Trust and Reliance on an Automated Target Recognition System for Underwater Mine Detection

by

Shala Knocton

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
June 2020

Table of Contents

List of Figures

Abstract

Automated target recognition (ATR) systems are typically designed to operate using a high sensitivity and a liberal decision criterion to reduce the risk of missing a target. The high number of false alarms that occur as a result of this design tend to lead to a decrease in operator trust and reliance. The purpose of this study was to determine how changing or informing a user of the false alarm rate of an ATR system affects the user's level of trust and reliance in the system and the user's performance during an underwater mine detection task. When not informed of the false alarm rate, the number of false alarms made by the system had a significant effect on the participants' response bias. In addition, when informed of the false alarm rate, the participants had greater trust in the system and a more consistent response bias. These results suggest that informing a user of the false alarm rate of an automated system may positively influence the level of trust and reliance the user has in the aid.

List of Abbreviations and Symbols Used

ANOVA – analysis of variance

ATR – automated target recognition

Auto – automation

β – response bias

DRDC – Defense Research and Development Canada

d' – sensitivity

F – F-statistic

FA – false alarm

HSD – honestly significant difference

I – informed

M – mean

MCM – mine countermeasure

MS_W – mean square within

NI – not informed

n_k – number of participants in each group

η_p^2 – effect size

p – p-value

q – studentized range distribution

REB – Research Ethics Board

SEM – standard error of the mean

SDT – Signal Detection Theory

SSHRC – Social Sciences and Humanities Research Council

UUV – unmanned underwater vehicle

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Heather Neyedli. The endless amount of knowledge, guidance, encouragement and support that you have provided to myself and other students is truly inspiring and I feel incredibly grateful to have gotten the chance to work with you over the past two years. To my committee members Dr. Lori Dithurbide and Dr. John Kozey, as well as my external examiner Dr. Aren Hunter, thank you for the invaluable feedback, insight, and unique perspectives you have given me throughout this process. To Warren Connors, thank you for taking the time to create the material required for this project and for sharing your expertise on mine detection and ATR systems. Finally, to my friends and family, thank you for your unwavering love and support.

Chapter 1: Introduction

Underwater mines continue to be a common threat in maritime warfare. The explosive devices are low in cost, easy to access and lay, provide a constant threat, and are difficult to detect, making them an ideal choice for a weapon (CF, 2011; Pavlovic et al., 2012). To result in maximum damage, mines are often strategically placed along routes where ships are expected to travel (Ho, Pavlovic, Arrabito & Abdalla, 2011). Since the ocean is significantly used for transportation, the threat of a mine could lead to major consequences such as ships being damaged or destroyed, people being injured, or disruptions in the trading of goods (CF, 2011). To reduce the risk of encountering a mine and the consequences that may arise, measures must be taken to locate and identify the explosive devices so they can be avoided.

Over the past few decades, unmanned underwater vehicles (UUV) have increasingly been used in mine countermeasure (MCM) operations (Ho et al., 2011). MCM operations often include the detection, classification, identification and sometimes neutralisation of underwater mines (Pavlovic et al., 2012). During the detection and classification process, UUVs are typically used to visualize the seafloor by using side-scan sonar (Ho et al., 2011; Kessel & Myers, 2005; Pavlovic et al., 2012). After retrieval of the sonar data, a sonar operator aboard a ship is responsible for scanning the seafloor imagery and classifying any foreign objects detected as mine-like or non-mine-like (Ho et al., 2011).

The process of manually detecting foreign objects in the sonar imagery and classifying the objects as potential targets can be mentally fatiguing for sonar operators, making the task difficult to complete over time. Along with the relatively rare presence of mines, the seafloor is home to rocks, seagrass and debris. The ability for mines to be hidden within the seafloor clutter or disguised as a seafloor object may make it more challenging for a mine to be located. In addition, the quality of the sonar imagery collected by the UUV may also make locating mines challenging. Mine-like objects are typically recognized by sonar operators by using a set of characteristics associated with the appearance of a mine when using sonar imagery. If the side-scan imagery is poor, the operator may not be able to distinguish a mine-like object from a non-mine-like object (Kessel & Myers, 2005).

Detecting or classifying mines in an image that is cluttered and/or of poor quality may lead to an increase in the workload experienced by the sonar operator. The operator may also be forced to operate under conditions of uncertainty resulting in a potential decrease in accuracy and performance. To help improve the accuracy of the detection and classification process, automated target recognition (ATR) algorithms are being created (Ho et al., 2011; Kessel & Myers, 2005; Myers, 2009). ATR algorithms developed for mine detection operate as an automated system that can be used to assist or replace a human operator in determining whether a foreign object detected in the sonar imagery is a potential target or not (Kessel & Myers, 2005).

Algorithms designed to aid with mine detection typically work in a similar manner as sonar operators by using known characteristics associated with the presence of a mine in sonar imagery to determine if an object is mine-like or non-mine-like. Some of the characteristics of interest that the algorithm may be designed to recognize include the shape, size and shadow of the object (Myers, 2009). If the ATR aid detects an object that appears to have the characteristics of a mine, the aid may classify the object as mine-like.

Depending on the level of automation programmed into the ATR system, the sonar operator may be notified that the system has detected a potential target in the side-scan sonar data. The notification could be received by the operator in the form of a visual cue to the area in the sonar imagery where the object of interest is present. Once being cued to the potential target, the sonar operator must decide if the object needs further examination before a final classification is made (Ho et al., 2011). Regardless of whether a manual investigation of the object is performed, the operator will be responsible for agreeing or disagreeing with the classification made by the algorithm.

In classifying an object as mine-like or non-mine-like, the ATR system may result in one of the following outcomes: hit, false alarm, miss or correct rejection (Wickens, Hollands, Banbury & Parasuraman, 2013). A hit is when the ATR algorithm correctly identifies a mine as present, a false alarm is when the algorithm states a mine is present when there is no mine, a miss is when the algorithm says a mine is not present when there is a mine and a correct rejection is when the algorithm correctly identifies a mine as not present. The sensitivity of the ATR system designed, as well as the decision criterion used, may alter the proportion of each outcome that occurs. In general, an ATR system with a greater sensitivity will result in more foreign objects being classified as mine-like, leading to a greater number of hits and fewer misses. In the case of an ATR system with a lower sensitivity, fewer foreign objects are expected to be classified as mine-like, leading to a fewer amount of hits and more misses. Since the algorithms are designed to aid in the detection of mines, a system with a higher sensitivity would allow for more mines to be identified and fewer missed. Although using the high sensitivity appears to be more beneficial, sonar operators tend to have low trust in the aids due to the high number of false alarms that occur as a result (Kessel, 2005; Kessel & Myers, 2005).

In addition to using a system with a higher sensitivity, using a more liberal decision criterion is also suggested to result in a high number of false alarms, but may also reduce the number of misses that occur (Allendoerfer, Pai & Friedman-Berg, 2008). The high number of false alarms typically result when using this type of criterion due to less evidence being required for the system to indicate that a signal is present (Wickens et al., 2013). In a mine detection task, this may result in the automated target recognition system classifying an object that only has some of the suggested mine-like characteristics as a potential target. Since the actual presence of mines along the seafloor is rare, a liberal criterion will result in a large number of objects incorrectly being classified as potential targets by the system (Wickens et al., 2013). To result in a fewer number of false alarms, the decision criterion could be altered to create a system with a similar overall sensitivity but requires more evidence to indicate that a signal is present. This type of decision criterion would operate more conservatively and may only classify an object that has all of the suggested mine-like characteristics as a potential target. Although the conservative criterion may result in fewer false alarms, the system may miss more potential targets than would be missed using the liberal criterion due to the objects not having all of the required mine-like characteristics (Allendoerfer et al., 2008).

To determine if it is the high number of false alarms specifically that is leading to the decrease in trust experienced by the sonar operators, an ATR system with different false alarm rates should be examined. It is important to understand why and under what conditions a human operator is expected to trust and rely on the automated system to achieve the benefits associated with using automation (Hollands & Neyedli, 2011). An assumption that is often made when developing automation to aid a human operator with a task is that the human-automation team will perform better than the operator or the automated system would on their own (Dzindolet, Peterson, Pomranky, Pierce & Beck, 2003; Dzindolet, Pierce, Beck & Dawe, 2002). The issue with this assumption is that once trust is broken between the automated system and the human operator, reliance on the automated system tends to decrease leading to misuse or disuse of the aid (Colebank, 2008; Wickens et al., 2013). These human performance issues have been suggested to arise when human operators are working with automated systems due to the devices being designed from a technology-centered perspective (Wickens et al., 2013). Therefore, instead of solely focusing on improving the algorithms or cues generated by the automated system, human performance patterns must also be considered.

Purpose

The purpose of this study was to determine if the false alarm rate of an automated target recognition system affects a user's trust in the system, confidence in their own abilities and mine detection performance during an underwater mine detection task, and whether informing the user of the false alarm rate of the system affects each of these factors. To test this purpose, two groups of participants performed an underwater mine detection task with the help of an ATR system. The reliability and sensitivity of the aid was held constant across groups and across the experimental sessions, but the false alarm rate of the system changed part way through the study. One group of participants was informed of the change in the false alarm rate and one group was not.

Hypotheses

It was hypothesized that the users' level of trust in the automated system would be higher during the low false alarm rate condition compared to their level of trust during the high false alarm rate condition. Furthermore, it was expected that the users' level of trust would be higher when they are informed of the automated system's false alarm rate compared to when they are not informed of the false alarm rate of the system. It was also expected that the false alarm rate condition and the group the participant was assigned to (informed vs. not informed) would affect the users' level of confidence in their own abilities and performance during the underwater mine detection task. Due to the limited research on the effect false alarm rate has on an individual's level of confidence or performance, non-directional hypotheses were used for these measures. Given the present research on trust and reliance on an automated target recognition system, the next chapter will discuss some of the literature on automation, factors suggested to influence trust and reliance, and research related to cueing and target detection.

Chapter 2: Review of Literature

Automation

The prevalence of automation within the world has significantly increased since the 1960s (Wickens et al., 2013). Today, automation can be found in all aspects of life from technology such as cell phones and cars to specialized systems in hospitals, aircraft and the military (Hollands & Neyedli, 2011; Madhavan, Wiegmann & Lacson, 2003; Wickens et al., 2013). The increasing desire for automated systems to perform tasks previously executed by humans may be attributed to the recent rise in advanced technology and operational challenges (Colebank, 2008; Parasuraman & Riley, 1997; Parasuraman, Sheridan & Wickens, 2000). It is proposed that automated systems have the ability to complete tasks that humans are unable to perform or do so poorly, to assist humans in completing tasks where human limitations may arise, to save money by not having to pay for human labor or training costs involved in the completion of tasks, as well as to increase productivity (Satchell, 1998; Wickens et al., 2013).

In completing a task, automation can be used to assist with four general processes including information acquisition, information analysis, decision and action selection, and action implementation (Balfe, Sharples & Wilson, 2015; Parasuraman et al., 2000). Within each of these processes, the level of automation applied can vary across a continuum from low (full manual control) to high (full automated control) (Parasuraman et al., 2000; Röttger, Bali & Manzey, 2009; Wickens et al., 2013). For information acquisition, a lower level of automation may involve the gathering of information (e.g., side scan sonar may be used to capture images of the sea floor) and a higher level may be associated with the organization or filtering of the information (e.g., the images captured may be prioritized according to whether or not objects appear to be present along the sea floor, as indicated by the strength of the return echoes or cast shadows) (Parasuraman et al., 2000; Wickens et al., 2013). A lower level of information analysis automation may involve the use of algorithms to process the information to make future predictions (e.g., an algorithm may be developed to scan the sonar images and highlight areas that may be suspicious) (Parasuraman et al., 2000; Yin, Wickens, Pang & Helander, 2011), where integration of the information may be included when using higher levels of information analysis (e.g., the algorithm may provide more distinct information to the human user such as whether it thinks a suspicious object is threatening or not). For decision and action selection, a level of automation from the lower to the higher end of the continuum may be associated with a decrease in the number of decision choices provided by the automated system (e.g., the automated system may provide a user with a full list of potential responses, such as "the object may be a rock, a log, debris or a mine", or the system may provide the user with a single best response, such as "the object may be a mine"). Similarly, an increase in the level of automation for action implementation may correspond to a decrease in the amount of manual activity used to execute a response (e.g., for a low level of automation a human may be fully responsible for carrying out a desired action or response, such as indicating that a suspicious object detected in the sonar data is a mine, where for a high level of automation the human may not take part in carrying out the action or response) (Parasuraman et al., 2000; Wickens et al., 2013).

In assisting with information acquisition, information analysis, decision and action selection and/or action implementation, automation use may be able to maintain the efficiency and efficacy of performing daily tasks (Colebank, 2008; Parasuraman &

Manzey, 2010; Wickens et al., 2013), improve safety (Parasuraman & Manzey, 2010; Wickens et al., 2013), reduce workload (Balfe et al., 2015; Parasuraman & Manzey, 2010; Röttger et al., 2009; Wickens & Hollands, 2000; Wickens et al., 2013), reduce costs (Wickens & Hollands, 2000), and compensate for human limitations (Re & Macchi, 2010; Wickens & Hollands, 2000). To achieve these benefits, it is important that the human users rely on the automation appropriately (Hollands & Neyedli, 2011), the design of the system encourages appropriate use, and proper training is received by the users (Parasuraman & Manzey, 2010). If the automation requires a high level of monitoring and cognitive effort by the user, the benefits of using the system will be diminished and errors will likely occur (Parasuraman & Riley, 1997). Some of these errors include automation complacency, automation bias and commission error (Balfe et al., 2015).

Automation complacency refers to a reduction in the user's ability to detect malfunctions with the automated system while under automated control (Parasuraman & Manzey, 2010). This typically occurs when the user is expected to attend to multiple tasks at the same time (Parasuraman & Manzey, 2010; Wickens et al., 2013), as well as in situations where the automation appears to have a consistently high level of reliability (Bahner, Huper & Manzey, 2008). Automation bias is the tendency for users to favor the information provided by an automated aid when different information is provided by a non-automated aid (Dzindolet et al., 2002; Parasuraman & Manzey, 2010; Parasuraman & Riley, 1997). The bias towards information provided by automated devices tends to occur in various settings regardless of any training or skills obtained by the user. Although training and user ability does not appear to have an effect, the literature suggests that automation bias may be influenced by the level of automation used, the reliability of the aid and how responsible the user feels for the outcome of the human-automation interaction (Parasuraman & Manzey, 2010). Commission error is when the user complies with incorrect information provided by the automated device (Bahner et al., 2008; Parasuraman & Manzey, 2010). Incorrect information may be used if the user does not attempt to determine if the information is correct or incorrect, or if the user is biased towards information provided by the automation (Bahner et al., 2008). Automation complacency, automation bias and commission errors are examples of situations where human users have developed inappropriate reliance on automated systems.

Reliance and Trust

According to the literature, the tendency for humans to develop inappropriate reliance on automation is fairly common (Dzindolet, Pierce, Beck, Dawe & Anderson, 2001). In general, users tend to initially have high expectations in automation regardless of the system's abilities (Bagheri & Jamieson, 2004). As a result of the high expectations, errors often occur due to the user relying on the automated aid when they should not rather than ignoring the aid when they should (Dzindolet et al., 2003; Dzindolet et al., 2001). The over reliance on imperfect automation is suggested to occur because the users want to save their resources (Wickens & Dixon, 2007) and use the least amount of cognitive effort when making decisions (Wickens & Hollands, 2000). Several factors suggested to influence reliance behaviors include trust in automation (Dzindolet et al., 2003; Lee & Moray, 1992; Lee & Moray, 1994; Lee & See, 2004), user self-confidence (Lee & Moray, 1994; Lewandowsky, Mundy & Tan, 2000; Madhavan & Wiegmann, 2004), attitudes toward or liking of automation (Merritt, Sinha, Curran & Ilgen, 2011), system reliability (Dzindolet et al., 2001; Parasuraman & Riley, 1997) and feedback (Lee & Moray, 1994; Wickens et al., 2013).

Trust is a multidimensional concept (Colebank, 2008) that can be influenced at the individual, organizational and cultural level. The individual level is concerned with an individual's past experiences with trust and reliance, and how these experiences affect the individual's specific tendency to trust and interpret information. Trust may be influenced at the organizational level through human interactions and discussion regarding the supposed trustworthiness of a specific agent. At the cultural level, the norms and expectations within society may impact an individual's propensity to trust (Lee & See, 2004). Lee and See define trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (2004). Therefore, trust can be identified within each level as situations where individuals allow themselves to be vulnerable around others with the assumption that a positive result will occur. The vulnerability seen in these situations can be toward other humans, as well as toward automated systems (Mayer, Davis & Schoorman, 1995).

Individuals that have a tendency to be more trusting in general are suggested to trust other humans and automation more appropriately than those who are less trusting (Lee & See, 2004; Mayer et al., 1995). The degree to which an individual will trust another human or automated aid is influenced by the traits of the trustor and how the incoming information is interpreted. Three characteristics identified in the literature that impact user trust include ability, benevolence and integrity. Ability refers to the characteristics or skills of the human or automated aid, benevolence is the aid's desire to do good and integrity is identified as the level of honesty or morality of the aid (Mayer et al., 1995).

The ability, benevolence and integrity of a human or automated aid can be observed or learned by the human user by taking a closer look at the performance, process and purpose of the aid. Performance is indicative of the ability of the aid and is concerned with aspects of the aid's behaviour, such as the aid's reliability and predictability. Benevolence can be determined by developing a deeper understanding of how the aid operates and the algorithms or underlying processes used when the aid is acquiring information, analyzing data, selecting an appropriate decision and/or implementing a response. The purpose of the aid can be used to help identify the aid's level of integrity by allowing the user to understand why the aid was developed and what the aid was designed to do (Lee & See, 2004). As the users gain more information about and work with the aid, an appropriate level of trust should occur (Mayer et al., 1995).

It is suggested that many commonalities exist between the trust a human has in another human and the trust a human has in an automated device (Hollands & Neyedli, 2011; Madhavan & Wiegmann, 2004). Although this may be true, it is important to note that slight differences in trust may exist between the two types of interactions as well (Madhavan & Wiegmann, 2004; Madhavan & Wiegmann, 2007). According to the literature, a user's level of trust appears to have a greater influence on the user's decision to rely on an automated system compared to the user's decision to rely on another human (Lee & See, 2004; Lewandowsky et al., 2000). This may be due to a user perceiving their responsibility in a human-automation relationship to lie within themselves and their responsibility in a human-human relationship to be dispersed between themselves and the other person (Lewandowsky et al., 2000). If a user believes they are solely responsible for the outcome of a human-automation interaction, it may explain why users tend to rely on automated systems they trust and disregard systems they do not (Lee & See, 2004).

Trust also has a tendency to breakdown more quickly when dealing with an automated aid compared to a human aid (Madhavan & Wiegmann, 2004). In general, when an automated aid is compared to a human aid, the mechanical counterpart is often perceived as more reliable (Dijkstra, 1999) and is expected to perform better than the human aid (Dzindolet et al., 2003). For these reasons, human users have a tendency to foster initial biases toward the automation (Bagheri & Jamieson, 2004; Dzindolet et al., 2002; Hollands & Neyedli, 2011; Lyons & Stokes, 2012; Merritt, Hall, Louis, Curran & Ilgen, 2015) and associate more realistic expectations with the human aid (Madhavan & Wiegmann, 2007). Therefore, when an error or fault occurs, the high expectations associated with the automated system result in the user's trust in the system to decrease faster than their trust in the other human (Lewandowsky et al., 2000; Madhavan & Wiegmann, 2004).

The decrease in trust humans experience toward automated devices appears to be exacerbated following the first failure. Prior to the first error made by an automated aid, users tend to be in a state of over-trust due to their high expectations in the aid's abilities (Wickens et al., 2013). These expectations appear to be violated upon the automation's first failure, sending the user into an instant state of mistrust (Dzindolet et al., 2003; Wickens et al., 2013). As a result, humans tend to ignore the automated system even if it is accurate and reliable (Dzindolet et al., 2002; Madhavan et al., 2003; Parasuraman & Riley, 1997). Over time, trust is expected to gradually recover to a level approximating appropriate trust in the automation's capabilities (Lee & Moray, 1992; Wickens et al., 2013; Yeh, Merlo, Wickens & Brandenburg, 2003).

Allowing human users to gain exposure to errors in training or practice may aid in reducing the rapid level of distrust users experience following the first automation failure (Parasuraman & Manzey, 2010). If humans are aware that automated aids are not perfect and are knowledgeable on when and why errors may occur, they may be more likely to trust and rely on the aid (Bagheri & Jamieson, 2004; Dzindolet et al., 2003; Dzindolet et al., 2002; Dzindolet et al., 2001; Parasuraman & Riley, 1997). In circumstances where the error made by the aid is obvious and detected by the user, trust and reliance may be undermined regardless of any information provided. This may occur due to the user believing that they can perform better than the automated system (Madhavan et al., 2003). In other words, the user's level of confidence in their own abilities may exceed the user's trust in the automated system.

The level of self-confidence possessed by a human user has a significant influence on whether the user trusts and relies on an automated aid (Lewandowsky et al., 2000; Madhavan & Wiegmann, 2004). In conditions where trust in automation exceeds the user's confidence in their own abilities, automation tends to be used. If instead the user's level of self-confidence is greater than their trust in automation, the task tends to be completed manually (Dzindolet et al., 2002; Lee & Moray, 1992; Lee & Moray, 1994; Parasuraman & Riley, 1997). Any changes that may occur in the level of trust in automation or user self-confidence is suggested to correspond to an associated change in automation use (Lee & Moray, 1994). Due to the ability of these changes to influence reliance on automation, designers should consider how different aspects of an automated system may influence the level of trust and/or self-confidence a user may experience and adjust these aspects accordingly to encourage appropriate use. Information should also be made available to the user regarding both automated and manual performance to prompt any adjustments in the user's level of trust and self-confidence (Lee & Moray, 1994).

In addition to the levels of self-confidence experienced, how much the user likes an automated aid is also proposed to influence trust and reliance. Whether or not the user likes an automated device can be determined by examining the level of positive attitude the user has toward the system. In the early stages of interaction, it is suggested that reliance is heavily dependent on whether the user likes the automation. As the user and the automated device continue to interact over time, reliance is more influenced by trust (Merritt et al., 2011). The tendency for one to like an automated aid is suggested to be influenced by individual differences as well as the various components that make up the device (Merritt et al., 2015). Knowledge of these individual and automated characteristics may be useful in designing an automated device that users will like and rely on.

Another factor suggested to influence trust and reliance on automation is the reliability of the system (Wickens et al., 2013). An increasing amount of research has shown that human operators are sensitive to slight changes in reliability levels and the types of errors made by an automated system (Madhavan & Wiegmann, 2007). In general, as the reliability of an automated aid increases there is a subsequent increase in user trust and performance. The opposite results in user trust and performance are to be expected when the reliability of the aid is decreased (Dzindolet et al., 2003; Hollands & Neyedli, 2011; Madhavan et al., 2003; Parasuraman et al., 2000). According to the literature, a decrease in the aid's reliability below a level of 70% results in a decrease in performance below what would be expected if no automation was used (Wickens & Dixon, 2007). Although using a higher reliability appears to be more beneficial for user trust and performance, it is also important to note that automated aids with a higher level of reliability tend to lead to a reduction in the amount of time the user monitors the aid as well as a decrease in the attention the user has toward the raw data the automation is processing (Parasuraman et al., 2000; Wickens et al., 2013). This decrease in effort toward overseeing the processes of the aid may also be seen in cases where the operator perceives the automated aid as a teammate and therefore believes the responsibility of the outcome is dispersed between themselves and the aid (Parasuraman & Manzey, 2010).

Since the majority of automated systems are implemented to aid human users with tasks that are performed under conditions of uncertainty, there will be situations when the automation is incorrect (Hollands & Neyedli, 2011). In these situations, imperfect automation tends to be underestimated resulting in a decrease in user trust and reliance (Madhavan et al., 2003; Parasuraman & Riley, 1997; Wickens & Hollands, 2000). Disclosing the level of reliability of the automated system to the user is proposed to increase performance (Bagheri & Jamieson, 2004; Hollands & Neyedli, 2011; Neyedli, Hollands & Jamieson, 2011; Wang, Jamieson & Hollands, 2009). By providing information regarding the reliability of the automation, the user may be able to adjust their expectations and develop appropriate trust and reliance in the system (Bagheri & Jamieson, 2004; Wickens & Dixon, 2007). This information may provide additional benefit in conditions where the reliability of the system changes over time in a context-sensitive manner. In these situations, users provided with real time reliability data (Neyedli et al., 2011) and knowledge regarding the sensitive nature of the system to different contexts, may improve performance and reliance when using automation that is imperfectly reliable (Bagheri & Jamieson, 2004). It is important to note that although some automated systems are not perfect, their assistance still may be useful throughout the process of completing a task. In complex environments or situations, the task components where automation use may be beneficial include signal detection, visual/attentional guidance and response selection (Madhavan & Wiegmann, 2004; Madhavan & Wiegmann, 2007; Parasuraman & Manzey, 2010; Wickens & Dixon, 2007; Wickens et al., 2013). Given the focus on mine detection, the role of automation in cueing and target detection will be discussed in further detail.

Cueing and Target Detection

In some situations, a signal or event may be so apparent that detection occurs almost instantaneously, while in other situations the ability to discern a signal or event from background noise may not be so easy. The process of detecting a signal or event during the latter situation is often modeled according to Signal Detection Theory (SDT) (Green & Swets, 1966; Phillips, Saks & Peterson, 2001; Wickens et al., 2013). SDT is a framework that can be used to measure the performance of human, automation or human-automation teams in detecting signals, as well as to measure user reliance patterns toward an automated device (Green & Swets, 1966; Hollands & Neyedli, 2011; Parasuraman, 1987; Sorkin & Woods, 1985; Swets, 1998). Two components of SDT that correspond to performance and reliance measures are sensitivity and response bias.

Sensitivity refers to the ability of a detector (human or otherwise) to identify an event or signal within background noise (Green & Swets, 1966; Neyedli et al., 2011; Wang et al., 2009). Several factors suggested to influence the ability to discriminate between signal and noise include the experience or skill obtained by the user, the level of technology programmed into the automated system, the quality or resolution of the stimuli being examined and the salience of the signal that the user or automated system is trying to detect (Phillips et al., 2001; Wickens et al., 2013). When dealing with SDT, the two available responses are that the signal is present or that the signal is not present and the four potential outcomes are a hit, miss, false alarm or correct rejection (Figure 1). If the user or automated system has a higher level of sensitivity, a greater amount of hits and correct rejections are expected to occur (Figure 2a). If instead the user or automated system has a lower sensitivity, more false alarms and misses are to be expected (Figure 2b) (Phillips et al., 2001; Wickens et al., 2013).

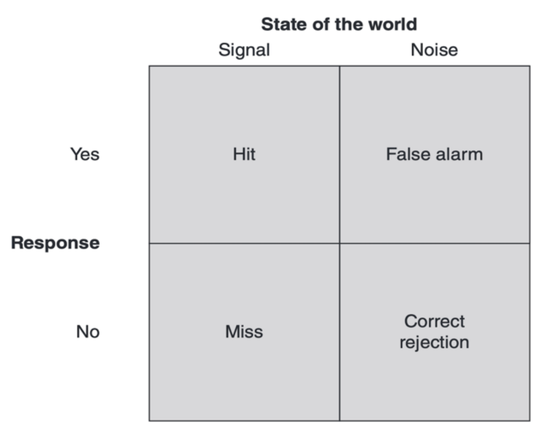


Figure 1. Four outcomes for Signal Detection Theory. Wickens, C. D., Hollands, J. G., Banbury, S. & Parasuraman, R. (2013). *Engineering Psychology and Human Performance* (4th ed.). Pearson Education. (Figure 2.1).

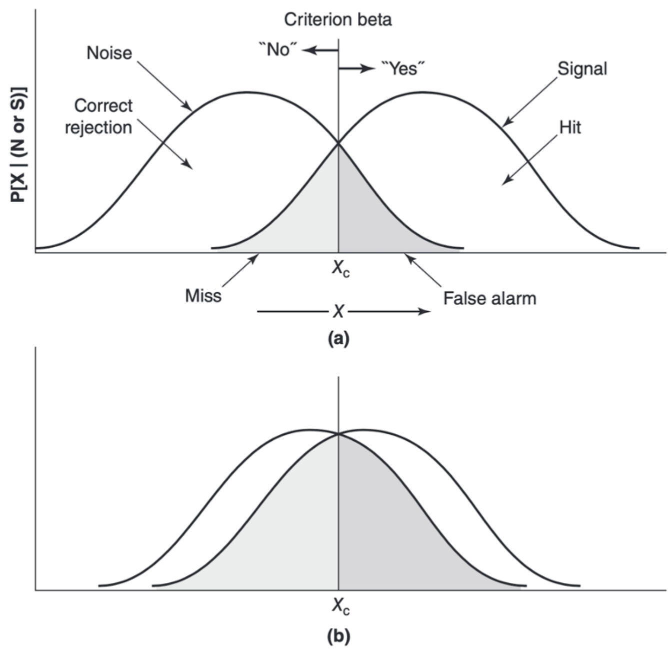


Figure 2. Sample Signal Detection Theory distributions for (a) high sensitivity and (b) low sensitivity. Wickens, C. D., Hollands, J. G., Banbury, S. & Parasuraman, R. (2013). *Engineering Psychology and Human Performance* (4th ed.). Pearson Education. (Figure 2.3).

In a detection task where two users have the same sensitivity, different outcomes may result due to the users having different response biases in determining whether a signal is present or not (i.e., the criterion beta line in Figure 2 shifts to the left or to the right) (Phillips et al., 2001). Response bias refers to the patterns in user response (Green & Swets, 1966; Neyedli et al., 2011; Wang et al., 2009). In general, users tend to vary in response bias by being more liberal (criterion beta line shifts to the left) or more conservative (criterion beta line shifts to the right) in indicating that a signal is present. If a user is more liberal, they may be more likely to indicate that a signal is present, leading to a greater amount of hits at the expense of making a lot of false alarms. If the user is more conservative, they may be less likely to indicate that a signal is present, resulting in less false alarms from occurring but increasing the number of potential misses

(Parasuraman & Masalonis, 2000; Wickens et al., 2013). Two factors suggested to influence whether a user responds more liberally or conservatively in indicating that a signal or event is present are the probability of a signal or event occurring and the costs and benefits associated with saying that a signal or event is present or absent (Parasuraman & Masalonis, 2000; Phillips et al., 2001; Wickens et al., 2013).

A classic example that demonstrates response bias is when a radiologist is examining an X-ray of a patient. If a patient is unhealthy, the probability of the radiologist finding an abnormality in the X-ray and classifying it as a tumor may be more likely than if the radiologist is examining an X-ray of a healthy patient (Swets & Pickett, 1982). In these circumstances, the user may be more likely to develop a liberal response bias in indicating that a tumor is present in the patient that is unhealthy and develop a conservative response bias in indicating that a tumor is present in the patient that is healthy. In addition to the probability of a tumor being present, the user may also weigh the costs of incorrectly indicating that a tumor is present and the patient undergoing unnecessary chemotherapy or incorrectly indicating that a tumor is absent and the patient's condition getting worse or the patient passing away. Depending on which cost the operator believes is worse, they may respond liberally to ensure if the patient does in fact have a tumor that it does not progress, or they may respond more conservatively to ensure the patient does not go through dangerous and costly treatment that they may not need.

SDT can be applied to visual search tasks where the goal is locating an object or target of interest. Searching for a target is a task that most humans do on a daily basis (Wickens et al., 2013). This is especially true for humans with specialized jobs such as radiologists, industrial inspectors and military operators, who are responsible for detecting indistinct targets and classifying the targets as signal or noise (Drury, 2006). In addition to trying to detect or classify a faint target, some humans are also forced to identify potential targets in a search field with a large amount of noise or distracting elements. The combination of the faint targets and the noisy background can make signals very difficult to detect. To aid in the detection process during these situations, automated systems can be used to cue or guide an individual's attention to a potential target or area of interest (Wickens et al., 2013).

Automated systems used for attentional guidance typically work by using an internal decision criterion set by the designers of the system to determine if a potential signal is present in the stimuli the system is being used to examine. If the system detects a potential target, according to the criterion programmed by the designers, the system may alert or cue the user to the location in the stimuli where it believes the target is present. Users are responsible for monitoring the alerts or cues made by the automated system and must decide upon receiving a response if they are going to trust the cue made by the automated system or if they are going to seek more evidence before indicating that a target is present (Allendoerfer et al., 2008). An important consideration when designing automated systems is the decision criterion that should be used. If a more liberal decision criterion is chosen, the automated system may be able to detect more targets but may also make a lot of false alarms. If a more conservative decision criterion is used, the automated system may make less false alarms but may not be able to detect as many targets (Wickens et al., 2013).

Most automated systems used for alerting or cueing a user to a potential target tend to use a more liberal decision criterion due to the cost of missing a target being perceived as more negative than the cost of being alerted to a target that is not present. This is often the case when dealing with fire alarms or airplane traffic, where alerting people about a fire that is not occurring or alerting pilots to a plane that is not flying on their route is better than the consequences that may result if a fire is actually occurring and nobody is alerted or a plane is about to collide with another plane and the pilot is not informed (Phillips et al., 2001; Wickens et al., 2013). Even though in some circumstances the probability of an event or signal occurring is low, the high cost associated with missing the event or signal when it does occur tends to result in designers using a more liberal decision criterion regardless of the high number of false alarms that result (Parasuraman, Hancock & Olofinboba, 1997; Wickens et al., 2013).

Although using an automated system that results in more hits and fewer misses has many benefits, several negative consequences are suggested to occur due to the high number of false alarms that the system also produces. In the case where a user is in charge of monitoring both the automated system and the raw data the system is processing, frequent false alarms may result in the user having to continuously check the raw data to determine if the alert or cue is correct in identifying a signal or event as present (Dixon, Wickens & McCarley, 2007). When the operator is in charge of performing multiple tasks at a time, or monitoring more than one system, the frequent disruptions may result in the benefits associated with using the automated systems to be reduced due to the operator experiencing an increased workload having to continuously respond to the false positive alerts (Wickens et al., 2013). In addition, a high false alarm rate also tends to lead to the “cry wolf” effect (Breznitz, 1983; Sorkin, 1989), where users ignore the alerts over time, even when they may be correct (Allendoerfer et al., 2008; Parasuraman & Riley, 1997; Satchell, 1993; Sorkin, 1988). In these situations the user may develop a more conservative response bias to compensate for the automated system’s liberal decision criterion (Wickens et al., 2013). The “cry wolf” effect has led to several tragic incidents including the injury or death of 21 percent of patients on long-term ventilators due to the ventilator alarms being ignored or checked too late (Joint Commission, 2002), and the death of 100 people on board a Korean Airlines flight that crashed due to a collision avoidance system being disabled (Wickens et al., 2013).

Another negative consequence that tends to occur when users are working with automated systems that result in a high number of hits and few misses is attention narrowing or tunneling. Attention narrowing or tunneling is a phenomenon demonstrated by users when an automated system consistently cues the user to the correct location of a target when a target is present. Due to the automation being perceived as reliable, the visual search that is performed by the user tends to decrease. If the user is not examining the search field, they may be more likely to miss a target if it is not cued by the automated system (Wickens et al., 2013). This phenomenon has been demonstrated in multiple studies that looked at target cueing with soldiers, where when the soldiers were cued to a target and a more threatening un-cued target was also present, the soldiers tended to miss the more dangerous un-cued target (Yeh et al., 2003; Yeh, Wickens & Seagull, 1999). Attention narrowing has also been seen in flight studies where skilled pilots failed to notice a dangerous event that was visible through the airplane windshield due to solely focusing on the information provided by the automated system and not examining the raw data in the search field (Wickens & Alexander, 2009).

The situations described above where the users had the tendency to ignore automated systems that had high false alarm rates and missed targets when working with automated systems with high hit rates are examples of issues that may occur when an automated aid is not 100% reliable. When a user is working with an automated system that generates cues that are imperfectly reliable, the benefits of when the cues are correct and the costs of when the cues are incorrect are pronounced as the reliability of the cues increases toward 100% (Wickens et al., 2013). This usually occurs due to the tendency of users to over trust automated systems that consistently produce hits and correct rejections, and therefore feel they do not need to monitor the system or the information the system is processing. Due to the complex nature of most detection tasks, the ability for an automated system to result in only hits and correct rejections is extremely rare (Parasuraman & Masalonis, 2000). Therefore, when an automated system makes a false alarm or misses a target and the user is not monitoring the system, the error often occurs without being detected (Bagheri & Jamieson, 2004). Depending on how many errors the system is expected to make, the costs of continuously missing an error may lead to negative consequences. Informing the user of the reliability of the system or the number of errors that are expected to occur may encourage the user to rely on the system more appropriately (Neyedli et al., 2011).

If a user is aware of the reliability level at which an automated system is expected to operate, they may be more likely to adjust their expectations in the system to lead to more appropriate reliance on and monitoring of the aid (Bagheri & Jamieson, 2004). In a study by Dzdinolet and colleagues (2003), participants who were not given information regarding the reliability of an automated system developed unrealistic expectations about the capabilities of the aid. Similar results were found in a study by Bagheri and Jamieson (2004), where participants who were given information regarding the reliability of an automated system were able to develop more appropriate levels of trust and expectations in the system. In a study by Wang and colleagues (2009), the researchers examined the effect of aid reliability and reliability disclosure on an automated identification system. Results from this study revealed a positive influence of reliability disclosure on the levels of trust and reliance the users had in the automated aid.

To measure user reliance patterns on an automated device, an optimal response bias difference can be computed. This approach, based on SDT, can be used to determine if a user is relying on an automated system too much, or if a user is not relying on the system enough (Wang et al., 2009). As discussed at the beginning of this section, whether a user indicates that a signal is present or not is influenced by sensitivity and response bias. Each of these measures can be determined based on the user's hit and false alarm rate when performing a signal detection task. The optimal response bias of a user for any signal detection task can be determined if the rate of signals, as well as the costs and benefits associated with indicating that a signal is present, are known.

By assessing how humans respond to automation under different automated or environmental conditions, designers may be able to develop systems from a more human-centered perspective that will encourage appropriate trust and reliance. This is the aim of the present research, which deals with an automated system designed to assist users in an underwater mine detection task. To summarize the content explored in the introduction section, underwater mine detection typically uses side-scan sonar to visualize the sea-floor. Although using sonar data in water tends to be more beneficial than using data collected using light or other methods, the complex nature of the seafloor may make the quality of any data obtained difficult to interpret (Ho et al., 2011). Since the operator's ability to view the seafloor is dependent on the sonar data obtained, the poor quality of the imagery may make identifying or classifying mines a complicated task to complete. This may result in the user operating under conditions of uncertainty in determining if a signal present in the sonar imagery is a potential mine or benign seafloor clutter. Automated systems have been developed to aid the operator during these situations by detecting signals within the sonar data, cueing or guiding the operator to areas of interest and/or providing suggestions regarding which response should be selected when determining if a signal is a mine or not (Madhavan & Wiegmann, 2004; Madhavan & Wiegmann, 2007; Parasuraman & Manzey, 2010; Wickens & Dixon, 2007; Wickens et al., 2013). A specific automated system that has been created to aid with each of these task components is the automated target recognition (ATR) system (Ho et al., 2011; Kessel & Myers, 2005; Myers, 2009).

ATR is an algorithm-based system developed to assist or replace a human operator in detecting and classifying underwater mines (Kessel & Myers, 2005). The algorithms designed for ATR systems work by identifying key characteristics associated with the presence of a mine in sonar imagery to detect and classify signals (Myers, 2009). Using an automated system to aid with these task components may reduce the frequency and amount of information the operator seeks from the external environment in order to make an accurate decision (Röttger et al., 2009). In addition to reducing the amount of workload experienced by the operator, the systems may also aid in improving the overall accuracy and performance in completing the mine detection task. Although ATR aids are on average accurate and reliable, they are not perfect and sometimes make mistakes. These mistakes may include false alarms and misses. Since ATR systems typically operate using a higher sensitivity and a more liberal decision criterion to reduce the risk of missing a mine, the number of false alarms that occur as a result tend to lead to a decrease in the trust and reliance the operators have in the systems (Kessel, 2005; Kessel & Myers, 2005).

As previously discussed, automated cues that result in a high number of hits or false alarms tends to lead to the user missing un-cued targets or ignoring the aid altogether. If researchers are aware of the underlying processes used by the operators when deciding whether to rely on the suggestions provided by the automated system or to rely on themselves, the design and implementation of these systems can be altered to encourage appropriate trust and reliance (Dzindolet et al., 2003). To develop and implement an ATR system that operators will rely on, the relationship between the false alarm rate and the operators' trust must be further examined, along with other aspects that may influence trust and reliance. To gain more insight into the effect of false alarms on the user, an ATR system with a more liberal (high false alarm rate) and a more conservative (low false alarm rate) decision criterion will be used. The purpose of this manipulation is to determine the effect a conservative decision criterion may have on user performance in a mine detection task compared to the more typically used liberal criterion. The study will also aim to explore how being informed of the high and low false alarms rates of the system may influence the user's performance compared to those that are not provided with this information. How the effect of different false alarm rates and disclosure patterns on human performance will be measured are explained in more detail in the next chapter.

Chapter 3: Methods

Participants

Seventy healthy adults (Age: M = 20.13 years, SD = 2.04; Gender: 55 female, 14 male, 1 no gender specified) were recruited from the Dalhousie University community using the Dalhousie Undergraduate Psychology Pool (Psychology Participant Pool (SONA) Contact Sheet – Appendix A). Those interested in the study were contacted and encouraged to ask questions to clarify information about the study as needed. To be eligible to participate, participants had to be between the ages of 18-55 and had to have normal or corrected to normal vision. The age range of 18-55 corresponds to the age of most sonar operators and the participants had to have normal or corrected to normal vision in order to detect the mines in the sonar images. Recruited participants were asked to read over an Informed Consent Form (Appendix B) and sign an Informed Consent Signature Page (Appendix C), SONA Signature Page (Appendix D) and an optional Age and Gender Form (Appendix E) before beginning the study.

In the consent form, participants were informed that they may withdraw from the study throughout the session at any given time and without any limitations by communicating to the experimenter that they no longer wished to continue. Participants were also notified in the consent form that they would receive \$10 if they achieved a significantly high level of performance in completing the study and that performance would be based on the total number of correct mines identified during the task. At the end of the study, all participants received \$10 regardless of performance so not to disadvantage participants who may have been assigned to a potentially less advantageous group. This performance bonus was implemented to provide an increased level of urgency and motivation to the participants while completing the task. The participants were also awarded two credit points to a psychology course.

Apparatus and Measures

Apparatus

A computer-based simulator was used for this study. Participants were seated in front of a monitor and interacted with a simulation using a computer mouse. The main simulation display showed sonar images of the sea floor. The task of the participants during the simulation was to identify whether a mine was present or absent in each of the sonar images. The participants entered each of their responses by clicking an area in the sonar image where they believed a mine was present or by clicking a black box in the top right-hand corner of the image labeled "NO MINE" if they believed a mine was not present.

During some blocks of trials an automated target recognition system aided the participants. For these blocks, a rectangle appeared around a region on the sonar image if the system believed a mine was present (Figure 3).

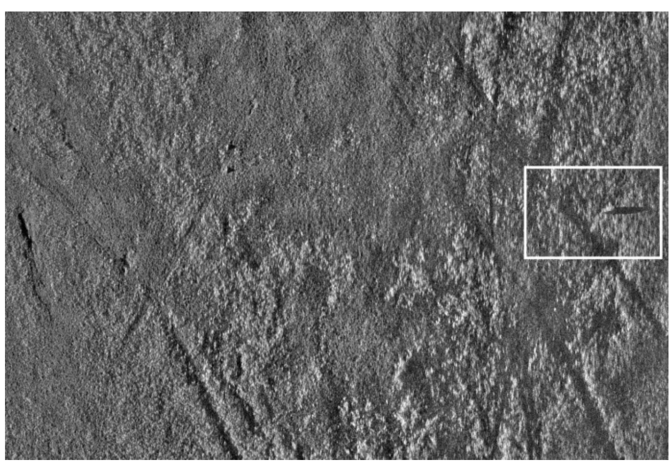


Figure 3. Example of a sonar image with a cue from the automated target recognition system to alert the user to a region on the image where the system believes a mine is present.

Trust

Participants were asked to complete a Trust in Automation Questionnaire (Jian, Bisantz & Drury, 2000; Appendix F) following each of the experimental blocks where automation was used. The questionnaire was comprised of 12 questions that the participants responded to using a 7-point Likert scale (1 = not at all : 7 = extremely) and was used to gather the participants' perceived level of trust in the automated system. Example questions included *"The system behaves in an underhanded manner"* and *"The system provides security"*.

Confidence

Participants were asked to complete a Confidence in Abilities Questionnaire (Appendix G) after the first training block and each of the experimental blocks. The questionnaire was an adapted version of the Trust in Automation Questionnaire and was used to assess the participants' level of confidence in their own abilities to detect underwater mines in the sonar images. The questions in the Confidence in Abilities Questionnaire were created using the questions from the Trust in Automation Questionnaire as a template. Each question was altered to assess the users' level of confidence in their mine detection abilities rather than their level of trust in the automated system (*e.g., I am confident in the system* (Trust in Automation Questionnaire) → *I am confident in my ability to identify mines* (Confidence in Abilities Questionnaire)). The adapted questionnaire had 10 questions that the participants responded to using a 7-point Likert scale (1 = completely disagree : 7 = agree). Example questions included *"I believe that my ability to identify mines on my own may lead to negative outcomes"* and *"I feel like others could depend on my ability to identify mines"*.

Performance

The coordinates on the sonar images where the participants clicked were recorded using the computer programming software MATLAB. Each click coordinate was used to determine whether the participant thought a mine was present (clicked an area on the sonar image) or not present (clicked the black "NO MINE' box), along with the exact location on the sonar image the participant believed the mine was if they thought there was a mine. Comparing the participants' click coordinates to the true coordinates of the mines in the sonar images, the number of hits, misses, false alarms and correct rejections could be obtained. Sensitivity and response bias for each block of trials could then be calculated using the number of hits and false alarms determined.

Response time, in seconds, was also recorded using MATLAB to determine the time it took the participants on each trial to select a response regarding whether they believed a mine was present or not following the image being displayed on the screen. Response time was measured from the time the imaged appeared to when the participant clicked on the image. Using the response time recorded on each trial, average response time was calculated for the trials where a mine was present and for the trials where a mine was absent.

Design

Participants completed four experimental blocks, two of which they completed with no automation and the other two with automation. The no automation experimental blocks were made up of 50 trials each, with 25 of the trials having sonar images with a mine present and the other 25 trials having sonar images with no mine present. The automation experimental blocks consisted of 100 trials each, with a mine present in the sonar images on 50 of the trials and no mine present in the sonar images on the remaining 50 trials.

In the automation experimental blocks, the reliability level of the automated system was set at either a low or a high false alarm rate. The order of the false alarm rate set for the automated experimental blocks was counterbalanced between participants. If the automated system was set at a low false alarm rate, 12% of the trials had false alarms. For the low false alarm rate automated experimental block, 6 of the trials where a mine was not present had false alarms and 11 of the trials where a mine was present had misses. If the automated system was set at a high false alarm rate, 24% of the trials had false alarms. For the high false alarm rate automated experimental block, 12 of the trials where a mine was not present had false alarms and 5 of the trials where a mine was present had misses. The misses were adjusted for each of the false alarm rate conditions to ensure that the overall sensitivity of the system was kept constant.

Participants were randomly assigned to one of two groups. The automated (not informed) group performed the automated experimental blocks without being informed of the false alarm rate/reliability of the system and the automated (informed) group performed the automated experimental blocks with being informed of the false alarm rate/reliability of the system. The false alarm rate/reliability information in the form of the percentage of false alarms that were expected to occur were given to the participants in the automated (informed) group via a script (Training and Automation Scripts – Appendix H) before starting the trials for the automation experimental blocks.

Procedure

Participants attended a single session lasting approximately 1.5-2 hours. For the duration of the study the participants were seated at a computer workstation. After providing informed consent, the participants read through a Mine Detection Training Script (Appendix H) explaining the simulation and the tasks that would be required of them. The participants were informed that they would play the role of a naval commander aboard a Navy Frigate about to sail in unchartered waters with the task of monitoring sonar images collected by an unmanned submarine to identify underwater mines. The participants were also provided an example of four sonar images and were given the opportunity to ask questions about any of the information presented to them that may have been unclear before continuing.

After going over the instructions, the participants completed six blocks of trials, two of which were training blocks (Figure 4 for study design). First participants completed a training block of 50 trials with no automation using images "collected from a previous mission" where they were given the opportunity to interact with the simulation and get comfortable with the mine detection task. On each trial a sonar image appeared and the participants were instructed to indicate using the computer mouse whether a mine was present or absent. The participants received feedback on the screen during this no automation training block indicating if they were correct in determining whether a mine was present or not in the sonar image (Figure 5).

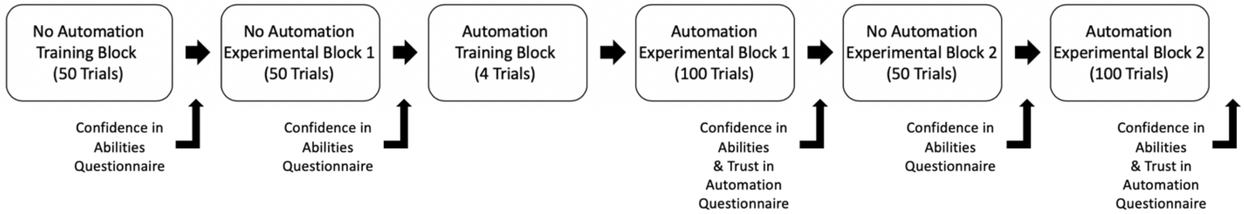


Figure 4. Study design for the training and experimental blocks of trials.

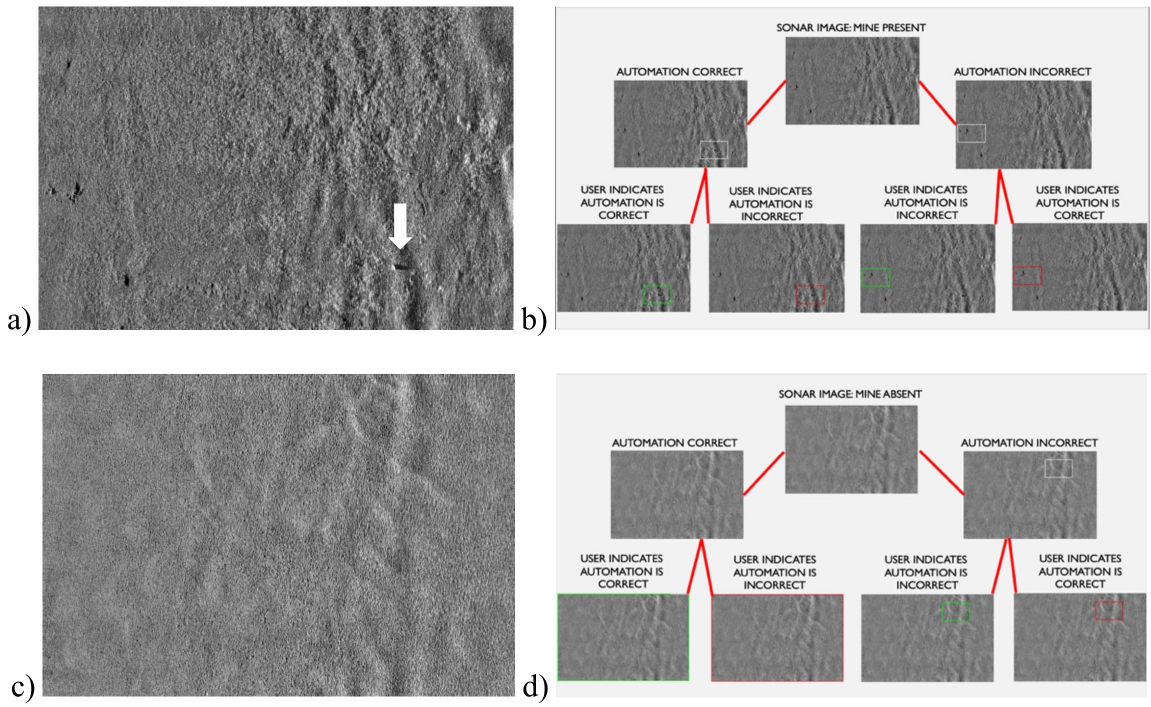


Figure 5. a) Sonar image with a mine present and b) the potential feedback that may be received for the image based on the automated system's and user's response. c) Sonar image with no mine present and d) the potential feedback for the image based on the response made by the automated system and the user.

After the first training block, the participants read Experimental Session Script – No Automation Block 1 (Appendix H) and completed an experimental block of 50 trials with no automation. The participants were told that the images for each trial were collected along their proposed route by the unmanned submarine and that it was important that all mines were detected in a timely fashion in order for the mission to progress. For this experimental block and those following, the participants had to determine whether a mine was present or not in each of the sonar images but did not receive feedback on whether their mine detection decision was correct.

Following completion of the first experimental block, the participants were informed that an automated target recognition system had been designed to assist them in completing their task. The participants were provided an Automation Training Script (Appendix H) that described the automated system and how it works. The script explained to the participants that the system uses computer vision to identify the presence of mines but is not entirely perfect. The script also informed the participants that the system could be altered to change the number of false alarms that occur. The participants were given an opportunity after reading the script to ask any questions they had about the information presented to them that may have been unclear. The participants then ran through a training block of 4 trials with automation to get familiarized with the appearance of the system. The participants received feedback during this training block regarding how the automated system works.

After completing the training block on automation, the participants completed an experimental block of 100 trials with assistance from the automated target recognition system. The participants were reminded through the Experimental Session Script – Automation Block 1 (Appendix H) how the automation works and that the system could be altered by the system designers to change the number of false alarms that occur. The participants in the informed group were given an additional portion of the script indicating whether the false alarm rate for the experimental block had been set to high (24% of the trials will have false alarms) or low (12% of the trials will have false alarms). The participants in the not informed group were not given any information regarding the false alarm rate set for the experimental block. As with the no automation experimental block, the participants had to determine whether a mine was present or not in each of the sonar images, but also received assistance from the automated target recognition system. The participants were informed that although the automated system may be used to assist them, they had to make the final decision regarding the potential presence of a mine. The participants were also reminded that they had to examine the images in a timely fashion.

Following completion of the automation experimental block, the participants completed an experimental block of 50 trials with no automation. The participants were told through Experimental Session Script – No Automation Block 2 (Appendix H) that the automated system was being re-calibrated and that they would not have access to it for this block of trials. The participants were reminded that they would be examining images collected by the unmanned submarine along their proposed route and that it was important that all mines were detected in a timely fashion. Similar to the other experimental blocks, the participants had to determine whether a mine was present or not in each of the images but would not receive assistance from the automated system.

After completion of the second no automation experimental block, the participants completed another experimental block of 100 trials with automation. The participants were told that the automated target recognition system was recalibrated and was now available to assist them with this block of trials. As with the other automation experimental block, the participants were reminded through a script, Experimental Session Script – Automation Block 2 (Appendix H), how the automation works and that the system could be altered by the system designers to change the number of false alarms that occur. The participants in the informed group were given an additional portion of the script indicating whether the false alarm rate for the experimental block was set to high (24% of the trials will have false alarms) or low (12% of the trials will have false alarms). The participants in the not informed group were not given any information regarding the false alarm rate set for the experimental block. The participants had to determine whether a mine was present or not in each of the sonar images and received assistance from the automated target recognition system. The only difference between this automation experimental block and the first automation experimental block was the false alarm rate that was set (i.e., if the first automation experimental block was set at a high false alarm rate this experimental block was set at a low false alarm rate). Time was allotted between each block to allow the participants to take a break if desired.

The participants were asked to complete the Confidence in Abilities Questionnaire following the first training block and each of the experimental blocks, as well as the Trust in Automation Questionnaire following the low and high false alarm rate automation experimental blocks. The participants did not receive the Trust in Automation Questionnaire following the no automation experimental blocks because they were not interacting with the automated system while completing the trials. At the end of the study the participants were debriefed about the experiment (Debriefing Form – Appendix I), the goals of the experiment, and were given an opportunity to ask any questions they had about the experiment or possible results of the experiment.

Data Analysis

Response time was calculated on each trial by taking the difference between when the sonar image appeared on the screen and when the participant clicked the mouse to select a response. Using the calculated response times for each trial, an average response time was determined for the trials when a mine was present and when a mine was absent. The response times for the mine present and mine absent trials were analyzed separately because visual search on trials where a target is absent typically takes longer when compared to trials where a target is present.

The proportion of hits and false alarms that occurred in each block of trials per participant were calculated by dividing the number of hits and false alarms determined by the number of trials present in each block. The hit and false alarm proportions were then converted to z-scores within MATLAB using an inverse complementary error function and inputted into the following equation to calculate sensitivity:

$$d' = z_{hit} - z_{false\ alarm} \qquad (1)$$

Using the calculated z-scores for hits and false alarms, response bias values were also determined for each block of trials per participant using the equation:

$$\beta = -0.5 * (z_{hit} + z_{false\ alarm}) \qquad (2)$$

The confidence scores collected from the Confidence in Abilities Questionnaire were inputted into an excel document per participant for each block of trials. To compute an average confidence score for each experimental block, the first four questions were reversed so a higher score indicated a higher confidence the participant had in their own abilities to identify the underwater mines (i.e., a score of 1 would change to a score of 7, a score of 2 would change to a score of 6, etc.). Using the reversed scores for questions one to four and the original scores for questions five to ten, an average confidence score was computed.

The trust scores obtained using the Trust in Automation Questionnaire were inputted into an excel document for the low and high false alarm rate automation blocks for each participant. To calculate a trust score for the two automation blocks, the first five questions were reversed so a higher score indicated a higher trust in automation. An average trust score was computed using the reversed scores for questions one to five and the original scores for questions six to twelve.

A 2 (Group: informed, not informed) by 4 (Automation Condition: no automation 1, no automation 2, low false alarm rate automation, high false alarm rate automation) mixed ANOVA was performed on response time when a mine was absent, response time when a mine was present, sensitivity, response bias and confidence and a 2 (Group: informed, not informed) by 2 (Automation Condition: low false alarm rate automation, high false alarm rate automation) mixed ANOVA was performed on response bias and trust. For measures where sphericity was violated ($p < .05$), Greenhouse-Geisser estimates were used. Effect size, η_p^2, was calculated for each main effect and Tukey's HSD post hoc comparisons were performed to follow up on any main effects involving Automation Condition or significant interactions. To determine which Automation Condition(s) or interaction(s) were significantly different from the others, a critical value was calculated for each measure using the following equation:

$$\text{critical value} = q\sqrt{\frac{MS_w}{n_k}} \tag{3}$$

If the difference between the means for two Automation Conditions or interactions was greater than the critical value calculated for that measure, Tukey's HSD was found to be significant or there was a significant difference between the two conditions/interactions. If instead the difference between the means for two Automation Conditions or interactions was less than the critical value calculated for that measure, a significant difference was not found between the two conditions/interactions.

Chapter 4: Results

Response Time

Response time when a mine was present was not significantly different between the Informed and Not Informed Groups, $F(1, 68) = .019$, $p = .89$, $\eta_p^2 = < .01$ (Figure 6 A). A significant difference was found between the response times for Automation Condition, $F(1.74, 118.6) = 147.9$, $p < .01$, $\eta_p^2 = .685$, where no automation 1 (M = 6.12s, SEM = .34) showed a significantly longer response time compared to the other conditions: no automation 2 (M = 2.76s, SEM = .14), low false alarm rate automation (M = 2.79s, SEM =.16) and high false alarm rate automation (M = 2.56, SEM =.16), (critical value = .51). No difference in response time was found between the three other conditions. There was no significant interaction between Group and Automation Condition for response time, $F(1.74, 118.6) = 1.18$, $p = .31$, $\eta_p^2 = .02$.

There was no significant difference in response time on trials where a mine was absent between the Informed and Not Informed Groups, $F(1, 68) = .093$, $p = .76$, $\eta_p^2 = < .01$ (Figure 6 B). There was a significant effect of Automation Condition for response time, $F(2.10, 142.5) = 68.8$, $p < .01$, $\eta_p^2 = .50$, where similar to the trials when a mine was absent, no automation 1 (M = 6.22s, SEM = .42) showed a significantly longer response time compared to the other conditions: no automation 2 (M = 3.67s, SEM = .24), low false alarm rate automation (M = 3.68s, SEM =.24) and high false alarm rate automation (M = 3.70s, SEM = .27), (critical value = .56). Again, no difference in response time was found between the three other conditions and the interaction between Group and Automation Condition was not significant, $F(2.10, 142.5) = .55$, $p = .59$, $\eta_p^2 = .01$.

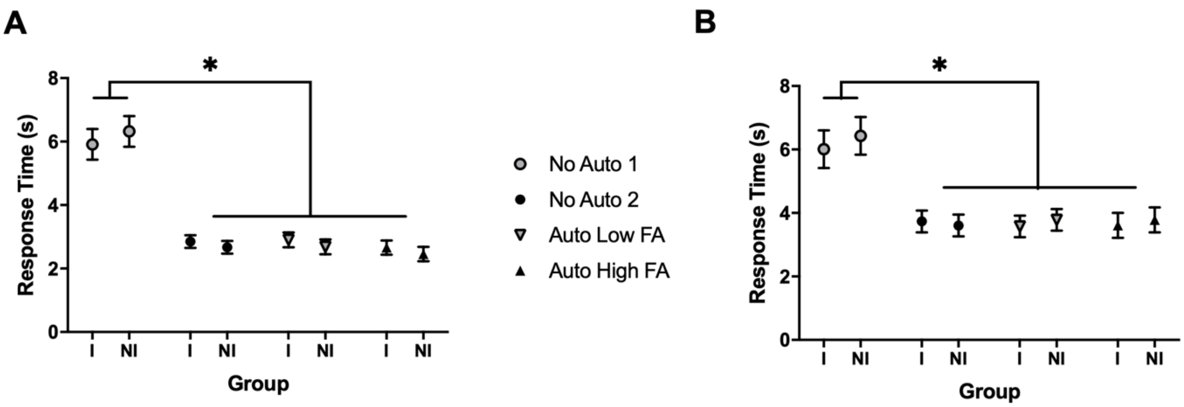


Figure 6. Average response time of participants in the informed (I) and not informed (NI) groups for each automation condition on trials where a mine was present (A) or absent (B).

Sensitivity

There was no significant difference in sensitivity between the Informed and Not Informed Groups, $F(1, 68) = 1.35$, $p = .25$, $\eta_p^2 = .02$, however there was a significant effect of Automation Condition, $F(2.60, 176.6) = 68.9$, $p < .01$, $\eta_p^2 = .50$ (Figure 7 A). Mean sensitivity for no automation 1 (M = .91, SEM = .04) was found to be significantly lower compared to the other three conditions: no automation 2 (M = 1.43, SEM = .04), low false alarm rate automation (M = 1.40, SEM = .03) and high false alarm rate automation (M = 1.31, SEM = .04), (critical value = .11). In addition, mean sensitivity was also found to be significantly lower in the high false alarm rate automation condition (M = 1.31, SEM = .04) compared to no automation 2 (M = 1.43, SEM = .04). The interaction between Group and Automation Condition for sensitivity was not significant, $F(2.60, 176.6) = .80$, $p = .48$, $\eta_p^2 = .01$.

Response Bias

Response Bias was not significantly different between the Informed and Not Informed Groups, $F(1, 68) = 1.53$, $p = .22$, $\eta_p^2 = .02$. A significant difference in response bias was found for Automation Condition, $F(2.63, 179.0) = 14.48$, $p < .01$, $\eta_p^2 = .18$, where no automation 2 (M = .99, SEM = .03) showed a significantly lower response bias (i.e., the participants were more likely to say a mine was present) compared to the other three conditions: no automation 1 (M = 1.08, SEM = .02), low false alarm rate automation (M = 1.12, SEM = .03) and high false alarm rate automation (M = 1.05, SEM = .03), (critical value = .05). The high false alarm rate automation condition (M = 1.05, SEM = .03) also showed a significantly lower response bias compared to the low false alarm rate automation condition (M = 1.12, SEM = .03). A significant effect for response bias was not found for the interaction between Group and Automation Condition, $F(2.63, 179.0) = 1.80$, $p = .16$, $\eta_p^2 = .03$.

To more specifically examine whether a difference in participant response bias was influenced by a change in the automation response bias, a separate ANOVA compared only the blocks of trials where the participants used the automation. In the automation present conditions there was no significant difference between the Informed and Not Informed Groups, $F(1, 68) = .79$, $p = .38$, $\eta_p^2 = .01$ (Figure 7 B). A significant main effect of Automation Condition was found, $F(1, 68) = 17.7$, $p < .01$, $\eta_p^2 = .21$, however this effect was superseded by the interaction between Group and Automation Condition, $F(1, 68) = 5.58$, $p = .021$, $\eta_p^2 = .08$. In the Not Informed Group, mean response bias was found to be significantly lower for the high false alarm rate automation condition (M = 1.01, SEM = .04) compared to the low false alarm rate automation condition (M = 1.12, SEM = .04) (critical value = .06). However, for the Informed Group no significant difference in mean response bias was found between the low false alarm rate automation (M = 1.12, SEM = .04) and high false alarm rate automation (M = 1.09, SEM = .04) conditions.

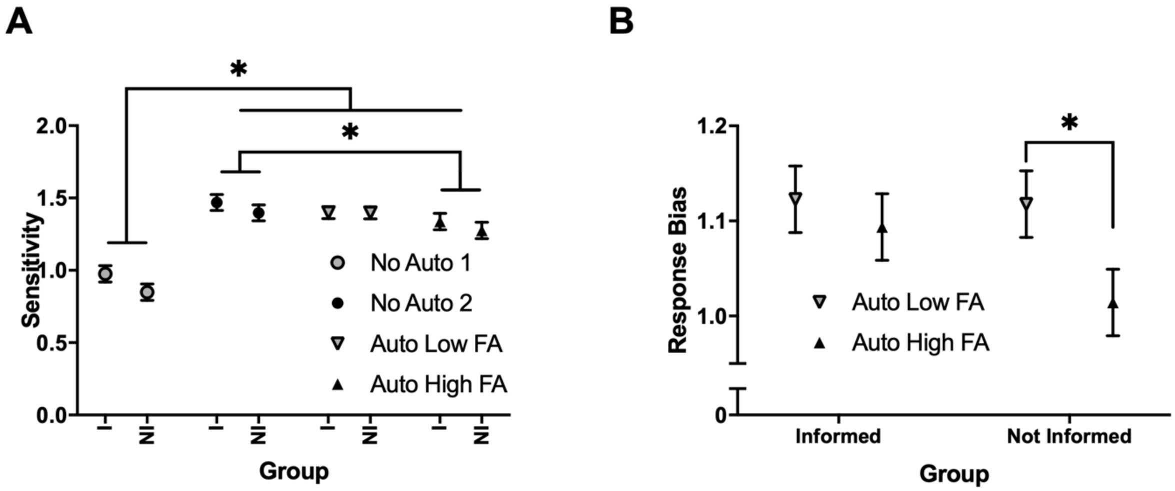


Figure 7. Average sensitivity of participants in the informed (I) and not informed (NI) groups for each automation condition (A) and average response bias of participants per group for the low false alarm rate (Auto Low FA) and high false alarm rate (Auto High FA) automation conditions (B).

Trust

Trust in automation was significantly different between the Informed and Not Informed Groups, $F(1, 68) = 4.37$, $p = .040$, $\eta_p^2 = .06$, where the Informed Group had greater trust in the automated system compared to the Not Informed Group (Figure 8 A). There was no significant difference in trust between Automation Conditions, $F(1, 68) = 3.39$, $p = .070$, $\eta_p^2 = .05$, nor was there a significant effect for the interaction between Group and Automation Condition, $F(1, 68) = .77$, $p = .39$, $\eta_p^2 = .01$.

Confidence

Confidence in mine detection ability was not significantly different between the Informed and Not Informed Groups, $F(1, 68) = 1.48$, $p = .23$, $\eta_p^2 = .02$ (Figure 8 B). A significant difference was found in confidence levels between Automation Conditions, $F(3, 204) = 24.63$, $p < .01$, $\eta_p^2 = .27$, where mean confidence for no automation 1 (M = 2.92, SEM = .10) was significantly lower compared to the other three conditions: no automation 2 (M = 3.44, SEM = .13), low false alarm rate automation (M = 3.37, SEM = .13) and high false alarm rate automation (M = 3.20, SEM = .13), (critical value = .17). In addition, mean confidence was also significantly lower in the high false alarm rate automation condition (M = 3.20, SEM = .13) compared to no automation 2 (M = 3.44, SEM = .13). The interaction between Group and Automation Condition was not significant for confidence, $F(3, 204) = .98$, $p = .40$, $\eta_p^2 = .01$.

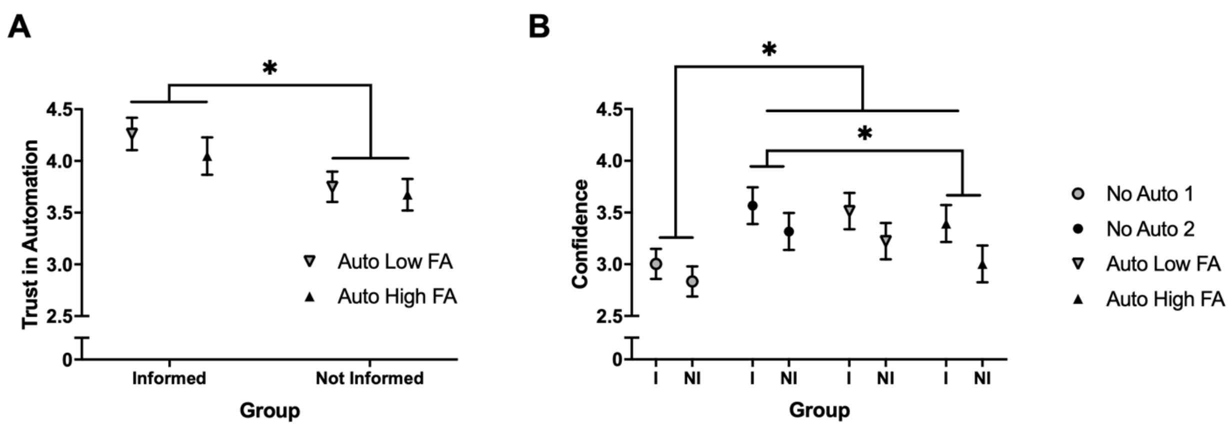


Figure 8. Average trust of participants in the informed (I) and not informed (NI) groups for the low false alarm rate (Auto Low FA) and high false alarm rate (Auto High FA) automation conditions (A) and average confidence of participants per group for each automation condition (B).

Chapter 5: Discussion

The purpose of this study was to determine if the false alarm rate of an automated target recognition system affects a user's level of trust in the automated system, level of confidence in their own abilities and level of performance during an underwater mine detection task, and whether these outcome variables are influenced by informing the user of the false alarm rate/reliability of the system. Trust in the automated system was greater for the participants who were informed of the false alarm rate compared to the participants who did not receive any false alarm rate information. In addition, the response bias of the participants in the informed group remained relatively unchanged between the low false alarm rate and high false alarm rate automation conditions, where the response bias of the participants in the not informed group appeared to be influenced by whether the system was set at a low or a high false alarm rate. Furthermore, sensitivity and confidence levels were found to be lower for the high false alarm rate automation condition compared to when no automation was used later in the session. It should be noted that participant's performance improved from the first no automation block to the second no automation block, indicating that they still may have been learning about the task during the first block. Therefore, the discussion will mainly focus on comparing the automation conditions to the second no automation block of trials.

Trust

Participants who were informed that the automated system was set at a low or a high false alarm rate had greater trust in the system compared to the participants who did not receive any false alarm rate information. This finding is consistent with our initial hypothesis that the users' level of trust would be higher when informed of the automated system's false alarm rate compared to when not informed of the false alarm rate of the system. This hypothesis was based on previous research that has indicated that disclosing the reliability level of an automated aid (Hollands & Neyedli, 2011; Neyedli et al., 2011; Wang et al., 2009), or providing information regarding the aid's behaviour or competence (Bagheri & Jamieson, 2004; Lee & See, 2004; Muir, 1994), should lead to more appropriate levels of trust and reliance. As discussed in Chapter 2, three characteristics of a human or automated aid that have been identified in the literature that impact user trust include the ability, benevolence and integrity of the aid. Each of these characteristics associated with the aid can be determined by a human user respectively by observing how the aid behaves or having knowledge pertaining to the reliability/predictability of the aid (performance), by developing a deeper understanding about how the aid operates or the algorithms used (process) and by learning information regarding why the aid was developed or what the aid was designed to do (purpose) (Lee & See, 2004).

All participants were provided information regarding why the automated target recognition system was designed (purpose) and how the system works (process), but information about the false alarm rate/reliability of the system (performance) was only given to participants who were placed in the informed group. Since the participants did not have previous interactions with the automated target recognition system, those who were not informed of the false alarm rate information may have had to determine the ability of the system by observing its actions and decisions (Bagheri & Jamieson, 2004). Due to errors made by an automated system being more prominent than correct decisions (Dzindolet et al., 2003), the ability or performance of the system may have been underestimated by the participants when highly salient false alarm errors occurred, resulting in the system being perceived as less trustworthy. This may especially be true if the participants had high expectations in the automated system (Bagheri & Jamieson, 2004; Dzindolet et al., 2002; Hollands & Neyedli, 2011; Lyons & Stokes, 2012; Merritt et al., 2015). Therefore, when the participants were informed of the false alarm rate of the automated system, they may have been able to adjust their expectations, resulting in the system being perceived as more trustworthy.

Although informing the participants of the false alarm rate of the automated system resulted in an increase in overall trust, it is interesting to note that a significant difference in trust was not found between the low false alarm rate and high false alarm rate automation conditions. These results are inconsistent with our initial hypothesis that the users' level of trust in the automated system would be higher during the low false alarm rate condition compared to their level of trust during the high false alarm rate condition. Trust was expected to be lower in the high false alarm rate condition because false alarm errors may be more noticeable than miss errors. According to the literature, as the reliability of an automated aid increases there is usually a subsequent increase in user trust and performance and as the reliability of an automated aid decreases a decrease in user trust and performance is expected (Dzindolet et al., 2003; Hollands & Neyedli, 2011; Madhavan et al., 2003; Parasuraman et al., 2000). Similarly, research has shown that human operators are sensitive to slight changes in reliability levels and the types of errors made by an automated system (Madhavan & Wiegmann, 2007). Since the overall sensitivity of the system and overall system reliability (i.e., total errors, both misses and false alarms) were kept constant for each false alarm rate condition, this may suggest that trust is less sensitive to changes in an automated system's response bias (i.e., whether the system makes more false alarms or misses).

Response Bias

Participants who were not informed of the false alarm rate of the automated system had a lower, more liberal response bias (i.e., the participants were more likely to say a mine was present) for the more liberal, high false alarm rate automation condition and a higher, more conservative response bias (i.e., the participants were less likely to say a mine was present) for the more conservative, low false alarm rate automation condition. On the other hand, the response bias for the participants who were informed of the false alarm rate remained relatively unchanged between the two automation conditions. These results suggest that when the participants were not informed of the system's false alarm rate, their responses were more likely to be influenced by the response bias of the automated system. This change in user response bias to be similar to that of the automated system may be attributed to automation bias, or the tendency for users to overly rely on information provided by an automated aid (Dzindolet et al., 2002; Dzindolet et al., 2001; Parasuraman & Manzey, 2010; Parasuraman & Riley, 1997). Automation bias has been demonstrated in multiple studies and scenarios such as when participants displayed lower diagnostic accuracy when making medical diagnoses under the aid of an imperfect decision support system compared to when the participants made the diagnoses without assistance from the aid (Goddard, Roudsari & Wyatt, 2012) and when 75% of pilots were found to make an error by shutting down an engine when incorrectly advised to do so by an automated aid compared to 25% when a manual checklist was used (Wickens et al., 2013).

Research suggests that human users often fail to examine all the necessary data required to verify if a diagnosis recommended by an automated aid is correct or not (Bahner et al., 2008; Parasuraman & Manzey, 2010; Skitka, Mosier & Burdick, 1999; Wickens & Dixon, 2007). This may be due to the tendency for users to select the path with the least amount of cognitive effort required to make a decision (Wickens et al., 2013), especially when the user perceives the aid as a teammate and therefore may feel that the responsibility for the outcome of the team's performance is dispersed between themselves and the aid (Dzindolet et al., 2002; Parasuraman & Manzey, 2010). In addition, as discussed in the previous section, those not informed of the false alarm rate would have been required to determine how much they should trust and rely on the system by observing the cues generated by the aid. Since no feedback was given regarding whether the cues made by the automated system were correct or not, and due to the difficulty of some of the sonar images, the participants may have perceived the reliability of the aid during the automation conditions to be greater than it was.

When the participants were informed of the false alarm rate of the automated system, their responses were not significantly influenced by the liberal and conservative response biases set for the aid. For the low false alarm rate condition the response biases of the participants from the informed and not informed groups were similar and for the high false alarm rate condition the response biases of the participants from the informed group were more conservative compared to the more liberal response biases seen for the participants from the not informed group (Figure 7 B). If a user is aware of the reliability level of an automated system, they may be more likely to monitor the cues or information generated by the aid appropriately (Bagheri & Jamieson, 2004) and interfere with the system when they do not believe the cue generated is an appropriate response (Colebank, 2008). Furthermore, in situations where an automated system has a high false alarm rate, the user may develop a more conservative response bias to compensate for the system's liberal decision criterion (Wickens et al., 2013). Therefore, if the participants are aware that in the high false alarm rate condition 24% of trials where a mine is not present will have false alarms (compared to 12% for the low false alarm rate condition) they may (1) pay more attention to the cues generated by the automated system compared to the participants who were not given this false alarm rate information, (2) be more cautious of the cues in the high false alarm rate condition compared to the cues in the low false alarm rate condition and (3) intervene when the automated cues deviate from what they believe to be the best response.

Sensitivity and Confidence

The sensitivity and confidence levels of the participants from both the informed and not informed groups were found to be lower for the high false alarm rate automation condition compared to the second experimental condition where no automation was used (No Automation 2). In other words, the participants performed significantly worse and had significantly less confidence in their own abilities when the mine detection task was completed with the aid of the automated target recognition system set at a 24% (high) false alarm rate compared to when the task was completed without the aid of the automated system. According to the literature, when trust in an automated system exceeds a user's level of confidence in their own abilities, automation tends to be used and when a user's confidence in their own abilities is greater than their trust in the automated system, the task tends to be completed manually (Dzindolet et al., 2002; Lee

& Moray, 1992; Lee & Moray, 1994; Parasuraman & Riley, 1997). When the automated system was set at a high false alarm rate, the system made 45 hits and 12 false alarms compared to 39 hits and 6 false alarms when the system was set at a low false alarm rate. Although during the high false alarm rate condition the automated system made more false alarms compared to the low false alarm rate condition, the system also made more hits. Therefore, depending on the difficulty of the false alarms and misses made by the aid, the participants may have perceived the system to be more capable than themselves of detecting the mines in the sonar images. This may have led to the participants developing lower confidence in their own ability to perform the mine detection task and as a result allocated more control to the automated system.

As discussed in Chapter 2, when a user is working with an automated system that results in a high number of hits and a low number of misses, performance decrements may occur due to attention narrowing or tunneling. Attention narrowing or tunneling is a phenomenon demonstrated by users when an automated system consistently cues the user to the correct location of a target when a target is present. Since the automated system is perceived as reliable, the visual search that is performed by the user tends to decrease. If the user is not examining the search field, they may be more likely to miss a target if it is not cued by the automated system (Wickens et al., 2013), such as when the soldiers missed dangerous un-cued targets when cued to a different location (Yeh et al., 2003; Yeh et al., 1999) or when the pilots failed to notice a dangerous event visible through the airplane windshield due to focusing on the information provided by the automated system (Wickens & Alexander, 2009). Therefore, due to the high number of hits performed by the automated system during the high false alarm rate condition, along with the reported decrease in confidence in their mine detection abilities, the participants may have decided to simply rely on the aid rather than monitor the images and cues appropriately.

It was hypothesized that the false alarm rate condition and the group to which the participant was assigned (informed vs. not informed) would affect the users' level of confidence in their own abilities and performance during the underwater mine detection task. While in the high false alarm rate condition users had lower confidence and sensitivity compared to when the task was completed without the aid, the difference between other conditions (i.e., low false alarm rate vs. no automation, low vs. high false alarm rate and main effect of Group) did not reach significance. Thus when the task was completed manually the users had similar sensitivity scores and confidence in their own abilities compared to when the task was completed with the automated system at a low false alarm rate. Similarly, the users' detection performance and confidence did not significantly differ between the high and low false alarm rate conditions. A significant difference may not have been found between the two false alarm rate conditions due to them having the same overall sensitivity and reliability (error rate). In addition to the users' confidence in their own abilities and mine detection performance being less affected by the slight changes in the decision criterion used by the automated system, the results also suggest that these measures were not affected by informing the users of the false alarm rate. While there appears to be a small numerical difference between the two groups, with sensitivity and confidence being higher for those in the informed group (Figure 7 A and Figure 8 B), knowledge of the number of false alarms that were expected to occur did not significantly improve the users' ability to detect the mines in the images or increase the users' confidence in their decisions.

For the first experimental condition where no automation was used (No Automation 1), sensitivity and confidence levels were found to be lower compared to the other three experimental conditions. This may suggest that the participants were still learning during the first experimental block. To aid in increasing performance during this condition, an additional training block could be implemented where the participants would complete the mine detection task manually without any feedback provided. In addition, a training block with automation could also be implemented, where the users could gain experience with the automated system and exposure to automated errors such as false alarms and misses. Since experiencing an automation failure is suggested to be more beneficial to a user compared to simply being informed that the aid may fail (Bahner et al., 2008; Parasuraman & Manzey, 2010; Wickens et al., 2013), the automation training condition may result in greater levels of confidence and performance.

Practical Implications

The aim of this study was to gain more insight into the effect false alarms have on a user. This was accomplished by having an automated target recognition system set at a more liberal (high false alarm rate) or a more conservative (low false alarm rate) decision criterion aid participants with an underwater mine detection task. Although the system was more likely to indicate that a mine was present in the high false alarm rate condition, or a mine was absent in the low false alarm rate condition, the overall sensitivity of the system was the same between the two conditions and the system produced the same amount of errors (17) within each block of trials. The effect of informing the user of the false alarms was also examined by providing the false alarm rate of the automated system to one group of participants and not to another.

Based on the results, when a user is informed of the false alarm rate of an automated system, the user has greater trust in the aid regardless of the number of false alarms the system is expected to produce. This may be beneficial for automated systems that tend to produce a high number of false alarms. When the signal to noise ratio is low, frequent false alarms may be made by an automated system, even when the system is designed to have a low false alarm rate (Ho et al., 2013; Parasuraman & Masalonis, 2000). This is frequently the case when ATR is used for mine detection, due to the rare presence of mines along the seafloor. In addition, since missing targets such as mines may be more costly than being alerted to a target that is not there, designing an ATR system with a liberal decision criterion (high false alarm rate) may be more beneficial than designing a system with a more conservative decision criterion (low false alarm rate).

If an automated system has a low false alarm rate, informing the user of the number of false alarms the system is expected to make does not significantly influence the user's response bias. If the automated system is expected to make a high number of false alarms, informing the user of the system's false alarm rate may encourage the user to adopt a more conservative response bias. As seen with trust, this may be beneficial in situations where the probability of encountering a true target is low. If an automated system makes a high number of false alarms regardless of the false alarm rate set, a conservative response bias adopted by the user may aid in reducing the amount of errors that may occur as the result of the false alarms. Furthermore, informing users that the false alarm rate of the system can be changed may improve trust by providing more process-based information on how the automation operates. Since the user is not able to observe the processes or algorithms used by the system to make a decision, knowledge that the system can be flexible and adopt different decision criterions to either detect more mines (liberal response bias/high false alarm rate) or miss fewer mines (conservative response bias/low false alarm rate) may make the decisions reached by the system appear more rational or understandable (Dzindolet et al., 2003; Wickens et al., 2013). Given that there is reported disuse of ATR systems (Kessel & Myers, 2005), increased trust from informing participants about false alarm rates may lead to increased use of the systems.

While informing the participants of the false alarm rate of the automated system had a significant effect on response bias and trust, sensitivity and confidence levels did not appear to be affected by this information. In addition, the participants' performance in the detection task and confidence in their own abilities were found to be similar for the low false alarm rate and high false alarm rate automation conditions. Based on these results, a user may be expected to have the same sensitivity and confidence levels during a detection task regardless of the number of false alarms made by the automated system or whether the false alarm rate was provided to the user.

Although the number of false alarms that occurred did not have a significant effect on the participants' sensitivity and confidence levels for the automation conditions, the participants did perform significantly worse and had lower confidence in their own abilities during the high false alarm rate condition compared to when the task was completed manually. Providing the users information during these situations regarding when and why the aid might make an error, as well as what factors the aid considers when making a decision, may reduce this performance decrement observed (Bagheri &

Jamieson, 2004; Dzindolet et al., 2003). In addition, allowing the users or the automated system to state their confidence in indicating that a signal is present or not, rather than providing a yes or no answer, may lead to an increase in confidence and performance (Sorkin et al., 1988).

Limitations and Future Directions

A limitation of the present study is the participant population used. Compared to trained sonar operators, the individuals who participated in this experiment were not familiar with ATR systems or the appearance of mines. Operators with experience examining sonar data may be able to easily detect errors made by an automated system where novices who are unaware of the defining features of a mine may mistake false alarms for hits or misses for correct rejections. In addition, since encountering a mine along the seafloor is a rare event, the mine detection task used for this study was designed to examine false alarms from more of a basic rather than applied research standpoint. In other words, the number of mines and false alarms that occurred during this study were inflated compared to what would be expected in an actual mine detection task in order to examine the specific effect different false alarm rates have on a user.

Overall, the results from this study suggest that informing a user of the false alarm rate of an automated system may positively influence the level of trust and reliance the user has in the aid. According to past research, users are just as likely to rely on an aid when it makes false alarms as they are when it makes misses (Dzindolet et al., 2003). This appears to be due to the users concluding that if it is reasonable for the aid to indicate that a target is present when it is not, it is just as reasonable to indicate that a target is not present when it is. Therefore, if reliance patterns are similar regardless of the type of error made by an automated aid, the results seen for this study may extrapolate to situations where the miss rate is provided rather than the false alarm rate. Due to missing a mine being more dangerous than being cued to a location where a mine is not present, future research may look at the effect misses have on trust and reliance compared to false alarms. In addition, researchers may want to determine if increasing the overall sensitivity of the automated system or adjusting the decision criterion will lead to an increase in confidence and performance. Furthermore, studies may also want to examine the effect false alarms have on a user's performance during a mine detection task where advice is provided from an automated system and a human teammate simultaneously. Finally, researchers may want to determine the effect false alarm rate has on experts in a more realistic scenario.

Chapter 6: Conclusion

In mine countermeasure operations, sonar operators are typically responsible for scanning seafloor imagery and classifying any foreign objects detected as mine-like or non-mine-like (Ho et al., 2011). Due to the high workload placed on the operators when tasked to manually examine the sonar data, as well as the decrease in accuracy and performance observed when users are forced to operate under conditions of uncertainty, automated target recognition systems are frequently designed and implemented to aid with identifying and classifying the underwater mines (Ho et al., 2011; Kessel & Myers, 2005; Myers, 2009). Although ATR aids are on average accurate and reliable, they are not perfect and sometimes make mistakes. These mistakes may include false alarms and misses. Since ATR systems typically operate using a more liberal decision criterion to reduce the risk of missing a mine, the number of false alarms that occur as a result tend to lead to a decrease in the trust and reliance the operators have in the system (Kessel, 2005; Kessel & Myers, 2005).

The purpose of this project was to determine how changing the false alarm rate in a mine detection task affects trust, reliance and performance and whether informing the user of the false alarm rate could mitigate the potentially detrimental effects of high false alarm rates on trust and reliance. When users were not informed of the false alarm rate of the automated system, the number of false alarms made by the system had a significant effect on the participants' response bias (reliance behaviour). Furthermore, informing the participants of the false alarm rate resulted in greater trust in the automated system and a more consistent response bias obtained by the user. Sensitivity and confidence were not influenced by disclosure of the false alarm rate, but the performance and confidence measures were significantly worse for the high false alarm rate condition compared to manual performance. To increase a user's level of trust in an automated system and to encourage the user to rely on the recommendations made by the system more appropriately, information should be provided regarding the number of false alarm errors that are expected to occur. In addition, designers should use caution when implementing systems with a high false alarm rate due to the detrimental effects it may have on user confidence or performance. Designers should also be aware that informing the user of the false alarm rate of an automated system may not result in an increase in each of these measures.

References

Allendoerfer, K. R., Pai, S. & Friedman-Berg, F. J. (2008). The Complexity of Signal Detection in Air Traffic Control Alert Situations. *PsycEXTRA Dataset*.

Bagheri, N. & Jamieson, G. A. (2004). The impact of context-related reliability on automation failure detection and scanning behaviour. *2004 IEEE International Conference on Systems, Man and Cybernetics,* Vol. 1, 212- 217.

Bahner, E., Huper, A. D. & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*, 66*,* 688–699.

Balfe, N., Sharples, S. & Wilson, J. R. (2015). Impact of automation: measurement of performance, workload and behaviour in a complex control environment. *Applied Ergonomics*. 47, 52–64.

Breznitz, S. (1983). Cry-wolf: The psychology of false alarms. Hillsdale, NJ: Lawrence Earlbaum.

Canadian Forces (CF) (2011). *Concept for Naval Mine Countermeasures*. Ottawa, ON.

Colebank, J. (2008). Automated intelligent agents: are they trusted members of military teams? *Monterey, California, Naval Postgraduate School,* Unpublished, 91.

Dijkstra, J. J. (1999). User agreement with incorrect expert system advice. *Behaviour and Information Technology, 18(6),* 399-411.

Dixon, S. R., Wickens, C. D. & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors, 49,* 564-572.

Drury, C. G. (2006). Inspection. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (Vol. 2). Boca Raton, FL

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G. & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58, 697– 718.

Dzindolet, M. T., Pierce, L. G., Beck, H. P. & Dawe, L. A. (2002). The Perceived Utility of Human and Automated Aids in a Visual Detection Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 44, 79-94.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A. & Anderson, B. W. (2001). Predicting Misuse and Disuse of Combat Identification Systems. *Military Psychology,* 13, 147–164.

Goddard, K., Roudsari, A. & Wyatt, J. C. (2012). Automation bias: a systematic review of frequency, effect mediators and mitigators. *Journal of the American Medical Informatics Association*, 19, 121-127.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley. (Reprinted 1988, Los Altos, CA: Peninsula).

Ho, G., Pavlovic, N. J., Arrabito, R. & Abdalla, R. (2011). *Human Factors Issues When Operating Unmanned Underwater Vehicles.* (Technical Report DRDC Toronto TM 2011-100). Toronto, ON: Defence R&D Canada.

Hollands, J. G. & Neyedli, H. F. (2011). *Trust in Military Teams* (ed. Stanton, N. A.) Ashgate Publishing Company. 151–182.

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

Joint Commission (2002). Sentinel Event alert. Preventing ventilator-related deaths and injuries. The joint commission American Association of Respiratory Care. Issue 25. Feb 26, 2002.

Kessel, R. T. (2005). *Apparent reliability*: *Conditions for reliance on supervised automation.* (Technical Memorandum DRDC Atlantic No. TM 2005-155). Halifax, NS: Defence R&D Canada.

Kessel, R. T. & Myers, V. (2005). Discriminating man-made and natural objects in sidescan sonar imagery: human versus computer recognition performance. *Proceedings of SPIE: Automatic Target Recognition XV*, Vol. 5807, Orlando, FL, 19 May 2005 (pp. 141-152). Society for Optics and Photonics.

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

Lee, J. D. & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, 40, 153–184.

Lee, J. D. & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors,* 46, 50–80.

Lewandowsky, S., Mundy, M. & Tan, G. P. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied, 6* (2), 104-123.

Lyons, J. B. & Stokes, C. K. (2012). Human-Human Reliance in the Context of Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 54, 112–121.

Madhavan, P. & Wiegmann, D. A. (2004). A new look at the dynamics of human-automation trust: Is trust in humans comparable to trust in machines? *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society. 581–585.

Madhavan, P. & Wiegmann, D. A. (2007). Similarities and differences between human-human and human-automation trust: An integrative review. *Theoretical Issues in Ergonomics Science,* 8, 277-301.

Madhavan, P., Wiegmann, D. A. & Lacson, F. C. (2003). Automation failures on tasks easily performed by operators undermines trust in automated aids. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica: CA.

Mayer, R. C., Davis, J. H. & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management, Rev.* 20, 709–734.

Merritt, S. M., Hall, S., Louis, S., Curran, P. G. & Ilgen, D. R. (2015). Attitudinal predictors of relative reliance on human vs . automated advisors. *International Journal of Human Factors and Ergonomics,* 3, 327–345.

Merritt, S. M., Sinha, R., Curran, P. & Ilgen, D. R. (2011). 'Comparing trust in humans and machines: development of a measure', Poster presented at *55th Annual Meeting of the Human Factors and Ergonomics Society*, Las Vegas, NV.

Muir, B. (1994). Trust in automation. Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37*, 1905-1922.

Myers, V. (2009). *Automatic target recognition for imaging sonar: A systematic survey* (No. DRDC Atlantic TN 2009-060). Halifax: DRDC Atlantic.

Neyedli, H. F., Hollands, J. G. & Jamieson, G. A. (2011). Beyond Identity: Incorporating System Reliability Information Into an Automated Combat Identification System. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 53, 338–355.

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

Parasuraman, R., Hancock, P. A. & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision-warning systems. *Ergonomics, 40,* 390-399.

Parasuraman, R. & Manzey, D. H. (2010). Complacency and bias in human use of automation: an attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 52 (3), 381–410.

Parasuraman, R. & Masalonis, A. J. (2000). Designing Automated Alerting Systems: Standard and Fuzzy Signal Detection Theory and Bayesian Analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 44*(13), 9-12.

Parasuraman, R. & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 39, 230–253.

Parasuraman, R., Sheridan, T. B. & Wickens, C. D. (2000). A Model for Types and Levels of Human Interaction with Automation. 30, 286–297.

Pavlovic, N. J., Ho, G., Smith, D., Maceda, E., LCdr Gowan, S., LCdr Holden, N., . . . Lt(N) Bures, G. (2012). *Scoping Study on DRDC Toronto Future Research Regarding Naval Mine Countermeasures.* (Technical Report DRDC Toronto TR 2011-178). Toronto, ON: Defence R&D Canada.

Phillips, V. L., Saks, M. J. & Peterson, J. L. (2001). The Application of Signal Detection Theory to Decision-Making in Forensic Science. *Journal of Forensic Sciences, 46*(2).

Re, A. & Macchi, L. (2010). From cognitive reliability to competence? An evolving approach to human factors and safety. *Cognition, Technology & Work,* 12 (2), 79–85.

Röttger, S., Bali, K. & Manzey, D. (2009). Impact of automated decision aids on performance, operator behaviour and workload in a simulated supervisory control task. *Ergonomics*, 52 (5), 512-523.

Satchell, P. (1993). *Cockpit monitoring and alerting systems.* Aldershot, U.K.: Ashgate.

Satchell, P. (1998). *Innovation and automation.* Brookfield, VT: Ashgate.

Skitka, L. J., Mosier, K. L. & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51, 991-1006.

Sorkin, R. D. (1988). Why are people turning off our alarms? *Journal of the Acoustical Society of America, 84*, 1107-1108.

Sorkin, R. D. (1989). Why are people turning off alarms? *Human Factors Society Bulletin, 32* (4), 3-4.

Sorkin, R. D. & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. *Human-Computer Interaction*, 1, 49-75.

Swets, J. A. (1998). Separating discrimination and decision in detection, recognition and matters of life and death. *An invitation to cognitive science: Methods, models and conceptual issues* (Vol. 4, D. Scarborough and S. Sternberg, Eds.) (2nd Ed., pp. 635-702). Cambridge, MA:MIT Press.

Swets, J. A. & Pickett, R. M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory.* New York: Academic Press.

Wang, L., Jamieson, G. A. & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human Factors,* 51, 281–91.

Wickens, C. D. & Alexander, A. (2009). Attentional tunneling and task management in synthetic vision displays. *International Journal of Aviation Psychology, 19,* 182-199.

Wickens, C. D. & Dixon, S. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8, 201–212.

Wickens, C. D. & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D., Hollands, J. G., Banbury, S. & Parasuraman, R. (2013). *Engineering Psychology and Human Performance* (4th ed.). Pearson Education.

Yeh, M., Merlo, J. L., Wickens, C. D. & Brandenburg, D. L. (2003). Head up versus head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors, 45,* 390-407.

Yeh, M., Wickens, C. D. & Seagull, F. J. (1999). Target cuing in visual search: The effects of conformality and display location on the allocation of visual attention. *Human Factors, 41,* 524-542.

Yin, S. Q., Wickens, C. D., Pang, H. & Helander, M. (2011). Comparing rate of change cues in trend displays for a process control system. In *Prooceedings of the 55th Annual Meeting of the Human Factors and Ergonomics Society.* Santa Monica, CA: Human Factors and Ergonomics Society.

Appendix A: Psychology Participant Pool (SONA) Contact Sheet

Study title: Trust and reliance on an automated target recognition system for underwater mine detection

Humanities Research Ethics Board approval code: 2019-4853

Principal Investigator: Dr. Heather Neyedli, School of Health and Human Performance

Co-investigator: Shala Knocton, Masters Student

Description: The study aims to look at performance, self-confidence and trust of participants performing an identification task with the support of an automated system. In the study, you will play the role of a naval commander monitoring sonar images of the sea floor. The task is to identify whether a mine is present or absent in each image. All together the experiment is expected to last 1.5-2 hours. The study will be conducted in the Dalplex in the Cognitive Motor and Performance Lab at Dalhousie University. Compensation from this experiment includes $10 if you perform the task well (rank within the top 25% of participants), as well as 2 (two) SONA credits if you are currently enrolled in a psychology class at Dalhousie University. The information gained form this experiment will contribute to our understanding of human-automation interactions and factors that affect performance, self-confidence and trust involving automated systems.

Inclusion criteria: Must be between the ages of 18 and 55 years or old, normal or corrected-to-normal vision.

Location: Cognitive and Motor Performance Lab, Dalplex 218C

Appendix B: Informed Consent Form

DALHOUSIE UNIVERSITY

Study Title: Trust and reliance on an automated target recognition system for underwater mine detection

Lead Researcher:
Dr. Heather Neyedli
Kinesiology, Faculty of Health
Department of Psychology
Dalhousie University
Phone: (902) 494-6786
Email: hneyedli@dal.ca

Other Researchers:
Shala Knocton
Masters Student
Division of Kinesiology, Health and Human Performance, Dalhousie University
Phone: (902) 870-2861
Email: sh846683@dal.ca

Affiliated Researchers:
Dr. Lori Dithurbide – Kinesiology – Dalhousie University
Dr. Aren Hunter – Defense Research and Development Canada

Funding:
This project has received funding from the Defense Research and Development Canada (DRDC) and the Social Sciences and Humanities Research Council (SSHRC).

Introduction
You are invited to take part in the research project described below. The Social Sciences and Humanities Research Ethics Board of Dalhousie University has reviewed the project and found it to conform to current ethical guidelines. These guidelines require:

1) That you be informed of the purpose of the research project and any possible inconveniences, risks, or benefits.
2) That the character of the task required be explained to you.
3) That you understand that participation is voluntary, and that you may decline to continue participation at any point throughout the course of the research project, without loss of expected compensation, nor any academic impact as a result of deciding whether or not to participate.
4) That you be assured that all information assembled is entirely confidential.

Should you have any further question after reading this informed consent form, please feel free to ask question about anything that may have been unclear.

Purpose of the Study

The purpose of this study is to 1) determine how the false alarm rate of an automated target recognition system affects your trust, self-confidence and performance during an underwater mine detection task and 2) to determine whether trust, self-confidence and performance in the automated target recognition system is influenced by informing you of the system's false alarm rate/reliability.

Who Can Take Part in the Research Study

You are eligible to participate in this study if you are between the ages of 18 and 55 years old and have normal or correct-to-normal vision.

What You Will Be Asked to Do

The experiment consists of a computer simulation of a mine detection task. In the simulation, you will play the role of a naval commander aboard a Navy Frigate. The simulation will consist of sonar images of the sea floor. Your task is to identify whether a mine is present or absent in each of the images.

The experiment is expected to take a total of 1.5-2 hours. First, the experimenter will provide you with detailed instructions about the task. After the instructions you will have the opportunity to ask the researcher about any questions or concerns you may have regarding the experiment or the simulation.

Following explanation of the experiment, you will complete a series of training and experimental sessions lasting between 5-15 minutes each. In all sessions, sonar images will appear on the screen in front of you and your task is to identify whether a mine is present or absent. After each session you will be provided an opportunity to take a break if needed.

The training sessions will allow you to get familiarized with the mine detection task and the automated target recognition system. For the experimental sessions, you will perform the mine detection task with and without assistance from the automated system.

Immediately after each of the experimental sessions you will be asked to complete a questionnaire measuring your self-confidence and if you were using the automation, one measuring your trust in the system.

Possible Benefits

Participation in this study may not benefit you directly, but this study will contribute to knowledge within the field of cognitive ergonomics as well as human-computer interactions.

Compensation / Reimbursement
If you achieve a high level of performance in the simulation you will receive $10 for participating in this study. Performance is based on the number of correctly identified mines. The top 25% of participant performances will receive the bonus $10. Participants who are students and are also registered in a psychology class at Dalhousie University will be granted 2 (two) credit points if you have signed up through SONA.

Possible Risks and Discomforts
Possible risks of participation in this study includes fatigue that may be caused by the mental effort required to perform the task in the experiment. The task will also be more challenging at some points. This could lead to stress similar to playing a more challenging level on a video game.

If You Decide to Stop Participating
You may choose not to continue your participation in the study at any time. If you decide not to take part in the study or if you leave the session early, your data will be automatically withdrawn from the study. Further, you may choose to withdraw your data after you have participated. However, once data has been analyzed, it will no longer be possible to withdraw from the study. We will hold off analyzing data for 1 week following collection from the final participant (i.e., study completion) to allow you to withdraw your data after you have participated.

How Your Information Will Be Protected
Every effort to protect your privacy will be made. No identifying information will be included in publications or presentations. Minimal information about you will be collected by the research team, ensuring only required information (such as age, and information from study questionnaires) is collected.

Confidentiality: In order to protect your privacy and keep your participation in the study confidential, you will be de-identified using a study code. For the purpose of data analyses, all participants will only be identified by their study code (e.g., P001) and all data will be stored on password protected computers and spreadsheets. You will be permitted to withdraw your data up to a week after completion of the study (i.e., one week after the final participant is collected), as data typically is analyzed one week after completion. However, once data is analyzed (i.e., after one week of study completion), it will not be possible to withdraw your data.

Data Retention: information that you provide to us will be kept private. Only the researchers will have access to this information. Only anonymized data will be sent to them through password protected files. Your name will not appear on any of these files. We will describe and share our findings in theses, presentations, public media, journal articles, etc. This means that you will not be identified in any way in our reports. The people who work with us have an obligation to keep all research information private. Also, we will use a participant number (not your name) in our written and computer records so that the information we have about you contains no names. All your identifying information contained on the consent will be securely stored separately from your data in a locked cabinet in the Cognitive and Motor Performance Lab.

Questions

We are happy to talk with you about any questions or concerns you may have about your participation in this research study. For further information about the study you may call the principal or co-investigator (Contact information has been provided on the first page of this Informed Consent Form).

If you have any ethical concerns about your participation in this research, you may also contact Research Ethics, Dalhousie University at (902) 494-1462, or email: ethics@dal.ca (and reference REB file 2019-4853).

Appendix C: Informed Consent Signature Page

DALHOUSIE UNIVERSITY

INFORMED CONSENT SIGNATURE PAGE

Study Title: Trust and reliance on an automated target recognition system for underwater mine detection

I have read the informed consent form and meet the requirements for participation as outlined on the screening form for this study. I have been given the opportunity to discuss the study and my questions have been answered to my satisfaction.

I agree that my study information may be used as described in this consent form.

I understand that my participation in this study is voluntary and that I may withdraw my consent from the study at any time, without penalty.

______________________	______________________	____________
Name (Please Print)	Signature	Date

Appendix D: SONA Signature Page

DALHOUSIE UNIVERSITY

SIGNATURE PAGE

(this page must be printed on a separate sheet)

- **Participants Must Read And Sign This Form To Confirm That They Understand And Accept Conditions Before Experiment Can Begin**
- **Participants Must Be Given A Copy Of This Form For Their Information And Records**

Feel free to address any questions you may have about the study to the Principal Investigator / Researcher either now, or after you have participated.

Study Title Trust and reliance on an automated target recognition system for underwater mine detection

Name of Principal Investigator Heather Neyedli

Research Supervisor (if different from PI)

Contact Person (if different from PI) Shala Knocton

Address 6260 South St.

Telephone 902-870-2861

Email sh846683@dal.ca

Psychology Department Subject Pool Policy

Individuals with specific ethical concerns should contact either the Research Supervisor or a member of the Human Research Participants & Ethics Committee of the Department of Psychology & Neuroscience, Tel: 902.494.1580, email psych.ethics@dal.ca.

Please sign below to confirm that you have had your questions answered to your satisfaction, that you are aware that all records are entirely confidential and that you may discontinue participation at any point in the study.

If you anticipate receiving educational credit points for assisting in this research, you may choose to do so as either a **Research Participant** or as an **Observer**.

If you choose to be a Research Participant, the researcher will keep your data and use it in the research project.

If you choose to be an Observer, the researcher will destroy any data that you may have provided, after you complete the study.

Please check one box below to indicate whether you choose to be a Research Participant or an Observer.

☐ **Research Participant (Use my data)**   ☐ **Observer (Destroy my data)**

Participant's Signature: Date:

Researcher's Signature: Date:

FACULTY OF SCIENCE | *Department of Psychology & Neuroscience* | Life Sciences Centre |1355 Oxford Street | PO Box 15000 | Halifax NS B3H 4R2 Canada | Tel. 902.494.1580 | Fax. 902.494.6585 | psych.ethics@dal.ca | dal.ca/psychandneuro

DAL.CA

Appendix E: Age and Gender Form

The following information is being collected to get a general idea of the sample group that took part in this study. Please fill out this information if you are comfortable disclosing it.

Gender: ________________________ Age: ____________

Appendix F: Trust in Automation Questionnaire

Questionnaire about Trust between People and Automation

Instructions: Mark an X at the location of your choice (in a space rather than on a line)
Note: 1 = not at all; 7 = extremely

1. The system is deceptive

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

2. The system behaves in an underhanded manner

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

3. I am suspicious of the system's intent, actions, or outputs

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

4. I am wary of the system

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

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

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

6. I am confident in the system

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

7. The system provides security

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

8. The system has integrity

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

9. The system is dependable

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

10. The system is reliable

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

11. I can trust the system

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

12. I am familiar with the system

|________|________|________|________|________|________|________|
1 2 3 4 5 6 7

Appendix G: Confidence in Abilities Questionnaire

In this section, please answer the questions regarding your confidence in your abilities *WITHOUT THE USE OF THE AUTOMATED SYSTEM.*
For each question, choose an answer between 1 (completely disagree) to 7 (agree).

1. I believe that my ability to identify mines on my own may lead to negative outcomes.
Completely Disagree *Agree*
1 2 3 4 5 6 7

2. I am not confident in my own ability to identify mines.
Completely Disagree *Agree*
1 2 3 4 5 6 7

3. I am wary of identifying mines on my own.
Completely Disagree *Agree*
1 2 3 4 5 6 7

4. My lack of confidence in my ability to identify mines worsens my performance.
Completely Disagree *Agree*
1 2 3 4 5 6 7

5. I am confident in my ability to identify mines.
Completely Disagree *Agree*
1 2 3 4 5 6 7

6. I feel secure in my ability to identify mines.
Completely Disagree *Agree*
1 2 3 4 5 6 7

7. I feel like others could depend on my ability to identify mines.
Completely Disagree *Agree*
1 2 3 4 5 6 7

8. My ability to identify mines is reliable.
Completely Disagree *Agree*
1 2 3 4 5 6 7

9. I can trust my ability to identify mines.
Completely Disagree *Agree*
1 2 3 4 5 6 7

10. My confidence in my ability to identify mines improves my performance.
Completely Disagree *Agree*
1 2 3 4 5 6 7

Appendix H: Training and Experimental Scripts

(*note that each script was given separately before each portion of the experiment)*

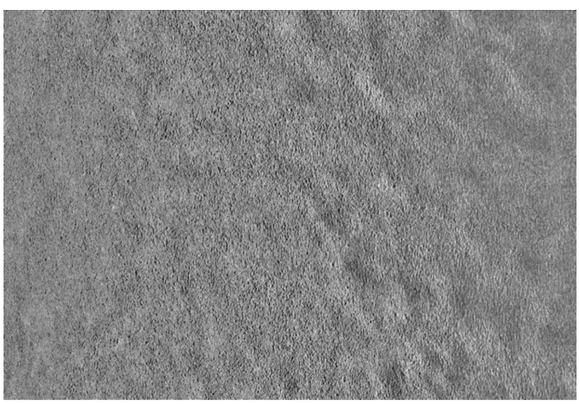
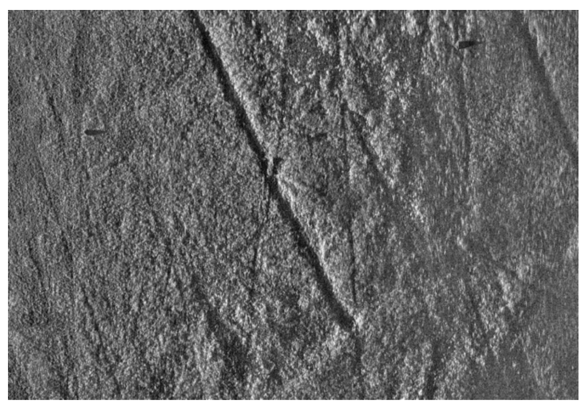
Mine Detection Training Script

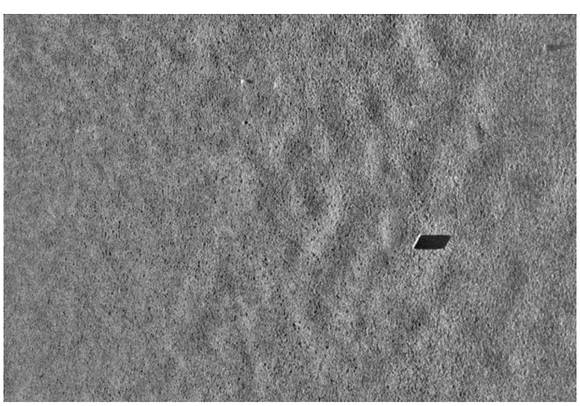
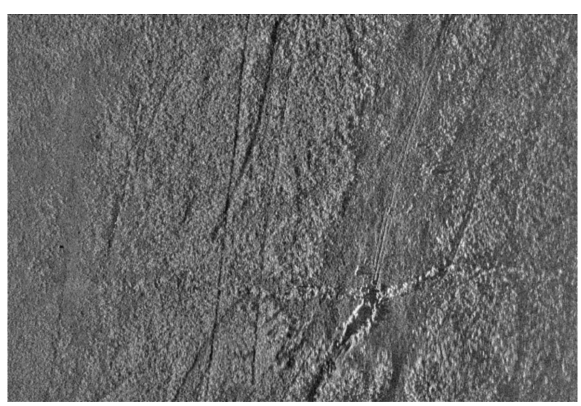
In this simulation, you will play the role of a naval commander aboard a Navy Frigate. You are about to sail in unchartered waters which could become dangerous due to the possibility of encountering underwater mines. A small unmanned submarine has been released into the waters to capture images of the sea floor. Your task is to examine the images to see if any mines are present. It is important that all mines are identified to ensure safe travels through the water. You must examine all of the images before the ship leaves.

You will complete two training sessions, and four mission blocks that range in length from approximately 5-15 minutes depending on the number of images you have to look at for each mission. We will provide you more information before each mission.

As with other sea floors, rocks, sea grass and debris may be present that could look like mines. Also, enemies will try to disguise the mines to appear like rocks or debris. This could make completing your task very difficult.

Here are some example images. The experimenter will now point out which contain mines and which contain rocks or other things that could be mistaken for mines. After the experimenter goes over this with you, you will have the opportunity to look at 50 different images from a previous mission and indicate whether a mine is present or absent. We will tell you if you were correct.

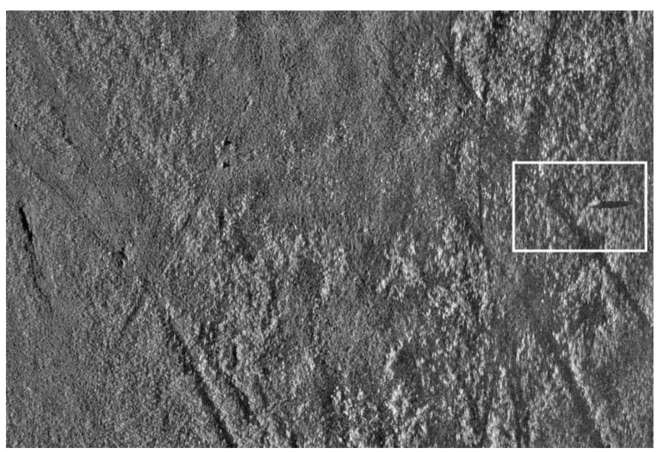
Experimental Session Script – No Automation Block 1

Now that you have completed training, you will examine 50 images that the unmanned submarine has collected of the sea floor along our proposed route. As mentioned before, it is important that all mines are detected; however, you cannot take too long with any one image because the mission needs to progress in a timely fashion. Looking at 50 images usually takes about 5 minutes and you will have a maximum of 8 minutes to examine them before we have to confirm our route to set sail. Therefore, please do not feel you have to rush for any particular image, but you will need to stay focused to make sure you get the task completed on time.

Automation Training Script

An Automated Target Recognition System has been designed to assist you to increase performance on the mine detection task. This system may help you make your decisions more quickly and help you detect more mines. The system uses computer vision to identify the presence of mines. The system isn't entirely perfect. If we want it to detect more mines (in other words, if we want it to miss detecting very few mines), often it can make a false alarm mistake and say a mine is present when it really isn't. The system designers can change the system to trade off how many mistakes of this type the automation makes.

Here is an example of what the Automated Target Recognition's response may look like on one of the images. A rectangle will appear around a region on the sonar image where the system believes a mine is present. You will now have the opportunity to look at 4 different images from a previous mission that display the 4 possible responses that you may receive from the automated system.



Experimental Session Script – Automation Block 1

Now that you have completed training on the automation, you can use the Automated Target Recognition System to help you with this block of trials. As discussed earlier, this system may help you make your decisions more quickly and help you detect more mines. The system uses computer vision to identify the presence of mines. If we want it to detect more mines (in other words, if we want it to miss detecting very few mines), often it can make a false alarm mistake and say a mine is present when it really isn't. The system designers can change the system to trade off how many mistakes of this type the automation makes.

For informed group only, High FA rate condition:

For this mission, your commander has set the sensitivity of the device high. **That means that it is expected that <u>24%</u> of the trials the automation is going to have a false alarm where it says a mine is present but in fact there is no mine. However, this also means that the system may detect more mines.**

For informed group only, Low FA rate condition:

For this mission, your commander has set the sensitivity of the device low. **That means that it is expected that <u>12%</u> of the trials the automation is going to have a false alarm where it says a mine is present but in fact there is no mine. However, this also means that the system may miss a few more mines.**

You can use the automation to help inform your decision, but you must make the final decision. This mission block will consist of 100 images collected from the unmanned submarine. You will have a maximum of 15 minutes to examine the images before we have to confirm our route to set sail though most times it should take only about 10 minutes to look through 100 images. Therefore, please do not feel you have to rush for any particular image because it is important to detect mines, but you will need to stay on task so we can confirm our route.

Experimental Session Script – No Automation Block 2

For this mission we are re-calibrating the automated system so you will not have access to it. You will examine 50 images that the unmanned submarine has collected of the sea floor along our proposed route. As mentioned before, it is important that all mines are detected; however, you cannot take too long with any one image because the mission needs to progress in a timely fashion. Looking at 50 images usually takes about 5 minutes and you will have a maximum of 8 minutes to examine them before we have to confirm our route to set sail. Therefore, please do not feel you have to rush for any particular image, but you will need to stay on task.

Experimental Session Script – Automation Block 2

The Automated Target Recognition System has been recalibrated and can help you with this block of trials. As mentioned before, this system may help you make your decisions more quickly and help you detect more mines. The system uses computer vision to identify the presence of mines. The system isn't entirely perfect. If we want it to detect more mines, often it can make mistakes and say a mine is present when it really isn't. The system designers can change the system to trade off how many mistakes of this type the automation makes.

For informed group only, High FA rate condition:

During the recalibration, your commander has set the sensitivity of the device high for this mission. **That means that it is expected that <u>24%</u> of the trials the automation is going to have a false alarm where it says a mine is present but in fact there is no mine. However, this also means that the system may detect more mines.**

For informed group only, Low FA rate condition:

During the recalibration, your commander has set the sensitivity of the device low for this mission. **That means that it is expected that 12% of the trials the automation is going to have a false alarm where it says a mine is present but in fact there is no mine. However, this also means that the system may miss a few more mines.**

You can use the automation to help inform your decision, but you must make the final decision. This mission block will consist of 100 images collected from the unmanned submarine. You will have a maximum of 15 minutes to examine the images before we have to confirm our route to set sail though most times it should take only about 10 minutes to look through 100 images. Therefore, please do not feel you have to rush for any particular image because it is important to detect mines, but you will need to stay on task so we can confirm our route.

Appendix I: Debriefing Form

Study Title: Trust and reliance on an automated target recognition system for underwater mine detection

Researchers: Heather Neyedli and Shala Knocton

Contact Information: sh846683@dal.ca

Debriefing:

The purpose of this experiment is to better understand how false alarms affect user self-confidence, trust and performance in an underwater mine detection task. Over the experimental blocks, two different false alarm rates were used. The experiment also consisted of 2 groups of participants. One group was informed of the automation's false alarm rates and one was not. Previous research has demonstrated that being informed of an automated system's false alarm rate may help the user have more appropriate levels of trust and reliance. Automated systems and tasks performed under uncertainty are present in a variety of fields (such as air traffic control, patient monitoring, nuclear power plant monitoring) and if appropriate self-confidence and trust in the automated system can be fostered, performance may increase leading to a reduction in the occurrence of errors. We told you at the start of the study that you would receive $10 compensation based on performance; however, all participants regardless of performance receive $10. We did this to increase your motivation on the task to more realistically mimic the motivation that a naval commander might have.

Questions

If after you leave today, you have questions or concerns about your participation in this research study please contact Dr. Heather Neyedli (902 494-6786, hneyedli@dal.ca) at any time with questions, comments, or concerns about the research study.

If you have any ethical concerns about your participation in this research, you may also contact Research Ethics, Dalhousie University at (902) 494-1462, or email: ethics@dal.ca.

Withdrawal of Data

As we stated in the consent form, you still have the opportunity to withdraw your data up until the point it is entered into analysis (typically 1 week after study completion). If you wish to do so after reading this debrief form please inform the experimenter verbally or you can also contact Dr. Heather Neyedli (hneyedli@dal.ca, 902-494-6786) at a later date.