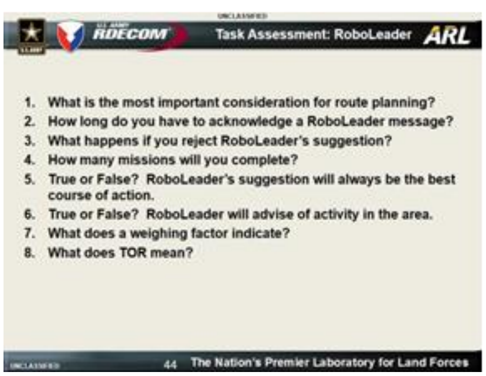




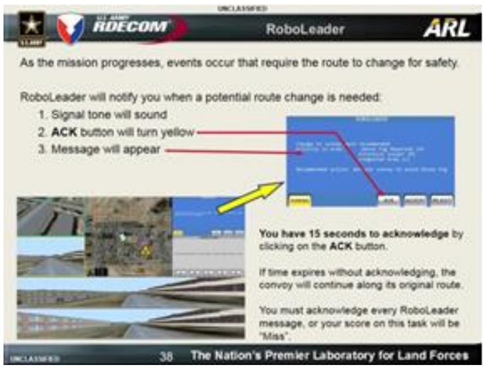


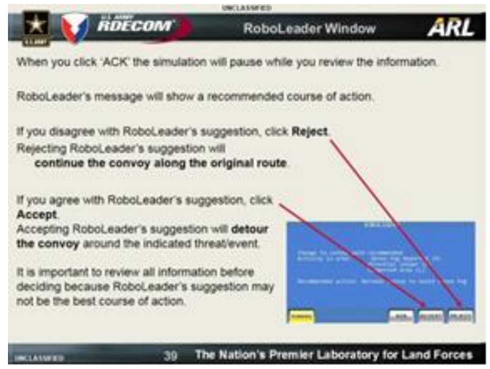


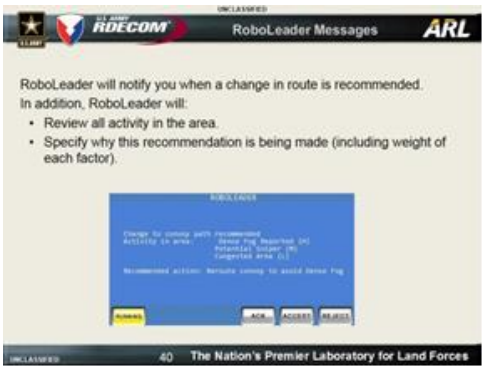


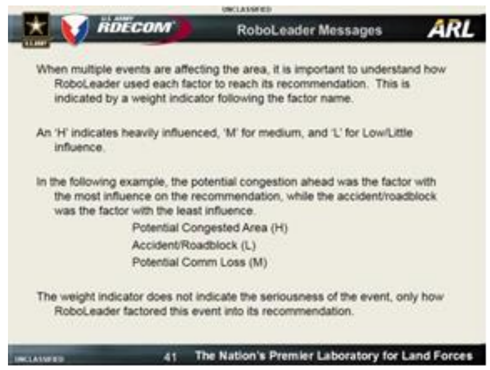


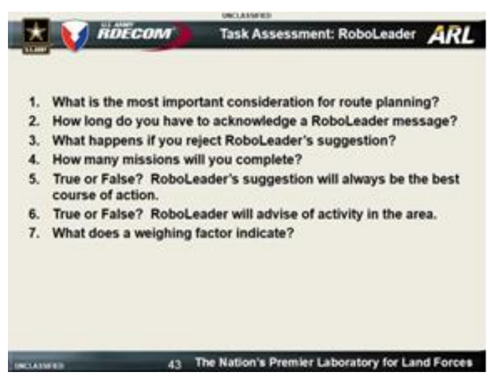
Route Supervision training slides, ART 2



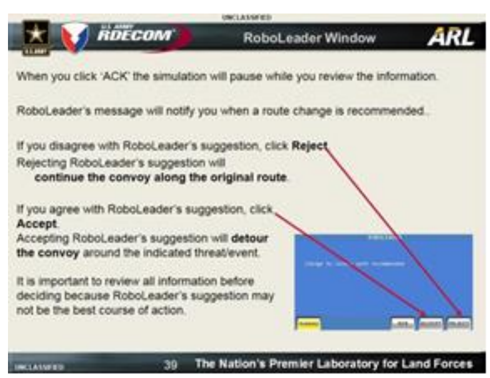


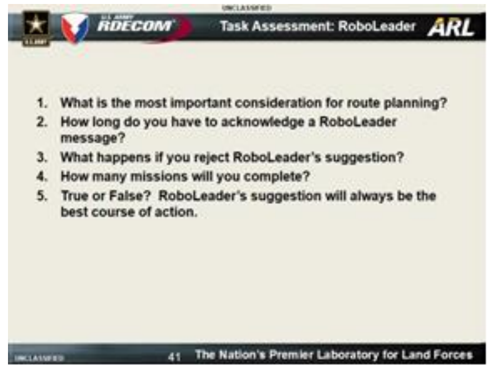




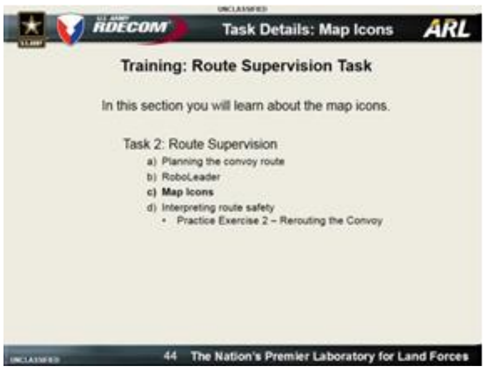


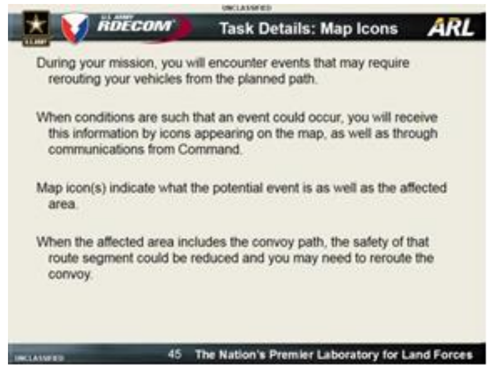
Route Supervision training slides, ART 1

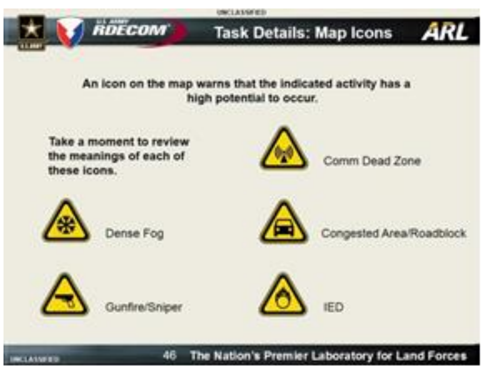


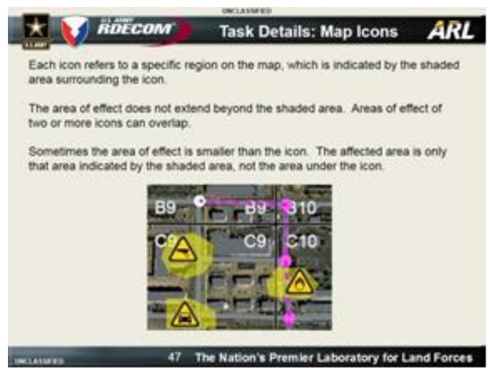


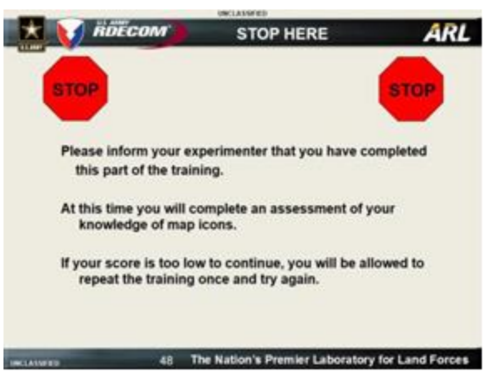
<u>The following slides are common to all ART levels.</u>

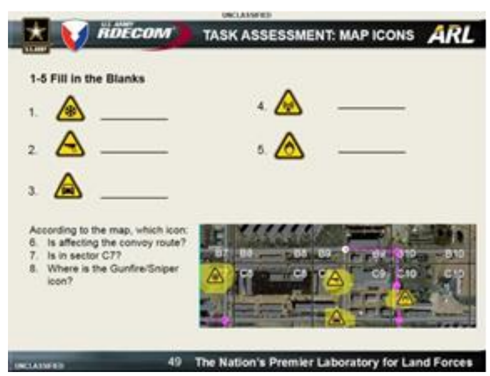


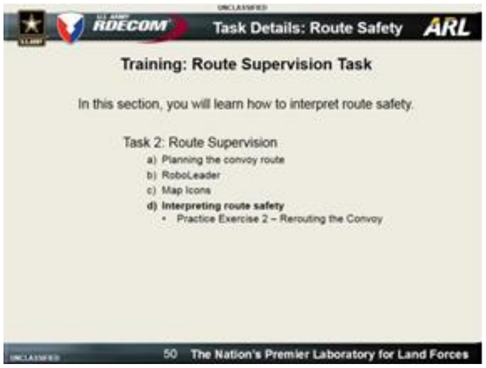




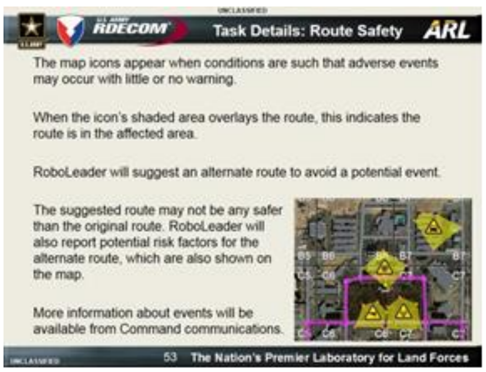


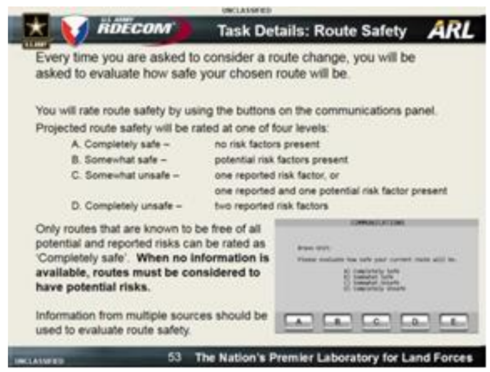




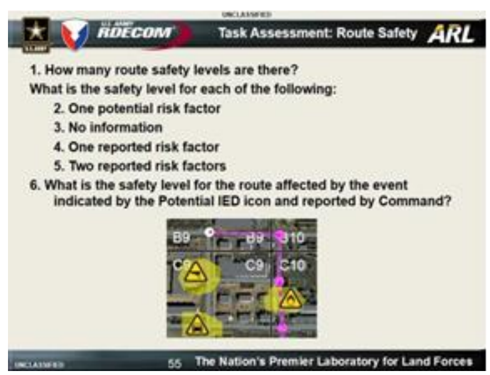

UNCLASSIFIED
RDECOM
Task Details: Route Safety
ARL
Training: Route Supervision Task
In this section, you will learn how to interpret route safety.
Task 2: Route Supervision
a) Planning the convoy route
b) RoboLeader
c) Map Icons
d) Interpreting route safety
• Practice Exercise 2 – Rerouting the Convoy
UNCLASSIFIED
50 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Task Details: Route Safety
ARL
Bravo Unit's primary objective is area reconnaissance. The convoy does not have offensive weaponry and has limited defensive abilities. Unit safety is a primary objective for mission success.
When RoboLeader detects situations that may threaten convoy safety, RoboLeader will evaluate and suggest an alternative route to bypass the danger.
RoboLeader does not always have the most recent information. Because of this, these alternative routes may not always be safer than the original route. Interpreting which route will be the safest is your responsibility.
Events indicated by icons on the map are potential risks until they are verified by Command. Then they become reported risks. Routes with reported risks are less safe than routes with only potential risks.
When Command announces an area is "all clear" that area is completely safe.
UNCLASSIFIED
51 The Nation's Premier Laboratory for Land Forces

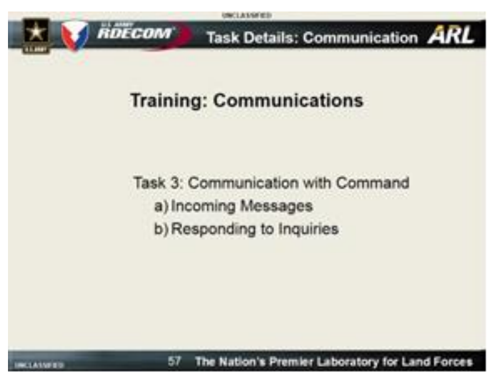

UNCLASSIFIED
RDECOM
Task Details: Route Safety
ARL
The map icons appear when conditions are such that adverse events may occur with little or no warning.
When the icon's shaded area overlays the route, this indicates the route is in the affected area.
RoboLeader will suggest an alternate route to avoid a potential event.
The suggested route may not be any safer than the original route. RoboLeader will also report potential risk factors for the alternate route, which are also shown on the map.
More information about events will be available from Command communications.
UNCLASSIFIED
53 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Task Details: Route Safety
ARL
Every time you are asked to consider a route change, you will be asked to evaluate how safe your chosen route will be.
You will rate route safety by using the buttons on the communications panel.
Projected route safety will be rated at one of four levels:
A. Completely safe – no risk factors present
B. Somewhat safe – potential risk factors present
C. Somewhat unsafe – one reported risk factor, or one reported and one potential risk factor present
D. Completely unsafe – two reported risk factors
Only routes that are known to be free of all potential and reported risks can be rated as 'Completely safe'. When no information is available, routes must be considered to have potential risks.
Information from multiple sources should be used to evaluate route safety.
A
B
C
D
E
UNCLASSIFIED
53 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed this part of the training.
At this time you will complete an assessment of your knowledge of assessing and reporting route safety.
If your score is too low to continue, you will be allowed to repeat the training once and try again.
UNCLASSIFIED
54 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Task Assessment: Route Safety
ARL
1. How many route safety levels are there?
What is the safety level for each of the following:
2. One potential risk factor
3. No information
4. One reported risk factor
5. Two reported risk factors
6. What is the safety level for the route affected by the event indicated by the Potential IED icon and reported by Command?
B9
C9
C10
UNCLASSIFIED
55 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed this part of the training.
At this time you will practice the route supervision tasks.
When you complete this practice mission, you will return to these training slides.
UNCLASSIFIED
56 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Task Details: Communication
ARL
Training: Communications
Task 3: Communication with Command
a) Incoming Messages
b) Responding to Inquiries
UNCLASSIFIED
57 The Nation's Premier Laboratory for Land Forces

Task Details: Communication

Throughout your mission, you will receive messages from Command.

There are **3 types of messages**:

1. Announcements (information for all units)
2. Communications with other units in your area
3. Requests for information

58 The Nation's Premier Laboratory for Land Forces

Task Details: Communication

Announcements are the most up-to-date information about events in the area and may impact your route selection choices.

Examples:

All Units:	Dense Fog Reported Sector C7
All Units:	Road Clear Sector C6

Announcements update mission information. They may modify existing map icons, or give you information about an area where there are no icons.

It is important to note announcements and use this information when conducting other tasks.

59 The Nation's Premier Laboratory for Land Forces

Task Details: Communication

Communications with other units **do not affect** your unit's mission.

Examples:

Alpha Unit:	Report status
Charlie Unit:	Return to Base
Victor Unit:	Rally at Checkpoint

60 The Nation's Premier Laboratory for Land Forces

Task Details: Communication

Requests for information should be answered promptly using the buttons beneath the communications window.

When a request is received, you have 15 seconds to respond.

There are **three types of requests** for information:

1. Safety assessments (discussed in previous section)
2. Reports regarding route selection
3. Requests for Environment Information (discussed in following section)

61 The Nation's Premier Laboratory for Land Forces

Task Details: Communication

Route Selection Reports: Every time you make a route selection you will be asked to report why you made the choice that you did.

These reports will ask why you are on your current route.

Respond using the buttons at the bottom of the communications window.

There may be multiple reasons for selecting the route, so select all that apply.

It is important that you select all of the applicable reasons.

62 The Nation's Premier Laboratory for Land Forces

STOP HERE

STOP STOP

Please inform your experimenter that you have completed this part of the training.

At this time you will complete an assessment of your knowledge of communications.

If your score is too low to continue, you will be allowed to repeat the training once and try again.

63 The Nation's Premier Laboratory for Land Forces

Task Assessment: Communications

1. **How many types of messages are there?**
2. **What are the types of messages?**
3. **Which type of messages do not affect your unit's mission?**
4. **How many types of requests for information are there?**
5. **Which message type updates mission information?**

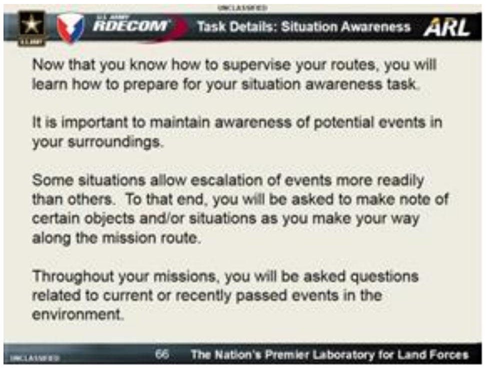
64 The Nation's Premier Laboratory for Land Forces

Task Details: Situation Awareness

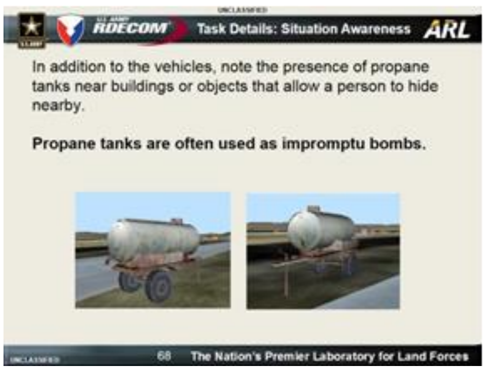
Training: Situation Awareness

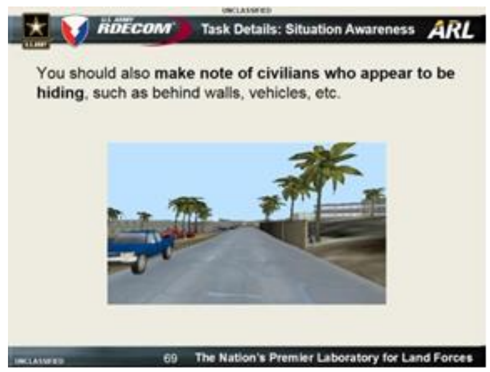
Task 4: **Situation Awareness**

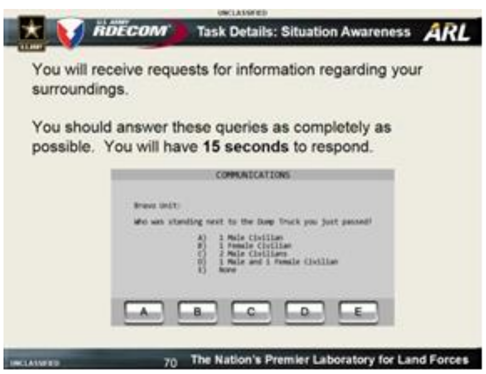
65 The Nation's Premier Laboratory for Land Forces


RDECOM
Task Details: Situation Awareness
ARL
Now that you know how to supervise your routes, you will learn how to prepare for your situation awareness task.
It is important to maintain awareness of potential events in your surroundings.
Some situations allow escalation of events more readily than others. To that end, you will be asked to make note of certain objects and/or situations as you make your way along the mission route.
Throughout your missions, you will be asked questions related to current or recently passed events in the environment.
66 The Nation's Premier Laboratory for Land Forces

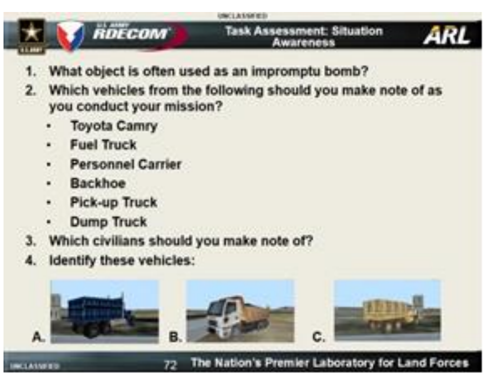

RDECOM
Task Details: Situation Awareness
ARL
Certain vehicles are used for enemy activity more than others. Make note of these vehicles and if people, particularly civilians, are hanging around them:
Personnel Carrier
Blue Garbage Truck
Fuel Truck
Dump Truck
67 The Nation's Premier Laboratory for Land Forces


RDECOM
Task Details: Situation Awareness
ARL
In addition to the vehicles, note the presence of propane tanks near buildings or objects that allow a person to hide nearby.
Propane tanks are often used as impromptu bombs.
68 The Nation's Premier Laboratory for Land Forces


RDECOM
Task Details: Situation Awareness
ARL
You should also make note of civilians who appear to be hiding, such as behind walls, vehicles, etc.
69 The Nation's Premier Laboratory for Land Forces


RDECOM
Task Details: Situation Awareness
ARL
You will receive requests for information regarding your surroundings.
You should answer these queries as completely as possible. You will have 15 seconds to respond.
COMMUNICATIONS
Bravo Unit:
Who was standing next to the Dump Truck you just passed?
A) 1 Male Civilian
B) 1 Female Civilian
C) 2 Male Civilians
D) 1 Male and 1 Female Civilian
E) None
A
B
C
D
E
70 The Nation's Premier Laboratory for Land Forces

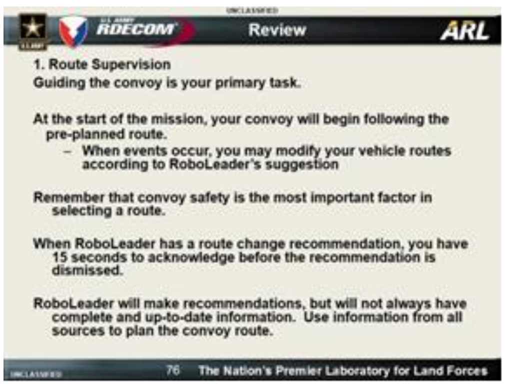

RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed this part of the training.
At this time you will complete an assessment of your knowledge of the situation awareness task.
If your score is too low to continue, you will be allowed to repeat the training once and try again.
71 The Nation's Premier Laboratory for Land Forces

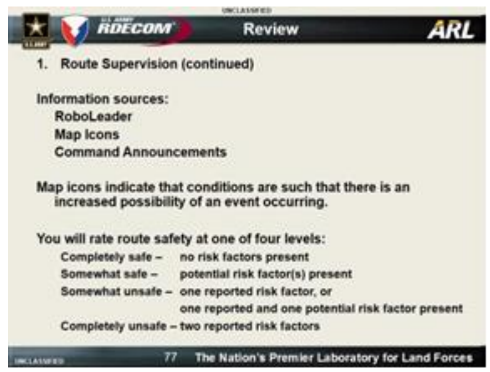

RDECOM
Task Assessment: Situation Awareness
ARL
1. What object is often used as an impromptu bomb?
2. Which vehicles from the following should you make note of as you conduct your mission?
• Toyota Camry
• Fuel Truck
• Personnel Carrier
• Backhoe
• Pick-up Truck
• Dump Truck
3. Which civilians should you make note of?
4. Identify these vehicles:
A.
B.
C.
72 The Nation's Premier Laboratory for Land Forces

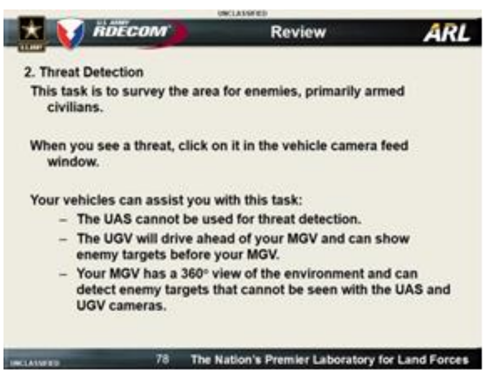

RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed this part of the training.
At this time you will practice the communication and situation awareness tasks.
When you complete this practice mission, you will return to these training slides.
73 The Nation's Premier Laboratory for Land Forces

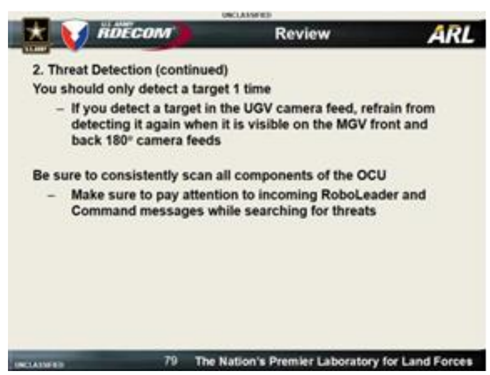

U.S. ARMY RDECOM
ARL
Review and Helpful Reminders
74 The Nation's Premier Laboratory for Land Forces

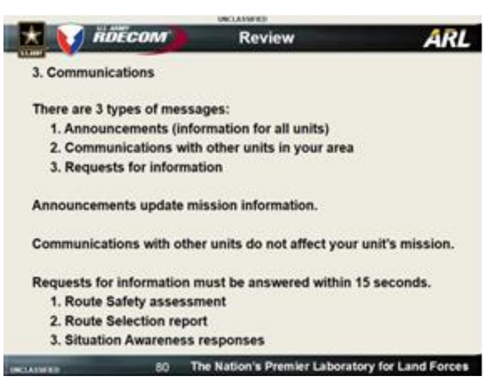

RDECOM
Review
ARL
You will be conducting 3 reconnaissance missions.
You have 4 tasks:
1. Route Supervision
2. Threat Detection
3. Communications
4. Situation Awareness
75 The Nation's Premier Laboratory for Land Forces

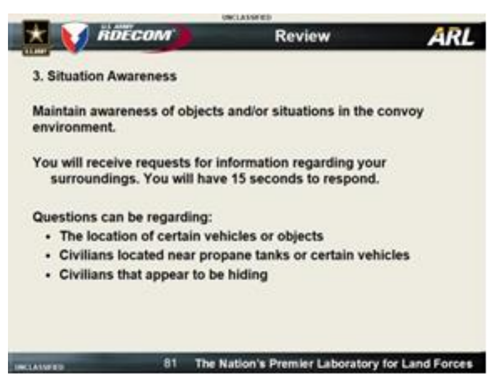

RDECOM
Review
ARL
1. Route Supervision
Guiding the convoy is your primary task.
At the start of the mission, your convoy will begin following the pre-planned route.
– When events occur, you may modify your vehicle routes according to RoboLeader's suggestion
Remember that convoy safety is the most important factor in selecting a route.
When RoboLeader has a route change recommendation, you have 15 seconds to acknowledge before the recommendation is dismissed.
RoboLeader will make recommendations, but will not always have complete and up-to-date information. Use information from all sources to plan the convoy route.
76 The Nation's Premier Laboratory for Land Forces


RDECOM
Review
ARL
1. Route Supervision (continued)
Information sources:
RoboLeader
Map Icons
Command Announcements
Map icons indicate that conditions are such that there is an increased possibility of an event occurring.
You will rate route safety at one of four levels:
Completely safe – no risk factors present
Somewhat safe – potential risk factor(s) present
Somewhat unsafe – one reported risk factor, or one reported and one potential risk factor present
Completely unsafe – two reported risk factors
77 The Nation's Premier Laboratory for Land Forces


RDECOM
Review
ARL
2. Threat Detection
This task is to survey the area for enemies, primarily armed civilians.
When you see a threat, click on it in the vehicle camera feed window.
Your vehicles can assist you with this task:
– The UAS cannot be used for threat detection.
– The UGV will drive ahead of your MGV and can show enemy targets before your MGV.
– Your MGV has a 360° view of the environment and can detect enemy targets that cannot be seen with the UAS and UGV cameras.
78 The Nation's Premier Laboratory for Land Forces


RDECOM
Review
ARL
2. Threat Detection (continued)
You should only detect a target 1 time
– If you detect a target in the UGV camera feed, refrain from detecting it again when it is visible on the MGV front and back 180° camera feeds
Be sure to consistently scan all components of the OCU
– Make sure to pay attention to incoming RoboLeader and Command messages while searching for threats
79 The Nation's Premier Laboratory for Land Forces


RDECOM
Review
ARL
3. Communications
There are 3 types of messages:
1. Announcements (information for all units)
2. Communications with other units in your area
3. Requests for information
Announcements update mission information.
Communications with other units do not affect your unit's mission.
Requests for information must be answered within 15 seconds.
1. Route Safety assessment
2. Route Selection report
3. Situation Awareness responses
80 The Nation's Premier Laboratory for Land Forces


RDECOM
Review
ARL
3. Situation Awareness
Maintain awareness of objects and/or situations in the convoy environment.
You will receive requests for information regarding your surroundings. You will have 15 seconds to respond.
Questions can be regarding:
• The location of certain vehicles or objects
• Civilians located near propane tanks or certain vehicles
• Civilians that appear to be hiding
81 The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
Review
ARL
For all Missions -
• Mission is complete when the vehicles arrive at the rally zone
• The mission will end automatically
UNCLASSIFIED
82
The Nation's Premier Laboratory for Land Forces


UNCLASSIFIED
RDECOM
STOP HERE
ARL
STOP
STOP
Please inform your experimenter that you have completed this part of the training.
At this point, you will perform one full practice scenario with all of the task components
• Route Supervision
• Threat Detection
• Communications
• Situation Awareness
When you have completed this practice mission, you have a short break, and then will begin your first mission
UNCLASSIFIED
83
The Nation's Premier Laboratory for Land Forces

REFERENCES

Agent. (n.d.). In Merriam-Webster dictionary online. Retrieved from http://www.merriam-webster.com/dictionary/agent

Ahmed, N., de Visser, E., Shaw, T., Mohamed-Ameen, A., Campbell, M., & Parasuraman, R. (2014). Statistical modelling of networked human-automation performance using working memory capacity. *Ergonomics, 57*(3), (pp. 295-318).

Automatones. (n.d.). In Theoi Greek mythology online. Retrieved from http://www.theoi.com/Ther/Automotones.html

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), (pp. 417-423).

Barber, D., Davis, L., Nicholson, D., Finkelstein, N, and Chen, J. Y. C. (2008). The mixed initiative experimental (MIX) testbed for human robot interactions with varied levels of automation. In *Proceedings of the 26th Army Science Conference,* Washington, DC: US Dept. of Army.

Barnes, M. J., Chen, J. Y. C., Jentsch, F., & Haas, E. (2006, October). Soldier interactions with aerial and ground robotic systems in future military environments. In *Proc. NATO Conf. Human Factors of Uninhabited Military Systems* (Biarritz, France).

Barnes, M. J., Chen, J. Y., Jentsch, F., Oron-Gilad, T., Redden, E., Elliott, L., & Evans III, A. W. (2014). *Designing for humans in autonomous systems: Military applications* (ARL-TR-6782). Aberdeen Proving Grounds, MD.: Army Research Laboratory (ARL), Human Research and Engineering Directorate.

Beatty, J., (1980). *Pupil dilation as an index of workload (TR-20)*. Office of Naval Research.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*(2), 276.

Bedny, G., & Meister, D. (1999). Theory of activity and situation awareness. *International Journal of Cognitive Ergonomics*, *3*(1), (pp. 63-72).

Bennett, S. (1996). A brief history of automatic control. *IEEE Control Systems Magazine, 16*(3), (pp. 17-25).

Billings, C. E., Lauber, J. K., Funkhouser, H., Lyman, G., & Huff, E. M. (1976). *NASA aviation safety reporting system.* Technical Report TM-X-3445. Moffett Field, CA: NASA Ames Research Center.

Bleckley, M. K., Durso, F. T., Crutchfield, J. M., Engle, R. W., & Khanna, M. M. (2003). Individual differences in working memory capacity predict visual attention allocation. *Psychonomic Bulletin & Review, 10*(4), (pp. 884-889).

Boy, G. A. (2014). From automation to tangible interactive objects. *Annual Reviews in Control*, *38*(1), (pp. 1-11).

Boyce, M.W., Chen, J.Y.C., Selkowitz, A.R., & Lahkmani, S.G. (in press*). Effects of agent transparency on operator trust.* Technical Report. Aberdeen Proving Grounds, MD.: Army Research Laboratory (ARL), Human Research and Engineering Directorate

Bradshaw, J. M., Sierhuis, M., Acquisti, A., Feltovich, P., Hoffman, R., Jeffers, R., . .. & Van Hoof, R. (2003). Adjustable autonomy and human-agent teamwork in practice: An interim report on space applications. In *Agent Autonomy* (pp. 243-280). Springer US.

Calhoun, G. L., Ruff, H. A., Draper, M. H., & Wright, E. J. (2011). Automation-level transference effects in simulated multiple unmanned aerial vehicle control. *Journal of Cognitive Engineering and Decision Making*, *5*(1), (pp. 55-82).

Carroll, J. B. (1993). Abilities in the domain of visual perception. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies. 1st Ed.* New York, NY: Cambridge University Press. (pp. 304-363).

Chapanis, A. (1965). On the allocation of functions between men and machines. *Occupational Psychology*, *39*(1), 1-11.

Chen, J. Y. C., & Barnes, M. J. (2010, September). Supervisory control of robots using RoboLeader. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54(19), (pp. 1483-1487).

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

Chen, J. Y. C., & Barnes, M. J. (2012b). Supervisory control of multiple robots in dynamic tasking environments. *Ergonomics, 55, (pp.* 1043–1058).

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), (pp. 13-29).

Chen, J. Y. C., Barnes, M. J., & Harper-Sciarini, M. (2011). Supervisory control of multiple robots: Human-performance issues and user-interface design. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 41*(4), (pp. 435-454).

Chen, J. Y. C., Barnes, M. J., & Qu, Z. (2010). *RoboLeader: A surrogate for enhancing the human control of a team of robots* (ARL-MR-0735). Aberdeen Proving Grounds, MD.: Army Research Laboratory (ARL), Human Research and Engineering Directorate

Chen, J. Y. C., Durlach, P. J., Sloan, J. A., & Bowens, L. D. (2008). Human robot interaction in the context of simulated route reconnaissance missions. *Military Psychology, 20*, (pp. 135-149).

Chen, J. Y. C., & Joyner, C. T. (2009). Concurrent performance of gunner's and robotic operator's tasks in a multi-tasking environment. *Military Psychology, 21*(1), (pp. 98-113).

Chen, J. Y., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). *Situation awareness-based agent transparency* (ARL-TR-6905). Aberdeen Proving Grounds, MD.: Army Research Laboratory (ARL), Human Research and Engineering Directorate.

Chen, J. Y. C., & Terrence, P. I. (2009). Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. *Ergonomics, 52*(8), (pp. 907-920).

Cheverst, K., Byun, H. E., Fitton, D., Sas, C., Kray, C., & Villar, N. (2005). Exploring issues of user model transparency and proactive behaviour in an office environment control system. *User Modeling and User-Adapted Interaction*, *15*(3-4), (pp. 235-273).

Chiappe, D., Conger, M., Liao, J., Caldwell, J. L., & Vu, K. P. L. (2013). Improving multi-tasking ability through action videogames. *Applied Ergonomics*, *44*(2), (pp. 278-284).

Chignell, M. H., Hancock, P. A., & Loewenthal, A. (1989). An introduction to intelligent interfaces. *Intelligent interfaces: Theory, research and design.* North Holland: Elsevier Science.

Christie, G. J., Cook, C. M., Ward, B. J., Tata, M. S., Sutherland, J., Sutherland, R. J., & Saucier, D. M. (2013). Mental rotational ability is correlated with spatial but not verbal working memory performance and P300 amplitude in males. *PloS one*, *8*(2), e57390.

Coovert, M. D., & Thompson, L. F. (Eds.). (2014). *The psychology of workplace technology*. New York: Routledge.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, (pp. 323-338).

Cramer, H., Wielinga, B. Ramlal, S., Evers, V., Rutledge, L., & Stash, N. (2008, April). The effects of transparency on perceived and actual competence of a content-based recommender. In *Semantic Web User Interaction Workshop, CHI*, *508*, (pp. 1-11).

Cring, E. A., & Lenfestey, A. G. (2009). *Architecting human operator trust in automation to improve system effectiveness in multiple unmanned aerial vehicles (UAV) control*. (AFIT/GSE/ENV/09-M06). Wright-Patterson AFB; OH: U.S. Air Force Institute of Technology, Graduate School of Engineering and Management.

Cuevas, H. M., Fiore, S. M., Caldwell, B. S., & Strater, L. (2007). Augmenting team cognition in human-automation teams performing in complex operational environments. *Aviation, Space, and Environmental Medicine*, *78*(5), B63-B70.

Cummings, M. L. (2004). The need for command and control instant message adaptive interfaces: Lessons learned from Tactical Tomahawk human-in-the-loop simulations. *CyberPsychology & Behavior*, *7*(6), (pp. 653-661).

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

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), (pp. 450-466).

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Dekker, S. W., Hummerdal, D. H., & Smith, K. (2010). Situation awareness: some remaining questions. *Theoretical Issues in Ergonomics Science*, *11*(1-2), (pp. 131-135).

Dekker, S. W., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human–automation co-ordination. *Cognition, Technology & Work, 4*(4), (pp. 240-244).

Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology, 111*(2), (pp. 225-236).

Descartes, R. (1984/1649). *Principles of philosophy.* (Vol. 24). London, England: Kluwer Academic Publishers.

Descartes, R. (1998/1664). *Descartes: The world and other writings*. Cambridge, UK: Cambridge University Press.

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), (pp. 697-718).

Ehmke, C., & Wilson, S. (2007, September). Identifying web usability problems from eye-tracking data. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1,* (pp. 119-128). British Computer Society.

Eliot, J., & Stumpf, H. (1987). *Models of psychological space: Psychometric, developmental, and experimental approaches*. New York: Springer-Verlag.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor referenced cognitive tests*. Princeton, N.J.: Educational Testing Service.

Endsley, M. R. (1988, October). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 32*(2), (pp. 97-101).

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

Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.) *Automation and human performance: Theory and applications,* (pp. 163-181). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in dynamic control task. *Ergonomics, 42*(3), (pp. 462–492).

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), (pp. 381-394).

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*, (pp. 102-134). Cambridge, UK: Cambridge University Press.

Fallon, C. K., Murphy, A. K., Zimmerman, L., & Mueller, S. T. (2010, May). The calibration of trust in an automated system: A sensemaking process. In *2010 International Symposium on Collaborative Technologies and Systems (CTS),* (pp. 390-395). IEEE.

Fincannon, T. D. (2013). Visio-spatial abilities in remote perception: A meta-analysis of empirical work. Dissertation. University of Central Florida, Orlando, FL

Fitts, P. M. (1951). *Human engineering for an effective air navigation and traffic control system*. Washington, DC: National Research Council.

Flach, J. M. (1995). Situation awareness: Proceed with caution. *Human Factors, 37*(1), (pp. 149-157).

Franklin, S., & Graesser, A. (1997). Is it an agent, or just a program?: A taxonomy for autonomous agents. In *Intelligent Agents III Agent Theories, Architectures, and Languages* (pp. 21-35). Berlin Heidelberg: Springer.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*(2), (pp. 172-179).

Fuld, R. B. (2000). The fiction of function allocation, revisited. *International Journal of Human-Computer Studies*, *52*(2), (pp. 217-233).

Gilson, R. D. (1995). Special issue preface. *Human Factors*, *37*(1), (pp. 3-4).

Gold, J. M., Fuller, R. L., Robinson, B. M., McMahon, R. P., Braun, E. L., & Luck, S. J. (2006). Intact attentional control of working memory encoding in schizophrenia. *Journal of Abnormal Psychology*, *115*(4), 658.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), (pp. 631-645).

Goodrich, M. A. (2010). “On maximizing fan-out: Towards controlling multiple unmanned vehicles,” In M. J. Barnes & F. Jentsch (Eds), *Human–Robot Interactions in Future Military Operations* (pp. 375-395). Surrey, U.K.: Ashgate.

Graves, R. (1960). *Greek gods and heroes*. Garden City, New York: Doubleday.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*, (pp. 534-537).

Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, *101*(1), 217.

Gugerty, L, & Brooks, J. (2004). Reference-frame misalignment and cardinal direction judgments: Group differences and strategies. *Journal of Experimental Psychology: Applied, 10*(2), (pp. 75-88).

Hancock, P. A. (2009). *Mind, machine and morality*. Surrey, England: Ashgate Publishing, Ltd.

Hancock, P. A., Billings, D. R., & Schaefer, K. E. (2011). Can you trust your robot?. *Ergonomics in Design*, *19*(3), (pp. 24-29).

Hancock, P. A., & Chignell, M. H. (1993, February). Adaptive function allocation by intelligent interfaces. In *Proceedings of the 1st international conference on Intelligent user interfaces* (pp. 227-229). ACM.

Hancock, P. A., & Diaz, D. D. (2002). Ergonomics as a foundation for a science of purpose. *Theoretical Issues in Ergonomics Science*, *3*(2), (pp. 115-123).

Hancock, P. A., & Szalma, J. L. (Eds.). (2008). *Performance under stress*. Hampshire, England: Ashgate Publishing, Ltd.

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

Hegarty, M., & Waller, D. (2005). Individual differences in spatial abilities. In P. Shah, & A. Miyake (Eds.), *Handbook of visuospatial thinking* (pp. 121–169). New York: Cambridge University Press.

Ho, G., Wheatley, D., & Scialfa, C. T. (2005). Age differences in trust and reliance of a medication management system. *Interacting With Computers, 17* (6), (pp. 690–710). doi:10.1016/j.intcom.2005.09.007

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* Oxford: Oxford University Press.

Hubert-Wallander, B., Green, C. S., & Bavelier, D. (2010). Stretching the limits of visual attention: the case of action video games. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(2), (pp. 222-230).

Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004, April). Task-evoked pupillary response to mental workload in human-computer interaction. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, (pp. 1477-1480).

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, *2*(3), 4.

Jameson, A., Baldes, S., Bauer, M., & Kroner, A. (2004). Resolving the tension between invisibility and transparency. In *Proceedings of 1st International Workshop on Invisible and Transparent Interfaces*, (pp. 29-33).

Joshi, A., Miller, S. P., & Heimdahl, M. P. (2003, October). Mode confusion analysis of a flight guidance system using formal methods. In *22nd Digital Avionics Systems Conference DASC'03* (p. 2).

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, 5(2), (pp. 113-153).

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T. Y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, *20*(8), (pp. 963-973).

Kilgore, R., & Voshell, M. (2014). Increasing the transparency of unmanned systems: Applications of ecological interface design. In *Virtual, Augmented and Mixed Reality*. *Applications of Virtual and Augmented Reality* (pp. 378-389). Springer International Publishing.

Kim, T., & Hinds, P. (2006, September). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In *The 15th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 80-85).

King, E. (2002). Clockwork prayer: A sixteenth-century mechanical monk. .*Blackbird: An Online Journal of Literature and the Arts*, *1*(1), (pp. 1-29).

Koetsier, T. (2001). On the prehistory of programmable machines: musical automata, looms, calculators. *Mechanism and Machine theory*, *36*(5), (pp. 589-603).

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition*, *29*(5), (pp. 745-756).

Langer, E. (1989). *Mindfulness.* Reading, MA: Addison-Wey.

Lathan, C., & Tracey, M. (2002). The effects of operator spatial perception and sensory feedback on human-robot teleoperation performance. *Presence, 11*, (pp. 368-377).

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

Lewis, M. (2013). Human interaction with multiple remote robots. *Reviews of Human Factors and Ergonomics*, *9*(1), (pp. 131-174).

Lewis, M., Wang, H., Chien, S. Y., Velagapudi, P., Scerri, P., & Sycara, K. (2010). Choosing autonomy modes for multirobot search. *Human Factors, 52*(2), (pp. 225-233).

Linegang, M. P., Stoner, H. A., Patterson, M. J., Seppelt, B. D., Hoffman, J. D., Crittendon, Z. B., & Lee, J. D. (2006, October). Human-automation collaboration in dynamic mission planning: A challenge requiring an ecological approach. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50(23), (pp. 2482-2486).

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (ed.), *Advances in the psychology of human intelligence* (pp. 181-248). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lohman, D. F. (1996). Spatial ability and g. In I. Dennis & P. Tapsfield (eds.) *Human abilities: Their nature and measurement*, (pp. 97-116). Mahwah: Lawrence Erlbaum Associates.

Lyons, J. B. (2013, March). Being transparent about transparency: A model for human-robot interaction. In *2013 AAAI Spring Symposium Series*.

Lyons, J. B., & Havig, P. R. (2014). Transparency in a human-machine context: Approaches for fostering shared awareness/intent. In *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments* (pp. 181-190). Springer International Publishing.

Manzey, D., Reichenbach, J., & Onnasch, L. (2008). Performance consequences of automated aids in supervisory control: The impact of function allocation. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 52*(4), (pp. 297-301).

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making, 6*(1), (pp. 57-87). DOI: 10.1177/1555343411433844.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. New York, NY: Lawrence Erlbaum Associates, Inc..

Marras, W. S., & Hancock, P. A. (2014). Putting mind and body back together: A human-systems approach to the integration of the physical and cognitive dimensions of task design and operations. *Applied Ergonomics, 45*(1), (pp. 55-60).

Masters, J. K., Miles, G., D'Souza, D., & Orr, J. P. (2004). Risk propensity, trust, and transaction costs in relational contracting. *Journal of Business Strategies, 21*(1), (pp. 47-67).

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*(1), (pp. 1-21).

Mercado, J. E., Rupp, M., 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.

Mercado, J. E., Reinerman-Jones, L., Barber, J., Szalma, J., & Hancock, P. A. (Submitted). Assessing the effectiveness of workload measures in the nuclear domain. *Safety Science.*

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors, 50*(2), (pp. 194-210).

Miller, C. A. (2014). Delegation and transparency: Coordinating interactions so information exchange is no surprise. In *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments* (pp. 191-202). Springer International Publishing.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), (pp. 134-140).

Murphy, R. & Burke, J. (2010). "The safe human–robot ratio," In M. J. Barnes & F. Jentsch (Eds.) *Human–Robot Interactions in Future Military Operations* (pp. 31–51). Surrey, U.K.: Ashgate.

Murphy, R., & Shields, J. (2012). *The role of autonomy in DoD systems*. Technical Report. Washington, D.C.: DoD Defense Science Board Task Force Report

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors, 56*, (pp. 476-488).

Ososky, S., Sanders, T., Jentsch, F., Hancock, P., & Chen, J. Y. (2014, June). Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. In *SPIE Defense+ Security* (pp. 90840E-90840E). International Society for Optics and Photonics.

Parasuraman, R., Cosenzo, K. A., & De Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology, 21*(2), (pp. 270-297).

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), (pp. 381-410).

Parasuraman, R., Molloy, R. & Singh, I. L. (1993). Performance consequences of automation-induced complacency. *The International Journal of Aviation Psychology, 3*(1), (pp. 1-23).

Parasuraman, R., & Mouloua, M. (1996). *Automation and human performance: Theory and applications*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), (pp. 230-253).

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 30*(3), (pp. 286-297).

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), (pp. 140-160).

Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology 11*(5), (pp. 559–566).

Pollard, J., & Reid, H. (2007). *The rise and fall of Alexandria: Birthplace of the modern world*. New York, NY: Penguin.

Pop, V. L., Shrewsbury, A., & Durso, F. T. (2014). Individual differences in the calibration of trust in automation. *Human Factors, 57*(4), pp. 545-556.

Pop, V. L. & Stearman, E. J. (2015). An updated automation induced complacency rated scale [measurement instrument].

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), (pp. 3-25).

Posner, M. I., & Petersen, S. E. (1989). *The attention system of the human brain* (No. TR-89-1). St. Louis, MO.: Washington University Department of Neurology.

Rao, A. S., & Georgeff, M. P. (1995, June). BDI agents: From theory to practice. In *Proceedings of the First International Conference on Multiagent Systems*, *95*, (pp. 312-319).

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*(3), 164.

Reichenbach, J., Onnasch, L., & Manzey, D. (2010). Misuse of automation: The impact of system experience on complacency and automation bias in interaction with automated aids. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(4), (pp. 374-378).

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological Science*, *12*(1), (pp. 48-55).

Riskin, J. (2003). The defecating duck, or, the ambiguous origins of artificial life. *Critical Inquiry*, *29*(4), (pp. 599-633).

Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 763.

Russell, S., & Norvig, P., (2003). *Artificial intelligence: A modern approach.* Upper Saddle River, New Jersey: Prentice-Hall.

Sarter, N. (2008). Investigating mode errors on automated flight decks: Illustrating the problem-driven, cumulative, and interdisciplinary nature of human factors research. *Human Factors*, *50*(3), (pp. 506-510).

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology*, *1*(1), (pp. 45-57).

Schaefer, K. E., Billings, D. R., Szalma, J. L., Adams, J. K., Sanders, T. L., Chen, J. Y., & Hancock, P. A. (2014). *A meta-analysis of factors influencing the development of trust in automation: Implications for human-robot interaction* (ARL-TR-6984). Aberdeen Proving Grounds, MD.: Army Research Laboratory (ARL).

Schaefer, K. E., Hancock, P. A., & Chen, J. Y. C. (2015). Autonomy of Future Systems: A meta-analysis of factors influencing the development of trust in automation. *Human Factors.* [in submission]

Schulte, A., Meitinger, C., & Onken, R. (2009). Human factors in the guidance of uninhabited vehicles: oxymoron or tautology? *Cognition, Technology & Work*, *11*(1), 71-86.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & cognition*, *21*(4), 546-556.

Sheridan, T. B. (2000). Function allocation: algorithm, alchemy or apostasy? *International Journal of Human-Computer Studies*, *52*(2), (pp. 203-216).

Sheridan, T. B. (2002). *Humans and automation: System design and research issues*. Santa Monica, CA: John Wiley & Sons, Inc.

Sheridan, T. B. (2006). Supervisory control. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics, Third Edition*, (pp. 1025-1052). Hoboken, N.J.: John Wiley & Sons, Inc.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology*, *3*(2), (pp. 111-122).

Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors*, *37*(1), (pp. 137-148).

Snyder, M., Qu, Z., Chen, J. Y. C., & Barnes, M. (2010). RoboLeader for reconnaissance by a team of robotic vehicles. In *Proceedings of the IEEE International Symposium on Collaborative Technologies and Systems* (Chicago). 522-530.

Stanton, N. A., Chambers, P. R. G., & Piggott, J. (2001). Situational awareness and safety. *Safety Science*, *39*(3), (pp. 189-204).

Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation: The five factor model of personality. *Journal of Experimental Psychology: Applied*, *17*(2), 71.

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: From fixations to object files. *The Quarterly Journal of Experimental Psychology Section A, 58*(5), (pp. 931-960).

Thurstone, L. L. (1951). *An analysis of mechanical aptitude*. Chicago, IL: Psychometric Laboratory, University of Chicago.

Transparent. (n.d.). In Merriam-Webster dictionary online. Retrieved from http://www.merriam-webster.com/dictionary/transparent

Truitt, E. R. (2010). The garden of earthly delights: Mahaut of artois and the automata at Hesdin. In *Medieval Feminist Forum, 46*(1), (pp. 74-79). Society for Medieval Feminist Scholarship.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, (pp. 498-505).

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*(6), (pp. 635-654).

Van Orden, K. F., Jung, T. P., & Makeig, S. (2000). Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological Psychology*, *52*(3), 221-240.

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors, 43*(1), 111-121.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system: The role of aid reliability and reliability disclosure. *Human Factors, 51*, (pp. 281-291).

Wang, H., Lewis, M., Velagapudi, P., Scerri, P., & Sycara, K. (2009, March). How search and its subtasks scale in N robots. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 141-148). ACM.

Wickens, C. D., & Holland, J. G. (2000). *Engineering psychology and human performance* (3rd Ed.). Upper Saddle River, NJ: Prentiss Hall.

Wright, J. L., Chen, J. Y., Quinn, S. A., & Barnes, M. J. (2013). *The effects of level of autonomy on human-agent teaming for multi-robot control and local security maintenance* (ARL-TR-6724). Aberdeen Proving Grounds, MD.: Army Research Laboratory (ARL).

Wright, J. L., Quinn, S. A., Chen, J. Y., & Barnes, M. J. (2014, September). Individual differences in human-agent teaming: An analysis of workload and situation awareness through eye movements. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), (pp. 1410-1414).

Zacks, J., Rypma, B., Gabrieli, J. D. E., Tversky, B., & Glover, G. H. (1999). Imagined transformations of bodies: an fMRI investigation. *Neuropsychologia*, *37*(9), (pp. 1029-1040).