Evaluating Reliance on Combat Identification Systems: The Role of Reliability Feedback

Greg A. Jamieson, Heather F. Neyedli, Lu Wang – University of Toronto;
Justin G. Hollands - Defence Research and Development Canada

Abstract

A pair of laboratory studies examined the effects of system reliability information and interface features on user reliance on an individual combat identification system. The first experiment showed that providing reliability information through instruction led to more appropriate reliance. In the second experiment, a human-machine interface displayed the reliability information. The results showed; 1) that the method of displaying reliability information affected the participants' sensitivity in discriminating the target from noise, and 2) that the display format (integrated vs. separated) affected the participants' reliance on the combat identification system. Taken together, the findings yield implications for the design of interfaces for individual combat identification systems and for the training of infantry soldiers. Finally, a new method of measuring reliance on automation was developed and employed across both experiments.

Introduction

A variety of technical solutions have been developed with the aim of improving soldiers' combat identification (ID) performance. Individual combat ID systems (Lowe, 2007), which operate similarly to identification friend or foe (IFF) systems, offer some support for this difficult task. In these systems, each soldier is equipped with a gun-mounted interrogator and a helmet-mounted transponder. The interrogator emits a laser inquiry, to which the transponder replies using an encrypted radio signal. If the signal is returned, the interrogator indicates (usually through a small light) that the target is friendly. The key drawback of this automated system is that it cannot positively identify a target without a working transponder (Sherman, 2000; Sherman, 2002; “SIMLAS,” 2006; Zari et al, 1997). If the interrogator receives no transponder signal, the target could be hostile, neutral or friendly.

Humans are prone to misuse and disuse imperfect automation (Parasuraman, Molloy, & Singh, 1993), and combat ID systems are no exception (Briggs & Goldberg, 1995). Misuse occurs when individuals rely on automation inappropriately, usually by over-relying on an imperfect aid. Disuse occurs when participants under-use or reject the capabilities of automation (Parasuramen & Riley, 1997). Informing individuals about the reliability of an aid can assist them in relying on an aid more appropriately (Lee & See, 2004), although several studies in the combat ID domain have shown that this is not always the case (Dzindolet et al., 2000, 2001a, 2001b; Karsh, et al., 1995). Moreover, little research has explored the possible means of providing this information.

The current research explores factors that affect reliance on combat ID systems, with the end goal of helping soldiers to use the systems more effectively and reduce fratricide incidents. To achieve this, two experiments were performed. Experiment I examined whether informing participants about the reliability of the aid affects participants’ reliance on the aid. Experiment II examined how different forms of displaying reliability information on an interface affect participants’ reliance on the aid.

Experiment I

The main goal of the first study was to examine how knowledge of aid reliability affects reliance on the automation. Previous studies suggest that participants tend to rely inappropriately on combat ID aids even when they are informed of the aid’s reliability (Dzindolet et al., 2000, 2001a, 2001b; Karsh, et al., 1995). For example, Dzindolet et al. (2001a) performed a study where the reliability of the feedback from a combat ID was 60%, 75% and 90% (in addition to a no aid group). Although the participants were informed of the aid reliability, the results showed no performance difference among the three aided groups and the unaided group. The researchers concluded that the participants over-relied on the aid feedback at all levels of reliability. However, Wang et al. (in press) questioned whether this conclusion may have been premature. A participant's decision to rely on the aid more often than ignoring it may be appropriate, especially in the case where the aid outperforms manual control. We anticipated that an alternative approach to the data analysis would reveal more insight into the reliance behavior.

We proposed a new method of measuring reliance on automation based on Signal Detection Theory (SDT) (Macmillan & Creelman, 1991; Wickens & Hollands, 2000). In SDT, two performance indicators—sensitivity (d') and response bias (C or ß)—characterize the participants' performance. When a soldier receives aid feedback, his/her expectation of the probability that the target is friendly or hostile should change. This change, according to SDT, should influence the setting of the response bias but not of the sensitivity. If this premise is true, then the reliance on the aid can be measured based on the change of the response bias. For this reason, we called our method the *response bias* approach (Wang, Jamieson & Hollands, in press)

In our first experiment, we revisited the question of reliance on imperfect combat ID aids. The objectives of the experiment were; 1) to examine the effectiveness of using aid reliability information to support appropriate reliance on a simulated, rifle mounted, combat ID system, and 2) to test the feasibility of using response bias as an indication of participants' reliance on the combat ID aid.

Experimental Design

A 3 (aid reliability: no aid, 67%, 80%)×2 (instruction of aid reliability: informed, uninformed) mixed design was employed. Aid reliability was manipulated as a within-subjects factor with a reliability of 67%, indicating that 67% of the time the target is actually an enemy. In the no aid condition, the participants conducted the combat ID task through visual identification alone. The instruction of aid reliability was a between-subjects factor.

The experiment was comprised of three mission blocks, one at each level of aid reliability. The order of conditions was counterbalanced separately across participants. Each block consisted of 120 trials, with one target appearing in each trial. For each block, half of the targets were friendly and half were hostile.

Combat ID Simulation

The IMMERSIVE (Instrumented Military Modeling Engine for Research using Simulation and Virtual Environments) synthetic task environment served as the test bed. Developed by Defence Research and Development Canada at Valcartier, IMMERSIVE uses the modules of a commercial first-person shooter game – Unreal Tournament 2004. Experimenters can create scenarios by setting terrains, combat activities, and characteristics of forces. Friendly and hostile forces are distinguished by differences in uniforms, weapons, actions, and feedback from the combat ID system.

Data Collection

Participants: 26 students from University of Toronto with normal visual acuity were recruited. Complete data were collected from 24 participants and only this data were used for analysis. Each participant was paid $30 CAD for his/her participation, and a bonus $10 CAD was given to the top performer.

Tasks and Procedures: The experiment took approximately 2.5 hours to complete. To take part in this experiment, each participant was required to pass a visual acuity test (measured with a Snellen eye chart), and an ocular dominance test (measured using the Porta Test (Roth et al., 2002)) was performed.

The participants were instructed to imagine themselves in a battlefield. Their primary task was to identify targets in the scene and shoot the enemies. Their score was determined by the accuracy and speed of their engagement decision. The participants were advised that they would have an aid to assist them in 2 of the 3 blocks. When the aid identified a friendly solider, it would respond with 'friend' feedback – a blue light. Otherwise it would respond with 'unknown' feedback - a red light. The participants were told that the 'unknown' feedback was set to be less than 100% reliable to mimic system failures. It was possible that a red light could be shown when a target was actually friendly. However, the blue light would never appear when a target was hostile.

Data Analysis: The participants' reliance on the aid was analyzed using the response bias method. SDT parameters d', C and ß were calculated using the values contained in Table 1.

Table 1. The outcome matrix in the condition that the aid gave 'unknown' feedback

<table>
<tr><td colspan="2" rowspan="2"></td><th colspan="2">States of the World</th></tr>
<tr><th>P(Enemy|Unknown)</th><th>P(Friend|Unknown)</th></tr>
<tr><th rowspan="2">Participant Response</th><th>Shoot</th><td>Hit (H)
P(H|Unknown)
Value=V(H)</td><td>False Alarm (FA)
P(FA|Unknown)
Cost=C(FA)</td></tr>
<tr><th>Not Shoot</th><td>Miss (M)
P(M|Unknown)
Cost=C(M)</td><td>Correct Rejection (CR)
P(CR|Unknown)
Value=V(CR)</td></tr>
</table>

Both C and ß were used as C has the simplest statistical properties, while ß can be compared to optimal ß values. According to Signal Detection Theory (SDT), response bias will vary with the expectation of the target probability, whereas their sensitivity will stay constant (Macmillan & Creelman, 1991; Wickens & Hollands, 2000).

The optimal ß values were calculated so they could be compared to the actual values. For the 67% reliability condition: $\beta_{optimal} = \frac{P(Friend \mid Unknown)}{P(Terrorist \mid Unknown)} = \frac{33\%}{67\%} = 0.50$;

for the 80% reliability condition:

$$\beta_{optimal} = \frac{P(Friend \mid Unknown)}{P(Terrorist \mid Unknown)} = \frac{20\%}{80\%} = 0.25 .$$

To increase the normality of the probability data, an arcsine transformation was applied to all probability data: Transformed Probability Data= 2 * arcsine $[\text{Probability Data}]^{1/2}$ (Dzindolet et al., 2001a; Howell, 1992; Winer, 1991). A 3 (aid reliability: no aid, 67%, 80%) X 2 (group: uninformed, informed) mixed ANOVA was used to examine the FA rate, miss rate, d', C and response time (measured from when the target appeared until the first shot was taken).

Results

False Alarm (Friendly Fire): There was a significant main effect of aid reliability, F(1.506[1],44)=10.752, p=.001. As seen in Figure 1, the participants made fewer false alarm errors when they had the combat ID aid, F(1,22)= 9.858, p=.005, r=.556, and fewer errors in the in the 80% reliability condition compared to the 67% reliability condition, F(1,22)=13.950, p=.001, r=.623.

Miss (Miss Hostile Target): There were no significant main effects of aid reliability or group on miss rate, and no significant interactions.

[1] The assumption of sphericity was violated, the degrees of freedom were corrected using the conservative Greenhouse-Geisser value.

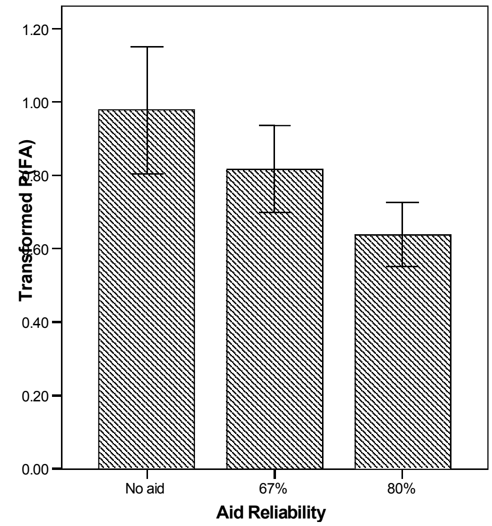

Figure 1 The effect of aid reliability on transformed P(FA)

Response Time: There were no significant main effects of aid reliability or group on response time, and no significant interactions.

Sensitivity, d': There were no significant main effects of aid reliability or group on sensitivity, and no significant interactions. Consistent with the hypothesis, sensitivity did not vary with the aid reliability or group assignment.

C: There was a significant main effect of aid reliability on response bias, $F(2, 34)=5.272$, $p=.010$ (Figure 2).

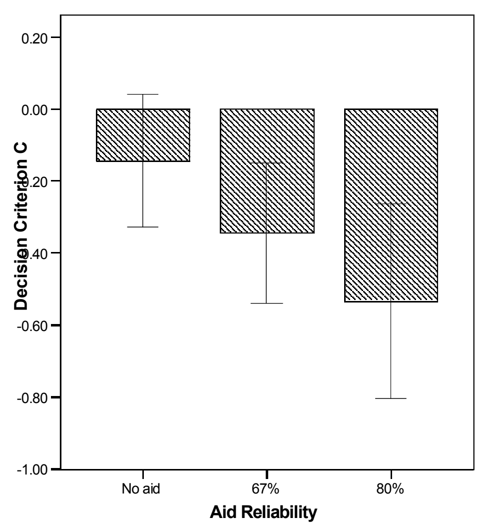

Figure 2 The main effect of aid reliability on decision criterion C

Contrasts reveal a significant difference between the no aid condition and two aided conditions, F(1,17)=5.475, p=.032, r=.494, and a significant difference between the 67% reliability condition and the 80% reliability condition, F(1,17)=4.657, p=.046, r=.464. Participants were more liberal (i.e., they shot at more targets) as the reliability level increased.

There was also a significant main effect of group, F(1, 17)=8.272, p=.010, r=.572. Regardless of aid reliability condition, the informed group (M=-.523) was more liberal in their decision to shoot than the uninformed group (M=-.138), which implied that the informed group relied on the 'unknown' feedback more than the uninformed group.

Appropriateness of Reliance: A Shapiro-Wilk's test revealed that the assumption of normality was violated for bias ß, so a natural log transformation was applied. A series of one-sample t-tests were conducted to test the difference between the ln ß and the optimal values (see Table 2).

Table 2. T-test comparing participants' response bias with the optimal value

Group	Aid reliability	ln(Optimal Beta)	t-value	p-value (2-tailed)
Uninform	No aid	ln(1.00)	t(8)=.630	.546
Inform	No aid	ln(1.00)	t(9)=-2.568	.030
Uninform	67%	ln(0.50)	t(8)=3.538	.008
Inform	67%	ln(0.50)	t(9)=-.544	.600
Uninform	80%	ln(0.25)	t(8)=5.937	.000
Inform	80%	ln(0.25)	t(9)=-.063	.951

For the two aided conditions, the response criterion of the uninformed group was significantly more conservative than optimal, while that of the informed group did not differ significantly from the optimal value. Thus, the informed group relied on the aid more appropriately in the aided condition than the uninformed group.

Discussion

Previous studies have found that combat identification speed and accuracy were not significantly improved by imperfect combat ID systems (Dzindolet et al., 2000, 2001a, 2001b; Karsh et al., 1995). In partial agreement with these findings, this experiment found no significant differences in the speed of engagement decision or the number of missed hostile targets among all test conditions. However, in contrast to previous studies, the combat ID aid in this study contributed to a significant reduction in the number of the friendly fire engagements. This improvement was found at two reliability levels, and increased with the aid's reliability.

In further contrast to previous studies, where reliance behavior was deemed to be suboptimal, there were two indicators that participants in this study generally relied on the aid reasonably. First, they almost always followed the 100% reliable 'friend' feedback. Second, they used the 'unknown' feedback to inform their identification decision, but did not blindly follow it. Although the informed group relied on the 'unknown' feedback more appropriately than the uninformed group, this difference in reliance did not lead to a difference in the accuracy of their engagement decision. One possible reason for this might be that the performance measure is less sensitive than the reliance measure, because the former is subject to individual differences in identification sensitivity.

Dzindolet et al. (2001b) found that instructions were not sufficient for the participants to rely on the feedback appropriately. The results in this experiment are inconsistent with Dzindolet et al.'s (2001b) findings. In this experiment, while the uninformed group did not rely on the aid enough, the informed group relied on the aid appropriately.

The target proportion and aid reliability in Dzindolet et al.'s (2001b) study were the same as the 67% reliability condition in this study, therefore, the observed differences in reliance is likely attributable to other factors. One possible factor is the participants' suspicion of the aid reliability instruction in Dzindolet et al.'s study (2001b). In Dzindolet et al.'s study, participants were able to view the stimulus a second time if they were unsure whether the target was present. Even for fully reliable feedback (similar to the 'friend' feedback in the present study), participants requested a second viewing of the stimulus, indicating that they might have been doubtful about the reliability information. A second possible factor is workload. In Dzindolet et al.'s study, the participants were required to respond to audio stimuli in addition to performing the detection task. Therefore, it is likely that the workload was higher in their study. Some research suggests that misuse of automation is more likely to happen when the participants are responsible for tasks in addition to the automated task (Parasuraman et al., 1993).

Experiment II

In Experiment I, providing participants with information regarding the aid's reliability was shown to improve reliance. The information was provided through instruction and not through a human-machine interface (HMI). Therefore, it is prudent to consider how the means of communicating reliability information might affect the appropriateness of reliance.

Reliability or uncertainty information is often probabilistic, and can be conveyed through numeric (e.g., 80 percent certain), or linguistic (e.g., unlikely, very likely etc.) indicators with little to no difference in judgment performance (Budescu et al, 1988). Numeric values can also be displayed in an analogue (e.g., a pie chart) or digital form. Finger & Bisantz (2002) proposed using an icon that degraded and became more pixilated as the reliability decreased. In experiments using the degraded icons, no difference in target identification performance was observed between seemingly more 'precise' methods (e.g., digital-numeric) and the degraded icon (Finger & Bisantz, 2002; Bisantz, Marsiglio & Munch, 2005).

The prospect of presenting reliability feedback along with the identification feedback raises a second question. Should these two sources of feedback be displayed separately or using an integrated multi-element display? The proximity compatibility principle (PCP) (Wickens & Carswell, 1995) suggests that integrating the two pieces of information may assist the user in using them simultaneously in a time-pressured environment. However, PCP also cautions that integrating information can inhibit the focused attention required to determine a precise value. Therefore the participant may not be able to glean the reliability information as accurately when using an integrated display. Our second experiment sought to tease out these potential advantages and disadvantages for the aided combat ID task.

Four prototypes were developed to display the inquiry feedback and the reliability of the feedback (see Figures 3). The prototypes differed in the method used to display reliability information as well as whether the feedback reliability was integrated or separated from the inquiry feedback. The first method to display reliability information, dubbed the grid display, used a degrading stimulus. The second method, the pie chart display, used an analogue graphical representation.

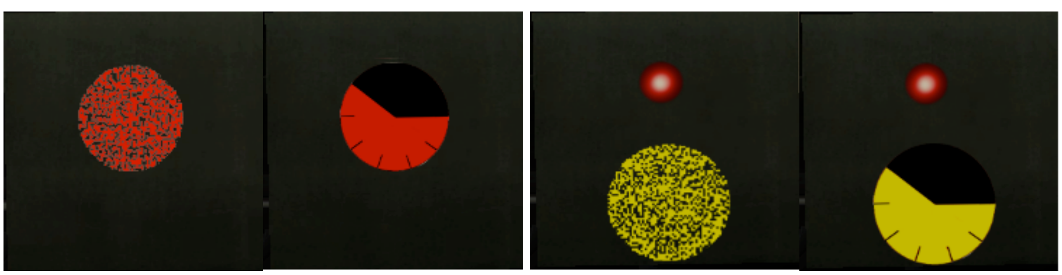

Figure 3. Integrated grid and pie displays and separated grid and pie displays as implemented in IMMERSIVE

The objectives of Experiment II were; 1) to determine which method of displaying feedback reliability information affords the best performance and most appropriate reliance on the combat ID system, and 2) to determine whether integrating or separating the feedback reliability and the feedback identification information affords the best performance and most appropriate reliance on the combat ID system.

Experimental Design

A 2 (reliability display type: pie method, grid method) x 2 (display format: integrated, separated) x 5 (reliability level: 50%, 60%, 70%, 80%, 90%) mixed design was used. The display formats and reliability levels were within subjects factors, while the reliability display type was a between subjects factor. In other words, each participant experienced one of the reliability display types in both an integrated and separated format at five different reliability levels. Each between-subjects group consisted of 14 participants.

The participants completed 840 trials, separated into eight blocks of 105 trials each. In four blocks the participants were presented with a separated HMI and in the other four blocks with an integrated HMI (the order was counterbalanced between participants). For each block, the targets in half of the trials were friendly and in the other half were hostile. The reliability levels varied within block and the order of presentation for each participant was randomized.

Data Collection

***Participants**:* 30 University of Toronto students with normal acuity vision were recruited. Complete data was collected from 28 participants and only this data was used in the analysis. Participants were paid $40 CAD for their participation with a bonus of $10 CAD for ‘good’ performance.

Task and Procedures: The procedure was similar to that in Experiment I with one difference; in Experiment II participants were required to kill (as opposed to shoot at) a target if they considered it hostile to receive a score. This change was made to mimic the time pressure of an actual engagement. The participants had a better chance of killing a target if they could make a decision earlier.

Data Analysis: As in Experiment I, probability data was transformed using an arcsine function. A 2 (display format: integrated, separated) x 2 (reliability display type: pie, grid) x 5 (reliability level: 50%, 60%, 70%, 80%, 90%) mixed ANOVA was conducted on the transformed P(FA), transformed P(miss), time to kill a target and d'. The time to kill the target was measured from when the target first appeared to when the participant successfully killed the target.

Results

False Alarm (Friendly Fire)*:* The main effects of display format and display type on false alarm rate were not significant. However, the main effect of reliability level was significant, $F(2.20, 5.67) = 21.3$, $p<0.001$, $r=0.41$, where P(FA) increased as the reliability level increased. That is, when the display indicated an increased probability of a terrorist when there was unknown feedback, participants killed more Canadian soldiers. None of the interactions were significant.

Miss (Miss Hostile Targets)*:* There was a main effect of display type on miss rate, $F(1, 26) =4.31$, $p<0.05$, $r=0.38$ (Figure 5).

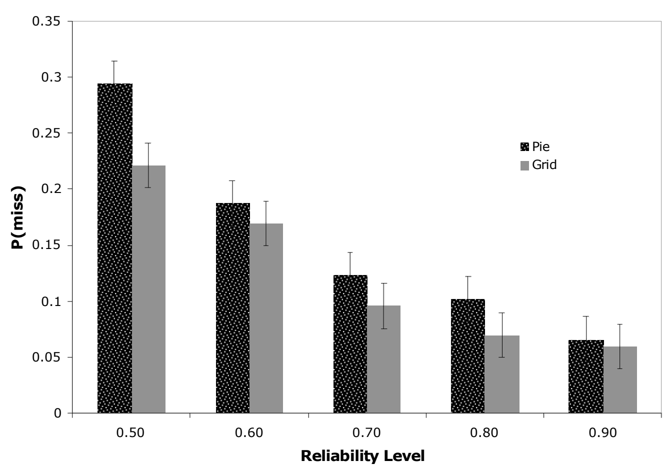

Figure 5. The main effects of display method and reliability level on P(Miss)

Participants who were shown the pie display missed more targets than participants who were shown the grid display. The main effect of reliability level was also significant, F(1.54, 40.2)=24.4, p<0.001, r=0.44, with P(miss) decreasing as reliability level increased. The main effect of display format was not significant, F<1, nor were any of the interactions.

Time to Kill Target: The main effect of reliability level had a significant effect on kill time, F(4, 104) =10.9, p<0.001, r=0.31, with the kill time decreasing as the reliability level increased. The main effects of display format and type were not significant.

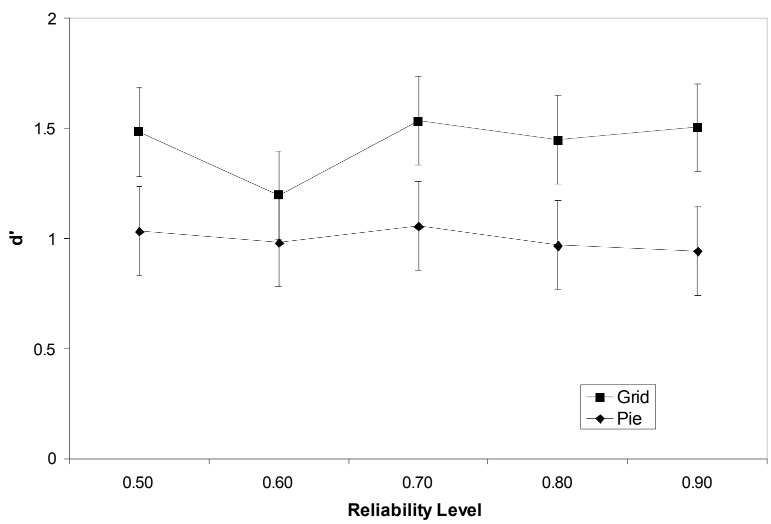

Figure 6: Main effect of display method on sensitivity, d'.

Sensitivity d': The effect of display type on sensitivity was significant F(1,22[2]) =6.62, p<0.05, r=0.48. The participants using the grid display were more sensitive in distinguishing the targets from noise (Figure 6). None of the other main effects (display format or reliability level) or interactions were significant.

Reliance: To analyze the appropriateness of the participant's reliance on the aid given different displays, we fit a function relating the reliability level to the $\beta_{optimal}$ to the participants' β_{actual} for

---

[2] If a participant had zero hits, misses, FAs or CRs for a reliability level, the SDT measures could not be calculated for that level. At high reliability levels, corrections (such as the one suggested by Macmillan and Creelman (2004)) produced unstable estimates of sensitivity and the decision criterion and therefore were not used.

each combination of display format and type. $\beta_{optimal}$ is equal to a cost function multiplied by the probability of noise divided by the probability of a signal occurring:

$$\beta_{optimal} = \frac{V(CR) + C(M)}{V(H) + C(FA)} * \frac{p(friend|unknown\ feedback)}{p(enemy|unknown\ feedback)}$$

In the present experiment the participants killed 90% of the targets they shot at, therefore even if the participants correctly identified a terrorist, only 90% of the time they could get credit, which changed the value of a hit (i.e., a correct identification of a terrorist).

$$\text{therefore: } V = \frac{V(CR) + C(M)}{V(H) + C(FA)} = \frac{1 + 0}{1*0.90 + 0} = 1.11$$

The probability that the target is an enemy given unknown feedback is equivalent to the reliability level (RL), and the probability that the target is a friend given unknown feedback is therefore *1 – RL.* The equation for $\beta_{optimal}$ can therefore be rearranged to:

$$\beta_{optimal} = V\left(\frac{1 - RL}{RL}\right) = V\left(\frac{1}{RL} - 1\right)$$

This non-linear inverse equation for $\beta_{optimal}$ was fit to the β_{actual} values for each combination of display type and format. The parameter V for each combination was compared to the optimal V of 1.11 and R^2 was calculated. Because this was a constrained model, it was possible for R^2 to be negative if a horizontal line was a better fit to the data (Table 3)

Display	V	R^2
Integrated-Pie	1.3	0.24
Integrated-Grid	1.2	0.19
Separated-Pie	1.3	-0.04
Separated-Grid	0.85	-0.27

Table 3*:* V and R^2 values for each display combination when the $\beta_{optimal}$ equation was fitted to β_{actual}.

Both of the integrated displays had positive R^2 values and a V parameter close to the optimal value of 1.11. Both of the separated displays had negative R^2 values, indicating that a horizontal line is a better fit to β_{actual} than the optimal equation. Therefore, participants relied on the separated displays less optimally (Figure 7).

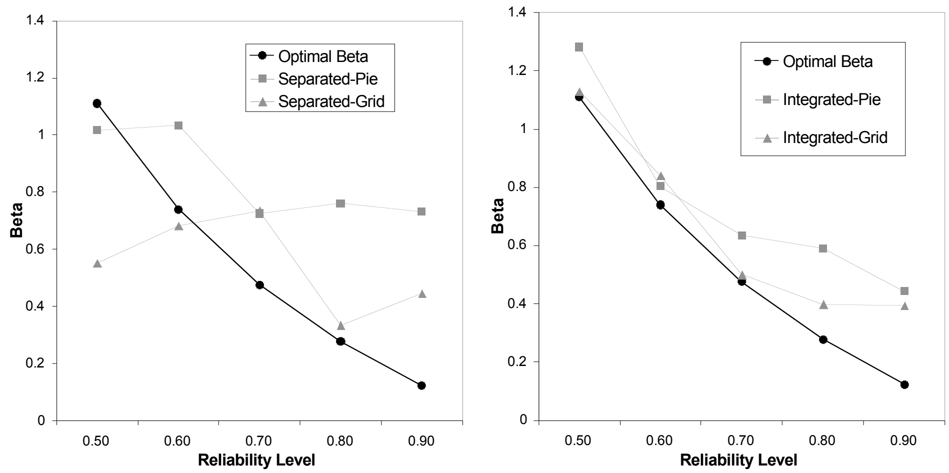

Figure 7: Comparison of $\beta_{optimal}$ to β_{actual} for separated and integrated displays

Discussion

Participants adjusted their reliance on the aid when the reliability information was provided through an HMI. However, the form of the HMI had an affect on how effectively this information was used to adjust reliance and identify targets.

Effect of Reliability Display Method: Participants using the grid display (whether integrated or separated) showed greater sensitivity in discriminating enemy soldiers from Canadian soldiers. This result was unexpected as displaying reliability information was expected to affect only the response criterion. Generally speaking, greater sensitivity can be achieved by increasing the signal strength (i.e., by making the target more salient) or by increasing the arousal of the observer (Wickens & Hollands. 2000); however, these variables were controlled in the study. We therefore look to the characteristics of the individual displays for alternative hypotheses that might explain this unexpected result.

One possibility we must examine is whether one of our displays (rather than the target set or participant arousal) increased or decreased the signal strength. The participants judged the presence of a target by examining both the display and the soldier in the environment. Because the targets remained identical, if we assume these form a cohesive signal, a change in the salience of the displays seems to offer a more likely source of the difference in sensitivity.

A more nuanced explanation for the observed difference in sensitivity considers that the participant samples both the reliability display and the target as separate channels. It is possible that, instead of these two information sources creating a cohesive signal, they actually compete for the participants' attention. Following this reasoning, a salient display could draw the participants' attention away from the target, thereby decreasing sensitivity. If the participants using the pie display tried to estimate the precise reliability level, they would have had less time to examine the stimulus to determine its identity. While intuitively it would seem more difficult to determine a precise reliability level using the grid, studies have found individuals perform just as well or better with such displays as compared with numeric or linguistic indicators (Finger & Bizantz, 2002; Bizantz, Marsiglio & Munch, 2005).

The grid display also contained an emergent feature of the graphical element dimming and degrading as the reliability level decreased which may have assisted the participants in gleaning the reliability information more quickly, leaving more time for them to examine the actual stimulus. Because kill time did not differ between conditions, it is possible that any additional time was spent determining the soldier's identity as increasing the stimulus duration has been shown to increase sensitivity (Wickens & Hollands, 2000). Further reinforcing this hypothesis, the group using the pie display has a significantly higher P(miss) without a corresponding decrease in P(FA), which may have been due to their spending more time sampling the display information. Other studies have shown the participants are significantly slower and less accurate in gleaning information from a pie chart than degraded indicators of probability information (Feldman-Stewart et al, 2007).

It is not completely clear why participants relied on the aid less optimally when shown the grid display. As hypothesized above, the pie display may have afforded more attempts at exact calculations of the reliability level, which may have mitigated the display format effect on decision criterion. Performing an eye-tracking study could produce the data required to confirm these hypotheses.

Effect of Reliability Display Format: Participants were particularly insensitive to changes in reliability level when they were shown the separated-grid display; failing to shift their decision criterion with the reliability level. It appears that these participants maintained a decision criterion that was slightly more liberal than optimal for a reliability level of 70%, regardless of the displayed reliability level. It is worth noting that the mean reliability level across trials was 70%. It may be that separating the information was so detrimental that the participants ignored the feedback reliability information altogether. It is clear that the participants still used the aid for the inquiry feedback even while ignoring the reliability information because they held fire during the 'blue light' friend trials and activated the aid on most if not all trials. As discussed above, the participants have to consult two sources of information, both the aid and the stimulus in the time pressured scenario. Separating the information may have effectively created another channel of information that the participants sometimes chose to disregard while under time pressure to kill the target. When the feedback reliability information was integrated with the feedback itself, it is possible the participant could more easily access the information while determining the results of the inquiry feedback.

Conclusion

Using the IMMERSIVE and the new reliability assessment method, we demonstrated that providing participants with a combat identification aid can reduce the number of friendly fire engagements. Providing participants with 'unknown' feedback reliability information allowed them to more appropriately rely on the automated decision aid, whether this information was provided through verbal instruction or displayed in an interface. The form of the interface influenced the effectiveness of the information conveyed by affecting both participants' sensitivity in detecting the targets and reliance on the aid.

References

Bisantz, A.M., Marsiglio, S.S., & Munch, J. (2005) Displaying Uncertainty: Investigating the Effects of Display Format and Specificity. Human Factors. 47(4). 777-796

Briggs, R. W., & Goldberg, J. H. (1995). Battlefield recognition of armored vehicles. Human factors, 37(3), 596-610.

Budescu, D.V., Weinberg, S., & Wallsten, T.S. (1988). Decisions based on numerically and verbally expressed uncertainties. Journal of Experimental Psychology: Human Perception an Performance. 14(1). 281-294

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2000). Misuse of an automated decision making system. In Conference on Human Interaction with Complex Systems 2000 (pp. 81-85). Urbana-Champaign, IL.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001a). Predicting misuse and disuse of combat identification systems. Military Psychology, 13(3), 147-164.

Dzindolet, M. T., Pierce, L., Pomranky, R., Peterson, S., & Beck, H. (2001b). Automation reliance on a combat identification system. In Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society (pp. 532-536). Minneapolis/St. Paul, MN: Human Factors and Ergonomics Society

Finger, R. & Bisantz, A.M., (2002) Utilizing graphical formats to convey uncertainty in a decision making task. Theoretical Issues in Ergonomics Science. 3(1). 1-25

Feldman-Stewart, D., Brundage, M.D. & Zotov, V. (2007) Further insight into the perception of quantitative information: Judgments of gist in treatment decisions. Medical Decision Making. 27(1). 34-43

Karsh, R., Walrath, J. D., Swoboda, J. C., & Pillalamarri, K. (1995). Effect of battlefield combat identification system information on target identification time and errors in a simulated tank engagement task. (Technical report ARL-TR-854). Aberdeen Proving Ground, MD, United States: Army Research Lab.

Howell, D. C. (1992). Statistical methods for psychology (3rd ed.). Belmont, CA: Duxbury Press.

Lowe, C. (2007). Cutting through the fog of war. Retrieved Aug 20th, 2007, from http://www.defensetech.org/archives/003496.html

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. New York, NY: Cambridge University Press.

Parasuraman, R., Molloy, R., & Singh, I. (1993). Performance consequences of automation-induced “complacency”. International Journal of Aviation Psychology, 3, 1-23.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39, 230-253.

Roth, H. L., Lora, A. N. , & Heilman, K. M. (2002). Effects of monocular viewing and eye dominance on spatial attention. Brain, 125(9), 2023-2035.

Sherman, K. (2000). Combat identification system for the dismounted soldier. In Proceedings of SPIE 2000: Digitization of the Battlespace V and Battlefield Biomedical Technologies II (pp.135-146). Orlando, FL: the International Society for Optical Engineering.

Sherman, K. B. (2002). Combat ID coming for individual soldiers. Journal of Electronic Defense, 25(3), 34-35.

SIMLAS: Manworn combat training and ID system. (2006). RUAG Electronics. Retrieved October 23, 2006, from http://www.ruag.com/ruag/binary?media=99587&open=true

Wang, L., & Jamieson, G. A., Hollands, J. G. (in press). “Selecting Methods for the Analysis of Reliance on Automation”. To appear in Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting.

Wickens, C.D., & Carswell, C.M. (1995) The proximity compatibility principle: Its psychological foundation and relevance to display design. Human Factors 37(3). 473-494.

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

Winer, B. J. (1991). In Brown D. R., Michels K. M. (Eds.), Statistical principles in experimental design (3rd ed.). New York: McGraw-Hill.

Zari, M.C., Zwilling, A.F., Fikes, J.W., Hess, D.A., Ward, R.N., Anderson, C.S., & Chiang, D.J., (1997) Personnel identification system utilizing low probability-of-intercept techniques: prototype development and testing, In Proceedings of IEEE 1997 on Security Technology (pp. 224-230). Canberra Australia: Institute of Electrical and Electronics Engineers.