Fall 2019

The Effect of Differing Degrees of Automation and Reliability on Simulated Luggage Screening Performance

Molly M. Liechty
Old Dominion University, mdacliechty@verizon.net

Recommended Citation

Liechty, Molly M.. "The Effect of Differing Degrees of Automation and Reliability on Simulated Luggage Screening Performance" (2019). Doctor of Philosophy (PhD), Dissertation, Psychology, Old Dominion University, DOI: 10.25777/as82-8503
https://digitalcommons.odu.edu/psychology_etds/333

THE EFFECT OF DIFFERING DEGREES OF AUTOMATION AND RELIABILITY ON SIMULATED LUGGAGE SCREENING PERFORMANCE

by

Molly M. Liechty
B.S. May 2000, Purdue University
M.S. May 2011, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

HUMAN FACTORS PSYCHOLOGY

OLD DOMINION UNIVERSITY
December 2019

Approved by:

Yusuke Yamani (Director)

Holly Handley (Member)

Jing Chen (Member)

ABSTRACT

THE EFFECT OF DIFFERING DEGREES OF AUTOMATION AND RELIABILITY ON SIMULATED LUGGAGE SCREENING PERFORMANCE

Molly M. Liechty
Old Dominion University, 2019
Director: Dr. Yusuke Yamani

The present work examined the effects of two types of decision support systems in a simulated luggage screening task: An input aid and an output aid. An input aid supports an operator's information gathering. An output aid supports decision making and action selection. A Time-Accuracy Function (TAF) analysis was applied to isolate processing time from performance asymptote, which conventional performance measures such as sensitivity and response time do not distinguish one from the other. Sixty participants performed a luggage screening task unaided (manual condition), with the assistance of an input aid (spatial aid), and with the assistance of an output aid (decision aid) across different stimulus exposure durations of 250 ms, 500 ms, 1000 ms, 2000 ms, or 3000 ms. Participants were asked to judge the presence of a knife in each of the bags and either "stop" the bag or "pass" the bag. Reliability of the automated aids was 90% in Experiment 1 and 60% in Experiment 2. Experiment 1 showed that sensitivity increased with the assistance of both the input and the output aids as the stimulus exposure duration increased. The performance improvement was greater for the input aid than the output aid condition. Though processing times did not differ across the conditions, asymptotic performance level was higher when participants had the assistance of the input aid compared to the unaided condition. Experiment 2 and cross-experimental analysis demonstrated that the unreliable aids eliminated the benefit of the reliable aids. TAF analysis further showed that, although asymptotic performance can differ, processing times can remain constant regardless of DOA. The results imply that the input aid elevates asymptotic performance without influencing processing times, perhaps allowing operators to crosscheck their decisions within the restricted area of the search field identified by the aid. The present findings are inconsistent with the lumberjack hypothesis (Onnasch et al., 2013) and future research directions are provided.

DEDICATION

I dedicate my dissertation work to my family and friends. A special feeling of gratitude to my loving husband, Dr. Derek Liechty whose positive words and push to keep going when I got tired were the encouragement I needed to get me through. I wish to share my feelings of gratefulness for my children, Amara and Charles who never complained when mom had to work. To my parents and sister, who were always supportive and encouraging.

I also dedicate this dissertation to my friends who have supported me throughout the long process. I will always appreciate all they have done, especially Peggy Kinard, Mary Boswell, Alex Proaps, Rachel Phillips, and Eric Chancey, who proof read, offered advice, and the chance to always bounce ideas off of them.

Finally, I dedicate this work and give special thanks to my best buddies, Evan and Georgia May. They were always there, through the long nights, early mornings, weekends, and holidays. They never left my side and were always ready to work with me no matter what. Both of them will be missed.

ACKNOWLEDGEMENTS

I wish to thank my committee members who were generous with their expertise and time. A special thanks to Dr. Yusuke Yamani, my committee chairman for his tireless hours of reading, editing, and patience during this process. Thank you to Dr. Holly Handley, Dr. Jing Chen, Dr. Jim Bliss, and Dr. Poornima Madhavan for agreeing to mentor me and serve on my committee. I would also like to acknowledge and thank my department for allowing me to conduct my research and providing me with the space and tools needed to complete it.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1: INTRODUCTION

Accurate detection of potential threat objects in passenger luggage is a critical task in aviation security. Agents in the Transportation Security Administration (TSA) must screen each bag for various banned objects such as firearms, flammables, and some daily necessities (TSA, 2017). Typical X-ray images of passenger luggage contain numerous objects. Suspicious or threatening objects are often infrequent targets embedded within high levels of visual clutter, making the visual search perceptually and cognitively challenging (Yamani & McCarley, 2011).

Automated decision support systems are being developed and tested to help improve TSA agents' visual search performance during a luggage screening task (Madhavan & Wiegmann, 2006). For example, Weigman and colleagues (2006) examined operator performance while interacting with different automated decision aids during a luggage screening task. One aid supported operators' decision-making (output aid) while the other assisted with perceptual processing (input aid). At a certain automation reliability level, the input aid provided a spatial cue to the location of a target while the output aid suggested either the presence or absence of the target. The results indicated that operators using an input aid showed greater levels of sensitivity than those using an output aid. However, conventional measures such as sensitivity and response times (RTs) can conflate perceptual processing and performance asymptote. That is, greater sensitivity using an input aid can arise due to better perceptual processing (i.e., enhanced filtering of distractors and information accumulation) and/or asymptotic performance level. Therefore, the present work aimed to explore the perceptual-cognitive mechanisms underlying the use of input and output automated aids during a luggage screening task. A *Time-Accuracy Function* (TAF) analysis using the parameters of intercept, slope, and asymptote were employed to isolate the perceptual processing of an operator's cognitive system from asymptote

(Wickelgren, 1977). Isolating perceptual processing from asymptotic performance allows for better identification of the cognitive locus of effect for input/output aids. This further helps automation designers improve automated decision aids used by future TSA.

CHAPTER 2: LITERATURE REVIEW

2.1 Automated aids

Modern automated aids can perform a variety of tasks including sensing and integrating information, making decisions, and/or executing actions (Parasuraman, Sheridan, & Wickens, 2000). Ideally, automated aids reduce operators' workload and optimize task performance of a human-machine system (Bainbridge, 1983; Bradshaw, Dignum, Jonker, & Sierhuis, 2012; Madhavan & Wiegmann, 2006; Miller & Parasuraman, 2007). Reliable automation can also reduce human errors in a variety of cognitive tasks (Parasuraman & Riley, 1997; Parasuraman & Wickens, 2008). Multiple taxonomies have been proposed to classify the functions an automated aid performs while assisting an operator during a cognitive task (Kaber & Endsley, 1999; Sheridan & Verplank, 1978, Wickens, Li, Santamaria, Sebok, & Sarter, 2010). Parasuraman, Sheridan, and Wickens (2000) distinguished different levels and stages of automation within the framework of the human information-processing model (Wickens, Hollands, Bandury, & Parasuraman, 2015). In their model, human performance is a product of a series of discrete processes from sensing, perceiving, decision-making, and selecting/executing a response, each supported by attention resources. Following these information-processing stages of human operators, automation can be designed to support each stage with information acquisition, information analysis, decision selection, and action implementation. Note that automation can differ in levels of automated capabilities within each stage of automation (Endsley & Kiris, 1995; Wickens et al., 2010).

Degrees of automation (DOA) introduced by Wickens, Li, Santamaria, Sebok, and Sarter (2010) stated that within each stage, automation can also differ by levels. A higher level of automation, within a later information-processing stage such as decision making and response selection denotes a higher “Degree of automation” (Manzey, Reichenback, & Onnasch, 2012). For example, an automated aid that presents an alarm to an operator would be considered a lower DOA than an aid that not only presents an alarm, but also identifies possible targets. Both aids are bringing information to the operator (representing the same stage) but the alarm also identifying the possible target is operating on a higher level within that stage (information gathering), therefore representing a higher DOA (Manzey, Reichenbach, & Onnasch, 2012). Few studies have addressed how DOA impacts an operator, with mixed results (Endsley, 1997; Li, Wickens, Sarter, & Sebock, 2014; Onnasch, Wickens, Li, & Manzey, 2013; Wickens, 2018; Vagia, Transeth, & Fjerdingen, 2016).

One way to further characterize the different functionality of automation in the current context is to distinguish between input and output aids. Input aids support the early stages of human information processing, such as sensation and perception, contributing to lower DOA. Within the framework of Parasuraman et al. (2000), input aids include automation that performs acquisition and analysis of incoming raw data. For example, aircraft radar systems help air traffic controllers (ATCs) to detect incoming aircraft, highlighting their presence through the use of a blinking icon and an auditory alert (Moray, 1997).

On the other hand, output aids support operators’ decision making and response selection or action implementation, supporting higher DOA. For example, the lane keeping function in some autonomous vehicles provides lane keeping assistance to drivers. The car continuously monitors its location in proximity to the lines painted on the road. When the car senses that it is drifting outside of the lane parameters, it will automatically correct itself, moving back within the constraints of the lane (Horowitz & Timmons, 2016). Output aids can function automatically or require that the operator approve the decision for executing an action. Output functions involve decision-making and action selection that occur after information has been acquired and filtered. Crocoll and Coury (1990) reported that when errors occur at this stage, recovery and the operator performance is poor. Operators when assisted by output aids making a recommendation will often follow the advice of the automation (Onnasch et al., 2013). Due to the opaque nature of the aid, operators often have a harder time processing raw information and render their own decisions when the automation fails (Parasuraman & Manzey, 2010).

When automation fails, operator performance can degrade often substantially (Bliss & Gilson, 1998; Manzey, Reichenbach, & Onnasch, 2012; Onnasch et al., 2013). In the event of automation error with an input or output aid, the literature suggests that the DOA can lead to differential impacts on human performance (Crocoll & Coury, 1990; Onnasch et al., 2013; Manzey, Reichenbach, & Onnasch, 2009; 2012). Onnasch and colleagues (2013) performed a meta-analysis comparing operator performance with differing types of automated aids across the continuum of input and output information. They reported that operator vulnerability to automation error was "amplified" when the threshold between input and output automation was crossed (line a on Figure 1). The greater the cognitive function an aid was performing (output aids), the more likely an operator was to be left out of the loop, degrading performance in the event of automation breakdown (Endsley & Kiris, 1995; Sarter & Woods, 1995) (see Figure 1).

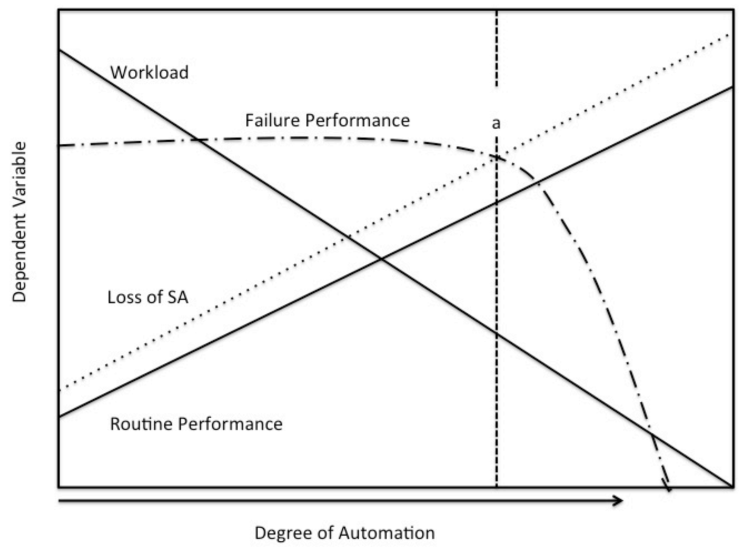

Figure 1. Failure performance drops substantially in the event of automation failure when the degree of automation is above a threshold, denoted as a in the figure (Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2013).

Crocoll and Coury (1990) provide empirical evidence for this point. Their participants performed a target identification task, identifying enemy aircraft. Half of the participants worked with an input aid that alerted them to aircraft status (friend or foe). The remaining participants worked with an output aid that told them the appropriate action to execute (fire or hold fire). Participants with the assistance of an automated aid, regardless of input or output, showed shorter RTs and greater accuracy than those without an aid. Crocoll and Coury stated that operator performance was distinctly different between the input and output aid groups when the aids failed. This performance decline was more evident with the output aid (fire or hold fire) when it was incorrect compared to participant performance when the input aid (friend or foe) was incorrect. Results from Crocoll and Coury implied that the functions of input and output aids have differential impacts on operator performance when automation failures occur (see also Manzey, Reichenbach, & Onnasch, 2009). Onnasch et al., (2013) proposed a lumberjack hypothesis that describes this relationship between differing automated aids and the effects on human performance. That is, an input aid provides operators with pre-processed information for their decisions. Because operators still have to perform decision making and action selection processes, impact of automation failure is limited to information acquisition and integration. Importantly, even when working with input aids, operators are expected to be in the loop (Endsley & Kiris, 1995), actively seeking information cues that guide their decision making. However, when provided with an output aid, operators can be left out of the loop because tasks with the assistance of output aids do not necessarily require them to actively seek further information for their decision making. Not only do they make the decision for the operator or assist in the decision making, limiting the operator's input into the actual decision being made but often output aids are by their nature opaque, lacking transparency (Lyons, 2013).

2.1.1 Transparency

Automation transparency means the degree to which an operator can access information that is processed by an automated aid (Lyons, 2013). An output aid may thus possess low levels of automation transparency if the aid does not reveal *how* it preprocesses, analyzes, and integrates information to derive its decisional recommendation. The output aid used in the current experiments (see below for the specific characteristics of the output aid) retains low automation transparency because operators are able to view only its decisional recommendations but not any of the preprocessing of available information to the aid. On the other hand, the input aid used in the current experiments maintains a higher level of transparency, compared to that of the output aid, because it shows which spatial area that the aid considers critical for operators' decision making.

Previous studies (indicate that operators interact with automated systems of varying levels of transparency differently (Lee, 2012; Lyons, 2013; Sarter and Woods, 1995). Operators working with highly transparent systems can remain more aware of how the system is functioning and therefore crosscheck the system's reliability. Remaining aware of how a system is using information allows the operator to better mitigate potential automation failures if and when they occur (Dao et al., 2009; Visser, Cohen, Freedy, & Parasuraman, 2014). But, operators working with systems with low transparency may have a diminished awareness of the system's ability to analyze information, because an operator is less able to predict when and how the automation malfunctions or needs the human operator to take over in the event of an automation failure (Bahner, Hüper, & Manzey, 2008; Lyons, 2013; Parasuraman et al., 2000; Parasuraman & Manzey, 2010). The ideal system should keep an operator informed of how and why the system is operating, while keeping the operator's workload low (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Lyons, 2013; Sarter and Woods, 1995).

In the context of luggage screening by the TSA, operators are required to scan bags for banned objects, a form of standard visual search (Wickens & McCarley, 2008). Along with knowledge of an automated aid's impact on operator performance, modern theories of visual search may lend empirical support for further development of automation that could effectively improve visual search performance in a luggage screening task. The next section considers this point.

2.2 Visual Search

Visual search is a behavior to look for an object of interest (target) among irrelevant objects (distractors), when the location and the presence of the target are unknown *a priori* (Wolfe, 1994). Ample empirical evidence suggests the number of display objects (set size) and elementary object features affect search performance (Treisman & Gelade, 1980; Treisman &

Gormican, 1988; Wolfe & Bennett, 1997; Wolfe, 1994; Quinlan, 2003; Wolfe & Horowitz, 2004). To characterize human visual search behavior, several theories have been proposed. One theory of visual search and object perception is Feature Integration Theory (FIT; Treisman & Gelade, 1980; Quinlan, 2003). FIT assumes that the visual system preattentively processes object features in parallel, followed by a serial attentive processing where focal attention integrates multiple features of an object. Consider a visual search display where a target is a red horizontal bar and distractors are blue or red vertical bars (Figure 2a). In this type of *feature search*, where a target is defined by a single feature dimension (e.g., orientation), visual search is often efficient. Now consider a display where a target is again a red horizontal bar but now distractors are either red vertical bars or blue horizontal bars (Figure 2b). In *conjunction search,* where more than one feature defines a target (color and orientation in this example), the visual system requires focal attention to *bind* the features for object perception. In this way, visual search is effortless in feature search but effortful in conjunction search.

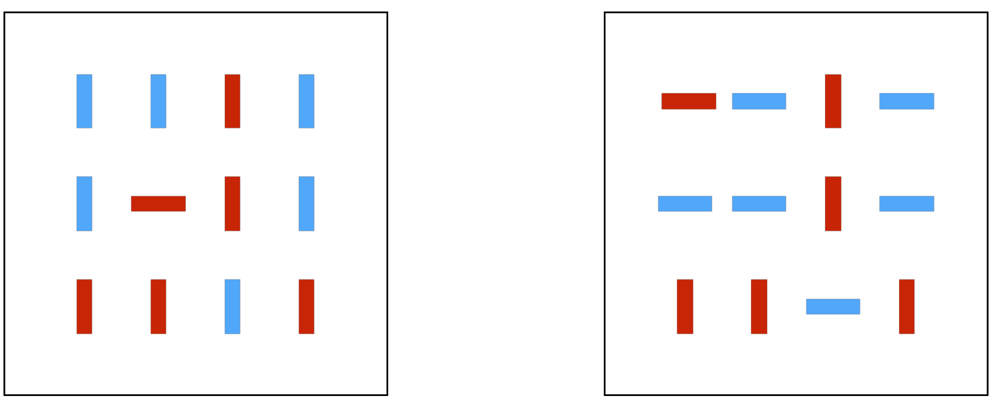

Figure 2. Examples of feature search (left) and conjunction search (right).

However, Wolfe, Cave, and Fanzel (1989) found that visual search was actually more efficient when more than two features defined a target. Because FIT predicts less efficient search for a target defined by multiple features, their result is inconsistent with predictions of the original

FIT. Wolfe (1998) later proposed the Guided Search (GS) model that assumes both bottom-up feature maps and top-down modulation of the activation map on which focal attention is guided. Similar with FIT, feature maps are created based on the output of sensory input channels in a parallel manner. For example, an individual might be searching for a green book. Therefore, items possessing the color green or the typical rectangular shape of a book would generate higher activation. The individual feature maps are combined to generate a master activation map guiding where visual attention will be deployed and directed to spatial locations of the highest to lowest activation. In other words, the GS model combines both bottom-up and top-down processes in visual search and offers insights to how object features "guide" attention in complex search fields (Wickens & McCarley, 2008).

Considering typical X-ray images of passenger luggage, the number of objects, overlap of the objects, and the number of banned items to search for can quickly overwhelm an operator's attention set during a visual search task. For example, salient features of a banned object, such as the edge of a knife, may be obscured by another overlapping object, which in turn reduces activation levels of the knife on the activation map. Further, when an operator is looking for a number of target objects differing in their target features (e.g., banned objects), top-down commands may not effectively activate relevant feature maps because features of the target objects can be diverse.

Previous research on visual search performance in the luggage screening domain have focused on a participant's ability to find targets among overlapping objects, varying categories of items, and varying degrees of training before screening bags (Evans, 2005; Fiore, Scielzo, & Jentshch, 2004; Smith, Redford, Gent, & Washburn, 2005). Consistent with the theoretical predictions on visual search performance, such research has confirmed that luggage screening is a challenging task and that screeners often fail to detect threat objects in an X-ray image of passenger luggage. According to Homeland security, for example, when tested by undercover inspectors in 2017, TSA agents failed to detect threats more than half the time with speculation that the failure rate was close to 95% (Kerley & Cook, 2017). With automation, if designed properly, operator performance in the difficult visual search task may become more efficient and the screener aided by an automated decision support more adept at locating suspicious targets more quickly than the screeners alone.

One way to use automation to improve screener performance is to have an automated aid direct visual attention through spatial cueing (Posner, Snyder, & Davidson, 1980; Christensen & Estepp, 2013). Wiegmann and his colleagues (2006) examined effects of spatial cueing on luggage screening performance. Specifically, they compared automated aids that provided either decisional cueing that consisted of a text message that would recommend an action ("stop" the bag or "pass" the bag), or spatial cueing that highlighted a spatial area of potential threat objects to exogenously guide operators' attention. The results indicated that operators' sensitivity was greater in the aided condition than the unaided condition. More importantly, the spatial cueing condition produced greater levels of sensitivity than the decisional cueing condition. As the spatial cue was of lower DOA than the decisional cue, this is contrary to results found in the meta-analysis (Onnasch et al., 2013).

This project aimed to explore the underlying information-processing mechanisms the differing DOAs impacted on operator performance during a luggage screening task. Presumably, the spatial cueing aid supports sensory and perceptual processes (input aid) while the decisional aid supports decision making (output aid). Time-Accuracy Function (TAF) analysis was applied to isolate operator *perceptual processing* from operator *performance asymptote*. As in previous studies (e.g., Onnasch et al., 2013; Verhaeghen, 2000; Wickelgren, 1977; Yamani & McCarley, 2011), operator sensitivity and RTs have been conflated. The input aid, for example, might have improved operators' performance by raising asymptote performance level without actually affecting the operator's perceptual-cognitive processes such as filtering of distractor and information accumulation or vice versa. It is possible that the output aid did not improve detection performance as much as the input aid because the output aid affected system dynamics. Operators needed more time to accumulate information to reach the asymptote performance level in the input aid condition. Unfortunately, analysis of mean RTs and performance accuracy alone do not allow distinguishing perceptual processing from asymptotic performance level; thus, the conventional measures of human performance conflate these characteristics (e.g., Crocoll and Coury, 1990; Wiegmann et al., 2006). The next section introduces the TAF analysis and its theoretical background.

2.3 Time-Accuracy Function Analysis

One may plot accuracy as a function of stimulus exposure duration, and the function can be characterized by a *delayed exponential function* (Verhaeghen, 2000). Formally, a delayed exponential function is defined as:

$$p = \begin{cases} 0 & t \leq I \\ A\left(1 - e^{\frac{(I-t)}{R}}\right) & t > I \end{cases} \tag{1}$$

where p represents performance accuracy such as signal detection sensitivity *(d')*, t represents stimulus exposure duration, and *I, R,* and *A* are the parameters of the delayed exponential function. Figure 3 presents an example of a delayed exponential function with specific parameters.

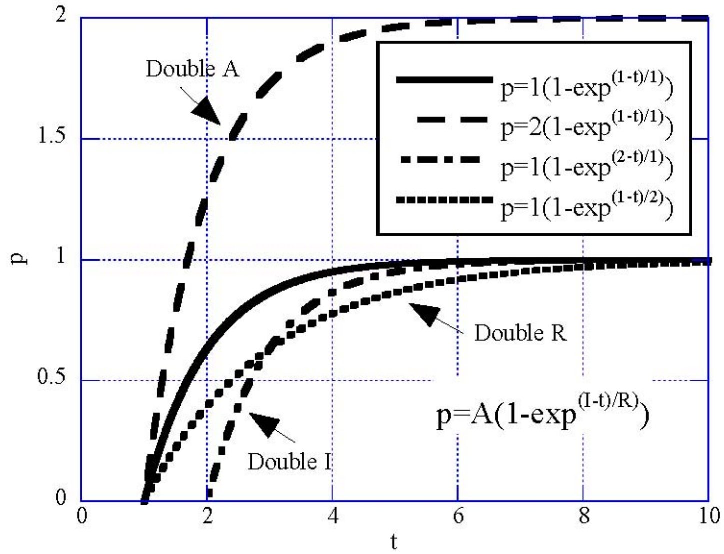

Figure 3. TAFs with different hypothetical parameter values. Note that performance and time variables are arbitrary in this figure.

The *I* parameter denotes an *intercept* that defines the time point at which performance measure exceeds the chance level performance. The *A* parameter represents the asymptotic performance level that participants achieve if unlimited time is allowed to perform a task. Finally, the *R* parameter refers to the rate at which performance grows from zero to the asymptotic performance level. In the context of human information processing, these three distinct parameters imply unique aspects of human performance. In a luggage screening task, the intercept reflects the minimal time that a participant needs to perform beyond the chance level such as sensory registration of raw data and early perceptual processes. The rate reflects the efficiency of information processing to make a correct response.

The hypothetical TAFs (in Figure 3) demonstrate what can occur as parameters *I*, *R*, and *A* change. The higher the value of *I,* the more time it takes for an operator to gather useful information to make an accurate response. Parameter *R* describes the degree of curvature of the function. The smaller the value of *R*, the steeper the curve, indicating a smaller amount of processing time needed to reach asymptote. Finally, the higher the parameter of A, the greater the performance accuracy of participants for the luggage screening task. Mean processing time is defined as:

$$Processing\ Time = I + R \tag{2}$$

Therefore, *I* and *R* parameters capture the *dynamics* of operator *perceptual processing* (Meyer et al., 1988; Yamani & McCarley, 2011).

The TAF analysis can isolate perceptual processing from performance asymptote unlike the conventional performance measures such as performance accuracy and mean RTs. That is, two distinct cognitive systems may have identical asymptotic performance levels but, due to differences in perceptual processing, the two systems may show distinctly different performance outcomes measured on mean RTs or performance accuracy.

CHAPTER 3: SUMMARY AND HYPOTHESES

The purpose of this study was to test the lumberjack hypothesis (Sebok & Wickens, 2017, Onnasch et al., 2014) in the context of luggage screening. Previous research demonstrates that human performance in a cognitive task improves with the assistance of an automated aid (Parasuraman & Riley, 1997). The lumberjack hypothesis poses that routine performance will increase dependent upon the DOA assisting the operator. The greater the DOA, the greater the performance an operator will have compared to the assistance of a lower DOA, if the automated system is reliable (Onnasch et al., 2013). As automation fails, according to the lumberjack hypothesis, the higher DOA will have a greater impact on operator performance compared to the lower DOA (Manzey, Reichenbach, & Onnasch, 2012), resulting in a greater performance decrement. However, results have been mixed regarding DOA and the improvement in performance or performance decrement based on it. For example, Crocoll and Coury (1990) found shorter RTs in an aircraft identification task when operators were assisted by an output aid (higher DOA) that issued a fire/hold fire recommendation than when an input aid (lower DOA) presented the friend/foe status of a target. In contrast, Wiegmann and colleagues (2006) found that, during a simulated luggage screening task, an input aid providing a spatial cue to a visual area of a potential threat produced greater sensitivity levels and faster RT, than an output aid providing a decisional cue. One potential account for this discrepancy in the empirical data was that the input and output aid affected processing time and asymptote, respectively. When assisted by the input aid, operators could attend to the highlighted spatial area to improve their judgment. Conversely, the output aid could not necessarily guide their spatial attention and only provided a recommendation for their response selection. Additionally, Wiegmann et al.'s (2006) data reflected performance levels at a single time point. At this time point, the output aid condition did not reach the participant's asymptote level, resulting in poorer performance than the input aid condition. In Experiment 1, participants in the luggage screening task were asked to perform a visual search task assisted by input and output aids with high reliability (90%). Based on the lumberjack hypothesis, the following hypotheses were generated for Experiment 1.

Hypothesis 1. Participants in the output aid condition will perform better than the input aid condition, followed by the manual condition, in the luggage screening task. This is supported by the lumberjack hypothesis (Onnasch et al., 2013).

Hypothesis 1.a. Participants' sensitivity in the output aid condition will be greater than the input aid condition, followed by the manual condition.

Hypothesis 1.b. Participants' response bias in the output aid condition will be lower than the input aid condition, followed by the manual condition (Meyer et al., 2013).

Hypothesis 1.c. Participants' mean RTs in the output aid condition will be shorter than the input aid condition, followed by the manual condition.

Hypothesis 2. Participants will perform the luggage screening task progressively better as the exposure duration increases.

Hypothesis 2.a. Participants' sensitivity will increase as the exposure duration increases.

Hypothesis 2.b. Participants mean RTs will decrease as the exposure duration increases.

Hypothesis 3. The input aid condition will produce shorter processing time than the manual condition.

Hypothesis 4. The output aid condition will produce shorter processing time and higher performance amplitude than the input aid and manual conditions.

Crocoll and Coury (1990) found that errors made by the output aid with greater DOA impaired operators' decision-making performance more pronouncedly than those by the input aid with lower DOA, when reliability of the aids was low (e.g., Onnasch et al., 2013). One account for the results was that errors by the output aid affected both processing time and asymptote while those by the input aid affected only processing time. In Experiment 2, we repeated Experiment 1 except that reliability of the aid was reduced to 60%. Following this account, the following hypotheses were generated for Experiment 2.

Hypothesis 5. Participants in the input aid condition will perform in the luggage screening task better than the output aid condition, followed by the manual condition. This is supported by the lumberjack hypothesis (Onnasch et al., 2013).

Hypothesis 5.a. Participants' sensitivity in the output aid condition will be lower than the input aid condition, followed by the manual condition.

Hypothesis 5.b. Participants' RTs in the output aid condition will be longer than the input aid condition, followed by the manual condition.

Hypothesis 6. Participants will perform the luggage screening task progressively better as the exposure duration increases.

Hypothesis 6.a. Participants' sensitivity will increase as the exposure duration increases.

Hypothesis 6.b. Participants mean RTs will decrease as the exposure duration increases.

Hypothesis 7. The input aid condition will produce longer processing time than the manual condition.

Hypothesis 8. The output aid condition will produce longer processing time and lower performance amplitude than the input aid and manual conditions.

CHAPTER 4: EXPERIMENT 1

Experiment 1 examined the effect of DOA on visual performance in a luggage screening task and tested Hypotheses 1-4.The participants were asked to judge whether the X-ray image of each luggage contained a predefined target—a knife at a random orientation—and stimulus exposure duration was manipulated (250 ms, 500 ms, 1,000 ms, 2,000 ms, 3,000 ms). The input aid provided a spatial cue selecting a visual area of a potential threat object. The output aid offered a decisional cue, a red box encompassing the image telling the participant to stop the bag or a green box telling the participant to pass the bag. The TAF analysis was applied to provide additional insights into the perceptual-cognitive mechanisms underlying visual performance of operators while interacting with different types of automated aids. The reliability of the aids was 90%.

4.1. Methodology

4.1.1. Participants

Thirty participants (22 females; mean age = 20 years, *SD* = 4.73 years; mean corrected far acuity = 20/21.36, *SD* = 6.57; mean corrected near acuity = 20/23.63, *SD* = 7.26) were recruited from the community of Old Dominion University. All participants were screened for normal color perception using the Ishihara color blindness test (1989).

4.1.2. Apparatus

Stimuli were presented on a Samsung T24C550 23.6” LED monitor with 1920 x 1080 resolution. The experiment was controlled by a Dell Optiplex 9020 running PsychoPy (v1.8) on Windows 7. Participants were seated at a distance of approximately 57 cm from the monitor. The experiment took place in a quiet room with dim lighting.

4.1.3. Stimuli

4.1.3.1. Target

The target was a standard military-styled knife obtained directly from the TSA. It was included in a set of knives encountered by the TSA in previous screenings and training (Madhavan, 2005). The image of the knife subtended 5.48° x 2.02°. The target appeared randomly within half of the X-ray luggage images (50% target rate). One hundred unique luggage images were used, both for the target present and target absent trials. The images were resized in Adobe PhotoShop, keeping the aspect ratio constant, such that the maximum image dimension angle was 15.95° (mean horizontal visual angle = 14.62°, *SD* = 1.34°; mean vertical visual angle = 15.64°, SD = 0.70°). The knife was presented in one of four varying orientations (0°, 90°, 180°, and 270°). PsychoPy 3 was used for rendering of the stimuli including superimposition of the image of the knife onto the luggage images. Transparency of the knife was set to 50% to blend onto the existing luggage images. The position of the center of the knife was determined by picking a random pixel location, measured from the center of the image, between -2.76° and 2.76° in both the horizontal and vertical directions (see Figure 4).

Figure 4. Target-present trial, with knife located in lower left quadrant.

4.1.3.2. Automated Aids

During the luggage screening task, participants received the assistance of either an input or an output aid when an automated aid was available.

4.1.3.2.1. Input Aid

The input aid was designed to direct attention to a suspicious area of the X-rayed luggage image. An input aid is an aid that acquires and/or filters information for the operator. The input aid's operations occur before decision making (Parasuraman et al., 2000). The input aid here provided a spatial cue designed to constrain the search space for the operator (Posner, 1980; Posner, Snyder, & Davidson, 1980) consisting of a yellow circle that highlighted portions of the X-rayed luggage image (see Figure 5). The radius of the spatial cue was 3.98°. The aid was centered on the target in the target-present trials. In the target-absent trials, the spatial cue appeared in a randomly chosen location in the X-ray luggage images. The aid was 90% reliable where it highlighted 90% of knives in the target present trials and incorrectly highlighting areas that did not encompass the knife in the remaining 10% of the target present trials). The spatial cue appeared in every image. The purpose of the aid appearing in every image was to ensure that it maintained its input status without influencing their decision-making stage of information processing. In other studies that have used a spatial cue (e.g., Weigmann, 2007; Madhavan & Weigmann, 2006), the cue appeared only when a target was present. By showing a decisional cue only when a target was present, the aid implies that a target is present, potentially biasing participants decision making (Deppe et al., 2005; Van't Wout & Sanfy, 2008). Regardless of the presence of the target, the input aid detected the places in the bag "most likely to contain a weapon." More specifically, the current participants were told that the input aid would not identify a knife, but instead provide cues as to areas where it was more likely to reside, encouraging them to search the area circled themselves.

Figure 5. Input Aid. The aid provides either correct (left) or incorrect (right) spatial cues.

4.1.3.2.2. Output Aid

The output automated aid assisted participants with decision making. This aid consisted of a simple colored frame encasing the luggage image, green to pass the bag and red if the decision was to stop it. It provided the participant with the decision cue indicating that they should "stop" the bag or "pass" it (see Figure 6). The output aid box (green & red) was set at 17.59° x 17.59° (to surround the entire image). An output aid box was presented in all the trials in the output aid condition.

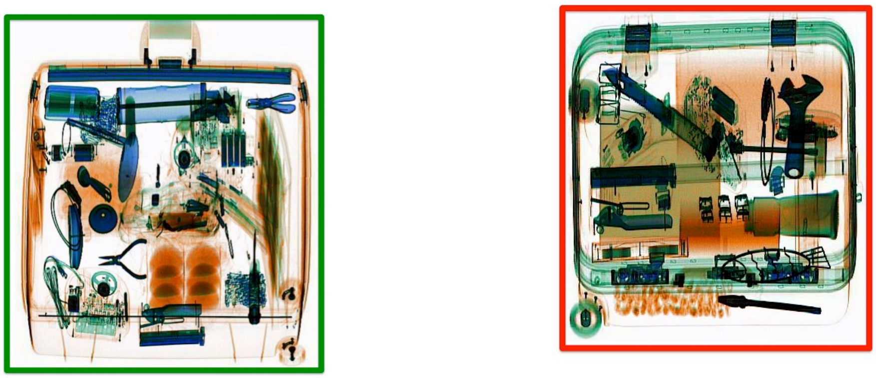

Figure 6. Output Automated Aids. The aid provides either "pass" the bag (left) or "stop" the bag (right) decisional cues.

4.1.3.2.3. Aid Reliability

Aid reliability levels were set at 90%. Participants were told ahead of time the reliability level of the aid they were working with.

4.1.4. Procedure

This study was approved by the Institutional Review Board of Old Dominion University. After completing the informed consent form and a short demographic questionnaire, participants were told that they would play the role of an airport luggage screener and that they must visually scan X-rayed luggage images to decide if the bag contained a target (knife). The instruction included a visual image of the target knife. They were asked to indicate whether they wanted to "stop" the bag (target present) or "pass" the bag (target absent). Participants completed 3 experimental blocks (the unaided, the input aid, and the output aid conditions) of 200 experimental trials each.

Each block began with 10 practice trials to help familiarize them with the task and the aids. Orders of the experimental blocks were counterbalanced across participants.

Each trial began with a 500 ms blank screen, followed by the imperative stimulus display presented for various stimulus exposure durations (250 ms, 500 ms, 1000 ms, 2000 ms, or 3000 ms), and then a mask stimulus for 500 ms. The mask stimulus was created by superimposing five randomly chosen X-rayed luggage images on top of one another. Immediately after the mask disappeared, participants were asked to make a forced-choice to either "stop" the bag or "pass" the bag by clicking the left or the right mouse button, respectively. After participants made their decision, a feedback display was presented for 750 ms presenting "+" for a correct response or "X" for an incorrect response. The next trial began automatically. The experiment lasted approximately 60 minutes. Participants were assigned credit, debriefed, and exited the lab.

4.1.4.1. Default Bayesian test.

Bayesian analyses were employed instead of conventional null-hypothesis significance tests (NHSTs). A Bayes factor is a likelihood ratio of the data favoring one statistical model to another. Bayes factors directly indicate a relative likelihood that data come from one statistical model over another. For example, when comparing the effect of using an automated aid on performance, a Bayes factor of 100 indicates that the observed data are 100 times more likely to come from a model that assumes the effect of the aid than a model that excludes the effect. On the other hand, a Bayes factor of .001 indicates that the data are 100 times more likely to come from the null model compared to the model with the effect of the aid. The nomemclature for describing Bayes factors comes from Jeffreys (1961; see Wetzels, Matzke, Lee, Rouder, Iverson, & Wagenmakers, 2011). Table 1 presents the specific descriptive terminologies.

Table 1. Bayes Factors. Taken from Wetzel et al., (2011).

Statistic	Interpretation for H_A	Statistic	Interpretation for H_0
>100	Decisive evidence for H_A	<.01	Decisive evidence for H_0
30-100	Very strong evidence for H_A	.01-.03	Very strong evidence for H_0
10-30	Strong evidence for H_A	.03-.1	Strong evidence for H_0
3-10	Substantial evidence for H_A	.1-.3	Substantial evidence for H_0
1-3	Anecdotal evidence for H_A	.3-1	Anecdotal evidence for H_0
1	No evidence	1	No evidence

4.2. Results

4.2.1. Data analyses

Values for *d', c,* and RTs were submitted to separate 3 x 5 repeated-measures Bayesian analyses with Condition (manual, input aid, vs. output aid) and Exposure Duration (250 ms, 500 ms, 1000 ms, 2000 ms, vs. 3000 ms) as within-subject factors. To explore the order effect, omnibus 3 x 5 x 6 mixed Bayesian analyses involving Order as an additional between-subject factor for the three dependent variables was performed. Data did not indicate substantial evidence for any of the effects involving Order as a factor (all $B_{10} < .42$) and, therefore, Order was excluded from the following analyses. The Greenhouse-Geisser correction was applied for correcting degrees of freedom where the assumption of sphericity was violated.

The delayed exponential function (Eq. 1) was fit to the data of each participant using the nonlinear least squares function, nls2, in R (R Core Team, 2014). Functions estimated using sum of squares residuals (RSS) accounted for 79.6% of the variance in the data. Each curve fit for every participant was examined. Estimated parameter values for processing time and asymptote were then submitted to separate one-way Bayesian analyses with Condition as the within-subject factor. The nls2 function was not able to calculate the parameters of the curve for the delayed exponential function for five participants, who were subsequently excluded from the analysis reported below. The results of the data with and without such participants did not differ.

4.2.1.1. Sensitivity

Figure 7 presents mean *d'* values as a function of Exposure Durations. As expected, data indicated decisive evidence for higher sensitivity as Exposure Durations increased, $F(4, 96) = 109.72$, $\eta^2_G = .40$, $B_{10} = 3.83 \times 10^{53}$, supporting Hypothesis 2.a. The data pattern did not substantially differ across the aid conditions, $F(8, 192) = 1.68$, $\eta^2_G = .02$, $B_{10} = 0.08$. The data indicated decisive evidence for the effect of Condition, $F(1.54, 36.96) = 11.93$, $\eta^2_G = .08$, $B_{10} = 1.51 \times 10^8$, partially supporting Hypothesis 1.a. Follow-up t-tests revealed that the sensitivity values in the manual condition were decisively lower than the input aid condition, paired-samples $t(24) = 5.12$, $B_{10} = 7.78 \times 10^2$, and substantially lower than the output aid, paired-samples $t(24) = 3.08$, $B_{10} = 8.52$. However, the sensitivity values did not differ between the two aided conditions, paired-samples $t(24) = .86$, $B_{10} = 0.30$.

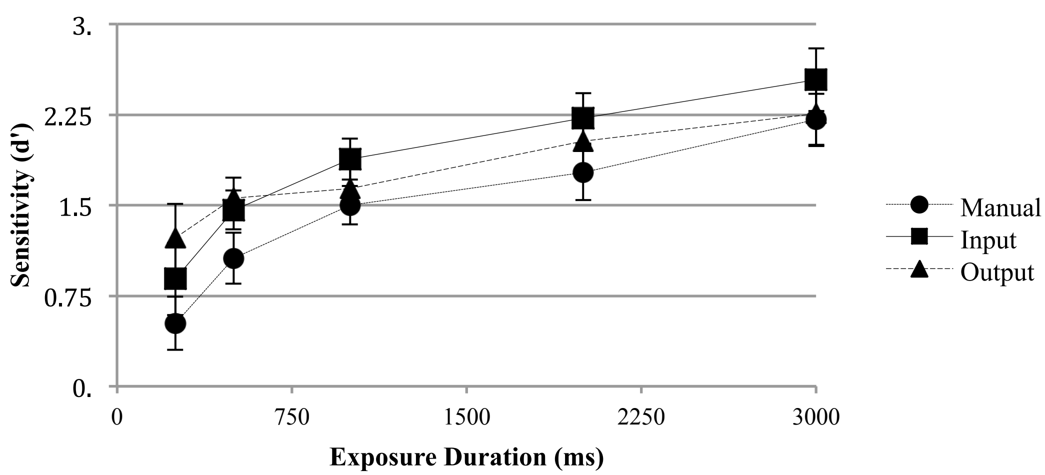

Figure 7. Mean *d'* scores across Exposure Durations for the three conditions. (Error bars represent 95% within-subjects confidence intervals).

4.2.1.2. Response Bias

The response bias of the participants in the three conditions across Exposure Duration is presented in Figure 8. Data did not provide substantial evidence for the effect of Condition on response bias, $F(2, 48) = 1.43$, $\eta^2_G = .01$, $B_{10} = 0.36$, failing to provide support for Hypothesis 1.b. However, unexpectedly, data demonstrated strong evidence that participants' responses became more conservative as Exposure Duration increased, $F(2.11, 50.77) = 3.43$, $\eta^2_G = .02$, $B_{10} = 16.21$. Furthermore, this main effect of Exposure Duration substantially interacted with Condition, $F(8, 192) = 5.04$, $\eta^2_G = .03$, $B_{10} = 8.40$. The interaction effect resulted from decisive and substantial differences between the manual and the input conditions at Exposure Durations of 1,000 ms, paired-samples $t(24) = 4.29$, $B_{10} = 118.44$, and 2,000 ms, paired-samples $t(24) = 3.29$, $B_{10} = 13.06$, partially supporting Hypothesis 1.b.

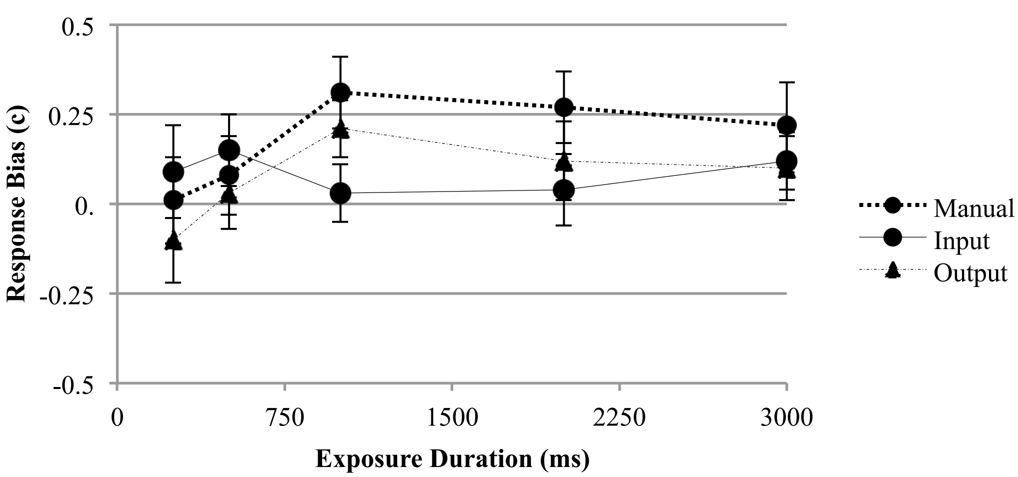

Figure 8. Mean response bias scores across Exposure Durations for the three conditions. (Error bars represent 95% within-subjects confidence intervals).

4.2.1.3. Mean RTs

Mean RTs as a function of Exposure Duration for the three conditions is shown in Figure 9. Data provided substantial evidence that RTs became progressively shorter as Exposure Duration increased, $F(2.76, 66.47) = 7.29$, $\eta^2_G = .02$, $B_{10} = 3.60$, supporting Hypothesis 2.b. The data also showed strong evidence for the effect of Condition, $F(2, 48) = 1.52$, $\eta^2_G = .01$, $B_{10} = 14.06$, supporting Hypothesis 1.c., and the effect was statistically comparable across different Exposure Durations, $F(5.02, 120.53) = 1.40$, $\eta^2_G = .01$, $B_{10} = 1.46 \times 10^{-2}$.

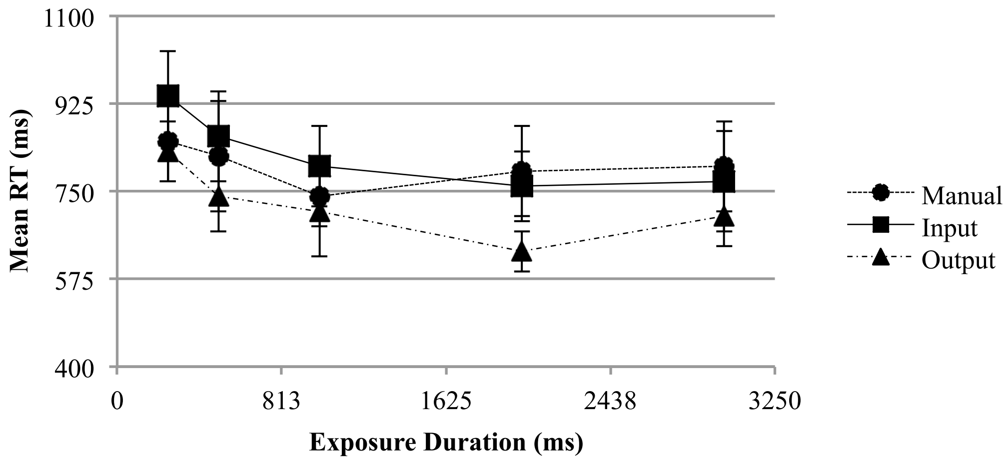

Figure 9. Mean RTs across Exposure Duration for the three conditions. (Error bars represent 95% within-subjects confidence intervals).

4.2.2. Time-Accuracy Function Analysis

4.2.2.1. Processing Time

Mean processing times for the separate conditions is shown in Figure 10. Data gave substantial evidence against the effect of Condition, $F(2, 48) = .07$, $\eta^2_G = .002$, $B_{10} = 0.16$, suggesting that the use of different types of automated aids did not influence processing time in the current detection task, failing to support Hypothesis 3 and 4.

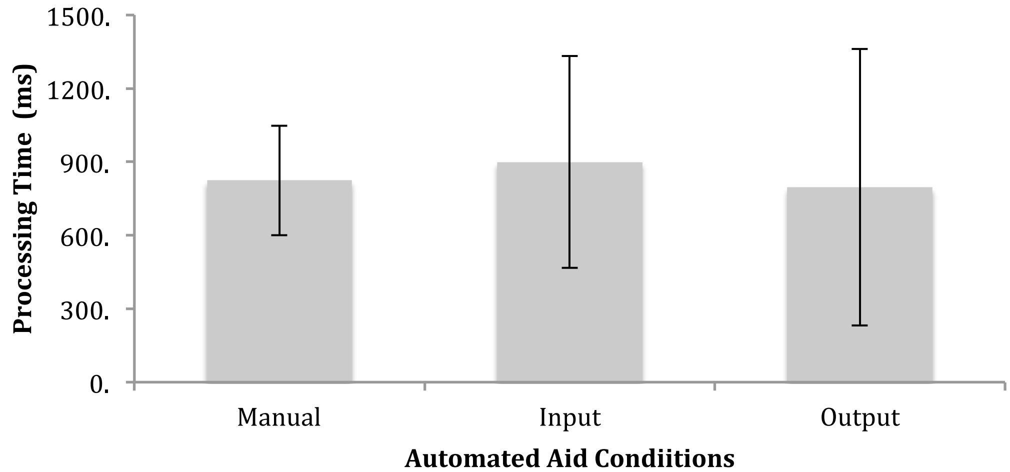

Figure 10. Mean processing time scores across conditions. (Error bars represent 95% within-subjects confidence intervals).

4.2.2.2. Asymptote

Mean asymptote values across the automated aid conditions is shown in Figure 11. The asymptote was numerically larger in the input condition than the output and manual conditions (M = 2.98 for the input condition vs M = 2.39 for the manual condition and M = 2.62 for the output condition), but the data fell short of producing substantial evidence for the effect of Condition, $F(2, 48) = 3.39$, $\eta^2_G = .07$, $B_{10} = 1.81$. To further explore the effect of the automated aids on asymptote, a series of follow-up t-tests were performed. The data demonstrated substantial evidence for improvement from the manual to the input condition, paired-samples $t(24) = 2.84$, $B_{10} = 5.17$, but not for a difference between the manual and output aid, paired-sample $t(24) = 1.07$, $B_{10} = 0.35$, and the input and output aid, paired-samples $t(24) = 1.39$, $B_{10} = 0.50$. Though there exist evidence that asymptote performance varied between the conditions, the data pattern was the opposite of that predicted by Hypothesis 4. Therefore, the results failed to support Hypothesis 4.

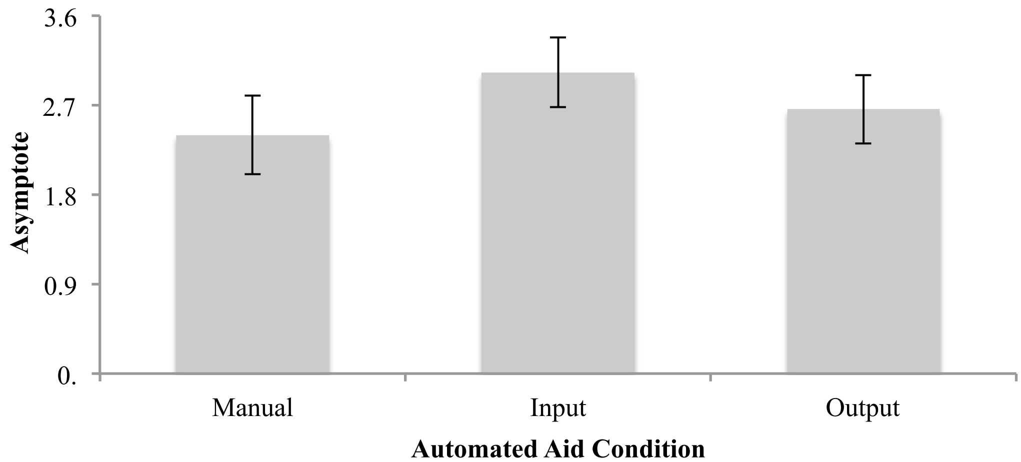

Figure 11. Asymptote scores across conditions. (Error bars represent 95% within-subjects confidence intervals).

4.3. Discussion

Participant performance with the input and output aids at a relatively high reliability level of 90% produced comparable benefit to operators' visual performance in the luggage screening task. This echoes the common findings that automated decision aids can generally improve human performance in applied visual tasks (e.g., Crocoll & Coury, 1990; Madhavan & Wiegmann, 2006; Parasuramn & Riley, 1997). The aids also produced similar improvement in visual search performance across the different levels of exposure durations. The greater performance improvement using the input aid compared to the output aid potentially argues against the lumberjack hypothesis that a higher DOA will yield greater performance benefits (Onnasch et al., 2013).

According to theories of spatial attention and object perception, spatial cueing enhances processing of objects within the spotlight of attention among surrounding objects (Mueller & Rabbitt, 1989, Posner, 1980; Posner & Cohen, 1984; Treisman & Gelade, 1980; Quinlan, 2003). By preferentially processing objects within the highlighted area, operators with the input aid may have achieved slightly better performance compared to their performance with the output aid. According to the theories, an operator's attention was focused within the area that the input aid highlighted, therefore effectively reducing the number of distractors. This allowed operators to better localize the target, making identification faster than when attention is not being directed to a specific area of a search field (Madhavan & Wiegmann, 2006; Muller & Rabbit, 1989).

Another potential reason for the increased sensitivity of the input aid is that participants underutilized the output aid, relying less on the aid's recommendation and more on their own search of the visual area (Dzindolet, Pierce, Beck, & Dawe, 2002; Parasuraman & Riley, 1997). In the present study, the X-rayed luggage images remained on-screen alongside the aid for the entirety of exposure durations. By retaining the visual stimulus (raw data) alongside the aid, participants had the opportunity to cross-check the accuracy of the aid. In Wiegmann et al., (2006), the decisional cue did not appear to help participants in their search task but the cue only consisted of a simple text message appearing following the offset of the X-ray image. Further research should examine whether simultaneous presentation of the raw data and the aid's recommendation influences visual search performance in the luggage screening task compared to sequential presentation.

Asymptote levels further support the benefit of the input aid. The results indicated that the input aid produced greater values of asymptote than the output aid. This implies that the spatial cue, although it did not explicitly recommend a decision, improved visual performance above and beyond what the input aid was designed for. Operator performance with the output aid remained lower than the input aid. Asymptote by definition means the maximum performance level that operators can achieve if unlimited time is allowed to perform the task. If more time were available to the operators, their sensitivity could have potentially reached beyond d' = 3, which is well above the aid's sensitivity of d' = 2.56.

Processing times did not substantially differ by DOA. I hypothesized that processing time would be greater for the input aid than the output aid because it provided the operator with support for earlier stages of information processing (Parasuraman et al., 2000). Even though their search field was presumably narrowed due to the spatial cueing, evidence suggests that operators aided by the input aid processed information for decision making at a similar rate of operators aided by the output aid and those unaided. Results imply that automated aids in the simulated luggage screening task, regardless of the automation modes (input vs. output aid), do not affect the rate at which an operator will reach asymptote.

Finally, the response bias data revealed that participants became less conservative with the input aid vs. the output aid than when no aid was available, especially at 1,000 ms and 2,000 ms following the onset of the display. This shift of response bias may indicate the participants' increased compliance. In the compliance-reliance framework of automation trust (Meyer, 2004; Dixon & Wickens, 2004; Dixon, Wickens, & McCarley, 2007; Chancey, Bliss, Yamani, & Handley, 2017), compliance refers to the extent to which an operator agrees with a machine's decision when a decisional cue is present, while reliance refers to the extent to which an operator agrees with a machine's decision when a cue is absent. More recently, Meyer and colleagues (2013) defined compliance as a difference of response biases between no aid condition and aided condition, or formally,

$$Compliance = C_{no\ cue} - C_{Alert} \tag{3}$$

Where $C_{\text{no cue}}$ is response bias for no aid condition and C_{Alert} is that for aided condition. In the current context, compliance for the input aid is roughly .3 at exposure durations of 1,000 ms and

2000 ms while that for the output aid is roughly .2 and .1. This observation indicates that operators were more likely to comply with the input aid's recommendations than the output aid.

The overall results highlight the benefit of the automated aids at different levels of automation and further demonstrate the fact that the input aid can elevate performance asymptote more than the output aid, exceeding the aid's own sensitivity. Thus, the use of the input aid or a spatial cue may be more beneficial to an operator during a visual search task than a decisional aid. TAF analysis showed that the input aid offering spatial cues enhances asymptotic performance when compared to the output aid or no aid without affecting the system dynamics measured by processing times. The results are inconsistent with predictions of the lumberjack hypothesis. In sum, the results partially supported Hypothesis 1 and 2 but failed to support Hypotheses 3 and 4.

CHAPTER 5: EXPERIMENT 2

Experiment 1 found that reliable aids could improve visual search performance for an operator in a luggage screening task. The TAF analysis revealed that the input aid improved visual search performance by increasing asymptote but not processing time. Experiment 2 asked whether participant performance would continue to benefit from the assistance of automated aids with lower levels of reliability and if performance with the input and output aids would differ pronouncedly. Experiment 2 tested Hypotheses 5 – 8 (see Chapter 3: Summary and Hypotheses). Participants performed the task identical to that of Experiment 1 except that the reliability of the aids was set at 60%.

5.1. Methodology

5.1.1. Participants

Thirty participants (26 females; mean age = 21 years, *SD* = 6.15 years; mean corrected far acuity = 20/21.92, *SD* = 7.75; mean corrected near acuity = 20/23.46, *SD* = 6.89) were recruited from the community of Old Dominion University. All participants were screened for normal color perception using the Ishihara color blindness test (1989). None of them had participated in Experiment 1.

5.1.2. Apparatus

The apparatus was identical to that in Experiment 1.

5.1.3. Stimuli and Procedure

The stimuli and procedure were identical to those in Experiment 1 except that the reliability of the automated aids was set to 60%.

5.2. Results

5.2.1. Data analyses

Data treatment and analyses were identical to those in Experiment 1. Functions estimated using sum of squares residuals (RSS) accounted for 72.9% of the variance in the data for the 60% condition. The delayed exponential function could not be fit to the data of nine participants using the nls2 function. These were excluded from the analysis reported below. The results of the data with and without such participants did not differ. Again, an omnibus ANOVA involving Order as the between-subject factor was conducted, and the results did not indicate substantial evidence for any effects involving Order as a factor, all $B_{10} < .39$. Thus, to simplify the exposition, the analyses without Order as a factor are reported below.

5.2.2. Sensitivity

Figure 12 illustrates mean sensitivity as a function of Exposure Duration for the three aid conditions. As expected, the data indicated that participants' sensitivity decisively increased with longer Exposure Durations, $F(2.47, 49.46) = 80.73$, $\eta^2_G = .39$, $B_{10} = 6.91 \times 10^{43}$, supporting Hypothesis 6.a. However, this data pattern was statistically similar across the three conditions, $F(8, 160) = .37$, $\eta^2_G = .004$, $B_{10} = 0.01$. Neither the input nor output aid elevated sensitivity when compared to the manual condition, $F(2, 40) = .57$, $\eta^2_G = .004$, $B_{10} = 0.08$, failing to support Hypothesis 5.a.

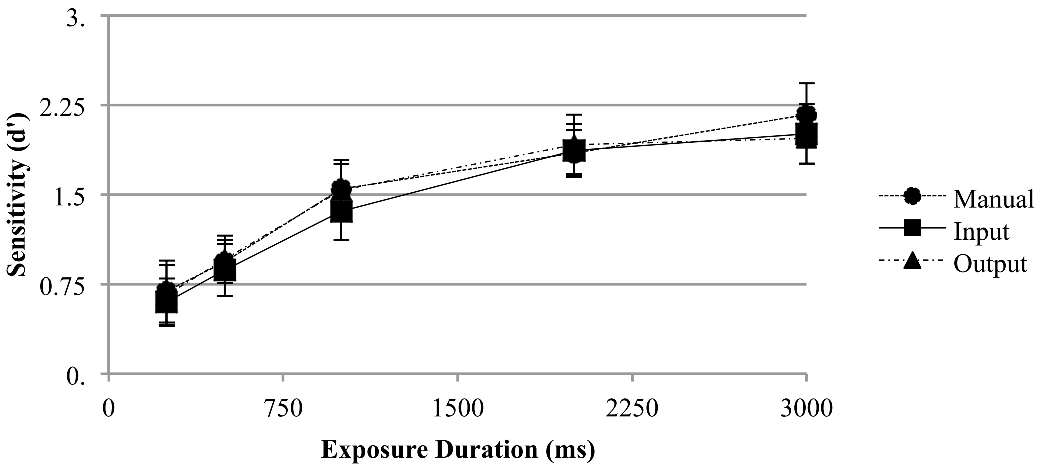

Figure 12. Mean d' scores across Exposure Durations for the conditions. (Error bars represent 95% within-subjects confidence intervals).

5.2.3. Response Bias

Figure 13 illustrates mean Response Bias as a function of Exposure Durations for the aid conditions. Response bias in Experiment 2 was analyzed for an exploratory purpose. Data indicate substantial evidence for the main effect of Exposure Duration on Response Bias, suggesting that participants made generally more conservative decisions with longer Exposure Durations of the stimuli, $F(4, 80) = 5.15$, $\eta^2_G = .02$, $B_{10} = 3.58$. Neither the main effect of Condition nor the interaction effect were substantial, $F(2, 40) = .49$, $\eta^2_G = .004$, $B_{10} = 1.72$ x 0.10, and $F(8, 160) = 1.05$, $\eta^2_G = .005$, $B_{10} = 0.01$, respectively.

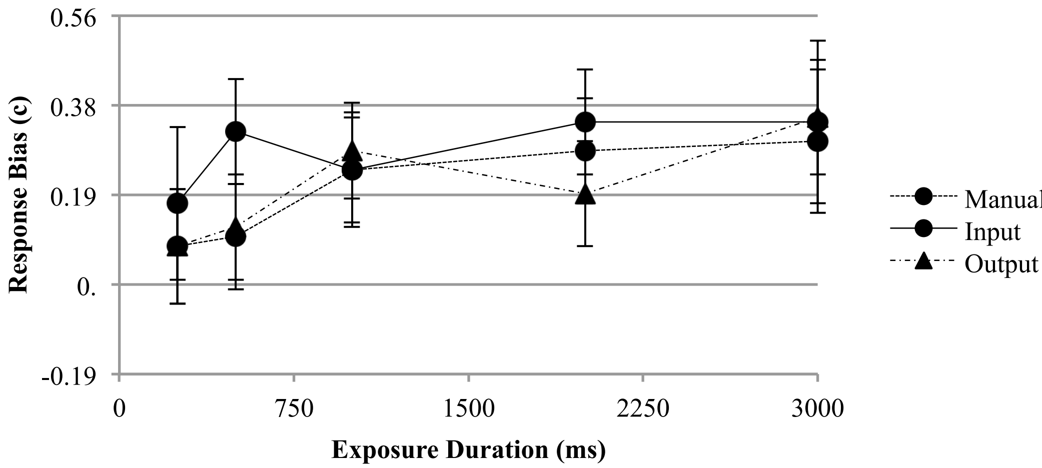

Figure 13. Mean Response Bias scores across Exposure Durations for the conditions. (Error bars represent 95% within-subjects confidence intervals).

5.2.4. Mean RT

Mean RTs as a function of Exposure Durations for the three aid conditions is shown in Figure 14. Analysis of the data indicated no substantial evidence for all the effects, all $B_{10} < 0.17$, failing to support Hypotheses 5.b. and 6.b.

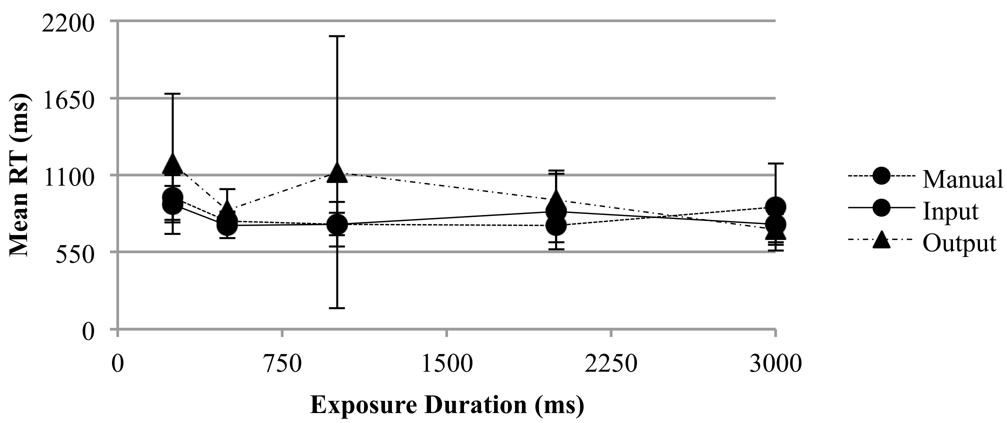

Figure 14. Mean RTs across Exposure Duration in the conditions. (Error bars represent 95% within-subjects confidence intervals).

5.2.5. Time-Accuracy Function Analysis

5.2.5.1. Processing Time

Figure 15 illustrates mean processing time across the conditions. The unreliable automated aids numerically shortened processing times (*M* = 1487.74 ms for the manual condition, *M* = 1236.86 for the input aid condition, and *M* = 1069.62 ms for the output aid condition), but data provided substantial evidence for the null model of Condition on processing time, $F(2, 40) = .37$, $\eta^2_G = .01$, $B_{10} = 0.18$, failing to support Hypothesis 7 and 8.

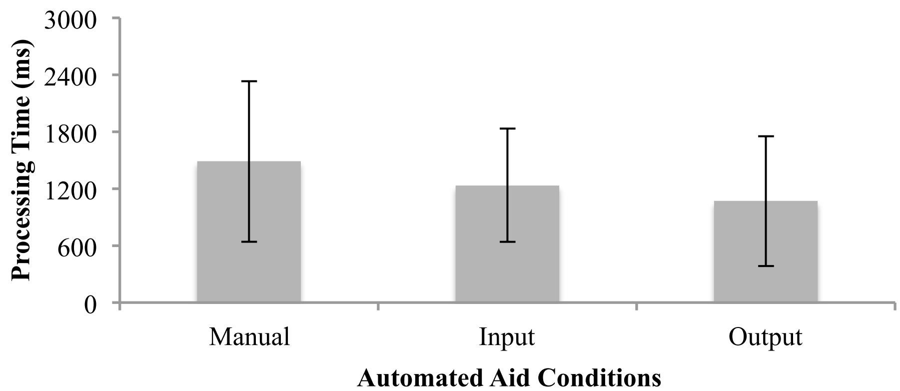

Figure 15. Mean processing time across automated aid conditions. (Error bars represent 95% within-subjects confidence intervals).

5.2.5.2. Asymptote

Figure 16 illustrates mean asymptotic performance levels across the conditions. Asymptote performance levels tended to be lower in the aided conditions than the manual condition. However, data did not support either the null or the model involving the effect of Condition, $F(2, 40) = 2.31$, $\eta^2_G = .07$, $B_{10} = 1.03$, failing to support Hypothesis 8.

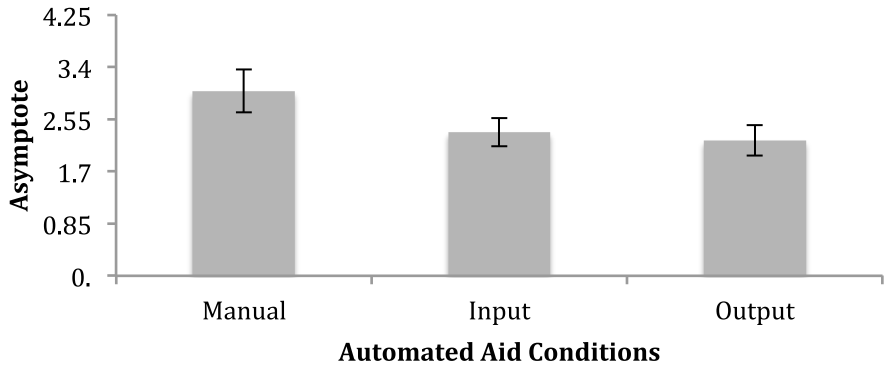

Figure 16. Asymptote scores across automated aid conditions. (Error bars represent 95% within-subjects confidence intervals).

5.3. Discussion

Similar to Experiment 1, participants continued to improve their visual performance as longer viewing time of the stimulus luggage was permitted, despite frequent inaccurate decisional recommendations of the automated aids. This improvement over exposure duration implies that the participants accumulated more sensory evidence as the stimulus display remained for a longer period of time to reach their decision of the target presence and improve their visual performance, while ignoring wrong decisional cues presented simultaneously with the raw stimulus image. This result supports Hypothesis 6. Additionally, participants' decisions became more conservative as exposure durations increased, similar with that of the manual condition in Experiment 1. However, unlike Experiment 1, the unreliable aids did not offset this pattern. This indicates that when an automated decision aid is not reliable, operators may ignore the recommendations and still improve visual performance as more time is allowed for them to view stimulus displays. These results fail to support the remaining hypotheses, likely because operators ignored the recommendations partially or fully to perform the task.

The less reliable aids trended to lower asymptotic performance levels when compared to the manual condition, but the magnitude of the decline was comparable between the input and output aids. Performance with the input aid may have declined because participants needed to ignore the aid's recommendations and reallocate attention to the stimuli when the aid miscued them. Posner described performance declines when cues are invalid as a cuing cost (Posner et al., 1980). In a typical cueing paradigm (e.g., Posner et al., 1980), participants are asked to fixate at the center of the screen during a visual target detection task. Then, an endogenous cue (an arrow, for example) appears at the display center, predicting a location of the target prior to the actual onset of the target. During a valid cueing trial, the target appears where the cue directed. On invalid cue trials, the target appears at an un-cued location. Posner and colleagues (1980) found that target detection accuracy declined when an invalid cue trial occurred when compared to the neutral condition where the cue was not predictive, suggesting that participants had to disengage their attention from the cued area and reengage to the target (Posner et al., 1980). Connecting to the present study, their performance might have declined because the input aid miscued to the location of a target and the participants needed to disengage from the cued location and reengage attention to the target following visual search.

In studies such as Crocoll and Coury (1990) and Wiegmann (2006) when aid reliability was low (reliability = 50 %), participants had greater performance decrements with the output aid compared to the input aid. Crocoll and Coury's findings support the lumberjack hypothesis that a higher DOA can leave an operator out of the loop (Manzey, Reichenbach, & Onnasch, 2009; Onnasch et al., 2013). However, such differences between the input and output aids were not seen in this study. One explanation for this may have been that our participants had access to the raw data, allowing them to incorporate the aid's recommendation into their decision-making process. In a previous study by Wiegmann (2006), participants were given a decision recommendation by the automated aid, but when that recommendation appeared, the image of the luggage disappeared. However, in the current study, the automated aids used to assist participants were both visual cues, and the raw data (the luggage image) were displayed for the entire duration that the aid's recommendation was present. This allowed participants the ability to cross-check their own judgment with the aid's recommendation. Another possible account for the lack of differences in processing time and asymptotes is a floor effect. The frequent invalid cues, combined with the need to verify them, might have imposed additional task load to the luggage screening task, producing the floor effect seen for both aids, regardless of DOA.

CHAPTER 6: CROSS-EXPERIMENT ANALYSIS

To provide quantitative comparison of the effects the aid reliability on visual performance between the two reliability levels, cross-experimental analyses were conducted involving the data from Experiment 1 and 2. The cross-experimental analyses were exploratory, testing whether the reliability manipulation interacted with the other independent variables.

6.1. Data analysis

d', *c*, and RTs were submitted to 2 x 3 x 5 mixed Bayesian analyses with Reliability (90% vs. 60%) as a between-subject factor and Condition (manual, input, vs. output) and Exposure Duration (250 ms, 500 ms, 1,000 ms, 2,000 ms, vs. 3,000 ms) as within-subject factors. Processing time and asymptote were submitted to 2 x 3 mixed Bayesian analyses with Reliability (90% vs. 60%) as a between-subject factor and Condition (manual, input, vs. output) as a within-subject factor.

6.1.1. Sensitivity

The data demonstrated decisive evidence that sensitivity levels increased with longer exposure durations, $F(3.26, 143.63) = 185.88$, $\eta^2_G = .40$, $B_{10} = 1.95 \times 10^{100}$. Figures 17, 18, and 19 present each condition separately for the differing reliabilities across Exposure Duration. There was very strong evidence that the input and output aid led to greater sensitivity levels than without them, $F(2, 88) = 4.33$, $\eta^2_G = .02$, $B_{10} = 59.18$.

The participants performing the task in the 90% reliability condition performed substantially better than those in the 60% condition, $F(1, 44) = 11.65$, $\eta^2_G = .10$, $B_{10} = 23.20$, the effect decisively interacting with Condition, $F(2, 88) = 8.58$, $\eta^2_G = .03$, $B_{10} = 2.22 \times 10^5$. The remaining effects were not reliable, $1.05 \times 10^{-2} < B_{10} < 1.57 \times 10^{-1}$. A follow-up mixed Bayesian analysis involving Reliability and Condition was conducted to explore the interaction effect. The data no longer provided evidence for the main effect of Condition, $F(2, 88) = 4.33$, $\eta^2_G = .03$, $B_{10} = 8.22 \times 10^{-1}$. However, the effect of Reliability on sensitivity remained strong, $F(1, 44) = 11.65$, $\eta^2_G = .10$, $B_{10} = 23.20$. Interestingly, the strong Reliability x Condition interaction effect demonstrates that the input and output aids elevated sensitivity levels when the reliability was 90% but, with markedly lower reliability of the aids, the participants' sensitivity declined back to the level of the manual condition, $F(2, 88) = 8.58$, $\eta^2_G = .05$, $B_{10} = 47.61$. Figure 20 demonstrates this trend across the three aid conditions. Follow up *t*-tests demonstrated only anecdotal evidence for performance differences of the manual condition, independent-sample $t(44) = .82$, $B_{10} = 0.38$. A substantial to strong difference was found between reliabilities for the input and output aids, independent-sample $t(44) = 4.74$, $B_{10} = 8.08 \times 10^2$, and independent-sample $t(44) = 3.49$, $B_{10} = 26.70$. However, because the difference between reliability conditions was greater for the input aid than the output aid, our data fail to show the lumber-jack effect of automation (Onnasch et al., 2013).

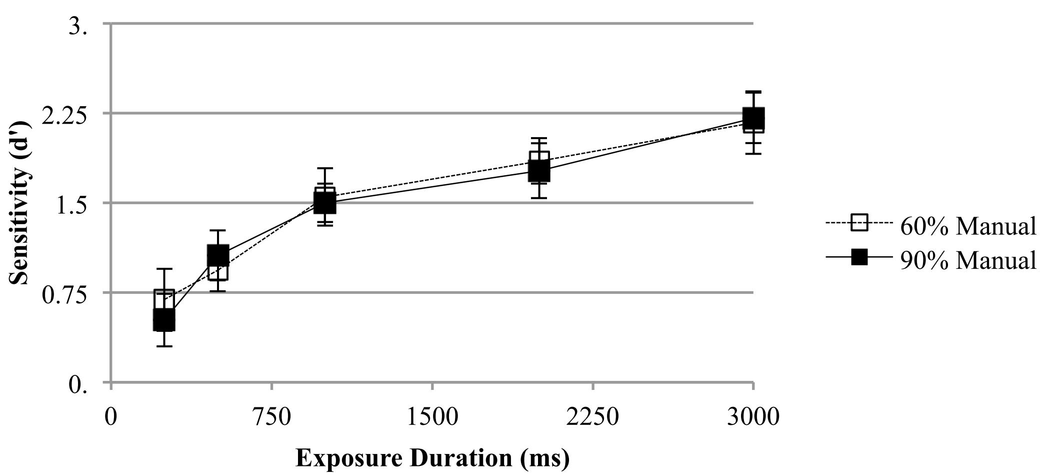

Figure 17. Mean d' scores for the manual condition, across exposure duration for both the 60% and 90% reliability groups. (Error bars represent 95% confidence intervals).

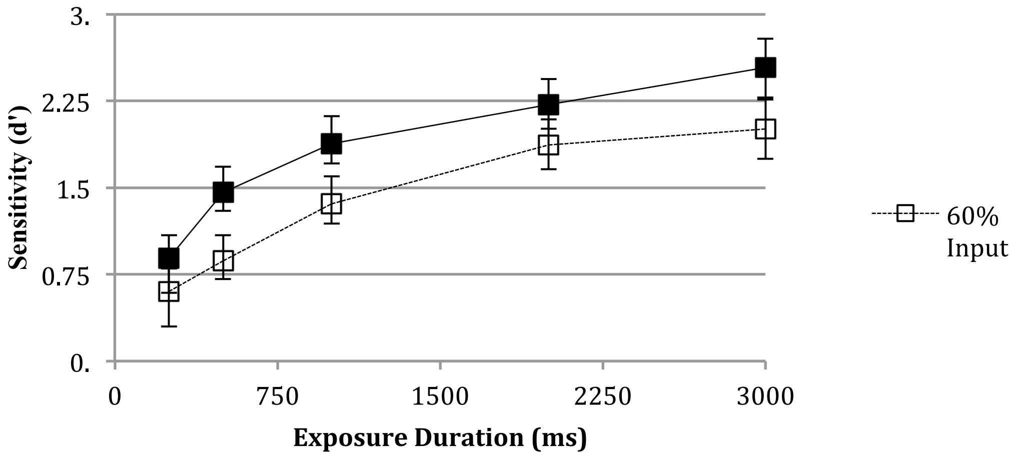

Figure 18. Mean d' scores for the input condition, across exposure duration for both the 60% and 90% reliability groups. (Error bars represent 95% confidence intervals).

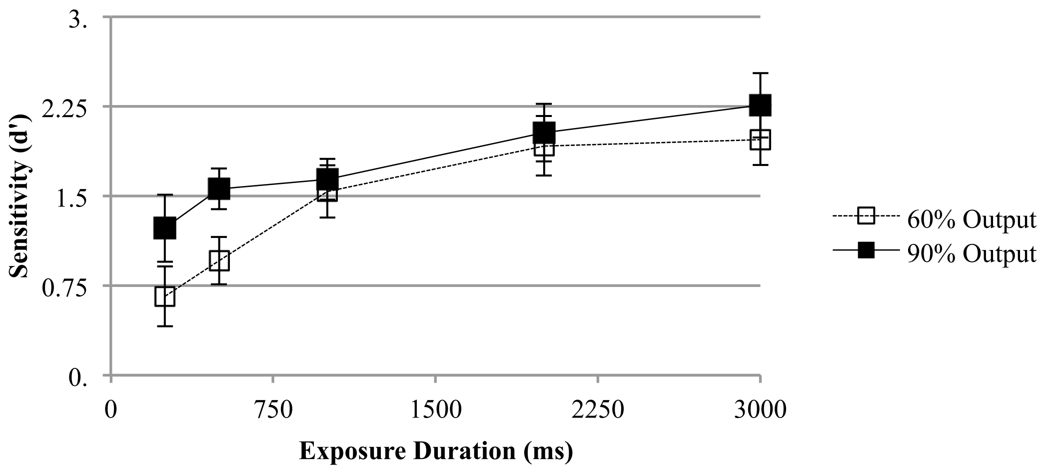

Figure 19. Mean d' scores for the output condition, across exposure duration for both the 60% and 90% reliability groups. (Error bars represent 95% confidence intervals).

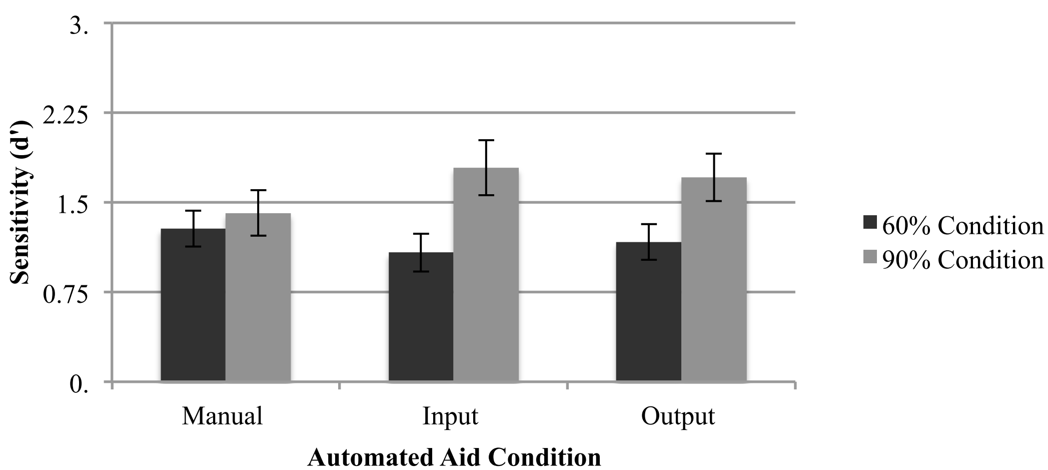

Figure 20. Mean d' scores across automated aid conditions for both the 60% and 90% reliability groups. (Error bars represent 95% confidence intervals).

6.1.2. Response Bias

Figure 21 presents response bias across Exposure Duration. Data again showed decisive evidence that the participants' response bias setting was progressively more conservative for stimulus exposure durations longer than 1,000 ms, $F(2.48, 109.46) = 7.49$, $\eta^2_G = .02$, $B_{10} = 1.94 \times 10^3$, this effect was comparable between the two reliability conditions, $F(2.48, 109.46) = .72$, $\eta^2_G = .002$, $B_{10} = 0.02$. The remaining effects were not substantial, $0.03 < B_{10} < 0.87$.

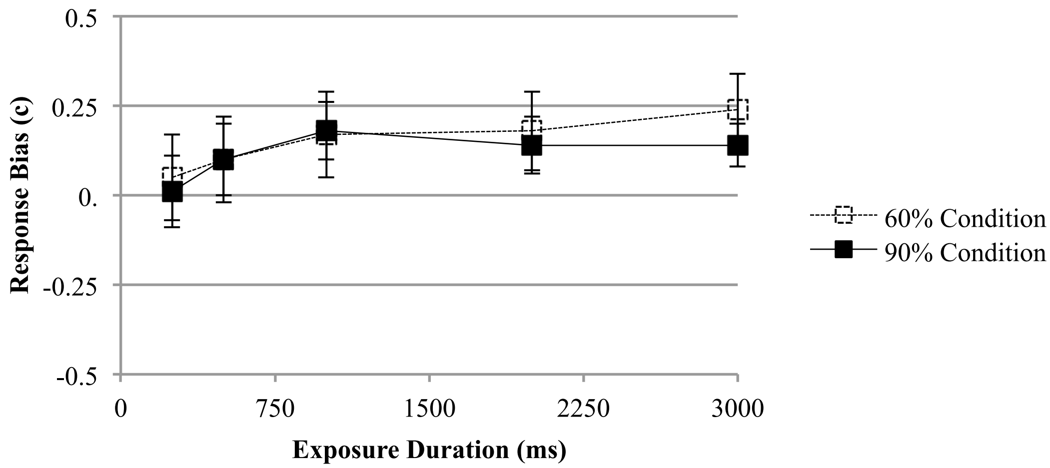

Figure 21. Mean response bias scores for 60% and 90% Reliability across Exposure Duration. (Error bars represent 95% confidence intervals).

6.1.3. Mean RT

Figure 22 represents mean RTs for the 60% and 90% experiments across Exposure Duration. Data indicate substantial evidence for the Reliability by Condition interaction, $F(2, 88) = 3.17$, $\eta^2_G = .04$, $B_{10} = 4.27$, suggesting that the output aid with 90% reliability produced RTs shorter than the two other conditions, while that with 60% elongated RTs. Data showed substantial to decisive evidence against the remaining effects, $0.004 < B_{10} < 0.26$.

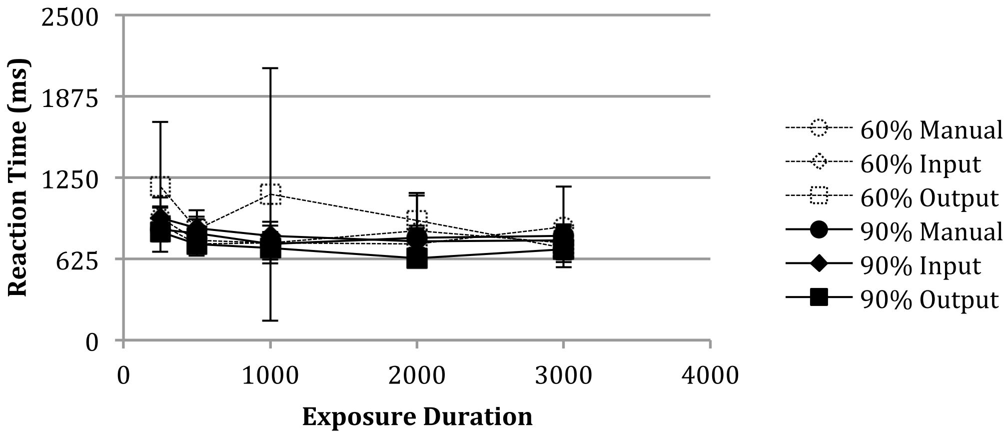

Figure 22. Mean RTs across Exposure Duration for the 60% and 90% Reliabilities for each condition. (Error bars represent 95% confidence intervals).

6.2. Time-Accuracy Function Analysis

6.2.1. Processing Time

Figure 23 shows processing time for the 60% and 90% Reliabilities across Conditions. From visual inspection of the figure, processing times tended to be shorter in the 90% reliability condition than 60%. However, data provided only anecdotal evidence for differences between the reliability levels, $F(1, 44) = 4.42$, $\eta^2_G = .03$, $B_{10} = 0.81$. The effect of conditions was strongly in favor of the null, $F(2, 88) = 0.38$, $\eta^2_G = .005$, $B_{10} = 0.10$. The data were indifferent to the interaction effect, $B_{10} = 0.14$.

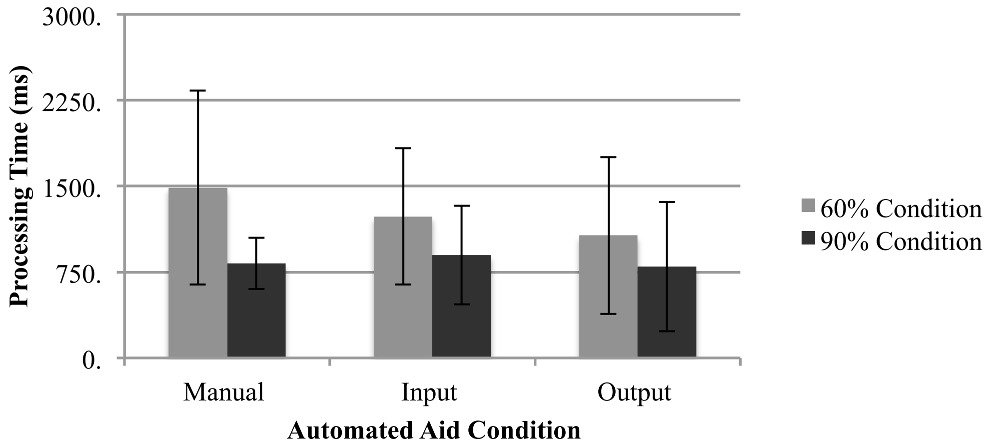

Figure 23. Mean processing times across 60% and 90% conditions. (Error bars represent 95% confidence intervals).

6.2.2. Asymptote

Figure 24 presents asymptote for the 60% and 90% Reliabilities across Conditions. Data gave substantial evidence for the interaction effect, $F(1.96, 86.11) = 4.61$, $\eta^2_G = .06$, $B_{10} = 6.33$. The remaining effects were not substantial, $B_{10} < 0.26$. Follow-up t-tests revealed that the difference between the 90% and 60% reliability levels for the input aid condition was close to substantial, independent-samples $t(44) = 2.21$, $B_{10} = 2.91$, but not for the manual and the output aid conditions, $B_{10} = 0.64$ for the manual and $B_{10} = .87$ for the output aid condition.

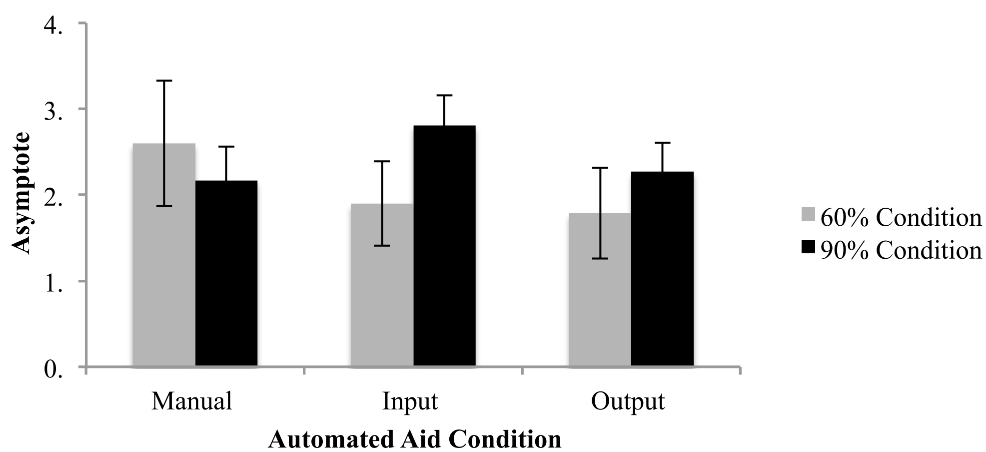

Figure 24. *Mean asymptote scores for 60% and 90% Reliability across conditions. (Error bars represent 95% confidence intervals).*

6.3. Discussion

The cross-experiment analysis confirmed the effect of reliability on human performance in an unspeeded visual search task, a common finding in the literature of human-automation interaction (Lee & Moray, 1994; Meyer, 2001; Mosier, Skitka, Heers, & Burdick, 1998; Riley, 1994; Sarter & Schroeder, 2001; Parasuraman & Riley, 1997; Yeh & Wickens, 2001b). Participants improved their visual performance when working with reliable aids but their performance suffered when working with less reliable aids compared to no assistance from automation. The increased performance when the aids were reliable was not surprising as the use of a spatial aid to direct attention can quickly help operators to isolate important objects within a highlighted area and ignore others in the same image when using the input aid (Posner, 1980; Madhavan & Wiegmann, 2006; Wiegmann, 2006). Notably, TAF analysis revealed that such performance improvement in the input aid condition was mainly driven by increases in asymptotic performance but not processing times, which conventional analyses of performance measures such as mean RTs and sensitivity do not reveal.

The input and output aids at low reliability did not produce performance differences. The invalid spatial aid misdirected attention. Presumably, requiring reengagement of spatial attention to the target, when present, could have ultimately impeded their performance. The invalid output aid was also problematic to operators. When it provided incorrect information, participants had to search the visual scene themselves to determine its accuracy. However, the availability of the raw data alongside the aid’s recommendation may have given participants an opportunity to confirm the accuracy of the aid.

The response bias in the aggregated form remained conservative ($c > 0$) although the response bias of the aid remained neutral. The results of the current study suggest that participant response bias can shift to a more conservative setting as stimulus exposure durations increase and operators are allowed more time to verify the aid's accuracy. Experiments did not detect measurable differences in processing times, suggesting that the automated aids tested in the current study did not affect operators' front-end processes regardless of aid reliability level such as sensory input and perceptual processes for accumulating information to make their judgments. Extending the previous research by Wiegmann and colleagues (2006), demonstrating that while the use of a spatial cue may be of great benefit to an operator's performance, it only affects operator asymptote.

The current results showed the effect of automation reliability on asymptotic performance only in the input aid condition, not the output aid condition. The input aid might have supported more accurate perception of a threat item by highlighting where the target was, if present and the aid is accurate, than the output aid. Given the limited exposure duration, the participants might have restricted their processing resources to perceive and compare their decision with the aid's decision efficiently in the input aid condition by ignoring the other unselected objects. On the other hand, the output aid condition did not allow such strategies because the decision cues were presented as the colored frame of each display. The aid's binary decision recommendation was randomly determined each trial, and the number of successive trials with the aids' correct recommendations might not have been large enough for the output aid to improve the participants' performance. At an event of automation failure, it is possible that operators do not show a large performance decrement because the aid has not yet sufficiently improved their performance. A frequent automation failure in the current experiments might have caused misperception of the reliability of the aid, which is partially supported by differing response bias setting across the experimental conditions. This can lead to less optimal human-automation joint decision making.

CHAPTER 7: GENERAL DISCUSSION

The current study examined the effects of automated decision aids with differing automated capabilities across different stimulus exposure durations in a luggage screening task. In the unspeeded visual search task, operators scanned a simulated X-ray image of simulated passenger luggage and were asked to find a threat object (a knife) among other objects without any automated aids (the manual condition) with an automated aid that spatially cued an area of a potentially threatening object (the input aid condition) and with an automated aid that only offered a binary decisional recommendation (the output aid condition) with reliability of either 90% (Experiment 1) or 60% (Experiment 2). This dissertation project extended the previous research on this topic by Wiegmann and colleagues (2006) by applying the TAF analysis to provide insights into the information-processing mechanisms that control visual performance when assisted by different degrees of automated aids. The TAF analysis allowed dissociating speed of information accumulation, or perceptual processing from asymptotic performance level, while the conventional analysis of performance such as ones using RTs or error rates conflates the two. That is, one may not know whether it is processing time and/or asymptote that affect visual performance when performance was measured by only RTs or error rates.

Experiment 1 showed operator's visual search performance improved with the assistance of both the input and output aid. Additionally, the spatial cue (input aid) was of greater benefit to operators' asymptotic performance compared to the decisional aid (output aid). Experiment 2 showed that when automated aid reliability was low, operator performance declined compared to when no aid assistance was offered at all for the task. Furthermore, cross-experiment analysis showed that aid reliability impacted operators' ability to successfully detect the target. In fact, compared to the high reliability condition (90%), the low reliability condition (60%) led to impaired visual performance. The magnitude of this performance decline caused by unreliable aids, however, did not differ between the input and the output aids.

Theoretically, the input aid and the output aid affect different stages of human information processing. The input aid should affect earlier stages of information processing such as aggregation of sensory information, perception, and stimulus organization. The output aid should affect integration for later stages such as interpretation, decision making, and response selection/execution (Parasuraman et al., 2000; Li et al., 2014; Onnasch et al., 2013). Following the lumberjack hypothesis, the further along the continuum of DOA that an automated aid operates, the greater the chance that an operator can become left out of the loop, when automation failures occur, leaving the potential for larger declines in performance when working with higher DOAs (Galster & Parasuraman, 2001; Merat & Lee, 2012; Metzger & Parasuraman, 2005; Sarter & Woods, 1995, 2000).

The current results, however, are not largely consistent with the predictions of the lumberjack hypothesis. Performance loss due to lower reliability level of the aids was not greater for the output than the input aid. One explanation for the lack of greater performance decrement with the higher DOA (output aid) may have been potentially high levels of automation transparency in the current study. The luggage image in the current study remained on the screen while the aid's recommendation was rendered. In previous studies such as Wiegmann (2006), the output aid was a simple text message given during the luggage screening task. As soon as their decision aid issued a recommendation, the luggage image disappeared, leaving the operator without the raw data in front of them to re-access. When an operator is left without raw data and only the recommendation of an opaque aid, their reliance on the aid can become unpredictable (Rice, 2009). Cross-checking the accuracy of the aid's recommendation presumably requires a comparison of the cue and the stimulus image in their memory. In the present study, the raw data remained available for operator access at the same time the output aid delivered a recommendation. The reliability level of the aid and availability of the raw data during each trial allowed the participants to inspect the luggage images themselves and verify the accuracy of the aid, especially as exposure duration increased. The prolonged availability of the raw stimulus information in the present study may have offset the cost of automation failure as seen in previous studies working with a high DOA at low reliability.

Relatedly, when the aid reliability was high, lower levels of sensitivity and asymptote levels in the output aid than in the input aid may suggest that different levels of information-processing created a cost of comparing the aid's recommendation to operator judgments. That is, operators in the input aid condition had a minimized search area to scan. Operators in the output aid condition had to scan a larger visual area, up to its entirety if the target was absent, to validate the aid's recommendation. Future efforts should extend this research by manipulating the effect of the presence of raw stimulus when an automation decision cue is rendered.

Participants performed better with the assistance of the input aid than the output aid, which was contrary to what was initially hypothesized (e.g., Hypothesis1, 5, 3, 4, and 7). According to the lumberjack hypothesis, automation operating at a lower DOA should not be as beneficial to operator performance as a higher DOA. Operators working with a lower DOA, such as the input aid in the present study, would need to devote more resources to evaluating the situation to make the correct decision or execute the correct action (Parasuraman et al., 2000; Rice, 2009). The literature on human-automation interaction, however, suggests that operators perform better with less automation or lower DOAs as they are encouraged to actively process information and not just passively accept it, keeping them more aware of the situation or the automated state

(Billings, 1997; Endsley & Kiris, 1995; Ferris, Wickens, & Sarter, 2010; Parasuraman & Riley, 1997; Parasuraman & Wickens, 2008; Sarter, Woods, & Billings, 1997). But, the effect of DOA on human performance when automation fails may depend on information processing demands of different professional tasks. The current results, as an example, seem to show that the input aid, offering visual-spatial information regarding a potential target location, can improve visual performance above and beyond what the input aid's framework of DOA does as classified in Onnasch et al. (2013).

Unlike previous studies (e.g. Wiegmann, 2006; Madhavan & Weignamm, 2005), the current input aid appeared every time a luggage image was shown. In this way, it was meant to avoid implying a decision regarding the presence or absence of a target by its appearance. Instead, the input aid was meant to direct attention to a visual area of the highest possibility of a threat object within each image. Operators were told that the input aid would not tell them if a weapon was present, only where it was likely to appear in an image. Operators still had to visually search the highlighted portions of the bag to determine if a knife was indeed present. It remains unknown whether a visual-spatial cue that also indicates the aid's decision of target presence can improve visual performance beyond what the current input aid can do. An aid that provides information both regarding a spatial area of higher threat and a decisional cue may offer greater levels of automation transparency, allowing operators to "see" how the automation has reached its decision.

Unexpectedly, operators' bias in the manual condition was conservative and became more neutral when the automated decision aids were available. Furthermore, their response bias became more neutral with the input aid than the output aid. When the input aid appeared to operators, it always highlighted a section of the image. The aid highlighted a quadrant even when no knife was present. It was limited to highlighting areas likely to contain a target, not to identify a threat object. This may have created a misperception that a knife was present more often than it really was. This potential misperception may have contributed to the less conservative response bias seen when participants were working with the input aid than the output aid. Madhavan and Wiegmann (2006) found similar results as participants in their study had a more conservative bias when working with the indirect aid (c =.47; similar to our output aid) vs. their direct aid (c =.39; similar to our input aid). Thus, operators working with the input aid viewed it as a FA-prone system, shifting their bias to respond "yes" more often to the presence of a target to counteract the perceived false alarms.

Target prevalence may also have contributed to the shift in operator response bias with the input aid. The target prevalence was set at 50% for both experiments, which is much higher than what would be seen for target presence in the real world. Wolfe and Van Wert (2010) noted that, when target prevalence was high in their study, participant response bias was affected. As target prevalence increased, it inflated the false alarm rate, shifting participant bias to become more liberal. Therefore, our tentative account for the response bias setting becoming less conservative at 1,000 ms of stimulus exposure duration (or longer when working with the input aid) is that participants may have misperceived it to be FA prone. Further research is necessary for examining relationships between operators' perception of the aid's response bias and their strategy to adjust their response bias for efficient human-automation collaboration.

7.1. Theoretical Contribution

The current study extends earlier research (e.g. Crocoll & Coury, 1990; Madhavan & Wiegmann, 2006; Wiegmann, 2006) with the TAF analysis complementing the conventional analysis of operator performance during computer-aided decision making. It is apparent from looking at participant levels of sensitivity, RT, and asymptote that operator performance improved when working with reliable automated aids, regardless of DOA. The TAF analysis showed that operators maintained comparable processing times between the input and output aid conditions to reach their asymptotic performance level. This was true when the aids were working reliably (90%) and unreliably (60%).

Parasuraman et al. (2000) proposed a theoretical framework based on the human information-processing model (Wickens & Hollands, 2000) for automation designers where automation can be designed to support distinct information-processing stages of human operators. The framework included four stages: sensory processing, perception/working memory, decision making, and action/execution, which are supported by information acquisition, information analysis, decision selection, and action implementation features of automation, respectively. The current study, by employing the TAF analysis, revealed that operators' sensory processing and perception (and potentially working memory) are not measurably affected by the input or output aids which were designed to offer automated support for information acquisition and analysis. Our tentative account for this null effect of the aid type on processing time is that the availability of the raw image when the decision cue was rendered made the cues of both the input and output aids a tool for operators to confirm their own decisions. The longer the stimulus display was available, the more opportunities operators could perform the task without actively using the aid's recommendations, diluting the effects of the different aids, and cross-check their own decision against the aid's recommendation.

Contrary to the predictions of the lumberjack hypothesis, the data did not indicate greater performance decrements in the output aid condition than the input aid condition. As stated above, the limitation of the current paradigm likely did not allow participants to observe successive and accurate decisional recommendations from the aids, which might have led to disuse of automation. This behavior might have been more severe when the stimulus exposure durations were longer where participants could perform the task accurately without interacting with the aids. One implicit assumption of the lumberjack hypothesis is that automation reliably supports human performance for a sufficiently long period of time so that operators establish their perception of behaviors and reliability of the automation, and the current experimental paradigm might not necessarily meet the assumption.

Yamani and Horrey (2018) recently offered a theoretical framework on human-automation interaction that explains relationships between attentional resource allocation dependent upon the DOA assisting them and human information processing. As DOA increases and the automation assists the operator with higher cognitive processes, the greater amount of attentional resources the operator should be able to devote to other tasks. However, when automation fails, based on the amount of raw data available, an operator may devote more attention to the raw data in order to maintain performance. The authors point out that operators allocate attention to monitor automation's behavior and performance to estimate its reliability, which in turn guides resource allocation policy. Applying this to the current context, participants in the current study might have misperceived the true reliability of the aids, and they may devoted the resources to actually perform the task by themselves and not optimally utilize the aids as designed.

Two points below are noteworthy. First, due to the fact that processing times did not differ between the input and output aids, the differing degrees of automation utilized in the present study did not necessarily influence operators' information-processing stages as designed. Therefore, the taxonomies created to help design these aids and explain how they should assist operators with a task may not best capture types of human-automation interactions that occurs with the designed automated aids; operators may not be utilizing the aids as they are described in the Parasuraman, Sheridan, and Wickens (2000) taxonomy or in the manner predicted by the lumberjack hypothesis.

Second, transparency of automated systems can assist operators to mitigate large performance decrements when the systems fail or behave unreliably (Lyons, 2013; Wickens, 2018). Implications that transparency can help an operator to maintain awareness provide an additional solution to that of simply lowering the DOA to maintain operator performance (Wickens, 2018). Both the input and the output aid in the present experiments afforded participants access to the raw information. According to the lumberjack hypothesis (Onnasch et al, 2013), performance, particularly with the unreliable output aid (higher DOA), should have shown a large performance decrement when the aid failed than properly functioned than the input aid. The present data are not consistent with the predictions. Additionally, if an operator does not perceive a benefit to relying on the automation, they are likely to revert to its disuse or to manual performance (Dzindolet, Pierce, & Beck, 1999; Dzindolet, Peterson, & Pomranky, 2003; Parasuraman & Riley, 1997). While the current data cannot determine that the participants in this study disused the automated aids, it is interesting to note that performance (RT, sensitivity, asymptote, and processing time) between the manual and output aid condition are similar, regardless of the reliability level of the aid. Finally, when looking at performance with the use of the input aid specifically, a larger performance decrement was seen, which may have been due to participants needing to disengage from one visual area and re-engage in another when the aid was unreliable. Therefore, the DOA will affect operator performance, however dependent upon other factors such as transparency of the aid and the raw stimulus that is salient to the operator.

Future research may further manipulate the presence of the raw stimuli when the aid's recommendation arrives in order to examine how operators' decision strategies and human-machine joint decision-making changes when interacting with automated decision aids with varying DOA and levels of transparency.

7.2. Practical Relevance

In practice, this study provides implications for luggage screening and human-automation teaming. First, TSA luggage screeners can benefit from spatial cues that assist in directing their attention to potential threat objects. The results of this study are consistent with previous literature on spatial cueing (Posner, 1980) and the benefit of directing operator attention in an applied visual workspace. Participants in this study showed a greater sensitivity and higher asymptote when working with the more reliable (90%) spatial aid than the higher DOA. Additionally, operators did not differ in their amount of processing time when using the higher DOA as opposed to the lower DOA or no automation at all. The rate at which they approached their asymptote did not change and this was true whether the automation was working at a high reliability or a low reliability level. The finding that the rate of processing did not change regardless of the DOA and that performance was higher with the input aid (lesser DOA) implies that the use of a higher DOA may not always be more beneficial or relieve a greater amount of the cognitive task load as initially suggested by Onnasch et al. (2013).

The data also suggest that when using aids that may not always be reliable, regardless of DOA, providing access to raw data or images may assist the operator in mitigating performance failures. By providing access to the raw luggage image, you allow the operator to decide where it might be most effective to invest their resources. As the TSA screener bears large responsibility for the safety of passengers and detecting threats, during human-automation teamwork, TSA operators would still need to maintain access to raw information in case automation fails so that they could invest attentional resources accordingly to mitigate potential disasters and recover.

The TAF analysis can be used to derive the time it takes for a TSA agent to reach a specific level of maximum performance (asymptote). Based on the results of this study, it would appear that given resources to invest in an output aid or an input aid, the input aid would be more likely to increase screener sensitivity, or allow for a shorter display time for the same level of sensitivity, if it were a reliable aid used. Additionally, the current results from the TAF analysis may suggest the amount of exposure time a TSA officer needs to have to an image to achieve a criterion level of sensitivity. Further, the TAF curve can be applied to TSA screeners, who self-pace, to assess the rate at which they view luggage images and suggest alternative pacing to improve performance.

7.3. Limitations

Several limitations exist in the current experiments. First, operators in the current study were asked to search for only one specific threat object (a knife). In reality, TSA screeners are asked to look for many different types of threat objects that come in many shapes and forms. This can make threat detection more difficult because a) operators need to keep a large set of threat objects in memory and, b) sometimes the threat objects can be disguised or broken into pieces, such as when a bomb has several components placed in different parts of luggage. This can create a more difficult search, as screeners must be able to search for multiple objects and combine them to identify possible threats. Second, operators were undergraduate students and lacked the experience that many TSA screeners with training and years of experience may possess, limiting generalizability of the present findings. Therefore, additional research is needed to further explore the consequences of these results to specific real-world scenarios. Third, due to the fitting of the TAF model to the obtained data, data of a number of participants were excluded from the analysis, potentially compromising experimental power. As the algorithm to fit the TAF curve was applied, it was found that data from several participants did not fit parameters set by the curve. For several participants performance was significantly higher at the 250 ms exposure duration compared to their performance in the longer exposure durations (e.g., 500 ms, 1,000 ms, 2,000 ms). This is contrary to previous research demonstrating greater accuracy with target detection as time to view the image increased (Madhavan & Gonzalez, 2006; Shapiro & Penrod, 1986; Wicklegreen, 1977; Yamani & McCarley, 2011). Therefore, the decision was made to exclude those participants from the analysis. However, for the measures of sensitivity, response bias, and RT, analyses were performed with and without those participants to assess if any significant differences in results existed and none were found. Another way the poor fitting curves could have been handled is to have dropped the 250 ms time point. However, to find an accurate estimate of intercept for our TAF curve, it was imperative that we provide a time point that would be at the lower bound of performance for our participants. Therefore the time point for 250 ms was retained. Future research could eliminate the 250 ms exposure duration and increase the number of participants to better estimate sizes of the effects of interest.

7.4. Conclusion

In conclusion, the reliable automated aids improved visual performance in a simulated luggage screening task. Operator performance was comparable across DOA, contradicting the previous research and theoretical framework of human-automaton interaction. The input aid, classified as low DOA in the Onnasch et al. (2013) framework, improved human performance more than the high DOA output aid when the aid reliability was high. Availability of the raw stimulus image throughout each trial could have affected how operators interacted with the aids by affording the ability to explicitly evaluate the accuracy of the aid's recommendation. Finally, the application of the TAF analysis separated the performance measures of asymptote and processing time, allowing us to determine more accurately the cognitive locus of effect of the input/output aids. The results demonstrated that asymptote differed between conditions but the processing time remained static regardless of whether automation was present, of the DOA it was operating at, or the reliability level of the automation. The results were inconsistent with the prediction of the lumberjack hypothesis. The discrepancy that operator performance was better with the input than the output aid may have arisen due to different strategies that operators took to incorporate the aid's decisional recommendation into their decision-making process. In general, the current taxonomies characterizing different types of automation based on operators' information-processing demands do not always match with how operators perform a task assisted by the automation.

REFERENCES

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779.

Bliss, J. P., & Gilson, R. D. (1998). Emergency signal failure: Implications and recommendations. *Ergonomics, 41*(1), 57-72.

Bliss, J. P., & Dunn, M. C. (2000). Behavioural implications of alarm mistrust as a function of task workload. *Ergonomics*, *43*(9), 1283-1300.

Bradshaw, J. M., Dignum, V., Jonker, C., & Sierhuis, M. (2012). Human-agent-robot teamwork. *IEEE Intelligent Systems, 27*(2), 8-13.

Christensen, J. C., Estepp, J. R., Wilson, G. F., & Russell, C. A. (2012). The effects of day-to-day variability of physiological data on operator functional state classification. *NeuroImage, 59*(1), 57-63.

Comstock Jr, J. R., & Arnegard, R. J. (1992). The multi-attribute task battery for human operator workload and strategic behavior research.

Crocoll, W. M., & Coury, B. G. (1990, October). Status or recommendation: Selecting the type of information for decision aiding. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 34, No. 19, pp. 1524-1528). SAGE Publications.

Deppe, M., Schwindt, W., Kugel, H., Plassmann, H., & Kenning, P. (2005). Nonlinear responses within the medial prefrontal cortex reveal when specific implicit information influences economic decision making. *Journal of Neuroimaging*, *15*(2), 171-182.

de Visser, E. J., Cohen, M., Freedy, A., & Parasuraman, R. (2014, June). A design methodology for trust cue calibration in cognitive agents. In *International conference on virtual, augmented and mixed reality* (pp. 251-262). Springer, Cham.

Endsley, M. R. (1997). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42*(3), 462-492.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(2), 381-394.

Endsley, M. R., Bolstad, C. A., Jones, D. G., & Riley, J. M. (2003, October). Situation awareness oriented design: from user's cognitive requirements to creating effective supporting technologies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 3, pp. 268-272). Sage CA: Los Angeles, CA: SAGE Publications.

Evans, P. (2005). Three-dimensional X-ray imaging for security screening. *Security Journal, 18*(1), 19-28.

Fiore, S. M., Scielzo, S., & Jentsch, F. (2004). Stimulus Competition During Perceptual Learning: Training and Aptitude Considerations in the X-ray Security Screening Process. *Cognitive technology.*

Godwin, H. J., Menneer, T., Cave, K. R., Thaibsyah, M., & Donnelly, N. (2015). The effects of increasing target prevalence on information processing during visual search. *Psychonomic bulletin & review*, *22*(2), 469-475.

Horwitz, J., & Timmons, H., (2016, September 20). There are some scary similarities between Tesla's deadly crashes linked to Autopilot. *Quartz, http://qz.com/783009/the-scary-similarities-between-teslas-tsla-deadly-autopilot-crashes/*

Kaber, D. B., & Endsley, M. R. (1999). 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), 113-153.

Kerley, D & Cook, J. (2017, Nov. 8th). TSA fails most tests in latest undercover operation at US airports. https://www.everythinglubbock.com/news/national/tsa-fails-most-tests-in-latest-undercover-operation-at-us-airports/853877236

Li, H., Wickens, C. D., Sarter, N., & Sebok, A. (2014). Stages and levels of automation in support of space teleoperations. Human Factors: *The Journal of the Human Factors and Ergonomics Society, 56*(6), 1050-1061

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Madhavan, P., & Gonzalez, C. (2006, October). Effects of sensitivity, criterion shifts, and subjective confidence on the development of automaticity in airline luggage screening. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 3, pp. 334-338). Sage CA: Los Angeles, CA: SAGE Publications.

Madhavan, P., & Wiegmann, D. A. (2006, October). Expertise levels of human versus automated decision aids influence response biases in a visual search task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 3, pp. 230-234). Sage CA: Los Angeles, CA: SAGE Publications.

Manzey, D., Reichenbach, J., & Onnasch, L. (2009, October). Human performance consequences of automated decision aids in states of fatigue. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 4, pp. 329-333). Sage CA: Los Angeles, CA: Sage Publications.

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*, 1555343411433844.

Mercado, J. E., Rupp, M. A., Chen, J. Y., 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.

Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. Human Factors: *The Journal of the Human Factors and Ergonomics Society, 47*(1), 35-49.

Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounois, J. (1988). The dynamics of cognition and action: mental processes inferred from speed-accuracy decomposition. *Psychological review, 95*(2), 183.

Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. Human Factors: *The Journal of the Human Factors and Ergonomics Society, 49*(1), 57-75.

Moray, N. (1997). Human factors in process control. *Handbook of human factors and ergonomics*, 1944-1971.

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2013). Human Performance Consequences of Stages and Levels of Automation An Integrated Meta-Analysis. Human Factors: *The Journal of the Human Factors and Ergonomics Society*, 0018720813501549.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors: *The Journal of the Human Factors and Ergonomics Society, 39*(2), 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), 286-297.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human factors*, *50*(3), 511-520.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology: General*, *109*(2), 160.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3-25.

Quinlan, P. T. (2003). Visual feature integration theory: past, present, and future. *Psychological bulletin*, *129*(5), 643.

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human factors*, *37*(1), 5-19.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. Handbook of human factors and ergonomics, *2*, 1926-1943.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. Psychological Bulletin, 100(2), 139.

Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators. MASSACHUSETTS INST OF TECH CAMBRIDGE MAN-MACHINE SYSTEMS LAB.

Smith, J. D., Redford, J. S., Gent, L. C., & Washburn, D. A. (2005). Visual search and the collapse of categorization. *Journal of Experimental Psychology: General, 134*(4), 443.

Transportation Security Administration. TSA Year in Review. (2017, Jan. 20th). Retrieved from: https://www.tsa.gov/blog/2017/01/12/tsa-year-review-record-amount-firearms-discovered-2016

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97-136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological review*, *95*(1), 15.

Van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796-803.

Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied ergonomics, 53*, 190-202.

Verhaeghen, P. (2000). The interplay of growth and decline: Theoretical and empirical aspects of plasticity of intellectual and memory performance in normal old age. *Cognitive rehabilitation in old age*, 3-22.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*(3), 291-298.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta psychologica*, *41*(1), 67-85.

Wickens, C. (2018). Automation stages & levels, 20 years after. *Journal of Cognitive Engineering and Decision Making*, *12*(1), 35-41.

Wickens, C. D., & Hollands, J. G. (2000). Attention, time-sharing, and workload. *Engineering psychology and human performance,* 439-479.

Wickens, C. D., & McCarley, J. S. (2008). Applied attention theory. CRC press.

Wickens, C. D., Li, H., Santamaria, A., Sebok, A., & Sarter, N. B. (2010, September). Stages and levels of automation: An integrated meta-analysis. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 4, pp. 389-393). SAGE Publications.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology & human performance*. Psychology Press.

Wiegmann, D., McCarley, J. S., Kramer, A. F., & Wickens, C. D. (2006). Age and automation interact to influence performance of a simulated luggage screening task. *Aviation, Space, and Environmental Medicine*, *77*(8), 825-831.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance*, *15*(3), 419.

Wolfe, J. M. (1994). Visual search in continuous, naturalistic stimuli. *Vision research*, *34*(9), 1187-1195.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision research*, *37*(1), 25-43.

Wolfe, J. M. (1998). Visual search. *Attention*, *1*, 13-73.

Wolfe, J. M., & Horowitz, T. S. (2004). Opinion: What attributes guide the deployment of visual attention and how do they do it?. *Nature Reviews. Neuroscience*, *5*(6), 495.

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of experimental psychology. General*, *136*(4), 623.

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*(2), 121-124.

Yamani, Y., & Horrey, W. J. (2018). A theoretical model of human-automation interaction grounded in resource allocation policy during automated driving. *International Journal of Human Factors and Ergonomics*, *5*(3), 225-239.

Yamani, Y., & McCarley, J. S. (2011). Visual Search Asymmetries in Heavy Clutter Implications for Display Design. Human Factors: *The Journal of the Human Factors and Ergonomics Society,* *53*(3), 299-307

APPENDIX A

OLD DOMINION UNIVERSITY INFORMED CONSENT DOCUMENT

PROJECT TITLE Automated Aid Effects During Visual Threat Detection

INTRODUCTION
The purpose of this form is to give you information that may affect your decision whether to say YES or NO to participation in this research, and to record the consent of those who say YES. The title of this project is "Automated Aid Effects During Visual Threat Detection". The experiment will be conducted on the ODU campus in Room # 336 Mills Godwin Building.

RESEARCHERS
Yusuke Yamani, Ph.D., Professor - **Responsible Project Investigator**
Department of Psychology, Old Dominion University
Molly Liechty, Doctoral student
Department of Psychology, Old Dominion University

DESCRIPTION OF RESEARCH STUDY
You will perform a luggage screening task where you will view X-ray images of luggage. Several of the bags will contain a knife (target) that you must try to identify. Sometimes you will have the assistance of an automated aid to help you in identifying if a knife is present. Bags will be screened on a computer screen while the automated aid assists in either highlighting potential areas in the luggage that look suspicious or tells you if it has detected a knife in a particular bag. Once you decide if a knife is present or absent in a bag you will be given feedback, on the computer, letting you know if you made a correct detection of a knife or not. You will complete a training session first, then you will be allowed a break before you begin the actual luggage screening task. After screening the series of bags you will be asked to complete a questionnaire regarding the luggage screening task you just completed before being released.

You will be seated in front of a computer for the entire duration of the task and given several breaks throughout the study. You have the option at any time to cease participation without penalty. If you say YES, then your participation will last for 1 hour in Room #336, Mills Godwin Building. Approximately 100 similarly situated undergraduate students will be participating in this study.

EXCLUSIONARY CRITERIA
You should be between the ages of 18 and 65 years, and have normal or corrected-to-normal vision. Also, to the best of your knowledge, you should not have any color blindness that would keep you from participating in this study.

RISKS AND BENEFITS
RISKS: If you decide to participate in this study, then you may face a risk of the common problems associated with computer usage such as eye strain or eye fatigue. The researcher has tried to reduce these risks by minimizing the amount of time in front of the computer and by allowing short breaks during the course of the experiment. Also, the researcher has removed all linking identifiers - data will be recorded under a participant number and will not be connected to your real identity in any way. As with any research, there is some possibility that you may be subject to risks that have not yet been identified.

BENEFITS: There are no direct benefits to participation. Indirectly, your participation will contribute to the development of better automated aid solutions for operators who work with automated teammates.

COSTS AND PAYMENTS

The researchers want your decision about participating in this study to be absolutely voluntary. There is no cost to participate and no monetary payment in this study. You will receive 1 research participation credit for participation.

NEW INFORMATION
If the researchers find new information during this study that would reasonably change your decision about participating, then they will give it to you.

CONFIDENTIALITY
All information obtained about you in this study is strictly confidential unless disclosure is required by law. The results of this study may be used in reports, presentations and publications, but the researcher will not identify you.

WITHDRAWAL PRIVILEGE
It is OK for you to say NO. Even if you say YES now, you are free to say NO later, and walk away or withdraw from the study at any time. Your decision will not affect your relationship with Old Dominion University, or otherwise cause a loss of benefits to which you might otherwise be entitled. The researchers reserve the right to withdraw your participation in this study, at any time, if they observe potential problems with your continued participation.

COMPENSATION FOR ILLNESS AND INJURY
If you say YES, then your consent in this document does not waive any of your legal rights. However, in the event of harm, injury or illness arising from this study, neither Old Dominion University nor the researchers are able to give you any money, insurance coverage, free medical care, or any other compensation for such injury. In the event that you suffer injury as a result of participation in this research project, you may contact Dr. James Bliss at 757-683-4051 or Dr. Tancy Vandecar-Burdin the current IRB chair at 757-683-3802 at Old Dominion University, who will be glad to review the matter with you. You may also contact the Office of Research (757) 683-3460.

VOLUNTARY CONSENT
By signing this form, you are saying several things. You are saying that you have read this form or have had it read to you, that you are satisfied that you understand this form, the research study, and its risks and benefits. The researchers should have answered any questions you may have had about the research. If you have any questions later on, then the researchers should be able to answer them.

And importantly, by signing below, you are telling the researcher YES, that you agree to participate in this study. The researcher should give you a copy of this form for your records.

Subject's Printed Name & Signature	**Date**

APPENDIX B

Demographic Survey

PARTICIPANT BACKGROUND INFORMATION FORM

Participant #________ Date:________ Time:________ Group: ________

The purpose of this questionnaire is to collect background information for participants in this experiment. This information will be used strictly for this experiment and for research purposes only. Please complete each item to the best of your ability.

1. Age _____

2. Sex: Male Female Other

3. Have you ever been diagnosed as having a deficiency in your visual acuity? _____(Y/N)

4. If yes, do you have correction with you? (i.e. glasses, contact lenses, etc.)? _____(Y/N)

5. Have you ever been diagnosed as color deficient or color blind?______(Y/N)

8. Indicate the average number of **hours per *week*** you spend using computers (personal and work combined): __

9. Circle the number that corresponds to how **confident** you are working with **computers**:

1 2 3 4 5 6 7

Low Average High

APPENDIX C

Instructions

In this task you will view a series of X-ray images of luggage. You will be shown a target (knife) that you must identify in the bags as you screen them. The X-ray image will appear for an unspecified amount of time and then it will disappear from the screen. Once the image disappears you will be asked to make a decision whether to "stop" the bag, indicating that a knife was present, or "pass" the bag if no knife was seen. If you saw a knife in the bag then you will need to click the left mouse button, indicating that you want to "stop" the bag. If you did not see a knife please click the right mouse button to indicate that you want to "pass" the bag. Once you make a decision you will be asked to rate your confidence in the decision you just made to "stop" or "pass" the bag.

You will screen three different sets of luggage and for two of these sets you will have the assistance of an automated aid to help you in detecting a target. Each set will begin with a training session first. You will be asked to screen 10 practice bags before beginning the actual luggage screening task to familiarize yourself with the visual search task and the automated aid that might be assisting you. The aids that you will be working with are 90% reliable.

Screening aid: As you perform the luggage screening task for this set of luggage you will have the help of an automated aid that will assist you in screening the bags for a target. When an X-ray image appears on the screen the aid will also appear in the form of a yellow circle. This yellow circle indicates where in the image a target is most likely to exist. You will still need to search that area and determine if a knife is present/absent. As before when the image disappears you will be asked to make a decision to "stop" the bag or "pass" it. Once you have screened all the bags in this set you will be allowed a short break to stand and stretch before continuing.

Decision aid: Now you will perform the luggage screening task but this time you will have the help of a decision aid. When an X-ray image appears on the screen a perimeter around the bag will also appear in the color of green or red. The perimeter of the image will be highlighted in red if a target has been detected by the aid and you need to “stop” the bag. It will be highlighted in green if no target has been detected and you need to “pass” the bag. Once the image disappears you will need to decide whether to click the left mouse button to “stop” the bag or the right mouse button to “pass” the bag. Once you have seen all the images in this section you will be allowed a short break to stand and stretch before continuing.

The experiment will last approximately an hour. During the experiment you will be seated in front of this monitor. Once you have seen all the images, you will be debriefed and allowed to leave. Pleas let the experimenter know now if you have any questions.

APPENDIX D

DEBRIEF INFORMATION

Thank you for your participation. You have just participated in a study that is attempting to find out how individuals rely on automated aids to make decisions and help them in a visual search task. Individuals are impacted by the aids that they work with; we are interested in assessing how completely they rely on these automated aids to make decisions. Specifically we are interested in examining whether individual performance is impacted to a greater degree when a decision aid is unreliable or when an aid that is simply directing attention is unreliable. The ultimate goal of this research is to help develop more effective automated aids to work with operators as they perform tasks such as the luggage screening search task you did today.

We don't know the results yet, but we hope that your participation will help us better understand how people make decisions when working with automated aids. If you have any questions or comments about this work or would like to be informed of the results, feel free to e-mail Molly Liechty (mcris005@odu.edu).

VITA

Molly M. Liechty

Research Scientist/ Leidos
2620 Q Street, Bldg 852, Area B
Wright Patterson AFB, OH 45433
Molly.Liechty@leidos.com
Old Dominion University
Human Factors Psychology
250 Mills Godwin Life Sciences Bldg
Norfolk, VA 23529

EDUCATION

Degree	Institution	Year	Field of Study
Ph.D.	Old Dominion University	2011- Dec 2019	Human Factors Psychology
M.S.	Old Dominion University	2009-2011	Psychology
B.A.	Purdue University	2000	Psychology

Dissertation: The Effect of Differing Degrees of Automation and Reliability on Simulated Luggage Screening Performance
M.S. Thesis: Contextual Cueing Effects in Visual Threat Detection

Advisor: Yusuke Yamani

Applied Skills

- Programming languages (Psychopy, E prime, R, SPSS)
- Statistical knowledge (Null Hypothesis Testing, Bayesian Analyses)
- Data Acquisition (Laboratory Research, Eyetracking, Surveys, User Testing, Field Research)

INTERNSHIPS

- NASA Langley Research Center, 1/2016 – 8/2016
- Newport News Shipyard, 4/2013 – 4/2015

PUBLICATIONS, POSTERS, AND PRESENTATIONS

Available upon request.