Effects of Automatic Target Detection on Detection and Identification Performance

by

Adam Joshua Reiner

A thesis submitted in conformity with the requirements for the degree of Masters of Applied Science

Department of Mechanical and Industrial Engineering
University of Toronto

© Copyright by Adam Joshua Reiner (2015)
Abstract

Title: Effects of Automatic Target Detection on Detection and Identification Performance

Student: Adam J. Reiner

Degree: Masters of Applied Science

Year: 2015

Department: Mechanical and Industrial Engineering

University: University of Toronto

Automatic target detection (ATD) is an aid being developed to support soldiers in combat. Intended to assist detection by highlighting human targets, ATD is imperfect and may cue different target affiliations at varying reliabilities, resulting in a cueing bias that could impact identification. To evaluate the effects of cueing biases in ATD on target detection and combat identification, an experiment was conducted in a simulated combat environment. Illumination within the simulated scenarios was varied as day and night conditions to assess automation reliance as task difficulty changed. Target detection rate was found to increase when ATD was available. ATD resulted in the greatest improvement relative to the no aid condition at night. Cueing bias affected identification as well. Participants shifted their decision criterion appropriately, though not optimally, for targets cued and missed by the ATD. Participants relied on ATD differently for detection and identification, but improved performance on both tasks.
Acknowledgements

I thank my supervisors, Dr. Greg Jamieson and Dr. Justin Hollands, for giving me the opportunity to work on this project. The guidance and feedback that they have provided throughout the duration of my work has been invaluable. I would also like to thank my committee members, Dr. Paul Milgram and Dr. Mark Chignell, for agreeing to review my thesis. I would like to acknowledge the Cognitive Engineering Laboratory (CEL) as well, especially the work conducted by previous members who partnered with Defence Research and Development Canada (DRDC).

I would like to acknowledge the staff at DRDC who allowed me to use their facilities and laid the groundwork for me to conduct my research. Dr. Mike Tombu, Ken Ueno, and Matt Lamb were of great help in understanding the military domain and developing this experiment in the Virtual Immersive Soldier Simulator. Acknowledgement for funding goes to DRDC.

Finally, I would like to thank my family and friends for providing me with support throughout my academic career.
# Table of Contents

Abstract.......................................................................................................................... ii
Acknowledgements....................................................................................................... iii
Table of Contents .......................................................................................................... iv
List of Figures ........................................................................................................... vi
List of Tables ............................................................................................................... vii
List of Appendices .................................................................................................. viii

1. Introduction ........................................................................................................ 1

2. Automatic Target Detection and Combat Identification ..................................... 1

3. Signal Detection Theory ...................................................................................... 10

4. Automation Reliance and Reliability .................................................................. 19

5. Hypotheses ......................................................................................................... 23

6. Method .............................................................................................................. 25

6.1. Participants .................................................................................................... 25

6.2. Apparatus ...................................................................................................... 26

6.3. Design ........................................................................................................... 29

6.4. Task ............................................................................................................... 33

6.4.1. Procedure ................................................................................................ 35

6.5. Measures ....................................................................................................... 36

7. Results ............................................................................................................. 37

7.1. Detection ....................................................................................................... 38

7.1.1. Cueing Reliance ....................................................................................... 41

7.2. Identification .................................................................................................. 42

7.2.1. Sensitivity (d’) ......................................................................................... 43

7.2.2. Criterion (lnβ) ......................................................................................... 44
List of Figures

Figure 1: Signal Detection Theory Distributions ................................................................. 12
Figure 2: Automation Cueing Bias (Hostile Bias) ................................................................. 15
Figure 3: Virtual Immersive Soldier Simulator ................................................................. 27
Figure 4: VISS Instrumented Weapon .................................................................................. 28
Figure 5: Target Affiliations (From Left to Right): Uncued Militia (Day), Cued Militia (Night), Cued Hostile (Day), Uncued Hostile (Night) ................................................................. 30
Figure 6: Participant Performing the Experiment ............................................................... 33
Figure 7: Post Trial Feedback ............................................................................................... 34
Figure 8: Probability of a detection hit, P(dH), as a function of cueing and illumination. Error bars indicate the 95% confidence interval .......................................................... 39
Figure 9: Probability of a detection hit, P(dH), as a function of cueing and ATD. Error bars indicate the 95% confidence interval ............................................................... 39
Figure 10: Probability of a detection hit, P(dH), for uncued targets as a function of illumination and ATD. Error bars indicate the 95% confidence interval ................................................. 40
Figure 11: Probability of a detection hit, P(dH), as a function of illumination and ATD, collapsed across cueing. Error bars indicate the 95% confidence interval .......................... 41
Figure 12: Decision criterion, lnβ, as a function of cueing and ATD. Error bars indicate the 95% confidence interval ........................................................................................................ 45
Figure 13: Decision criterion, lnβ, for uncued targets as a function of ATD. Error bars indicate the 95% confidence interval ........................................................................................................ 46
Figure 14: Shift in lnβ between cued and uncued targets, as a function of ATD. Error bars indicate the 95% confidence interval ........................................................................................................ 47
List of Tables

Table 1: Human Target Detection Outcomes ................................................................. 14
Table 2: Automation Target Detection Outcomes .......................................................... 14
Table 3: Target Identification Outcomes ........................................................................ 15
Table 4: Number of Targets per Condition ..................................................................... 31
Table 5: Criterion Shift Calculation ................................................................................ 48
Table 6: Correlations between \( P(d_H) \) and \( d' \) ............................................................ 49
List of Appendices

Appendix A: Participant Recruitment Poster................................................................. 72
Appendix B: Pre-Experiment Information Sheet............................................................ 73
Appendix C: Consent Form............................................................................................... 76
Appendix D: Pre-Experiment Questionnaire................................................................. 79
Appendix E: Post-Experiment Questionnaire............................................................... 80
Appendix F: Debriefing Sheet....................................................................................... 82
1. Introduction

Automatic target detection (ATD) systems aid users in detection tasks by cueing targets to facilitate visual search. As with most forms of automation, ATD is imperfect and may miscue or fail to cue targets. Depending on the context, those errors may have severe consequences. In a military setting, ATD automation can help soldiers detect human targets on the battlefield. After detecting a target, the soldier must identify it as friendly or hostile (or some other identity). This identification process can be described using signal detection theory (SDT; Green & Swets, 1966). SDT has been widely adopted by human factors practitioners to investigate performance on detection and identification tasks (Wickens, Hollands, Banbury, & Parasuraman, 2013). By considering the signal detection performance of both ATD and soldier under varying circumstances, it is possible to investigate how soldiers rely on ATD and whether the automation technology improves detection and identification.

2. Automatic Target Detection and Combat Identification

Automated systems—in a variety of forms—are being introduced to the dismounted soldier. These include: blue force tracking (BFT) systems that show the location and identity of friendly forces on a map background (e.g., Bryant & Smith, 2012; Ho, Hollands, Tombu, Ueno, & Lamb, 2013); automatic targeting systems that detect potential hostile targets (Tombu, Ueno, & Lamb, 2014); and combat identification (CID) systems that help the soldier identify friendly forces (Karsh, Walrath, Swoboda, & Pillalamarri, 1995; Neyedli, Hollands, & Jamieson, 2011; Wang, Jamieson, & Hollands, 2009). In combat, the consequences of incorrect identification include fratricide and failing to notice a hostile target. CID and BFT automation are intended to support
the soldier’s identification process by identifying friendly units (Karsh et al., 1995), while ATD is intended to direct the attention of the user to potential hostile targets in the environment.

Automated CID systems have received a good deal of attention in the human factors literature (e.g., Chancey & Bliss, 2012; Karsh et al., 1995; Neyedli et al., 2011; Wang et al., 2009). In contrast, ATD technology has received more limited consideration, typically focusing on in-vehicle cueing systems (Maltz & Shinar, 2003; Yeh, Wickens, & Seagull, 1999; Yeh & Wickens, 2001). Dzindolet, Pierce, Beck, Dawe, and Anderson (2001b) investigated the benefit of an aid that stated whether or not a soldier was present or absent within a series of images, without indicating target location. They found that participants tended to over-rely on or misuse (agree with the aid when it was incorrect) rather than disuse (disagree with the aid when it was correct) the automation’s binary response. Participant error rate and automation use were not affected by changes in the automation’s reliability.

Defence Research and Development Canada (DRDC) is investigating the potential of automated combat systems to support soldiers in the field, one of which is ATD. ATD is a detection aid that locates and highlights human targets in the environment and is intended to aid dismounted soldiers by overlaying additional information through an optical scope mounted on their weapon. The algorithms and target features used by the ATD for detection are currently in development (Defence Research and Development Canada, 2014). For instance, a target might be defined by shape or thermal signature (Se, 2013). The purpose of this investigation is to determine the potential benefits and risks of using less than perfectly reliable ATD in a simulated combat situation.
As noted earlier, ATD is not CID: ATD will highlight friendly or hostile targets under the same circumstances (i.e., at the same distance and in the same position having the same shape or thermal signature). However, certain factors may cause targets of one affiliation to be cued more frequently than another. For instance, friendly soldiers may be detected because they are closer to the ATD sensor, or other soldiers might wear clothing that conceals their heat signature, making them less detectable. The responsibility for target identification will be on the soldier using ATD. Nonetheless, if soldiers are aware of the types of errors that the automation is likely to make, they may adjust their expectations. If for example the ATD is more likely to detect friendly soldiers, the soldier may take that into consideration during the identification process.

Prior to identification, soldiers will first need to perform a visual search of their environment to detect targets. Many studies have investigated target cueing aids and how they facilitate detection (Maltz & Shinar, 2003; Yeh et al., 1999; Yeh & Wickens, 2001). Some of these studies have focused on target detection while others investigated identification as well. While there is a wealth of psychology studies focusing on the detection and identification of simple stimuli (e.g., Muller & Rabbitt, 1989; Murrell, 1977), this review will focus on more realistic, complex stimuli found in the military human factors context.

Yeh, Wickens, and Seagull (1999) investigated the benefit of target cueing in a visual search task. Their participants acted as scouts in an aircraft, detecting targets and communicating their location to a command centre using a radio. A secondary task required the participant to radio the direction the target was facing. Cues indicated only certain target types. Yeh et al. (1999) found that for target types with 100% reliable cueing, time to detect was reduced, though accuracy was unaffected. In the scenarios where a target was cued, participants were less likely
to detect the presence of an uncued object. Thus, cueing produced a higher miss rate for uncued targets.

Yeh et al. (1999) also investigated the relative benefit of a handheld versus head-mounted display (HMD). They found that the HMD produced slower detection times for uncued targets than the handheld display, and suggested that this result was due to HMD clutter. In a later study that was similar in experimental design to the visual search task of Yeh et al. (1999), Yeh, Merlo, Wickens, and Brandenburg (2003) found that a HMD with reduced clutter resulted in better performance than a handheld display in most scenarios. Current ATD is being designed for use within a weapon scope and will overlay automated visual cues over areas of interest, like a HMD. The Yeh et al. (2003) results suggest that as long as ATD does not have a cluttered visual presentation, it may result in a greater improvement to detection than other handheld aids, such as BFT.

Yeh and Wickens (2001) investigated the effects of automation reliability (100% and 75% reliable) and image realism (high and low) on target detection and identification performance. Yeh and Wickens used an experimental design similar to that of Yeh et al. (1999), but presented a box around the target as the cue (Yeh et al., 1999, used an arrow). Yeh and Wickens found that the reduced cue reliability resulted in performance no different from uncued targets in terms of detection accuracy for unexpected targets, or in terms of the distance at which targets were detected. Participants over-relied on the automation at 100% reliability, resulting in a high false alarm rate and poorer detection of uncued targets. The less reliable automation may have reduced reliance on the cues and improved awareness of uncued targets. However, automation reliance increased when more realistic images were used.
Studies conducted at Defence Research and Development Canada (DRDC) have also investigated the benefit of target detection aids. (Some of this work is classified and not publicly available.) Tombu, Ueno, and Lamb (2014) investigated the benefits and risks of ATD in terms of target detection and combat identification using a high fidelity simulator and participants with military training. Participants were shown a scene with between one and six partially obscured human targets and had 25 seconds to perform a visual search task to detect and identify all targets. ATD was simulated by drawing a yellow box around some targets. Tombu et al. compared the participants’ performance under conditions with imperfect cueing (ATD misses, false alarms, or both) to conditions with perfect cueing and without cueing. Tombu et al. found that ATD, even when not fully reliable, aided detection performance and resulted in more detections and correct identifications than without an aid. ATD misses had a more negative effect on performance than ATD false alarms since participants were typically able to quickly reject the ATD false alarms. As the number of targets within a trial increased, the benefit of ATD to human detection accuracy over no aid conditions also increased.

In another DRDC study, Glaholt (2014) also investigated the benefit of target cueing. The study had participants perform a visual search for, then identify, targets at different differences (near or far) that were moving or stationary. Glaholt found that cueing significantly reduced target detection time, a benefit that was greater for stationary targets and improved with distance. Though cueing improved detection, it had a negative impact on identification. In contrast to Tombu et al. (2014), Glaholt (2014) found an identification penalty to cueing where participants were more likely to misidentify a target if cued, particularly if that target was moving. Participants also required more time to identify cued targets. Glaholt also found that decreasing
cue salience reduced its negative effect on identification, and that increasing cue size eliminated the identification penalty.

Target cueing studies of this type have not considered the potential identity information that may be present if one target affiliation is cued at a disproportionate frequency to another. Cueing bias occurs when a cue - whose purpose is to aid detection of potential targets - highlights a particular target identity more frequently than others. Cueing bias refers to the automation itself and is distinct from automation bias (Mosier, Skitka, Heers, & Burdick, 1998), which describes the degree to which a human operator relies on automation. Cueing bias, however, may impact automation bias and reliance. For example, soldiers who are aware of cueing bias might take it into account when making an identification decision. Alternatively, they might neglect the cueing bias and continue to make decisions as if the automation detected different target identities at the same rate.

Target detection is necessary but not sufficient for target identification. Factors like visibility can reduce the likelihood that a target will be detected; if it is not detected, it cannot be identified. Cueing bias in ATD can affect target identification in an indirect manner: automation reliability affects target detection, which then affects identification. Moreover, target detection and identification are affected by reliance and trust on the automated system. Consider ATD that detects friendly soldiers at an 80% rate and hostiles at 60%. While this automation does not provide the target’s identity, the soldier may make more correct identifications if the cueing bias is understood. If soldiers are uncertain about a cued target’s identity, they should be more willing to hold fire given that the automation is known to cue friendly soldiers more frequently. Optimal decision making performance for targets that are missed by the automation would be achieved if soldiers using the aid were twice as likely to engage an uncued target, since two thirds of targets
that are not cued are expected to be hostile (assuming that the actual ratio of hostiles to friendlies is equal, 40% of hostiles and 20% of friendlies would be missed, resulting in a 2:1 ratio of hostiles to friendlies that are uncued). If the automation hit and false alarm rates are the same regardless of target identity, then the soldier should not rely on cueing bias when making a decision.

Soldiers may be more likely to consider the potential identity information in the cueing when they encounter targets that are harder to identify. Maltz and Shinar (2003) investigated the effect of cueing reliability on a visual search target detection task by manipulating task difficulty (within subjects) and automation hit and false alarm rates (between subjects). Using nine different levels of automation reliability (three cueing hit rates by three false alarm rates), Maltz and Shinar found that for low task difficulty, cueing did not aid performance. The task was easy enough to perform unaided that even automation that was nearly completely reliable did not improve sensitivity, but error prone automation decreased performance relative to the control (no automation) condition. For high task difficulty, participants that saw automation with few false alarms approached optimal reliance on the cueing system and had a significant improvement relative to the control condition. Cue dependency was also greater in the more difficult condition.

Though Maltz and Shinar did not investigate the effects of cueing on identification, an increased reliance on detection automation may transfer to identification, especially in conditions where the task difficulty is increased. Kogler (2003) investigated the effects of CID automation reliability and visual transmissivity (through ballistic goggles) on target identification and found that as visual ability was reduced, participants relied more on 60% reliable automation than their own ability. How reliance on one feature or purpose of the automation extends to other features
will depend on how users trust and rely on the system as a whole, and if that trust is calibrated appropriately (Lee & See, 2004).

Prior to describing methods of assessing reliance on ATD and cueing bias, it is important to consider the target identification task that must be carried out in combat. In the current study, a target will refer to an individual on the battlefield. Targets can have friendly or hostile affiliation, which will be associated with the correct ‘hold fire’ and ‘engage target’ responses, respectively. In this context, neutral targets or non-threats may be considered friendly since soldiers should hold fire upon identifying them.

Even in a simulated environment, participants with military training may differ from civilians in identifying targets that are considered a threat and the degree of certainty required to make a combat decision. In combat, Canadian Armed Forces (CAF) soldiers are required to make a positive identification before discharging their weapon and are legally responsible for each shot that they take. Positive identification here refers to visual confirmation of target characteristics or behaviour about which the soldier must be certain prior to classification as a threat (Department of National Defence, 2009). While the positive identification requirements discussed refer to Canadian Armed Forces, other allied countries will have their own similar regulations and requirements for procedures regarding identification and soldiers would likely react similarly to CAF in a given situation. Bryant and Smith (2009) investigated the impact of salience and uncertainty of visual and behavioural target features. They found that participants, who had received prior military training, were more likely to engage targets and had an increased false alarm rate when simulated targets were more likely to exhibit salient visual and behavioural cues that were predictive of enemy targets relative to non-salient cues and the baseline condition.
Positive identification depends on the rules of engagement, which differ for given combat situations. For example, for a soldier to justify engaging a target, that target would typically need to perform some hostile action, such as pointing a weapon towards the soldier. However, in a firefight in an operational theatre, the positive identification requirement may depend only on uniform/clothing or even weapon style. For example, if an individual is carrying an AK47 then they might be judged as hostile, since that weapon is not used by the Canadian Armed Forces or the armed forces of allied nations. In those cases, the soldier may assume that the target is hostile regardless of behaviour. In general, the soldier must make decisions that have life and death consequences under extreme time pressure.

The perceptual cues, also referred to as information channels, available to soldiers that are useful for identification include – but are not necessarily limited to – target uniform, weapon characteristics, and target behaviour (Bryant & Smith, 2009). A soldier can therefore be considered to be sampling from a range of information sources, varying in diagnosticity (Hollands & Neyedli, 2011). The uniform, for example, is typically a highly reliable cue, so much so that in the event that hostiles were able to obtain friendly uniforms, they might be incorrectly identified and pass as friendlies.

Hollands and Neyedli (2011) consider a sampling model for CID, where an observer views multiple information channels from the target to be identified, referred to as the manual channels, as well as some information from an automated aid. Each cue has a different reliability level and produces some output. The sum of all outputs is compared to a criterion to determine if the target is a threat. The CID sampling model considers automation as another cue amongst multiple features used to identify a target. The individual weighting of target cueing likelihood will vary from soldier to soldier and situation to situation. For example, under reduced visibility soldiers
may rely on less reliable automation (Kogler, 2003). Disclosing the automation reliability level to users has been shown to lead to more appropriate reliance (Wang et al., 2009). Appropriate reliance will depend on the discriminability of the levels of each channel. For example, manual channels may be difficult to distinguish with reduced visibility and the user may give a greater weight to unreliable automation.

Depending on the strength of cueing bias, a soldier may not find it helpful to rely on that discrepancy when making a positive identification. If the detection discrepancy between friendly and hostile targets is slight, (e.g., a 5% difference in detection likelihood), the bias may go unnoticed. In addition, if the difference is larger, a soldier might overestimate the detection bias and associate cueing with friendliness or hostility. As the cueing bias increases, uncertainty about the identity of a cued or uncued target would decrease and ATD would become primarily an identification aid. While this may result in higher reliance on ATD as an identification aid, it would undermine its purpose as a detection aid. Furthermore, given that the system is designed to detect targets regardless of identity, it is unlikely that the cueing bias will be large. For these reasons, the cueing bias examined in this study will tend towards a slight difference between detection reliabilities. (For example, Neyedli et al. (2011) used intervals of 10% differences in reliability, which would be considered small relative to Yeh and Wickens (2001) or Kogler (2003), who used differences of 25% and 40%, respectively, between reliability levels.) In order to investigate if cueing bias has an impact on identification, signal detection theory is considered.

3. Signal Detection Theory

SDT (Green & Swets, 1966; Macmillan & Creelman, 2004) classifies responses in a detection or identification task into four possible outcomes: hits, false alarms, misses, and correct
rejections. Based on these results, SDT provides a measure of the observer’s sensitivity, $d'$, and subjective decision threshold or criterion, $\beta$, which reflects the observer’s willingness to respond ‘yes’ or ‘no’. These measures are domain independent and can be used to characterize any type of detection or identification performance. Gaussian distributions of equal variance are used to represent background noise and presence of a signal with variability due to noise. In terms of detection, the noise distribution can refer to background activity that is not a signal, but may appear to be one.

For identification, the observer must first detect something, correctly or not. He or she must then make a judgement about whether or not the detected object meets the criteria of the object that is being looked for. The further apart the distributions, the greater the sensitivity, or ability to differentiate between a signal and noise. The sensitivity for detection is independent of identification; an observer may be able to detect an object with little difficulty, but determining whether that object matches what he or she is looking for may be much more difficult. An observer’s decision threshold can be represented as a line separating ‘yes’ from ‘no’ responses and can change independent of the sensitivity. The observer will have a more liberal criterion as he or she becomes more willing to say ‘yes’, which will result in a greater number of hits, but also more false alarms. A conservative bias would result in fewer false alarms and more correct rejections, but also fewer hits and more misses.
In this experiment, $\ln \beta$ was selected as the measure of an observer’s decision criterion. The value of $\beta$ is unbounded, ranging from zero to infinity. When bias is liberal, the range of values is between zero and one; when bias is conservative the range is between one and infinity. Thus, the range of conservative values is much greater than the range of liberal values. By taking the natural logarithm, we obtain a linear scale that is conceptually easier to interpret in the sense that a positive value is as conservative as a negative value is liberal (Macmillan & Creelman, 2004; Murrell, 1977; See, Warm, Dember, & Howe, 1997). The value $\ln \beta$ has also been used as a measure of criterion in other studies that have evaluated reliance on detection and identification automation (e.g. Maltz & Shinar, 2003; Wang et al., 2009).

An observer needs to weigh the value of hits and correct rejections, $V(H)$ and $V(CR)$, against the costs of false alarms and misses, $C(FA)$ and $C(M)$, when setting the decision criterion, for both detection and identification. Signal likelihood is an important element of criterion setting. If a signal is unlikely, the observer will be less likely to respond positively, even if the value of detecting a signal is high (Wickens et al., 2013).

Signal detection measures $d'$ and $\ln \beta$ can be calculated empirically based on the following formulas, where $z(H)$ and $z(FA)$ are calculated as the inverse of the cumulative

![Figure 1: Signal Detection Theory Distributions](image)

In this experiment, $\ln \beta$ was selected as the measure of an observer’s decision criterion.
standard normal distribution for the probability of getting a hit, \( P(H) \), and a false alarm, \( P(FA) \), respectively (Macmillan & Creelman, 2004):

\[
d' = z(H) - z(FA) \tag{1}
\]

\[
\ln \beta = -d' \frac{z(H) + z(FA)}{2} \tag{2}
\]

Sorkin and Woods (1985) extended SDT to describe human-automation interaction. In their description, humans and automation \((h \text{ and } a)\) each receive an input from their environment, \(X_h\) and \(X_a\) respectively, which are translated through their respective detection systems, each having its own sensitivity and criterion, to produce a response. If the automation’s response is available to the human monitor, it can be used in the decision making process. The automation detection parameters can therefore affect human decision making. For example, if the automation is prone to false alarm, the human may choose a more conservative criterion than without the automation.

The decision weight that the human assigns to the automation’s response will depend on its reliability as well as the human’s ability to perform the task. Dzindolet, Pierce, Beck, and Dawe (2001a) refer to this gap as the perceived utility of automation. The human may rely on unreliable automation if their ability to perform the task is limited, or neglect reliable automation if they feel that they are able to perform the task well enough (Dzindolet, et al., 2001a).

For a soldier, detection consists of a visual search for human targets in an environment. Detection correct rejections, dCR, will happen constantly, rejecting objects in the environment that are not human. Detection false alarms, dFA, would typically be rare, but soldiers may be distracted by an object in the environment that appears human, such as an animal, a shadow, or
foliage, and mistake it for a target. False alarm rate may increase under visually demanding circumstances such as during a firefight or a night operation. Detection hits and misses (dH and dM respectively) are better defined since there is a human target present that is either detected or not. Table 1 summarizes the four possible outcomes for target detection.

Table 1: Human Target Detection Outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Target Present</th>
<th>Target Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Detected</td>
<td>dH: Detect target</td>
<td>dFA: Mistakenly respond that target was present</td>
</tr>
<tr>
<td>No Response</td>
<td>dM: Miss target</td>
<td>dCR: Provide no response when no target is present</td>
</tr>
</tbody>
</table>

The detection outcomes for ATD can be defined similarly to the human’s: hits (aH), misses (aM), false alarms (aFA), and correct rejections (aCR) (see Table 2). Cueing bias considers that different types of targets may be easier or more difficult to detect by the ATD. That is, the ATD could have a different detection sensitivity for different target identities (where friendly and hostile soldiers are signals and anything else is noise). If the ATD is more likely to detect hostile soldiers (hostile bias), the detection $d'$ is larger for hostile targets, but smaller for friendlies (Figure 2). Assuming that the ATD has the same detection criterion for either affiliation, the ATD will miss more friendly targets than hostiles. The larger the difference in $d'$ between friendly and hostile targets, the stronger the cueing bias.

Table 2: Automation Target Detection Outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Target Present</th>
<th>Target Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cue</td>
<td>aH: Cue target</td>
<td>aFA: Cue without target present</td>
</tr>
<tr>
<td>No Cue</td>
<td>aM: Uncued target</td>
<td>aCR: No cue for no target</td>
</tr>
</tbody>
</table>
Identification outcomes need to be defined differently than detection (Table 3). Identification hits (iH) refer to correctly identifying a human target as a threat, and correct rejections (iCR) refer to correctly determining that a human target is not a threat. The cost of an identification false alarm (iFA) can be devastating, resulting in a friendly fire incident. But an identification miss (iM), letting a hostile pass as friendly, can also have severe consequences. A conservative bias means that the shooter tends to hold fire, resulting in fewer engagements and a larger lnβ, whereas a liberal bias has a smaller or negative lnβ, meaning the shooter will say yes, or engage targets more frequently. Individual soldiers will likely weigh the identification costs and values differently based on expectations, training, and previous combat experience, leading to different criterion settings. For that reason, looking at criterion setting under different conditions may help in understanding the impact of various factors.

Table 3: Target Identification Outcomes

<table>
<thead>
<tr>
<th></th>
<th>Hostile Target</th>
<th>Friendly Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fire Weapon</td>
<td>iH: Engage a hostile target</td>
<td>iFA: Engage a friendly target</td>
</tr>
<tr>
<td>Hold Fire</td>
<td>iM: Detect and clear a hostile target</td>
<td>iCR: Detect and clear a friendly target</td>
</tr>
</tbody>
</table>

Figure 2: Automation Cueing Bias (Hostile Bias)
Wang et al. (2009) used the difference in criterion, referred to as $\ln \beta$ difference, between automation settings to determine if participants adjusted their criterion between conditions with and without automation as well as between automation reliabilities. If participants shift their criterion based on the automation, they have adjusted their decision making and can be said to have relied on the automation’s reliability in some capacity (regardless of if this shift was in the appropriate direction). The criterion shift is the ratio of $\beta$ in one condition to another (i.e. $\beta_{\text{manual}} : \beta_{\text{automation}}$). When $\ln \beta$ is used, this ratio becomes the difference in criterion (i.e. $\ln \beta_{\text{manual}} - \ln \beta_{\text{automation}}$), and since $\ln \beta$ uses a linear scale, an equal difference between two sets of points means a criterion shift of the same magnitude has occurred regardless of how conservative or liberal the initial criterion were. The equation for a soldier’s optimal identification decision criterion would be as follows (Macmillan & Creelman, 2004; Wang et al., 2009):

$$
\ln \beta = \ln \left[ \frac{P(\text{Friendly})}{P(\text{Hostile})} \right] + \ln \left[ \frac{\text{value of friendly not hit} + \text{cost of friendly hit}}{\text{value of a hostile hit} + \text{cost of hostile missed}} \right]
$$

Wang et al. (2009) consider that if the optimal decision criterion is calculated for different conditions, then taking the difference between them should result in the optimal shift in decision making between those conditions. It is then possible to compare the observed performance in decision making to the calculated optimal difference in order to determine if participant adjustment to decision criterion matched the difference between the optimal criterion values.

Wang et al. (2009) had participants perform a combat identification task with the help of automation that correctly identified friendly targets at a given reliability level, and returned a target “unknown” response for targets (hostile and friendly) that the automation could not
identify. As the identification reliability for friendly targets increased, more hostiles would be expected when “unknown” feedback was shown. This outcome should ideally result in a more liberal criterion for the user. In order to control for individual bias, Wang et al. calculated the criterion shift between automation reliability conditions as the difference between a participant’s $\ln \beta$ for each condition pair among the three levels of automation reliability they examined. This shift in $\ln \beta$ was used as a measure of reliance to determine if participants changed their criterion based on enemy likelihood when receiving the “unknown” response, thereby using the “unknown” response to aid identification.

The calculation for the optimal shift in $\ln \beta$ between cued and uncued targets for ATD can be calculated as follows:

$$\ln \beta \text{ Shift} = \ln \beta_{\text{Cued}} - \ln \beta_{\text{Uncued}} = \ln \left[ \frac{P(\text{friendly|cued})}{P(\text{hostile|cued})} \right] - \ln \left[ \frac{P(\text{friendly|uncued})}{P(\text{hostile|uncued})} \right]$$

(4)

This formula is derived from Wang et al.’s calculation of the criterion shift between the optimal manual and “unknown” responses. Since hostile and friendly targets may have different probabilities of being present based on the cueing bias, $\ln \beta$ can be calculated for a target that is cued or uncued. When the difference between the $\ln \beta$ for each cueing conditions is taken, the costs and values of each signal detection outcome (seen in equation 3), if assumed to be constant regardless of cueing (i.e. engaging a friendly target should have the same cost whether they were cued or not), will cancel out. How much a soldier should change his or her criterion between cued and uncued targets regardless of individual outcome costs and values results.

Wang et al. (2009) found that participants who were informed about the automation’s reliability could adjust their criterion such that it was not different from the optimal shift for
certain conditions. When automation reliability was not disclosed, participants did not adjust the criterion optimally – indeed, it did not differ from zero. Thus, participants did not appear to adjust their criterion based on trial-to-trial experience, but rather based it on disclosed reliability levels. The experiment conducted by Neyedlı et al. (2011) also showed that participants adjust their criterion in the right direction based on automation reliability information (all participants were informed of the automation’s reliability), but again not optimally. Maltz and Shinar (2003) found that participants also adjusted their criterion based on cueing; participants were more liberal in making detections with cues and more conservative without them. As cue reliability increased, the difference between lnβ with and without cues also increased.

Criterion adjustment falling short of the optimal value has been seen in other automation interaction research (Meyer, 2001) and is often referred to as ‘sluggish beta’ (Wickens et al., 2013). ‘Sluggish beta’ refers to the insufficient adjustment in human response to changes in and/or extreme probability. Humans have been shown to underestimate high probability events and overestimate low probabilities (Wickens et al., 2013). Due to its primary function as a detection aid, cueing bias is expected not to reach extreme probabilities (i.e. otherwise it is not viable as a detection aid), but soldiers using the automation may have difficulty adjusting to the changes in probability between cued and uncued targets, especially if they are detecting and identifying targets in quick succession. The duration of experience with a particular cueing bias may cause a sluggish adjustment if participants require time to adjust to different target cueing biases.

Signal detection theory provides a method of describing performance on detection and identification tasks using two independent measures, sensitivity and criterion. Reliance on automation can be assessed though comparing decision making under various settings. This
difference can then be compared to an optimal value. For the current problem, reliance on cueing bias as participants switch between cued and uncued targets can be assessed to determine if and how they react to ATD’s potential identity information.

4. Automation Reliance and Reliability

With signal detection theory and criterion shift having been discussed as a method of describing the effects of cueing bias on reliance, we now consider how reliance has been addressed in previous military human factors and human automation interaction literature, and how their findings may impact reliance on ATD. The effects of automation errors and reliability are also considered.

Reliance on automation has been shown to depend on the user’s perception of his or her ability to perform the task (Lee & Moray, 1994). Similarly, Dzindolet et al. (2001a) argue that one of the factors that influences automation use is the gap between the perceived reliability of the automation and of the user, defined as the automation’s perceived utility. If the user feels that they are able to competently perform the detection and identification tasks, they may disuse cueing bias for identification or even the ATD as a whole as a detection aid.

One aspect that may affect the automation’s perceived utility is the timeliness of the automation’s feedback. Dzindolet, Pierce, Beck, and Dawe (2002) investigated the effects of CID automation on detection where participants received the automation’s response after the trial. Participants were shown images of a wooded area that either did or did not have a soldier present. After viewing the image for 750 milliseconds, participants were asked whether or not they detected a target, after which they were told whether or not an imperfect automation aid had detected a target. After 200 trials, participants were told that a random sample of the completed
trials would be taken and that for every correct response, they would receive additional compensation. For this random sample, participants could have used either their responses or the automations’. Dzindolet et al. (2002) found that participants disused the automation and tended to rely on their responses, even when the automation was superior (made half as many errors as participant average). Even when the automation outperformed participants, they tended towards self-reliance.

In another CID study, Karsh et al. (1995) found that participants disused an identification aid when available, interrogating less than 15% of targets. Despite its disuse, participants who had access to the CID aid made fewer errors than those who did not, principally resulting from a decrease in identification misses. Participants with the CID aid had greater response times than those without. This was likely due to a delay when the automation was activated; the aid took 750 milliseconds to respond to a target.

The results of Dzindolet et al. (2002) and Karsh (1995), who found that participants tended towards self-reliance, are inconsistent with those of Dzindolet et al. (2001b), who found that misuse was more prevalent than disuse regardless of automation reliability. One key difference between these studies is when the automation feedback was presented; the beginning of the trial for Dzindolet et al. (2001b), after the participant’s response for Dzindolet et al. (2002), and after 750 milliseconds for Karsh et al. (1995). The participants in Dzindolet et al.’s experiment (2002) had already performed the task and generally chose to rely on the decisions that they had already made. When considered within the framework of automation use by Dzindolet et al. (2001a), this may have been due to a lack of perceived utility of the automation in contrast to their skill and the time and effort that they had already exerted to complete the task. In general, it seems that advice from automated systems needs to be available to the human at the time of decision to
avoid disuse, though immediate feedback may encourage misuse instead (Dzindolet, 2001b). ATD is designed as a detection aid so the cues are present throughout the duration of its use. When a soldier spots a target, any cueing bias information will be available immediately by virtue of the target being cued or not.

While cueing bias refers to a specific type of error, the overall reliability of the ATD and the types of errors that it makes (i.e. detection misses and false alarms) will impact how soldiers use the aid in combat. Meyer (2001, 2004) considered automation use in terms of compliance and reliance, and defined *compliance* as an operator’s response when the automation indicates that a signal is present and *reliance* as the operator’s assumption that no signal is present when the automation is silent. Dixon, Wickens and McCarley (2007) investigated the relationship between these types of automation use and the effect of automation misses and false alarms. Dixon et al. found that false-alarm prone automation affected both compliance and reliance, whereas miss-prone automation only affected operator reliance. Using a signal detection analysis, Maltz and Shinar (2003) also found that the automation false alarm rate negatively affected participant detection performance and decision criterion more significantly than automation miss rate and that at high false alarm rates participants disused the automation. Dixon and Wickens (2006) found that error prone automation at 80% reliability (with equal miss and FA rates) improved performance over a baseline condition, which had no automation assistance. This supports the conclusion that imperfect automation has been shown to improve performance (St. John & Manes, 2002).

Multiple other studies have included automation reliability as a factor and have often shown that more reliable automation can lead to improved performance and greater automation reliance. In a meta-analysis of automation reliability literature, Wickens and Dixon (2007) concluded that
at reliabilities below 70%, the automation tends to provide little or no improvement to performance in detection and identification tasks. Tombu et al. (2014) found that low automation detection reliability conditions (up to 33% miss rate and 20% false alarm rate) still provided some benefit over a no-aid condition and in many cases was not significantly different from a 100% reliable aid in terms of detection and identification rate.

Reliance on automation and the performance of the human-automation system has been shown to depend on presentation (i.e. Handheld vs. HMD in Yeh et al., 2003), timeliness (Dzindolet et al., 2002), and reliability (Wickens & Dixon, 2007) of feedback. Given the proposed design of ATD, reliance in the current study is investigated in terms of cueing bias. Target cueing has shown benefits as a detection aid (Yeh et al., 1999, Maltz & Shinar, 2003, Tombu et al., 2014), but its effects on identification are often overlooked. If ATD misses different target identities at different rates, and this fact is relayed to the soldier, there may be a net benefit to target identification. Biased ATD can thereby function as an identification aid implicitly due to its detection discrepancies and change the user’s identification sensitivity or decision criterion for cued and uncued targets. Though previous studies have considered the impact of varying automation error rates for detection and cueing aids (Maltz & Shinar, 2003; Tombu et al., 2014), none have considered how the automation may make errors that have direct implications for identification. This study investigated the impact of cueing bias on detection and identification performance. In order to determine how participants change their decision making when a detection aid can impact identification, reliance was described in terms of their criterion shift.

Automation reliance can also be affected by conditions that make the human’s task more difficult. As users are less able to detect or identify targets, they may rely more on automation,
regardless of its reliability. Other studies that have considered some measure of reduction to participant visual ability have focused on either detection (Maltz & Shinar, 2003) or identification (Kogler, 2003). This experiment investigated the impact on both detection and identification in terms of performance and automation reliance under day and night conditions. Illumination affects both the visual ability to detect and identify a target and has a realistic implication for combat performance and automation use since soldiers are required to conduct operations throughout the day.

To investigate how users of ATD would react to cueing bias, this experiment used a high fidelity combat simulator and military participants to determine ATD’s effects on target detection rate and identification sensitivity and criterion in a visual search and identification task. To investigate the shift in $\ln \beta$ based on cueing likelihood each participant performed all ATD settings, including cueing bias conditions that were more likely to detect hostile or friendly targets as well as baseline conditions with and without ATD. Participants saw each ATD condition under both day and night illumination. Since imperfect automation that made false alarms and missed targets was used, targets were cued at a rate based on their identity and the ATD’s cueing bias.

5. **Hypotheses**

5.1. Detection: Participants should detect more targets in conditions where ATD is available (Tombu et al, 2014). With unreliable automation, participants should also detect cued targets at a higher rate than uncued targets (Yeh et al., 1999). ATD is expected to have a greater detection benefit for the night illumination condition (Maltz & Shinar, 2003).
5.1.1. Automation misses: aMs, are expected to negatively affect participant detection hit rate, dH (Tombu et al., 2014). Participants may detect fewer uncued targets in conditions with ATD than they do without it since the visually salient cues will stand out relative to the uncued targets and therefore be easier to detect (Maltz & Shinar, 2003).

5.1.2. Automation false alarms: aFAs, are expected to act as distractors, but may not result in actual detection false alarms. Based on previous experimental results (Tombu et al., 2014) and military combat identification procedures (Department of National Defence, 2009), the dFA rate may be near zero since participants are not expected to engage a target unless they can positively identify it as hostile. Because of this positive identification requirement, aMs should have a greater effect on performance than aFAs (Tombu et al., 2014).

5.2. Identification Sensitivity: Participants should show poorer identification and reduced sensitivity during night conditions, since target features will be harder to identify (Kogler, 2003). Conditions with cueing bias may produce improved identification over unbiased ATD if participants use the cueing biases appropriately in their decision making.

5.3. Decision Criterion: If participants take cueing bias into account, they should engage a greater or smaller number of targets depending on the cueing bias, and whether or not the target is cued. Thus, an interaction between ATD bias and target cueing (cued, uncued) is expected. ATD bias and cueing may also interact with illumination. For the night condition participants may exhibit a larger criterion shift between cued and uncued targets if they rely more on the ATD bias when targets may be more difficult to identify (Kogler, 2003; Maltz & Shinar, 2003). Participants may also become more conservative at night
since their ability to make a positive identification should be reduced, resulting in fewer target engagements (Department of National Defence, 2009).

5.4. Criterion Shift: Participants may disuse the cueing bias information or underestimate the identification value of the cues and not achieve the optimal shift in criterion (Wang et al., 2009) between cued and uncued targets. A sluggish beta between cueing conditions may result. Participants may also rely on the cueing bias for one type of target or cue and not another. For example, participants may have a criterion shift for cued targets, but not consider cueing bias for uncued targets.

5.5. Detection vs. Identification: Due to the limited amount of time per trial, participants may experience a detection-identification trade-off, sacrificing speed (fewer detections) for accuracy (more correct identifications), or vice versa. If so, we may expect that \( P(dH) \) will be negatively correlated with identification \( d' \). This will vary between individuals and depend on their approach to the task and how they value detection and identification. As part of this trade-off participants may also change their criterion, though the nature of the shift is uncertain. Some participants may hold fire due to the uncertainty of a target’s identity, or rely more on less reliable cues, which could result in more iMs or iFAs, respectively.

6. Method

6.1. Participants

Thirty-two Regular and Reserve Force members of the Canadian Armed Forces (CAF) participated in the experiment. Half of the participants were recruited through a poster sent to Reserve units and posted at Department of National Defence facilities (Appendix A). The rest of
the participants were recruited through an annual experimentation campaign that took place at the DRDC Toronto Research Centre. All participants received prior training in both weapons handling and positive identification procedures as part of their duty. A questionnaire (Appendix D) solicited data about demographic factors. Participants were all male with an average age of 30.9 ($SD=10.1$) and an average of 9.8 years of service ($SD=7.9$).

Participants received a half-day of pay at their regular salary, as well as stress remuneration according to DRDC guidelines (Duncan et al., 2008). Stress remuneration is a sum of money given to participants as compensation for participation in studies that are not part of their regular duties. This amount was calculated as $25.44 for the current experiment based on the requirements for minor physical activity and cognitive decision making.

### 6.2. Apparatus

This experiment was conducted using the Virtual Immersive Soldier Simulator (VISS), which is a simulated combat environment developed at DRDC for scientific study. The VISS consists of three main elements: a projection system, a simulated weapon system, and a camera tracking system. The Virtual BattleSpace 2 (VBS2) simulation engine (Bohemia Interactive Solutions, 2007) was used to render a virtual world that simulates the task of searching for targets within a particular section of terrain. VBS2 contains a variety of environments, landscapes, and character models upon which a scenario can be based. The Motive software package (OptiTrack, Motive [Software], 2013) tracked weapon position and orientation. Additional custom software allowed communication among the three main elements.

The virtual environment was displayed across three adjacent projector screens, each measuring 3 meters diagonally (Figure 3). The outer screens were angled inward to approximate
a 130-degree visual arc from the participant’s position in front of the centre screen at a distance of 4 meters. Three rear projectors each displayed the virtual world at a resolution of 1920 x 1080 pixels and aspect ratio of 16:9, resulting in a total resolution of 5760 x 1080.

![Figure 3: Virtual Immersive Soldier Simulator](image)

The participant controlled an M4A1 airsoft rifle instrumented with two lightweight tracking bars mounted on the top rail that allowed the camera system to track the weapon’s position and orientation in real time (Figure 4). A twelve-camera OptiTrack infrared optical tracking system (OptiTrack, Prime 13 [Hardware], 2013) was used to track the weapon. Each camera had a resolution of 832 x 832 and a 250 frames-per-second sample rate. The cameras were oriented to capture a virtual box (approx. 2 m³) centred on the shooting location.
The weapon was equipped with a micro-display mounted on the top rail to simulate an optical scope. The micro-display presented a magnified (3.4x) version of the virtual environment that continually updated based on the weapon’s current position and orientation to mimic what would be seen through an optical scope. The various cues and overlays used in the experiment are displayed only through the scope, including ATD which highlighted human targets with varying reliability. Cueing was represented by a yellow box around a target with a height to width ratio of 10:3, resulting in an approximate size of 2 x 0.6 metres at the target plane, scaling with distance to appear as the same size relative to target height. The colour yellow for the cue box was initially used by Tombu et al. (2014) and provides contrast against the environmental scene while not introducing colours associated with friendly (blue) or hostile (red) forces (Neyedli et al., 2011). The weapon was equipped with a custom trigger sensor. When fired, the simulation
depicted auditory feedback as simulated weapons fire sounds and visual feedback in the form of dirt and dust clouds as bullets hit the ground.

6.3. Design

The experiment was developed as a 2 x 2 x 4 design with the following factors: Cueing (Cued or Uncued), Illumination (Day or Night), and ATD (No ATD, No Bias, Hostile Bias, or Militia Bias). All independent variables were carried out as repeated measures, with each participant performing trials in all conditions. Levels of ATD and illumination were manipulated across experimental blocks, whereas cueing varied from target to target within each trial.

For the identification task, participants were shown two target identities that were labeled hostiles, or sometimes referred to as threats, and a group of non-threats labeled militia. Militia targets were used instead of Canadian or allied forces or civilians due to the reluctance of CAF to engage a target when there was some possibility that the target was friendly or civilian (Tombu et al., 2014). In pilot testing, it was found that cues such as uniform or clothing were too discriminable for participants and led to identification performance at ceiling. Thus threats and militia had the same uniform and the target’s weapon was used to show identity. Hostile targets carried Soviet style weapons with light coloured wood on the stock and barrel (AK47, RPK, SVD) while the militia carried NATO style weapons that were black (M16, FAL, M60). Two other visual features varied between targets, a secondary weapon was carried over the shoulder (M72, RPG7, or none) for some targets and some wore body armor (armor or uniform only). None of these cues were diagnostic and they appeared for targets from both affiliations. Although the militia targets were armed, participants were told that they were not considered a threat and should not be engaged. Participants were informed that the militia was operating in the area but were not considered friendly or allied.
In the simulation, each affiliation was represented by five possible character models, making ten visually distinct targets in total. Each character model varied slightly by carrying either a different weapon or kit and was presented only once per scene.

For 62 scenes, a predetermined viewpoint (i.e. a position from the first person perspective overlooking a unique section of terrain) within the simulated environment was selected based on its ability to support target placement (e.g. foliage that was not dense, sufficient distance to horizon – at least 100 meters, some visual variety from other viewpoints). Each viewpoint was seen four times (each time with different target placement making $62 \times 4 = 248$ unique scenes in total) but was not seen more than once in a given block, appearing twice under each illumination condition and once under each ATD condition Scene order was randomized within each block. Across the eight blocks, no participant saw the same target locations within a scene twice, even though they saw each viewpoint four times.
Each scene contained either four or five targets. The number of targets per scene was varied to introduce uncertainty so that participants would continue to search for targets for the trial duration. There were 31 trials per block, 16 trials with 5 targets and 15 trials with 4 targets. These numbers were selected to obtain an even number of total targets (140) that could be divided for each cueing condition (Table 4). 140 targets were presented in each block (16 x 5 = 80, 15 x 4 = 60, 80 + 60 = 140).

Table 4: Number of Targets per Condition

<table>
<thead>
<tr>
<th>Illumination</th>
<th>No ATD</th>
<th>No Bias</th>
<th>Hostile Bias</th>
<th>Militia Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cueing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day</td>
<td>Cued</td>
<td>-</td>
<td>49</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>49</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Uncued</td>
<td>70</td>
<td>70</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>Night</td>
<td>Cued</td>
<td>-</td>
<td>49</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>49</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Uncued</td>
<td>70</td>
<td>70</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>28</td>
<td>28</td>
</tr>
</tbody>
</table>

For each block, half the targets presented were militia and half were threats. When automation was available, 98 targets were cued (56 + 42 for cueing bias conditions or 49 + 49 for no bias) and 42 were uncued (28 + 14 for cueing bias conditions or 21 + 21 for no bias). Automation detection reliability, considered here as aH, was set at to 70%. This means that P(cue | target) = 0.70. The false alarm rate was set to 22.22%, defined as P(no target | cue) = 0.2222. There was an average of 0.9 aFAs per scene in conditions with ATD. The automation error rates used in this experiment are based on those used by Tombu et al. (2014) that were found not to be significantly different from performance with 100% reliable cueing. For the ATD conditions where the automation was more likely to detect hostile targets, P(cue | hostile) = 0.80 (56 cued hostiles / 70 total hostile targets) and P(cue | militia) = 0.60 (42 cued militia / 70 total militia targets), resulting in a consistent automation detection hit rate of 70%. If a target was cued, it
was more likely to be hostile and if uncued, more likely militia, \( P(\text{hostile} \mid \text{cue}) = 0.571 \) (56 cued hostiles / 98 cued targets) and \( P(\text{militia} \mid \text{no cue}) = 0.667 \) (28 uncued militia / 42 uncued targets). For the militia bias condition, the probabilities were switched between target affiliations, \( P(\text{militia} \mid \text{cue}) = 0.571 \) (56 cued militia / 98 cued targets) and \( P(\text{hostile} \mid \text{no cue}) = 0.667 \) (28 uncued hostiles / 42 uncued targets). The same aFA rate was used for all ATD conditions.

Targets were placed randomly on the virtual terrain within the field of view of the participant and at a simulated range of 50-150 meters for each scene. The distance and position of each target was then adjusted as necessary to an appropriate distance and cover position within 10 degrees of the visual angle. For about 10% of targets, the initial location was unsuitable and a suitable position could not be found after adjustment (this was due to factors like foliage or high target density). Another random position within the scene was chosen in these cases. Based on scene and viewpoint characteristics, target distance varied. The mean target distance across scenes was 93.73 meters (SD = 33.32). To control for target placement effects across conditions, different participants were shown each scene under each automation-illumination condition.

The order of ATD and illumination conditions was counterbalanced. Participants were shown both illumination blocks consecutively for a given ATD condition. An illumination level was not presented for more than two consecutive blocks. Each block required between 15 and 20 minutes to complete and the entire experiment took approximately 3 hours. Since the experiment took place in a dark room, participants were able to partially adapt to night conditions during the time between blocks (typically 2-3 minutes).
6.4. Task

Participants performed a visual search task for simulated human targets and provided verbal confirmation when they detected a target. Once detected, participants identified the target as militia or hostile. Participants were given “a wolf in sheep’s clothing” scenario, and were told that a terrorist group was planning an attack by mimicking militia uniforms. Participants were told that their mission objective was to detect all targets and engage any threats present in each scene. They were told that the positive identification cue was the target’s weapon. They were also informed that target behaviour was not always realistic: targets did not react to the sound of weapons fire or to a nearby target being downed. Targets did not engage in any behaviour that would be indicative of their identity either, such as pointing a weapon at or moving towards the participant. This aspect was based on the experimental design used by Tombu et al. (2014), where targets were also stationary and identification was based on the target’s weapon (though in that study identification was based on whether or not a weapon was present).

Figure 6: Participant Performing the Experiment

33
For each trial, participants were shown the scene from the fixed viewpoint, with the terrain displayed continuously across the three projector screens (Figure 6). Multiple targets were present at varying distances within the scene and were sometimes obscured. Each scene was presented for 25 seconds, after which participants were given feedback on their performance (Figure 7). The goal of the feedback was to allow participants to determine whether a particular target was a threat or not. It also allowed them to determine the number of detection misses they made by subtraction.

![Figure 7: Post Trial Feedback](image)

While conducting the experiment, participants sat in an office chair with adjustable armrests that could support their arms when holding the weapon. This chair was located behind a 0.5 metre high sandbag wall and could rotate 360°. Participants were asked to wear corrective eyewear if it was prescribed to them.
6.4.1. Procedure

Upon arrival, participants completed a consent form (Appendix C) and demographics questionnaire (Appendix D). They were then briefed on the experiment and the simulated combat scenario. They were informed about rules of engagement and what was required to make a positive identification.

A practice block allowed participants to learn to identify each target. The practice block consisted of 16 trials, 8 day and 8 night illumination scenes. For the practice trials, each scene contained 8 targets in obvious locations. For each illumination level, participants saw two scenes without cueing, two with fully reliable cueing, two with aMs, and two with aMs and aFAs. Practice trials had no time limit and participants were informed that the experimental trials would be more difficult due to target position and trial time limit, though there would be fewer targets. The specific number of targets per experimental scene was not given.

During the experiment, participants were asked to indicate when they had detected a target by putting the scope on target and saying the word “contact” (a military term for detection of a target), or verbally identifying the target. After the verbal confirmation, the experimenter pressed a button to register the target in the scope as “seen”.

After the practice trials, participants performed eight blocks of 31 trials. The blocks were defined by the combination of the four ATD and two illumination conditions. Participants were offered a short break before the next block began. Prior to each block, participants were informed of the automation’s reliability and cueing bias for that block. Participants were told that the automation would detect an average of 7 out of 10 targets throughout the block and that 1 out of 5 cues would be an empty box (an automation false alarm). For the hostile bias ATD
condition, participants were told that the automation would detect 8 out of 10 threats but only 6 out of 10 militia, which meant that if the ATD missed a target it was twice as likely to be a militia (i.e. since there were 28 uncued militia and 14 uncued hostiles). For the militia bias condition, this description was relayed as the reverse, that the automation would detect 8 out of 10 militia but only 6 out of 10 hostiles, which meant that if the ATD missed a target it was twice as likely to be a threat.

After the eight blocks, participants were asked to complete a post-experiment questionnaire (Appendix E) soliciting feedback about the ATD.

6.5. Measures

The measures collected for this experiment were the participant’s verbal indication of target detection, whether or not they engaged a target (identification), the time during a trial when a target was engaged, and weapon position and orientation throughout the trial.

For this experiment, the participants’ dFA rate was extremely low. To make a dFA, they would first need to detect something that is not a human, consider it as a target and identify it. Hit rate, \( dH \), was selected instead as the measure for detection. \( P(dH) \), defined as \( P(\text{"contact" } | \text{ target}) \), was calculated by dividing the number of verbally confirmed targets by the total number of cued and uncued targets in each block.

For target identification, the signal detection parameters \( d' \) and \( \ln \beta \) were calculated for each combination of ATD, illumination, and cueing. \( P(iH) \) was calculated by dividing the number of threats engaged by the overall number of threats detected. \( P(iFA) \) was calculated as the number of militia engaged divided by the total number of militia detected. The inverse of the standard normal cumulative distribution was then calculated for \( P(iH) \) and \( P(iFA) \), and \( z(iH) \) was
subtracted from z(iFA) (equation 1, p. 13) to obtain $d'$. The value of $\ln \beta$ was then calculated as equation 2 (p. 13).

Criterion was compared across cueing levels to determine the shift in $\ln \beta$ between cued and uncued targets. The observed criterion shift was compared to zero and the optimal difference to determine if participants changed their decision criterion between cued and uncued targets, and if that difference matched the shift between the optimal criterion values.

7. Results

Three repeated measures analyses of variance (ANOVAs) were performed on each dependent variable. A 2 x 2 x 3 analysis was used to investigate the effects of cueing, illumination, and ATD for automation available conditions (i.e., excluding the no ATD condition). In addition, two 2 x 4 repeated measures ANOVAs used illumination and ATD as factors. The first of these used data for uncued targets only, which allowed for comparison of all targets for the no ATD condition and the uncued targets in conditions when automation was available. The second 2 x 4 ANOVA collapsed across levels of cueing in the ATD conditions, calculating detection and identification measures regardless of whether a target was cued. Since there were no cued targets for the no ATD condition, the measures used for the uncued targets and all targets ANOVAs were the same for no ATD.

Data were checked for normality prior to each repeated measures ANOVA and did not appear to deviate significantly. Kolmogorov-Smirnov tests were also used for each combination of cueing, illumination, and ATD and no deviations at $\alpha = .05$ were found. For each repeated measures ANOVA, Mauchly’s test for sphericity was conducted. Any violations were corrected
with the Greenhouse-Geisser adjustment to the degrees of freedom. Only violations and their adjustments were reported; if not specifically mentioned then sphericity was not violated.

Since the repeated measures ANOVAs were testing effects based on specific hypotheses, no correction was made to control for Type I error. Main effects and interactions were considered significant at $\alpha = .05$. Tests for Pearson’s correlation coefficient were also considered significant at $\alpha = .05$.

### 7.1. Detection

The measure used for target detection was the probability of a detection hit, $P(dH)$. This value was calculated for each cueing, illumination, and ATD condition by dividing the number of targets detected (targets that were engaged or verbally confirmed as present) by the number of total targets in that condition. A larger $P(dH)$ value means a greater target detection rate, with a $P(dH) = 1$ meaning that all targets were detected for that condition.

The $P(dH)$ values for each condition were submitted to the three-way repeated measures ANOVA investigating cueing, illumination, and ATD. There was a main effect for cueing, $F(1, 31) = 216.63, p < .0001$; participants had a higher $P(dH)$ for cued targets ($M = .82$) than uncued targets ($M = .56$). A main effect for illumination was present, $F(1, 31) = 469.04, p < .0001$; participants detected more targets during the day ($M = .79$) than at night ($M = .59$). There were two interactions: cueing x illumination, $F(2, 62) = 305.33, p < .0001$, and cueing x ATD, $F(2,62) = 5.78, p = 0.0049$. The cueing x illumination interaction showed that cueing increased $P(dH)$ (i.e., improved detection) more in the night condition than during the day (Figure 8). The cueing x ATD interaction (Figure 9) showed that the detection advantage for cued targets was greater for the hostile bias than for the other conditions.
Figure 8: Probability of a detection hit, \( P(dH) \), as a function of cueing and illumination. Error bars indicate the 95% confidence interval.

Figure 9: Probability of a detection hit, \( P(dH) \), as a function of cueing and ATD. Error bars indicate the 95% confidence interval.
The $P(dH)$ values for uncued targets in each illumination and ATD condition were submitted to a 2 x 4 repeated measures ANOVA. Participants detected more uncued targets during the day ($M = 0.75$) than at night ($M = 0.42$), $F(1, 31) = 588.48, p < .0001$. A main effect of ATD showed that $P(dH)$ was higher without ATD ($M = .65$) than any ATD condition (no bias, $M = .56$; hostile bias, $M = .54$; militia bias, $M = .58$), $F(3, 93) = 19.861, p < .0001$. An illumination x ATD interaction (Figure 10) showed that the detection decrement for uncued targets in conditions with ATD was greater at night, $F(3, 93) = 4.721, p < .005$, though planned comparisons showed that $P(dH)$ was greater for no ATD ($M = .78$) than conditions with ATD ($M = .73$) during the day as well, $t(31) = 4.758, p < .0001$.

![Figure 10: Probability of a detection hit, $P(dH)$, for uncued targets as a function of illumination and ATD. Error bars indicate the 95% confidence interval.](image)

The $P(dH)$ values for each condition were submitted to a second 2 x 4 (illumination x ATD) repeated measures ANOVA that collapsed across cueing levels. Illumination had an effect on $P(dH)$, $F(1, 31) = 344.49, p < .0001$. Day illumination resulted in an larger $P(dH)$ ($M = .80$)
relative to night conditions ($M = .63$). There was a main effect for ATD as well, $F(3, 93) = 43.240, p < .0001$, with poorer detection for the no ATD condition ($M = .65$) than any condition with the automation (no bias, $M = .75$; hostile bias, $M = .74$; militia bias, $M = .74$). An illumination x ATD interaction was found, $F(3, 93) = 35.781, p < .0001$ (Figure 11). Participants detected fewer targets at night for the no ATD condition than when ATD was available, an effect that was not as large during the day, though still present. Planned comparisons showed that, during the day, the no ATD condition ($M = .78$) resulted in poorer detection than conditions with ATD ($M = .81$), $t(31) = 3.909, p < .0005$.

![Figure 11: Probability of a detection hit, $P(dH)$, as a function of illumination and ATD, collapsed across cueing. Error bars indicate the 95% confidence interval.](image)

### 7.1.1. Cueing Reliance

Pearson’s correlation coefficient was used to test for a relationship between detection of cued and uncued targets for each illumination-ATD condition. The purpose for these correlations was to determine whether the detection of cued targets was related to detection for uncued targets. A
positive correlation would imply that there were some participants who were better at the
detection task regardless of cueing, while a negative correlation would imply that as detection of
cued targets improved, fewer uncued targets were detected and a trade-off between looking for
cues and targets may have occurred. The only significant correlation was for militia bias during
the day, $r(30) = 0.675, p < .001$. A negative relationship between detection of cued and uncued
targets at night was suggested in the results, but the effect is not large enough to be significant.

7.2. Identification

Identification hit rates, $P(iH)$, were calculated by dividing the number of hits (engaging an
eemy target) by the number of detected eemy targets. Identification false alarm rates, $P(iFA)$
were computed by dividing the number of false alarms (engaging a friendly target) by the
number of detected friendly targets. For some conditions, a participant may have had a perfect
hit rate or may not have made an identification false alarm. In these cases, it is impossible to
calculate $d'$ or $\ln \beta$ since the z transformation of the probabilities of 0 or 1 would result in values
of $-\infty$ and $\infty$ respectively. In order to calculate the SDT parameters, if no identification false
alarms occurred, the $P(iFA)$ was calculated as $1/(2N)$ (Macmillan & Kaplan, 1985, Hautus, 1995)
where $N$ is the number of detections. If no identification misses occurred resulting in a perfect hit
rate, $P(iH)$ was calculated as $1 – 1/(2N)$. There were no cases of a participant not getting an iH. If
participants had a detection rate that was too low for a certain condition (at least three hostiles
and three militia had to be detected so that hit and FA rates could be calculated such that
participants could show performance beyond minimum chance levels, i.e. one hit and one miss
or one FA and one CR), there were too few trials to calculate the signal detection parameters and
$d'$ and $\ln \beta$ were set equal to zero (only three incidents of insufficient detection in a given
condition occurred across the 32 participants). $P(iH)$ and $P(iFA)$ were z-transformed according to
the inverse of the normal distribution, resulting in $z(iH)$ and $z(iFA)$. $d'$ and $\ln \beta$ were calculated according to equations 1 and 2 (p. 13) respectively.

Three ANOVAs were conducted on the signal detection parameters $d'$ and $\ln \beta$ that were computed for each cueing, illumination, and ATD condition (e.g., for cued targets with hostile biased ATD at night). For the ANOVA that collapsed cueing conditions, the cued and uncued hits and FAs were combined before calculating $d'$ and $\ln \beta$. For the no ATD condition, parameters for the uncued targets and all targets ANOVAs were the same since there were no cued targets.

### 7.2.1. Sensitivity ($d'$)

For the cueing, illumination, and ATD 2 x 2 x 3 ANOVA, there were two main effects and no interactions. Participants showed greater identification sensitivity for cued ($M = 1.98$) than uncued targets ($M = 1.67$), $F(1, 31) = 27.489, p < .0001$. Sensitivity was greater under day ($M = 2.22$) than night ($M = 1.43$) illumination, $F(1, 31) = 36.691, p < .0001$.

For uncued targets, the illumination and ATD 2 x 4 ANOVA showed only one significant main effect, for illumination $F(1, 31) = 33.672, p < .0001$, and no interactions. Participants had greater sensitivity under day ($M = 2.10$) than night illumination ($M = 1.36$).

The 2 x 4 illumination and ATD ANOVA that collapsed across cueing conditions showed a main effect for illumination, $F(1, 31) = 25.569, p < .0001$. This again showed that participants had a larger $d'$ under day ($M = 2.26$) than night illumination ($M = 1.59$). No other effect was significant.
7.2.2. Criterion (lnβ)

In the 2 x 2 x 3 ANOVA conducted on lnβ, the main effect of ATD violated sphericity based on Mauchly’s test, $\chi^2(2) = 9.090, p = .0106$. The Greenhouse-Geisser adjustment ($\epsilon = .793$) was used to correct the degrees of freedom for all effects involving this factor. There was no main effect for ATD, $p > .05$, but the interaction between cueing and ATD was significant, $F(1.848, 57.285) = 15.417, p < .0001$. As Figure 12 shows, in the condition where a cue was more likely to indicate a hostile, and cueing occurred, the participants’ criterion shifted in a liberal direction (i.e., more likely to engage a target) relative to the situation where the cue did not occur. In contrast, when a cue was more likely to detect militia targets, the participants’ bias shifted in the conservative direction relative to targets not highlighted by a cue. Finally, in a condition when the automation was unbiased (equally likely to cue hostile and militia targets) and the cue occurred, the participant was more conservative relative to a situation where the cue did not occur. Illumination had a main effect on lnβ, $F(1, 31) = 22.499, p < .0001$, with participants having a more conservative criterion at night ($M = .271$) than during day conditions ($M = -.208$).
**Figure 12:** Decision criterion, $\ln \beta$, as a function of cueing and ATD. Error bars indicate the 95% confidence interval.

The 2 x 4 ANOVA for uncued targets showed main effects for illumination and ATD on $\ln \beta$, though no interactions. The main effect for ATD, $F(3, 93) = 11.794, p < .0001$, showed that participants adopted a more conservative criterion for uncued targets when the ATD was biased towards detecting hostiles, a more liberal criterion when the ATD was biased towards detecting militia, and a neutral criterion for no bias and no ATD conditions (Figure 13). Planned comparisons showed that hostile bias was different from no bias and no ATD conditions, $t(31) = 4.117, p = .0003$, and $t(31) = 2.557, p = .0157$, respectively, and that militia bias differed from no bias and no ATD conditions, $t(31) = 2.042, p = .0497$, and $t(31) = 2.989, p = .0054$, respectively. There was a main effect for illumination, $F(1, 31) = 22.499, p < .0001$; participants adopted a conservative $\ln \beta$ at night ($M = .209$) and a more liberal criterion during day conditions ($M = -.237$).
Figure 13: Decision criterion, lnβ, for uncued targets as a function of ATD. Error bars indicate the 95% confidence interval.

The 2 x 4 ANOVA that collapsed across cued and uncued targets showed only a main effect for illumination, $F(1, 31) = 28.480, p < .0001$. Participants used a more conservative criterion under night illumination ($M = .345$) than during the day ($M = -.221$). No other effect was significant.

7.2.3. Criterion Shift

To further examine the criterion shift between cued and uncued targets, the difference in lnβ between cued and uncued targets was calculated for each participant for the three conditions with ATD (no bias, hostile bias, militia bias). A negative difference in lnβ between cued and uncued targets indicated that a participant held a more liberal criterion for a cued target than an uncued one, whereas a positive shift in lnβ referred to a more conservative criterion for cued than uncued targets.
These differences were submitted to a 2 x 3 (illumination and ATD) repeated measures ANOVA. There was only a main effect for ATD, $F(2, 62) = 15.417, p < .0001$. This effect (Figure 14) showed that participants had a different criterion shift when the automation had different detection biases. Planned comparisons were used to determine differences among the ATD conditions. No bias ATD differed from hostile bias, $t(31) = 5.061, p < .0001$, but was not different from the militia bias conditions, $p > .05$. Hostile and militia biases also resulted in significantly different criterion shifts, $t(31) = 4.666, p < .0001$.

![Figure 14: Shift in lnβ between cued and uncued targets, as a function of ATD. Error bars indicate the 95% confidence interval.](image)

If participants rely on the aid appropriately, the shift in lnβ will reach an optimal value that maximizes the number of correct responses for each cueing condition. The optimal shift in lnβ was calculated according to equation 4 (p. 17) for comparison to the observed criterion shift for each ATD condition. Even though participants had a different shift between certain ATD conditions, they were not able reach the optimal shift based on the 95% confidence interval. The
empirically calculated 95% CIs calculations (Jarmasz & Hollands, 2009) for no bias, hostile bias, and militia bias were [0.07579, 0.50792], [-0.57432, -0.14218], and [0.20032, 0.63244] respectively (Table 5). Based on these confidence intervals, no observed shift in \( \ln \beta \) achieved the optimal values of 0, -0.9808, and 0.9808 respectively. All shifts in \( \ln \beta \) were significantly different from zero, meaning that participants changed their decision criterion between cued and uncued targets for all conditions (Figure 14).

Table 5: Criterion Shift Calculation

|                  | P(Hostile|Cued) | P(Hostile|Uncued) | Optimal Shift in \( \ln \beta \) | Observed Mean | Observed 95% Confidence Interval |
|------------------|-----------|----------------|-------------------------------|--------------|---------------------------------|
| No Bias          | 0.5       | 0.5            | 0                             | 0.29185      | 0.07579, 0.50792                |
| Hostile Bias     | 0.5714    | 0.3333         | -0.9808                       | -0.35824     | -0.57430, -0.14218             |
| Militia Bias     | 0.4286    | 0.6667         | 0.9808                        | 0.41638      | 0.20032, 0.63244               |

7.3. Detection - Identification Relationship

To examine the relationship between detection and identification, \( P(dH) \) was tested for correlation with \( d' \) for each cueing, illumination and ATD condition. A significant correlation indicated that higher rates of detection were associated with greater identification sensitivity. Some correlations were significant between \( P(dH) \) and \( d' \) (Table 6). \( P(dH) \) and \( d' \) were correlated at \( \alpha = .05 \) for all militia bias and no ATD conditions but none of the hostile bias conditions.

The same analysis was conducted using \( P(dH) \) and \( \ln \beta \), but no correlation was significant in that case, \( p > .05 \).
Table 6: Correlations between P(dH) and d’

<table>
<thead>
<tr>
<th>Cueing</th>
<th>Illumination</th>
<th>ATD</th>
<th>r</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cued</td>
<td>Day</td>
<td>No Bias</td>
<td>.266</td>
<td>.144</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hostile Bias</td>
<td>.086</td>
<td>.641</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Militia Bias</td>
<td>.413</td>
<td>.019*</td>
</tr>
<tr>
<td></td>
<td>Night</td>
<td>No Bias</td>
<td>.370</td>
<td>.037*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hostile Bias</td>
<td>.305</td>
<td>.090</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Militia Bias</td>
<td>.378</td>
<td>.033*</td>
</tr>
<tr>
<td>Uncued</td>
<td>Day</td>
<td>No ATD</td>
<td>.413</td>
<td>.019*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No Bias</td>
<td>.568</td>
<td>.001**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hostile Bias</td>
<td>.336</td>
<td>.060</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Militia Bias</td>
<td>.435</td>
<td>.013*</td>
</tr>
<tr>
<td></td>
<td>Night</td>
<td>No ATD</td>
<td>.420</td>
<td>.017*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No Bias</td>
<td>.231</td>
<td>.203</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hostile Bias</td>
<td>.343</td>
<td>.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Militia Bias</td>
<td>.485</td>
<td>.005**</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

7.4. Subjective Responses

In order to determine how participants perceived the task and whether or not their strategies and goals when performing the task were related to their actual performance, dependent measures were tested for correlation with questionnaire responses (Appendix E).

7.4.1. Detection

Participants were asked to rate how helpful the ATD aid was to detection using a 7-point Likert scale (from not helpful [1] to extremely helpful [7]), for both day and night conditions. This resulted in two cueing utility ratings per participant. The ratings were tested for correlation with P(dH) for each ATD condition. No correlation was significant under day illumination, but under night illumination the cueing utility rating was correlated with the P(dH) for all conditions where ATD was available, regardless of ATD bias: no bias, \( r(30) = .564, p = .001 \); hostile bias, \( r(30) = .390, p = .027 \); and militia bias, \( r(30) = .440, p = .012 \). The no ATD condition was tested for correlation with cue utility rating since a positive correlation between cue utility rating and
detection performance without cueing could suggest that participants who rated the cue utility higher may have just performed better overall at detection and that cueing was not responsible for improvement. The correlation was not significant for the no ATD condition, \( r(30) = .107, p = .558 \).

7.4.2. Sensitivity (d’)

Participants were also asked to rate how helpful cueing was to identification using a 7-point Likert scale in both day and night conditions resulting in two values per participant. Cueing utility rating was tested for correlation with the calculated \( d’ \) for each cueing and ATD condition in both day and night illuminations. No correlation was found at \( \alpha = .05 \).

7.4.3. Criterion (ln\( \beta \))

Cueing utility rating was tested for correlation with \( \ln\beta \) for each illumination condition. No significant correlations were found.

Participants were asked to rate their approach to the task based on importance of getting (or not getting) each outcome at the end of a trial (value of threat hit, value of militia not hit, cost of threat missed, and cost of militia hit) on a 10 point scale. These ratings were used to calculate an approximate reported \( \ln\beta \) for each participant based on equation 3 (p. 16). In order to determine whether the reported \( \ln\beta \) was similar to the empirically calculated value, a test for correlation between the participant’s average empirical \( \ln\beta \) across all conditions and their reported \( \ln\beta \) was conducted. The test was significant, \( r(30) = .475, p = .006 \). Participants’ ratings of costs and values of each outcome were related to empirical adjustment of response bias.

Participants were asked to list a weighting for each target feature when making an identification decision (weapon colour, weapon style, cueing likelihood, etc.) where all weights
summed to 100 across all features. If they did not consider a feature, it was given a weighting of 0. Their weightings for cueing likelihood were checked for correlation with their criterion under ATD conditions with hostile and militia bias for cued and uncued targets at each illumination. For the day condition, cueing likelihood weight was correlated only with $\ln \beta$ for cued targets in the hostile bias condition, $r(30) = -.447, p = .01$. This negative correlation suggests that as the subjective weighting for cueing likelihood increased, the participant was more likely to have a liberal criterion.

In the night condition, a negative correlation was also found between cueing likelihood weighting and $\ln \beta$ for cued targets for hostile bias, $r(30) = -.549, p = .001$. Furthermore a negative correlation was found for uncued targets under militia bias, $r(30) = -.436, p = .013$. Participants who relied on the cueing adjusted their criterion to be more liberal when they were informed that hostiles were more likely present. There were no correlations significant at $\alpha = .05$ for conditions where militia targets were more likely to be present.

### 7.4.4. Target Feature Weighting

A 2 x 3 repeated measures ANOVA was conducted to examine participant weighting of target features (weapon colour, weapon style, and cueing likelihood) and their interaction with illumination. Feature, $F(2, 62) = 18.825, p < .0001$, and the feature x illumination interaction, $F(2, 62) = 12.892, p < .0001$, were both significant. During day illumination, participants relied more on weapon colour ($M = 49.69$) than weapon style ($M = 3.31$), $t(31) = 3.052, p = .0046$, a difference that was not significant at night (colour: $M = 34.78$, style: $M = 34.28$). Both weapon colour and style weightings were greater than cueing likelihood, $t(31) = 6.611, p < .0001$. Cueing likelihood weighting at night ($M = 15.16$) was rated higher than during the day ($M = 7.97$), $t(31)$
Thus, participants were less likely to rely on weapon colour at night, and did not rely more on weapon style, deferring to other cues such as cueing likelihood.

7.4.5. **Cueing Utility Rating**

Paired t-tests were used to evaluate the relationship between cueing utility ratings between day and night. Participants found the ATD more helpful for detection at night ($M = 5.81, SD = 1.18$) than during the day ($M = 4.91, SD = 1.06$), $t(31) = 3.604, p = .0011$. Illumination did not affect cue utility rating for identification, $p > .05$.

7.4.6. **Target Engagement Certainty**

In the post-experiment questionnaire, participants were asked to rate how certain they needed to be that a target was hostile in order to engage it. They used a continuous scale from 0 (no evidence either way) to 100 (no doubt at all). Participants were asked to mark this certainty rating from when they started the experiment, when they finished, and during a real combat situation. Paired t-tests were used to compare the certainty ratings. The ratings at the beginning ($M = 68.59, SD = 21.71$) and the end of the experiment ($M = 74.69, SD = 19.43$) were not significantly different, $p > .05$. However, participants said that they needed to be more certain about a target’s identity in a real combat situation ($M = 83.69, SD = 18.64$) than at the beginning, $t(31) = 3.921, p = .0005$, or end, $t(31) = 2.830, p = .0081$, of the experiment.

8. **Discussion**

Automatic target detection aided both detection and identification performance on the target search and combat identification task.
8.1. Detection

Participants had a higher detection rate for ATD conditions where cueing was available, even though it was imperfect. This result is consistent with previous studies that have investigated the benefit of detection aids and unreliable automation (Maltz & Shinar, 2003; St. John & Maines, 2002; Tombu et al., 2014). With the automation hit rate, P(aH), set at 70% and false alarm rate, P(aFA), at 22.22%, participants still used the automation to improve target detection, P(dH). The improvement was greatest for night illumination. Targets cued at night were detected at a rate that almost matched the detection rate for cued targets during the day.

However, ATD conditions with cueing saw a decreased P(dH) for uncued targets compared to the no ATD condition, a difference that was larger at night. This discrepancy was due to an increased reliance on cueing at night. Illumination had a significant impact on both detection and identification measures, and participants subjectively found the night condition more difficult than day trials. This suggests that there was a change in the perceived utility of the automation based on illumination (Dzindolet et al., 2001a). During day conditions, participants were able to perform the task sufficiently well such that the discrepancy between the automation’s ability and the human’s was low. Human detection for the no ATD condition was actually better during the day than the automation’s detection hit rate. But when the task became more difficult, the ATD was correctly perceived to perform better than the human and participants depended upon it for detection. Increased reliance on automation under more challenging visual conditions is consistent with other studies (e.g., Kogler, 2003; Maltz & Shinar, 2003).

When P(dH) for cued and uncued targets was tested for correlation for each block, hit rate between cued and uncued targets was found to be uncorrelated for most conditions. Participants detected cued and uncued targets differently; better detection for cued targets did not transfer to
uncued targets. Detection performance under night conditions was also correlated with a participant rating of how helpful the ATD was at night, but uncorrelated with the no ATD condition. This shows that participants who performed better at target detection using ATD in night conditions found the cueing more useful subjectively and were not necessarily better at detecting targets in general. However, the cueing utility rating was uncorrelated with detection performance using ATD during the day. It appears that cue reliance benefited those who relied on ATD at night but not during the day.

Participants performed better in the no ATD condition than conditions with ATD for uncued targets, but worse in the no ATD condition when collapsing across cueing conditions. ATD did improve detection performance, even if participants relied on it heavily at night. Depending on the automation’s reliability, almost complete reliance on ATD for detection might be warranted under reduced illumination.

Participants detected more cued targets and fewer uncued targets when the ATD had hostile bias, as compared to the no bias and militia bias conditions. If participants were searching primarily for hostile targets, they may have relied more on the cueing for detection if they expected the ATD to detect more threats.

Since participant detection false alarms were extremely rare (there were only three instances of a participant attempting to engage something that was not a target), a signal detection analysis could not be performed on target detection. Due to the lack of observed consequences of automation FAs (almost zero resulting dFAs), it appears that aFAs did not have a direct effect on performance. Additionally, based on participant feedback, automation misses seemed to have a more significant impact on performance than aFAs. In terms of the compliance and reliance
distinction discussed by Dixon et al. (2007), participants seemed to rely on the automation (i.e., agree when it did not indicate a target), more than they complied with it (i.e., agreed with the automation when it suggested a target was present). The lack of dFAs suggests that participants did not necessarily expect a target to be present when a cue was shown, whereas the high miss rate for uncued targets at night suggests that participants were more likely to concur that a target was not present if they did not see a cue. These results are consistent with the findings of Tombu et al. (2014) but stand in contrast with those of Dixon et al. (2007), who concluded that aFAs had a more negative impact on performance than aMs. One reason for this difference may be due to the nature of the false alarms. Dixon et al. considered a tracking and system monitoring task using a simple interface, while the current study had participants performing a visual search task using a high fidelity simulator. Since the aFAs were integrated with the current study’s detection task, they were likely easier to reject than aFAs in Dixon et al., which used different modalities (auditory and visual) for the alerts and monitoring tasks, respectively. This suggests that automation errors can affect compliance and reliance differently based on the design of the automation and the context of its use.

8.2. Identification

8.2.1. Sensitivity

While designed as a detection aid, ATD also affected target identification with respect to both sensitivity and criterion. Identification sensitivity was greater for cued than uncued targets. By aiding participants in detecting targets, ATD cues also improved the ability of participants to discriminate between target identities. Tombu et al. (2014) obtained similar findings, with improved identification accuracy for completely reliable cueing compared to no cueing conditions. In contrast, Glaholt (2014) found an identification penalty to cueing, where cued
targets were incorrectly identified more frequently than uncued targets, a difference that was greater for moving targets. Glaholt also found that when cue salience was reduced, the identification penalty decreased as well. This suggests that the type of cueing used does impact identification, meaning that the results from these studies are not necessarily contradictory. Indeed, the experimental design may be another reason for the discrepancies between these studies. Like the current study, the experiment in Tombu et al. was conducted using the VISS, which required participants to search for targets more actively than Glaholt, who used a computer monitor. Participants likely regarded the actions taken to identify a target differently between the studies. Using the VISS, participants were required to pull the trigger to engage a target belonging to one of two affiliations whereas Glaholt required a button press for each of the four target types that appeared at further distances, half of whom were moving. If a cue is distracting, it may have a negative effect on identification that may scale with distance and movement, two effects that were not directly assessed in the current study or by Tombu et al. Depending on the context of the automated cueing system’s use, detection benefit and identification costs may need to be traded off.

Illumination had a significant effect on identification sensitivity. Participants were better able to discriminate between hostile and militia targets during the day than during night conditions. As with detection, participants subjectively referred to the night condition as more difficult for identification than the day condition. The night illumination reduced the contrast and brightness of the scene, making the visual cues more difficult to distinguish. Participants relied less on the weapon colour cue, for example, and likely used other less reliable cues to compensate. This was demonstrated by an increase in target feature weighting for cueing likelihood at night.
Neither the presence nor type of ATD had an effect on identification sensitivity. Sensitivity was no different with or without ATD, and did not vary with ATD bias. Since participants can adjust their criterion independently of sensitivity, target likelihood did not affect the overall sensitivity of cued or uncued targets. ATD differs from the automation used by Wang et al. (2009) which cued only one type of target. They found that the participant error rate was very low for that cue since it was never incorrect.

### 8.2.2. Criterion

The decision criterion was affected by illumination and the interaction between cueing and ATD. Participants were less likely to engage targets (i.e., they were more conservative) at night, likely due to the reduced ability to discriminate between targets. Since participants were often unsure about a target’s identity, they were usually less willing to engage targets that might not be hostile. The trained requirement for positive identification may have resulted in participants being more reluctant to engage targets when their ability to identify them was hindered.

The interaction between cueing and ATD condition confirms that the detection bias of the ATD did have an impact on participant decision making. Participants generally adopted a more liberal criterion (i.e., were more likely to engage a target) for conditions where hostiles were more likely present, and a more conservative criterion when militia were more likely. The use of target likelihood information to adjust the decision criterion is consistent with the findings of Wang et al. (2009) and Neyedli et al. (2011). Similar to these studies, participants showed a criterion shift between automation conditions.

Even though participants adopted a different criterion for cued and uncued targets under different ATD conditions, similar to the results of Wang et al. (2009) and Neyedli et al. (2011)
they also showed a sluggish beta effect. Participants were required to switch between their criterion for cued and uncued targets while moving though the scene and may have felt that the pressure was too substantial to consistently take cueing likelihood into account. Though participants showed some evidence of disuse, they didn’t disregard the cues completely.

Participants had a different shift in $\ln\beta$ for hostile-biased ATD than no bias and militia bias conditions, which did not differ. Criterion for conditions with more hostile targets (i.e., cued targets for hostile bias and uncued targets for militia bias) was also negatively correlated with the target feature weighting of cueing likelihood; $\ln\beta$ became more liberal as participants placed greater weight on cueing for identification. No correlation was significant for conditions with increased militia likelihood. Participants adjusted their decision criterion liberally under conditions where they expected a heavier hostile presence, but did not appear to make a conservative adjustment for increased militia presence. This may be due to the perceived objective of the task, since participants were told to “engage all threats without hitting any militia.” Participants may have held different attitudes towards the automation based on its cueing bias and adjusted their reliance differently (Wang et al., 2009).

The prediction that participants may have a larger criterion shift at night was not directly supported. The level of illumination did not produce a greater difference in $\ln\beta$ for cueing bias conditions. However, in the post-experiment questionnaire participants weighted the target feature of cueing likelihood higher at night than during the day. The increased weighting may have been to compensate for visual cues that were more difficult to discriminate at night, which lead to a lower identification sensitivity.
8.3. Detection and Identification

The relationship between detection and identification performance was investigated through an examination of correlations among the dependent measures \( P(dH), d', \) and \( \ln \beta \) for each cueing, ATD, and illumination combination. While detection and decision criterion were uncorrelated, correlations between detection and identification sensitivity were observed. All correlations were positive, which did not support a trade-off between detection and identification. Illumination and cueing did not have a consistent effect on the relationship between \( P(dH) \) and \( d' \), but the ATD factor did. Detection rate and sensitivity were correlated for all militia bias and no ATD conditions, half of the no bias, and none of the hostile bias conditions. These correlations further suggest that participants treated hostile bias automation differently from the other ATD conditions.

In accordance with its purpose, ATD was used primarily as a detection aid. Participants used the cueing to help detect more targets. Participants often used the weapon scope to scan the scene for targets, especially under night conditions. Some participants were observed to miss uncued targets that passed through the scope, especially at night, or if the uncued targets were partially hidden. These incidents suggest that some participants often over-relied on the ATD for detection under night conditions at the expense of manual channels. This conclusion is supported by the result that uncued targets were more likely to be missed in conditions with ATD than without the aid.

Participants reacted to the ATD’s impact on target identification by changing their decision criterion based on cueing likelihood, but the sluggish criterion adjustment shows that they underused cueing biases for identification. Reliance on ATD as an identification aid therefore differed from how participants used ATD for detection. This may have also been due to the
nature of each task. When detecting targets, the consequences of an error are not as severe and may have been seen as more of an inconvenience whereas identification errors are considered much more serious. In their framework for automation use, Dzindolet et al. (2001a) suggest that automation use may be affected by processes such as feelings of control and moral obligation for self-reliance. The current results showed a tendency towards self-reliance for identification (i.e. the ability to discriminate based on manual channels) and because of the consequences of incorrect combat identification, participants may have felt obligated to rely on themselves instead of the ATD (Dzindolet et al., 2001a).

Additional analyses were conducted on between-subjects factors such as block order and demographics. The majority of these factors were not significant and the few significant effects did not constrain or contradict any of the main findings and were of little interest.

9. Limitations

For the participants who had been in combat, a lack of more detailed intelligence about the scenario (i.e. what kind of threats were present, the militia’s standing relative to the CAF, troop movement and locations, friendlies in the area, etc.) made the simulation less realistic or believable. Typically soldiers would be provided more information in an actual military operation. Threats mixed with militia targets while wearing similar uniforms was also seen as unrealistic by most participants. In addition to the deviation from a typical combat scenario, a high level of difficulty may have led to participants adopt a more liberal criterion than would be expected in combat. Without adopting a more liberal criterion, most participants would have rarely fired the weapon during the night conditions.
Reluctance to engage targets was an initial concern, but ended up not being problematic. The simulation (i.e. targets did not move or react to weapons fire) and scenario (i.e. having hostile and friendly targets mixed together, moving to a new scenario after 25 seconds) were far enough from a true combat situation that participants did not perform the experiment with a decision criterion set as if it were an actual battlefield scenario. Participants rated their identification requirement as more conservative in a real life combat scenario than during the experiment. That being said, this experiment was conducted using an apparatus and participants that were more reflective of a military combat scenario than previous published work in human factors literature and, although it is difficult to predict whether a criterion shift based on cueing likelihood information would be as large during field use, validation in a combat scenario is not feasible.

The identification task was designed to be difficult, so if the identification task were easier then participants may have relied less on the cueing bias. The relationship between using cues other than a target’s weapon for identification and cueing bias is unknown since other identification cues were not tested. However, the adjustment in difficulty due to illumination did not result in any significant interactions with cueing bias, which suggests that task difficulty may not change how cueing bias affects identification. If other cues, such as movement or hostile actions were used as well, then the effects of cueing bias may have been different since soldiers are trained to react differently to movement than other visual cues, though it is difficult to predict how the results would be affected.

Based on the findings of Wang et al. (2009), all participants in this experiment were informed of the automation’s cueing bias. We therefore did not look at the effects of cueing bias on participants who were unaware of its presence. Since the amount of time that participants were available for the experiment was relatively short and the differences in cueing bias were
relatively small, we decided that informing participants of the automation’s reliability and bias would provide them with sufficient information about the aid that they could then decide whether or not to integrate that information into their identification process. This can also be considered a limitation in the sense that some participants may have understood and considered the information differently than others.

A limitation of the experimental design was the need for verbal confirmation from participants when they detected a target. Since no direct action would be taken for half of the targets otherwise, verbal confirmation from the participant was determined to be a relatively accurate method of counting detections that had little subjective experimenter impact. This issue was really only a limitation when participants neglected to provide a verbal confirmation, possibly due to fatigue, frustration (due to difficulty), or forgetfulness. After a trial where a participant did not vocalize target detection, he was reminded to confirm when he had seen a target. Though rare, this led to some instances where the experimenter subjectively judged when detection had occurred, using the following criterion: the target had to be in the participant’s weapon scope for at least one second after which the experimenter checked to ensure that the participant was looking through the scope for at least one second.

Initially, data regarding when a target was in scope for a predetermined amount of time was planned to be used as the measure for detection, but through pilot testing it was found that targets would sometimes pass the scope but were not actually detected by the participant. This typically happened in the night condition when the target was uncued and near foliage. There were also instances of a target being in scope when the participant was not looking through it, if they were looking over the weapon scope for example. Since scanning behaviour and speed differed
between participants, there was not an appropriate in scope duration that could be consistently used for all participants across all conditions.

When calculating the shift in $\ln(\beta)$, the costs and values of each outcome were set as equal, based on the calculations of Wang et al. (2009). In the military context though, certain outcomes would likely be considered to have different costs, where identification false alarms would be considered the worst outcome since they can directly result in friendly casualties. Though the calculation of the optimal shift in criterion cancelled out the costs and values of each outcome, the negative effects of an iFA may have had additional effects beyond what was considered.

One other limitation was the inability to test the assumptions of the signal detection measures due to the experimental design. Since the experiment was conducted as a binary response task, testing the equal variance assumption of the signal and noise distributions was not possible (Macmillan & Creelman, 2004; Stanislaw & Todorov, 1999). Violations of this assumption would primarily affect $d'$, where unequal variances can cause sensitivity to vary with response bias since the ratio of signal to noise likelihood has changed relative to the distance between the two distributions (Stanislaw & Todorov, 1999). Decision criterion, which is a more important measure to the analysis and conclusions in this experiment, is less affected.

10. Contributions

In this experiment, we have shown that cueing bias can impact the human’s identification performance. How a system detects targets and the errors that can result unintended consequences on human decision making and need to be considered in design. Though the aid improved performance in the intended way (i.e. better detection), it affected target identification. While the effect was positive in this instance, cueing bias in a different automated system could
have unintended negative consequences. In terms of automation errors, this study also supported the perspective that the type of automation error, i.e., misses and false alarms, will impact performance differently depending on the task being performed and the design of the automation being used.

This study also continued to show that criterion shift based on automation reliability, initially used by Wang et al. (2009), is a viable way to measure reliance on automation. Following from Wang et al. and Neyedli et al. (2011), the results suggest that informing participants of the automation’s reliability can encourage appropriate adjustments to decision making.

11. Future Work

The results show that automatic target detection aids improved soldier performance in simulated detection and identification tasks. Future work could investigate additional factors that may impact ATD use, such as increases or decreases in automation reliabilities and cueing biases beyond what was used in this study. Understanding how reliance is affected as ATD bias increases or decreases will be important for the development of a cueing system that supports soldiers as they detect and identify targets. If the detection biases occur in the actual technology, it is important to understand how soldiers may react to them as situations and biases change. Future studies should investigate methods of disclosing cueing reliability in the case of cueing bias. If cueing bias proves to be an issue with the ATD, it would be important to have a method of relaying this information to the soldiers using it, including how the bias might change under different circumstances.

Since other forms of automation are also being developed for soldiers, future studies should consider the interaction between the multiple forms of automation that a soldier would use in
combat, such as BFT and CID. When given additional technology, certain features may be disused or the aids could be combined to improve performance. To describe this performance, future studies will need to develop an advanced model of reliance that can account for multiple information channels of differing reliability at once. This model should be able to explain performance on multiple automation channels as well as from manual channels.

12. Conclusion

Automated systems have been shown to improve performance in both detection and identification tasks, and the results from the current study support that conclusion. ATD resulted in improved target detection, especially under conditions where the participants’ ability to perform the task was poor. As a detection aid, ATD proved to be successful. Even during day conditions where the participants performed the task better than the automation, detection hit rate was better when targets were cued. ATD had the greatest benefit under night conditions, where participants had the most difficulty performing the task. Participants were also able to reject the automation’s false alarms and rely on the ATD to improve detection performance.

ATD improved detection sensitivity for cued targets compared to uncued ones. Participants were also able to adjust their decision criterion due to changes in the automation detection reliability in appropriate directions. However, this adjustment did not reach the optimal performance, another demonstration of sluggish beta. In summary, ATD cueing bias had a significant impact on identification and participants were able to use that bias to adjust their criterion liberally and engage more targets when they expected more threats to be present.
References


OptiTrack. (2013). Motive [Computer Software]. Corvallis, OR: NaturalPoint, Inc.


Appendix A: Participant Recruitment Poster

Volunteers Needed

Effects of Bias in Automatic Target Cueing on Human Target Identification

**Purpose:** This study will evaluate the effectiveness, costs and benefits of automatic target cueing in the dismounted infantry context. Can cueing likely target locations improve shooting performance?

**Volunteers:** We need thirty-two military participants (male or female) between the ages of 18 and 60 years with normal or correct-to-normal vision.

**Procedures:** You will be asked to perform dismounted infantry engagements (similar to a shooting gallery) in the simulated military environment Virtual Battle Space 2 (a first-person shooter game-like environment). A modified airsoft weapon will be used to shoot at simulated targets on a screen who will react when shot. An automatic target cueing capability has been added to the scope to help detect targets in the environment.

**Location:** DRDC Toronto Research Centre, 1133 Sheppard Ave. W., Toronto, Ontario

**Duration:** The experiment will require one session, lasting approximately 3 hours.

**Risks:** This is a minimal risk study. There are no anticipated physical, social, psychological, emotional, economic, or other risks associated with the research proposed herein. You will be able to take breaks as needed to avoid any undue physical or cognitive strain.

**Benefits:** There are no direct benefits to the participant. However, this study will contribute to the design of an effective automatic target cueing capability for dismounted infantry soldiers that will provide improved battlefield situation awareness.

**Compensation:** Stress remuneration will be provided according to Government of Canada guidelines, in addition to regular salary.

**Points of Contact:**

Adam Reiner (adam.reiner@drdc-rddc.gc.ca, adam.reiner@utoronto.ca)

**Principal Investigator:** Dr. Justin Hollands (416-635-2073, justin.hollands@drdc-rddc.gc.ca)
Appendix B: Pre-Experiment Information Sheet

Protocol Number: 2014-033
Research Project Title: Effects of Bias in Automatic Target Cueing on Human Target Identification
Principal Investigator: Dr. Justin Hollands
Co-Investigators: Mr. Adam Reiner (University of Toronto, Department of Mechanical and Industrial Engineering), Dr. Greg Jamieson (University of Toronto, Department of Mechanical and Industrial Engineering), Dr. Mike Tombu, Mr. Matt Lamb, Mr. Ken Ueno
Project: Future Small Arms Research (FSAR) 02aa

Background
The experiment investigates the utility of automatic target cueing as a shooting aid. Automatic target cueing is a technology that analyzes sensor data looking for patterns with the same signature as human forms. Patterns matching this signature are then cued in the scope display with a yellow box around the potential target. We are interested in the impact of this technology on the shooting process, with a special emphasis on the effects of cueing reliability, bias and environmental conditions.

Survey Overview
The experiment lasts approximately three hours and will be carried out using Virtual Battle Space 2 (VBS2), a first-person shooter, simulated military environment, using the Virtual Immersive Soldier Simulator (VISS). The VISS is a first-person shooter test bed where the participant sits in front of a set of projection screens onto which VBS2 is displayed. The participant wears a helmet and ballistic eyewear (or glasses) and uses an instrumented mock weapon equipped with a scope to interact with the environment. The scope display can be augmented with automatic target cues of varying reliability. A set of optical tracking cameras is used to track the location of the weapon in space to insure that the image displayed in the scope is properly aligned with the weapon point of aim.

Before the experiment begins, the participant will be given training on how the automatic target cueing technology works as well as how to use the VISS. A short practice block of trials will then be performed to help the participant get the hang of the task to be performed. Given that the bias of the automatic target cueing towards detecting certain types of targets is being examined, the participant will be told how reliable the cues for each type of target will be for that block of trials before the participant begins said block. In addition to the practice block, four experimental blocks, with varying levels of automatic target cueing bias will be performed.

Each trial lasts 25 seconds. On each trial the participant will be presented with a scene containing a number of virtual targets (people). The participant’s task is to engage all hostile combatants in the scene, while not engaging non-threats. Sample hostile combatants and non-threats will be provided during the practice trials to familiarize participants with the categories. The participant will attempt to clear all of the threats prior to the end of the 25 second time limit.
We are interested in how long it takes participants to detect and identify targets and how accurate they are in doing so. Participants are requested to complete the task of detecting all targets and engaging all threats as quickly as possible without engaging non-threats or missing hostile combatants. Rest breaks can be taken between trials.

This study is for scientific purposes only and serves to gain a better understanding of automatic target cueing technologies. Therefore, there will be no career implications due to the findings of the study.

**Risks:** There are no anticipated physical, social, psychological, emotional, economic, or other risks associated with the proposed research. To reduce long term strain or fatigue during the experiment, there will be regular breaks after each 30 minute trial block.

**Benefits:** There are no direct benefits to participants. However, this study will contribute to the design of an effective automatic target cueing capability for dismounted infantry soldiers that will provide improved battlefield situation awareness. At a personal level, participants may benefit from the personal satisfaction of knowing that they have contributed to the generation of a potentially important future technology.

Following completion of the experiment participants will be fully debriefed regarding its specific aims and given the opportunity to ask additional questions.

**Compensation:** Participants will receive a stress remuneration of $25.44 in addition to their regular salary.

**Confidentiality**

Data will be kept in an anonymous fashion, names will not be recorded and participants will only be identified by an ID number. I understand that my experimental data will be protected under the Government Security Policy (GSP) at the appropriate designation and not revealed to anyone other than the Principal or Co-Investigator(s) without my consent except as data unidentified as to source.

I understand that my name will not be identified or attached in any manner to any publication arising from this study. Moreover, I understand that the experimental data may be reviewed by an internal or external audit committee with the understanding that any summary information resulting from such a review will not identify me personally.

I understand that, as a Government Institution, DRDC is committed to protecting my personal information. However, under the Access to Information Act, copies of research reports and research data (including the database pertaining to this project) held in Federal government files, may be disclosed. I understand that prior to releasing the requested information, the Directorate of Access to Information and Privacy (DAIP) screens the data in accordance with the Privacy Act in order to ensure that individual identities (including indirect identification due to the collection of unique identifiers such as rank, occupation, and deployment information of military personnel) are not disclosed.
Your Rights as a Participant
I understand that I am free to refuse to participate and may withdraw my consent without prejudice or hard feelings at any time. I also understand that the Investigator(s) or their designee responsible for the research project may terminate my participation at any time, regardless of my wishes.

Contact Information
Should I have any questions or concerns regarding this project before, during or after participation, I understand that I am encouraged to contact the appropriate DRDC research centre cited below. This contact can be made by surface mail at this address or by phone or email to any of the DRDC numbers and addresses of individuals listed below:

Defence R&D Canada – Toronto Research Centre
1133 Sheppard Avenue West
PO Box 2000
Toronto, Ontario, M3M 3B9

Principal Investigator or Principal DRDC Investigator:
Dr. Justin Hollands (justin.hollands@drdc-rddc.gc.ca)

Chair, DRDC Human Research Ethics Committee (HREC):
DRDC HREC Chair (416) 635-2098, HREC-CEESH-toronto@drdc-rddc.gc.ca

University of Toronto Office of Research Ethics:
416-946-3273, ethics.review@utoronto.ca
Appendix C: Consent Form

Protocol Number: 2014-033
Research Project Title: Effects of Bias in Automatic Target Cueing on Human Target Identification
Principal Investigator: Dr. Justin Hollands
Co-Investigators: Mr. Adam Reiner (University of Toronto, Department of Mechanical and Industrial Engineering), Dr. Greg Jamieson (University of Toronto, Department of Mechanical and Industrial Engineering), Dr. Mike Tombu, Mr. Matt Lamb, Mr. Ken Ueno
Project: Future Small Arms Research (FSAR) 02aa

I, ______________________ (name) hereby volunteer to be a participant in the study. I have read the information package, and have had the opportunity to ask questions of the Investigator(s). All of my questions concerning this study have been fully answered to my satisfaction. However, I may obtain additional information about the research project and have any questions about this study answered by contacting Justin Hollands (416-635-2073, justin.hollands@drdc-rddc.gc.ca).

This experiment spans one 3-hour session. Before the experiment begins, I will be given training on how the automatic target cueing technology used in this experiment works. I will then learn how to use VBS2, the first-person shooter simulated military environment in which the experiment takes place and learn how to use the Virtual Immersive Soldier Simulator (VISS). I will then take part in a series of shooting engagements within VBS2. I will be briefed before each block of engagements on the reliability of the automatic target cueing technology in use in that block.

I have been assured that participation in this experiment involves minimal risk. Also, I acknowledge that my participation in this study, or indeed any research, may involve risks that are currently unforeseen by Defence Research and Development Canada (DRDC).

I understand that this study has no direct benefits to participants, though its results will help contribute to the design of an automatic target cueing capability for future dismounted infantry soldiers to help provide improved battlefield situation awareness.

I understand that my experimental data will be protected under the Government Security Policy (GSP) at the appropriate designation and not revealed to anyone other than the DRDC-affiliated Investigator(s) or external investigators from the sponsoring agency without my consent except as data unidentified as to source.

I understand that my name will not be identified or attached in any manner to any publication arising from this study. Data will be kept in an anonymous fashion, names will not be recorded and participants will only be identified by an ID number. Moreover, I understand that the experimental data may be reviewed by an internal or external audit committee with the understanding that any summary information resulting from such a review will not identify me personally.
I understand that, as a Government Institution, DRDC is committed to protecting my personal information. However, under the Access to Information Act, copies of research reports and research data (including the database pertaining to this project) held in Federal government files, may be disclosed. I understand that prior to releasing the requested information, the Directorate of Access to Information and Privacy (DAIP) screens the data in accordance with the Privacy Act in order to ensure that individual identities (including indirect identification due to the collection of unique identifiers such as rank, occupation, and deployment information of military personnel) are not disclosed.

I understand that I am free to refuse to participate and may withdraw my consent without prejudice or hard feelings at any time. Should I withdraw my consent, my participation in this research will cease immediately. I also understand that the Investigator(s) or whom they designate responsible for the research project may terminate my participation at any time, regardless of my wishes.

I understand that as a CAF member participating in this research project during work hours, I am entitled to a remuneration in the form of a stress allowance for each completed session ($25.44 for a CAF member) if I complete the entire research project as set out in the protocol. I also understand that I am entitled to partial remuneration if I do not complete the entire session. Stress remuneration is income and is subject to income tax. As a CAF member, my Service Number (SN) is required for remuneration.

I understand that I am considered to be on duty for disciplinary, administrative and Pension Act purposes during my participation in this study and I understand that in the unlikely event that my participation in this study results in a medical condition rendering me unfit for service, I may be released from the CAF and my military benefits apply. This duty status has no effect on my right to withdraw from the study at any time I wish and I understand that no action will be taken against me for exercising this right.

I understand that by signing this consent form I have not waived any legal rights I may have as a result of any harm to me occasioned by my participation in this research project beyond the risks I have assumed. Also, I understand that I will be given a copy of this consent form so that I may contact any of the individuals mentioned below at some time in the future should that be required.

Volunteer’s Name:____________________________________________________

Signature:__________________________ Date:___________________________

Section Head/Commanding Officer (see Notes below)

Signature:__________________________ Date:___________________________

Commanding Officer’s Unit:_____________________________________________
Notes:

For Military personnel on permanent strength of CFEME:
Approval in principle by Commanding Officer; however, members must still obtain their Section Head’s signature designating approval to participate in this particular research project.

For other military personnel:
All other military personnel must obtain their Commanding Officer’s signature designating approval to participate in this research project.

FOR PARTICIPANT ENQUIRY IF REQUIRED:
Should I have any questions or concerns regarding this project before, during or after participation, I understand that I am encouraged to contact the appropriate DRDC research centre cited below. This contact can be made by surface mail at this address or by phone or email to any of the DRDC numbers and addresses of individuals listed below:

Defence R&D Canada – Toronto Research Centre
1133 Sheppard Avenue West
PO Box 2000
Toronto, Ontario, M3M 3B9

Principal Investigator or Principal DRDC Investigator:
Dr. Justin Hollands, 416-635-2073, justin.hollands@drdc-rddc.gc.ca

Chair, DRDC Human Research Ethics Committee (HREC):
DRDC HREC Chair, (416) 635-2098, HREC-CEESH-toronto@drdc-rddc.gc.ca

University of Toronto Office of Research Ethics:
416-946-3273, ethics.review@utoronto.ca
Appendix D: Pre-Experiment Questionnaire

Automatic Target Cueing Experiment Questionnaire

Age: ________________________________

Gender: ______________________________

Rank: ________________________________

Unit: ________________________________

Trade: ________________________________

Reg. Force or Reserves: ________________________________

Years of Service: ________________________________

Number of Overseas Deployments: ________________________________

Handedness?

Right

Left

Which way do you shoot?

Right

Left

Dominant Eye?

Right

Left

Do you require corrective vision?

Yes

No

If so, what is your prescription? ________________________________

Do you play video games?

Never

Rarely

Sometimes

Often
Appendix E: Post-Experiment Questionnaire

Automatic Target Cueing Post Experiment Questionnaire

How helpful did you find the Automatic Target Cueing for:

Detecting targets during the day?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Helpful</td>
<td>Somewhat Helpful</td>
<td>Extremely Helpful</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Detecting targets at night?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Helpful</td>
<td>Somewhat Helpful</td>
<td>Extremely Helpful</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Identification during the day? (when it was more likely to detect hostiles or militia)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Helpful</td>
<td>Somewhat Helpful</td>
<td>Extremely Helpful</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Identification at night? (when it was more likely to detect hostiles or militia)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Helpful</td>
<td>Somewhat Helpful</td>
<td>Extremely Helpful</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Is there anything that you particularly liked or disliked about the Automatic Target Cueing?

What are some recommendations you would make to improve the ATC?
Did your evidence requirement for positive identification of a threat change throughout the experiment? If so how?

Rate how certain you needed to be about a target in order to make a positive identification:

At the beginning of the experiment: Use an X
By the end of the experiment: Use a +
In a real-life combat situation: Use a O

0% Confident
(No evidence either way)

50% Confident
target is a threat

100% Confident
target is a threat
(No doubt at all)

Please rate the importance of getting (or not getting) each outcome listed below by using a number from 1-10 (1 being the least important to 10 being the most important):

Threat Hit: ________________________________ Militia Not Hit: ________________________________

Threat Missed: ________________________________ Militia Hit: ________________________________

Please list an approximate weighting that you considered for each target feature when making an overall identification decision (added together should total 100):

Day:


Body Armor: _______________ Cueing Likelihood: ___________ Other: ___________________________

Night:


Body Armor: _______________ Cueing Likelihood: ___________ Other: ___________________________
Appendix F: Debriefing Sheet

Protocol Number: 2014-033
Research Project Title: Effects of Bias in Automatic Target Cueing on Human Target Identification
Principal Investigator: Dr. Justin Hollands
Co-Investigators: Mr. Adam Reiner (University of Toronto, Department of Mechanical and Industrial Engineering), Dr. Greg Jamieson (University of Toronto, Department of Mechanical and Industrial Engineering), Dr. Mike Tombu, Mr. Matt Lamb, Mr. Ken Ueno
Project: Future Small Arms Research (FSAR) 02aa

Dear participant:

Thank you for having completed this experiment.

The Canadian Armed Forces is currently investigating technologies that could be incorporated into a family of future small arms. One such technology is automatic target cueing. As currently implemented, automatic target cueing is not 100% reliable. Depending on the parameters of the algorithm and the difficulty of detecting targets, the system sometimes misses potential targets, sometimes cues non-existent targets (false alarms), and sometimes does a mix of both. The algorithm may also tend towards detecting a specific type of target due to a number of factors. The goal of this experiment was to establish the benefit to be gained from automatic target cueing and the costs associated with automatic target cueing detection biases. In addition, we were also interested in how these factors interacted with different visibility conditions. By gaining a better understanding of how automatic target cueing affects detection, identification and the shooting process, we can make recommendations about the development of future automatic target cueing technologies and how to evaluate their effectiveness.

Through this sort of experimentation, we hope to identify the conditions under which automatic target cueing is optimally effective and incorporate these findings into future requirements for automatic target cueing technologies. By doing so, we hope to improve the shooting effectiveness of Canadian soldiers by providing them with an automatic target cueing system that is optimized to alert them to potential targets in the environment.

You may obtain additional information about the research project by contacting Adam Reiner (adam.reiner@drdc-rddc.gc.ca) or Dr. Justin Hollands (416-635-2073, justin.hollands@drdc-rddc.gc.ca).