The Effects of Design Features on Users' Trust in and Reliance on a Combat Identification System

Lu Wang, Greg A. Jamieson and Justin G. Hollands

Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2011 55: 375
DOI: 10.1177/1071181311551077

The Effects of Design Features on Users' Trust in and Reliance on a Combat Identification System

Lu Wang
Medtronic, Inc

Greg A. Jamieson
University of Toronto

Justin G. Hollands
Defence Research and Development Canada

In a previous study, we found that users' trust in and reliance on an individual combat identification system is influenced by the system's reliability as well as users' awareness of the reliability. In this exploratory study we test the effects of design features of the same system on users' target identification performance as well as their trust in and reliance on the system. In a simulated task environment, we varied the automation activation mode (i.e., automatic vs. manual) and the presentation of the "unknown" feedback (i.e., explicit vs. implicit). Participants responded fastest when the "unknown" feedback was provided automatically with explicit indication. In addition, participants trusted the explicit "unknown" feedback more than the implicit feedback. However, neither reliance behavior nor identification accuracy changed significantly across conditions. This study has implications for the design of combat identification systems to achieve appropriate trust. In addition, the results suggest that when studying trust in automation using a simulation, it is important to simulate the design features.

INTRODUCTION

Design features of an automated system, like the interface aesthetics and the automation etiquette (e.g., interruptiveness and politeness), do not affect real automation capabilities. However, such features might affect the users' beliefs about the automation's capability and their trust in and reliance on the automation (Corritore et al., 2003; Lee & See, 2004; Parasuraman & Miller, 2004). As a result, users may trust and rely on an automated system differently based on its distinctive design features.

Individual combat identification (CID) systems work through an interrogation/response process to confirm whether an unknown target is friendly or not ("Identification friend or foe", 2011; Ivtsenkov et al., 2009; Sherman, 2000). In such systems, soldiers manually activate the process to interrogate a target. If a correct electronic reply is received from the target, their CID systems will give "friend" feedback. Otherwise, the system will not give any explicit feedback and the target will be deemed "unknown". Because the electronic communication can occasionally fail in the chaotic battlefield, the system cannot always recognize friendly soldiers.

Multiple experiments have been conducted to examine users' trust in and reliance on imperfect CID systems (Dzindolet et al., 2000, 2001a, 2001b; Karsh et al., 1995; Kogler, 2003; Neyedli et al., 2010, in press; Wang, et al., 2008a, 2009). However, the automation activation mode and the representation of "unknown" feedback in these simulated systems were dissimilar in some ways to real system prototypes. One difference is that the simulated systems responded automatically after the appearance of the target (Dzindolet et al., 2000, 2001a, 2001b; Wang, et al., 2008a, 2009). Another distinction is that there was explicit audio or visual feedback when the automation could not recognize the stimuli (Dzindolet et al., 2000, 2001a, 2001b; Karsh et al., 1995; Kogler, 2003; Wang, et al., 2008a, 2009). Since users' beliefs about, trust in and reliance on automation can be influenced by design features (Corritore et al., 2003; Lee & See, 2004; Parasuraman & Miller, 2004), the differences in design features between the prototype individual CID systems and the experimental automation simulation raises a question: Does the way that the interface provides feedback to users affect their belief about, trust in, and reliance on the automation?

We conducted an exploratory, design-oriented study to test whether differences in the design features of an individual CID system, such as the activation mode and the feedback type, would cause participants to trust in and rely on the system differently, and consequently lead to differences in the accuracy and speed of the target identification task.

METHODS

Participants

Twenty-five students from the University of Toronto with normal visual acuity were recruited. Each participant was paid $30 CAD for their participation, and a bonus of $10 CAD was given to the participant with highest target identification accuracy.

Experimental Design

A 2 (automation activation mode: automatic vs. manual) x 2 ("unknown" feedback type: red light/explicit vs. no light/implicit) repeated-measures design was employed. In the automatic mode, the simulated CID aid was always turned on and it responded automatically whenever the weapon was pointed at a target; in the manual mode, the system was off unless the participants pressed an activation button. For the "red light" feedback type condition, the aid responded with a red light to "unknown" targets; for the "no light" condition, the aid did not send out any signal when it considered a target "unknown".

Apparatus

The experimental simulation was modified from a commercial shooter game (Figure 1).

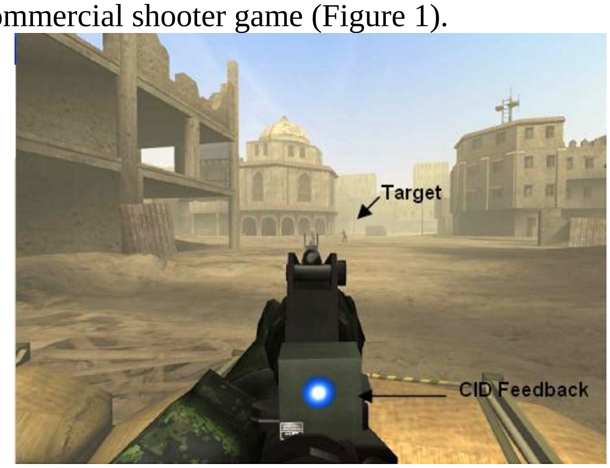

Figure 1. Participant's view in simulation

The simulation was installed on two Dell OptiPlex GX270 desktop computers with 20-in UltraSharp 2000FP flat panel monitors. Each participant only used one computer and one monitor.

Procedures

The experiment took approximately 2.5 hours to complete. Each participant first took a vision test, signed a consent form and filled out a demographic survey. Participants read the instruction sheet to familiarize themselves with the experimental procedure. Following that, an experimenter showed participants pictures of friendly and hostile targets. Then participants were asked to identify targets and shoot at hostile targets in a training session consisting of 120 trials. After each training trial, the experimenter informed the participant whether a target was friendly or hostile. After training, participants completed four mission blocks, one for each combination of activation mode and type of "unknown" feedback. Each mission block consisted of 60 trials – 30 friendly targets and 30 hostile targets. The order of the conditions was counterbalanced across participants.

The experimental task was to shoot a target if it was hostile and hold fire if it was friendly. A CID aid was provided to participants. They were told that the "friend" feedback (blue light) was 100% accurate, while the "unknown" feedback was fallible and only corresponded to hostile targets 67% of the time. Before each block, the automation activation mode and the type of "unknown" feedback were disclosed to participants. At the end of each block participants were asked to fill out questionnaires about their trust in the aid, the aid's usability, and their workload (Hart & Staveland, 1988; Jian et al., 2000; Lund, 2001). After the participant completed all four mission blocks, they were asked to indicate their preferences for the activation mode and the "unknown" feedback type and give reasons to support their choices.

RESULTS

Performance

Response time (RT) was defined as the elapsed time between when a target appeared in the scene and when the first shot was fired. A RT value was recorded for each trial in which the participants shot at a target, regardless of whether the decision is correct or not. The mean of these values was computed for each participant in each condition. The mean RT data were submitted to a 2 (activation mode:

automatic vs. manual) x 2 ("unknown" feedback type: red light vs. no light) repeated-measures ANOVA. There was a significant main effect of activation mode, $F(1,24)=9.96$, $p<.01$, $r=.54$, and a significant interaction between activation mode and feedback type, $F(1,24)=5.49$, $p=.03$, $r=.43$. As shown in Figure 2, participants responded more quickly in the automatic mode than the manual mode. In addition, the benefit for automatic activation was greater when the "unknown" feedback was "red light" than when it was "no light".

Each trial was scored, and false alarm and miss rates were computed. These data were submitted to a 2 x 2 ANOVA. Neither type of error was affected by the design features, and there was no statistical signification interaction, p > .05 in each case.

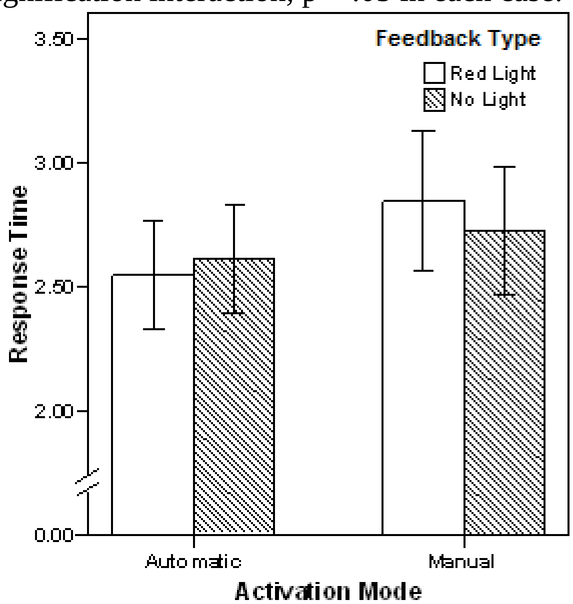

Figure 2. RT as a function of design features. Error bars indicate standard deviation (SD).

Belief about Automation Characteristics

To measure participants' beliefs about automation characteristics, participants were asked to pick their preferred design features and provide reasons. In addition, they were also asked to provide ratings of automation usability and workload under each experimental condition.

Eighteen out of twenty-five participants preferred the automatic mode. They claimed that the automatic mode was faster, required less effort, and enabled them to focus more on the identification task than the manual mode. Six participants favored the manual mode, claiming it allowed them to use the aid only when necessary. Since the aid feedback could be misleading, they wanted to avoid using it when they were confident. One participant did not have a clear preference for one activation mode.

Sixteen participants preferred the "no light" feedback type. They felt that it was dangerous to use the salient "red light" to convey fallible information. Upon seeing the red light, their first impulse was to shoot the likely hostile target immediately. In addition, participants pointed out that the "red light" feedback could draw their eyes away from the target and disturb their visual identification. Nine participants picked the "red light" feedback because they felt that they reacted faster to this more salient feedback type. Moreover, the "red light" could also confirm that they had successfully activated the aid. When the "unknown" feedback was "no light", participants were often confused about whether they had activated the automation or not.

Pearson's chi-square analysis indicated that participants' preference for neither the activation mode nor the "unknown" feedback type was significantly different from equal distribution, p > .05 in both cases.

The NASA TLX workload scores were submitted to a 2 x 2 repeated-measures ANOVA. The main effect for activation mode approached conventional significance levels, $F(1,24)=3.79$, $p=.06$, $r=.37$. The mean workload scores was higher in the manual mode ($M=.59$) than the automatic mode ($M=.55$). Neither the effect for "unknown" feedback type nor the interaction was significant, $p > .05$ in each case.

The participants' ratings of four usability criteria (i.e., the aid usefulness, ease of use, ease of learning, and satisfaction) did not significantlyby design features, p > .05 in each case.

Trust in Automation

Trust was measured using 7-point scales (Jian, et al., 2000). Regardless of experimental conditions, almost all of the participants indicated complete trust in the fully reliable "friend" feedback. Their trust ratings of the "unknown" feedback (i.e., "red light" and "no light") were not normally distributed, and therefore a non-parametric test was used. The Friedman's ANOVA revealed that trust in the "unknown" feedback was significantly different among the four conditions, $\chi(3)=9.85$, $p=.02$. As shown in Figure 3, participants appeared to trust the "unknown" feedback more when it was "red light" than when it was "no light". This effect of feedback type was significant in the automatic mode, $z=-2.18$, $p=.03$, $r=-.44$, but not in the manual mode, $z=-1.63$, $p=.10$, $r=-.33$. The effect of activation mode was not significant regardless of feedback type.

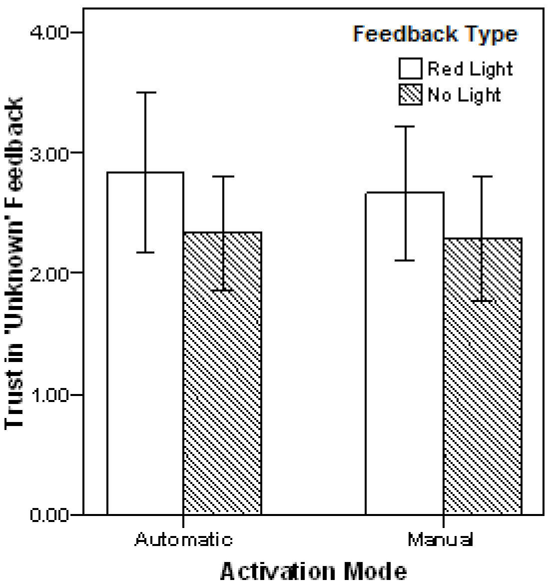

Figure 3. Trust in "unknown" feedback as a function of design features. Error bars indicate SD.

Reliance on Automation

Regardless of the "unknown" feedback type, participants activated the CID aid almost all the time in the manual activation mode (M=92.6%). In addition, a paired t-test indicated that participants were twice as likely to activate the aid when the "unknown" feedback was "no light" (M=22.9%) than when it was "red light" (M=11.4%), $t(24)=-3.63$, $p<.01$. Participants reported that they activated the aid more than once to verify that the "no light" feedback was not attributable to an unsuccessful activation.

Participants almost always complied with the 100% correct "friend" feedback. However, they did not always follow the fallible "unknown" feedback. We calculated participants' identification sensitivity and response bias for those trials when they received "unknown" feedback. The more participants relied on the "unknown" feedback, the more liberal they should have become when they received it (Meyer, 2001; Wang et al., 2008b, 2009). However, neither sensitivity nor response bias differed significantly among the four conditions.

Belief, Trust and Reliance

The belief scores (i.e., TLX workload scores as well as ratings of aid usefulness, ease of use, ease of learning, and satisfaction), the trust ratings for "unknown" feedback, and reliance scores (i.e., response bias in the "unknown" feedback trials) were calculated for each participant in each condition. These scores were submitted to a set of one-tailed Kendall's τ correlations to examine the relationships among belief in, trust in, and reliance on the CID aid.

Belief and Trust. Trust in the "unknown" feedback was negatively correlated with the workload, $\tau(100)=-.21$, $p<.01$, and positively correlated with satisfaction with the aid, $\tau(100)=.22$, $p<.01$.

Trust and Reliance. Participants' reliance on the "unknown" feedback, as indicated by the response bias, was positively correlated with their trust in the "unknown" feedback, $\tau(100)=-.15$, $p=.03$.

Belief and Reliance. Reliance on the "unknown" feedback was also positively correlated with the usefulness ratings, $\tau(100)=.15$, $p=.02$. However, it was negatively correlated with the ease-of-learning ratings, $\tau(100)=.13$, $p=.05$.

DISCUSSION AND CONCLUSION

Participants trusted the explicit "unknown" feedback more than the implicit feedback. In addition, participants also responded quickest when receiving explicit feedback under automatic mode. These results suggest design features can alter users' trust in the system and their performance. Therefore, experimental studies should simulate design features of automation systems to increase the confidence of generalizing study results to the field.

The results shed light on four aspects of the interaction design of CID systems. First, the lower the workload and the higher the satisfaction with the automation, the more the users trusted the automation feedback. This result suggests that users' trust in automation can change in response to variations in automation usability. Second, when the "unknown" feedback was implicit, participants tended to activate the aid more than once. Without the feedback, participants appeared to be concerned about a failure to activate the system. When there is no explicit identity feedback, a CID system should provide a clear indication of successful activation (e.g., an 'automation on' light). Third, the feedback representation should not interfere with the visual identification task. Participants' responses suggest that using subtle visual feedback, tactile or audio feedback might garner greater user acceptance. Fourth, the automatic activation mode was preferred to the manual mode for reducing the users' workload and response time under stressful task conditions. However, special caution must be taken in designing a system that can reveal an interrogating soldier's presence to enemies equipped with signal detection equipment.

REFERENCES

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). Online trust: Concepts, evolving themes, a model. *International Journal of Human-Computer Studies, 58*, 737-758.

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*, 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). Santa Monica, CA: Human Factors and Ergonomics Society.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: North Holland Press.

Identification friend or foe. (2011). In *Wikipedia*. Retrieved Jan 20, 2011, from http://en.wikipedia.org/wiki/Identification_friend_or_foe

Ivtsenkov, G., Mantsvetov, A, & Berik, E. (2009). *CA Patent No. 2549727*. Quebec, Canada: Canadian Intellectual Property Office.

Jian, J., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*, 53-71.

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: Army Research Lab.

Kogler, T. M. (2003). *The effects of degraded vision and automatic combat identification reliability on infantry friendly fire engagements*. Unpublished master's thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

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

Lund, A. M. (2001). Measuring usability with the USE questionnaire. In *Usability Interface: Usability SIG Newsletter*, October. Retrieved Jan 20th, 2011, from http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.html

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors, 43*, 563-572.

Neyedli, H. F., Hollands, J. G., & Jamieson, G. A. (in press). Beyond identity: Incorporating system reliability information into an automated combat identification system. *Human Factors*.

Neyedli, H. F., Wang, L., Jamieson, G. A., & Hollands, J. G. (2010). Evaluating reliance on combat identification systems: The role of reliability feedback. In D. H. Andrews, R. P. Herz, & M. B. Wolf (Eds.), *Human factors issues in combat identification* (pp. 247-262). Farnham, England: Ashgate.

Parasuraman, R., & Miller, C.A. (2004). Trust and etiquette in high-criticality automated systems. *Communications of the ACM, 47*, p. 51-55.

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.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2008a). Improving reliability awareness to support appropriate trust and reliance on individual combat identification systems. In *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society* (pp.292-296). Santa Monica, CA: Human Factors and Ergonomics Society.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2008b). Selecting methods for the analysis of reliance on automation. In *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society* (pp.287-291). Santa Monica, CA: Human Factors and Ergonomics Society.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human Factors, 51*, 281-291.