2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)(pp.212-17 vol.1). Piscataway, NJ: IEEE. 7 vol. (lxxx+6496)pp.

The Impact of Context-related Reliability on Automation Failure Detection and Scanning Behaviour*

Nasrine Bagheri
Cognitive Engineering Laboratory
University of Toronto, Canada
nbagheri@mie.utoronto.ca

Greg A. Jamieson
Cognitive Engineering Laboratory
University of Toronto, Canada
jamieson@mie.utoronto.ca

Abstract - *This article presents an experiment that examined the effect of providing operators with information about the context-related nature of automation reliability. Results are compared against those from previous experiments where no information about automation reliability was provided to participants. Providing context led to a significant increase in participants' detection performance of automation failures in situations known to trigger poor detection performance (i.e., under constant and highly reliable automation). This improvement in performance seems to be the result of a more efficient attention allocation strategy which, in turn, appeared to be the consequence of participants' better understanding of the automation behaviour. Implications for future research on human-automation interaction and the development of human-centered automation are drawn.*

Keywords: Human-automation interaction, monitoring behaviour, context-related automation reliability.

1 Introduction

Several authors have noted that the role of the human operator has gradually changed from that of an active manual controller to a passive supervisor ([11], [12]). As supervisors of automated systems, operators are expected to monitor the automation, and to intervene and manually perform the task if the automation fails. Unfortunately, human operators are not well-suited to be passive monitors [8]. Indeed, a monitoring task can be tedious while the automation is functioning correctly, and overwhelming when the automation fails. Parasuraman and Riley [8] noted several factors that directly or indirectly affect automation use, noting the central importance of automation reliability. Several studies have confirmed that the automation reliability level affects joint human-automation failure detection performance (e.g., [9], [5], [6]).

Studies of human-automation interaction typically model automation failures as unpredictable events, neglecting their context-sensitive nature. That is, the conditions that might affect automation performance, and make automation failures more likely to happen are not considered [14]. In addition, participants are usually not informed of the automation reliability and are offered no rationale (i.e., no explanation) for the automation behaviour. However, automation reliability is rarely context-free [14], and field studies (e.g., [3]) have shown that operators are sometimes able to recognize certain phases of operation as being more or less likely to be associated with automation failure. This highlights the importance of considering more representative automation reliability models. The present study introduced context-related automation reliability and informed participants about the automation behaviour. The goal of this study was to investigate the effect of context-related automation reliability on participants' monitoring behaviour and on their trust in automation.

2 Method

In this paper, we compare the results of two studies of human detection of automation failure. In a previous study [1], participants received no information about the automation reliability and the external factors that may affect its behaviour. In a new study, participants were provided with information that made the automation behaviour quasi-predictable, and its reliability context-related. In both experiments the same apparatus and experimental protocol were used.

2.1 Participants and apparatus

Twenty-four participants were recruited and paid $40 to complete the study. Participants had no prior experience with the Multi-Attribute Task Battery (MAT; Comstock and Arnegard, 1992), which was used in the study. The MAT Battery is a multi-task flight simulation that requires participants to perform three equally important tasks: (1) tracking, (2) fuel management, and (3) system-monitoring. The goal of the tracking task was to keep the aircraft within a central rectangular area using a joystick (first-order control). The goal of the fuel management task was to compensate for fuel depletion by pumping fuel from the supply tanks to the main tanks. The system-monitoring task consisted of four engine gauges that participants had to monitor for occasional abnormal values that represented system malfunctions. The monitoring task was automated so that a gauge showing an abnormal value would normally reset itself without participant intervention. However, participants were advised that the automated system would sometimes fail to correct these malfunctions. In such a situation, participants were required to correct malfunctions manually. If they did not detect the automation failure within 10 seconds, the event was scored as a "miss" and the pointer was automatically reset. Participants were not informed that they missed a failure.

An Eye-gaze Response Interface Computer Aid (ERICA) system, using infrared light technology, was also used to track the eye movements of the participants. The equipment was non-invasive to participants. Gaze location samples were taken 30 times per second.

2.2 Procedure

Following a 10-minute training session, participants completed four 30-minute sessions on the MAT battery for a total of 12 10-minute blocks. At the end of each session, participants rated their trust in the automated system and their self-confidence in performing each task on a 10-point scale similar to the one used by Lee and Moray [5], [6].

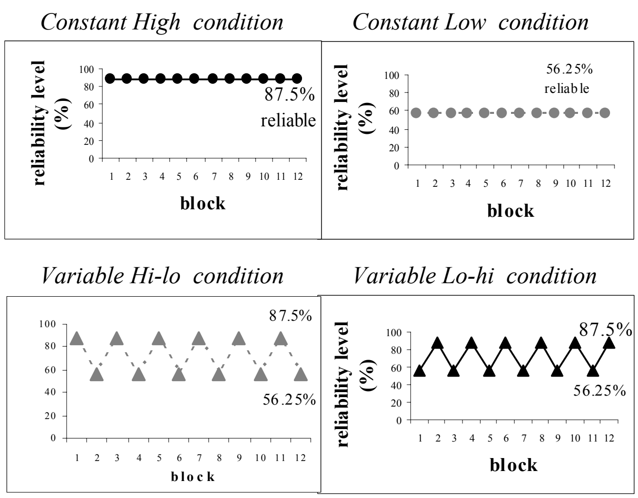

Figure 1. Graphical representation of the automation reliability condition.

A 4 (reliability) by 12 (blocks) mixed factorial design was used. Automation reliability was varied as a between-subjects factor with four levels (see Figure 1). There were 16 malfunctions in each 10-minute block. Automation reliability was defined as the percentage of malfunctions successfully corrected by the automation in each block. Six participants were randomly assigned to each of the four reliability conditions.

In the current study, before the experimental sessions began, participants received information about how the automation would behave during the experiment and a rationale to substantiate this behaviour. Participants in both constant reliability conditions were informed that the automation reliability would not vary. Constant High participants were informed that the reliability would be "slightly under 100%", while Constant Low participants were told that the automation reliability would be "slightly more than 50%". Participants in both reliability conditions were also told that the automation would detect the same percentage of automation failure in each 10-minute block. In the Variable Reliability conditions, participants were told that the reliability of the automated system would change from high ("slightly" under 100%") to low ("slightly more than 50%") every ten minutes. After 20 minutes, the automated system would undergo maintenance, thereby returning the reliability to a high level for ten minutes. The reliability would then decrease to a low level for another 10 minutes before the next scheduled maintenance. Time was displayed on the MAT Battery screen. Participants were provided with the above information in both textual and graphical forms. Before commencing with the experimental sessions, participants completed a comprehension test on key aspects of the reliability tutorial. As part of the test, they were also asked to reproduce graphically the evolution of the automation reliability through time. Feedback was provided by the experimenter until participants showed complete understanding of the training material.

In a previous study [1], participants were only told that the automation was fallible. In that study [1], participants' automation failure detection rate, performance on the tracking and fuel management task, visual scanning behaviour, and subjective ratings were measured. The same dependent variables were assessed in the current study.

3 Results

3.1 Task Performance

In the current study, the effect of automation reliability on participants' performance on the monitoring task, assessed by their detection rate of automation failures, was not significant $F(3, 20) = 2.13$, $p > .05$ (see Figure 2).

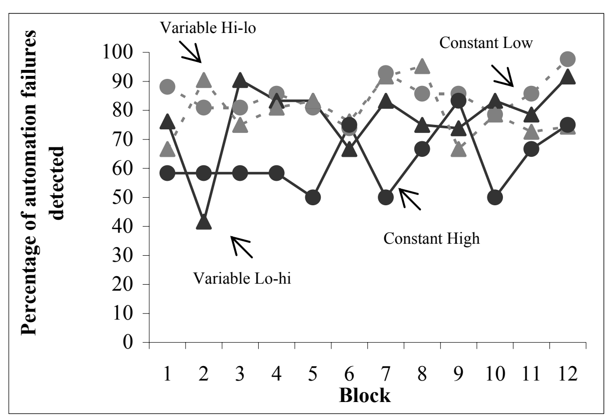

Figure 2. Detection of automation failures as a function of reliability and block

This null result is valuable when compared to previous studies which used the same simulation and reliability conditions, but provided no rationale to participants as to when and why the automation was most likely to fail. These studies have consistently found a significant effect of automation reliability on participants' detection performance (e.g., [1], [9], [13]). In particular, participants in one or both of the Constant reliability conditions exhibited poorer detection performance than those in the variable reliability conditions.

No significant differences were observed between reliability conditions for performance on the tracking, $F(3, 20) = 0.507$, $p > .05$, or fuel management tasks, $F(3, 20) = 0.747$, $p > .05$. This result is consistent with previous findings. Thus, providing information about the automation seemed to enhance participants' detection performance without affecting their performance on the other tasks.

3.2 Attention and sampling rate

Assessments of participants' eye point of gaze behaviour revealed a significant effect of automation reliability on the Mean Time Between Fixation (MTBF) of the automated monitoring task $F(3, 20) = 21.23$, $p < .001$ (see Figure 3). Post-hoc analysis revealed that Constant High participants had the highest MTBF.

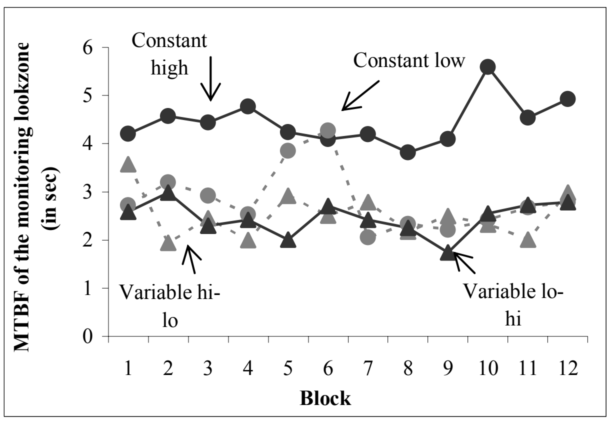

Figure 3. Mean Time Between Fixation of the Monitoring Lookzone per reliability and block for participants receiving information about the automation failure context.

More importantly, the attention allocation patterns of Constant High participants who were provided with information about the automation behaviour differed significantly from those who received no information (see Figure 4, which integrates data from the present study with data from Bagheri and Jamieson [1]). Indeed, the effect of context on participants' MTBF of the monitoring lookzone was found to be significant $F(1, 40) = 55.06$, $p < .001$, and post-hoc analysis revealed that the MTBF of the monitoring task was significantly smaller for Constant High participants who received context information. The difference, 4.45 sec on average for those provided with context information vs. 12.5 sec on average for those who were not, reflects a non-trivial effect. Further, Constant High participants provided with a context for failure demonstrated a relatively consistent monitoring behaviour across blocks (see Figure 3 and 4). In contrast, when Constant High participants did not receive any information about the automation reliability, their monitoring behaviour changed markedly across blocks, and it took them more time to reach a MTBF similar to that of other reliability conditions (see Figure 4). From these results, it appears that providing context led Constant High participants to develop a more skeptical monitoring behaviour right from the beginning. Constant High participants who were not provided with information about the automation behaviour, and who could thus only rely on their experience with the automation, eventually appeared to develop such a behaviour, but with some delay.

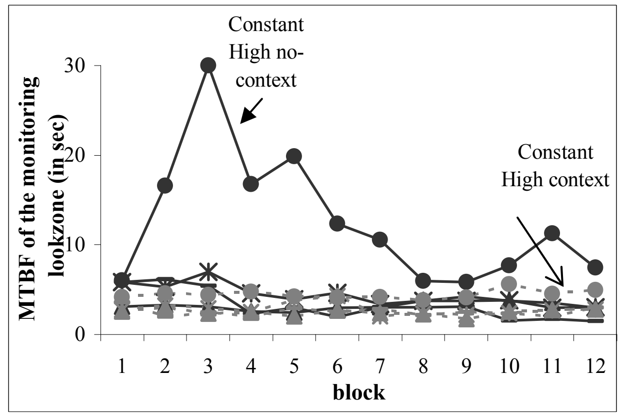

Figure 4. Mean Time Between Fixation of the Monitoring Lookzone per reliability and block for participants receiving (grey dashed lines) or not receiving (black lines) information about the automation failure context.

It should be noted that participants' detection rate of automation failures was found to be significantly correlated with the MTBF of the monitoring lookzone $r = -0.586$, $p < .001$, $n = 90$. The smaller the MTBF, the better the detection rate. Participants' detection rate was also found to be significantly correlated with the percentage of time allocated to the monitoring task $r = 0.512$, $p < .001$, $n = 90$. Thus, shorter MTBF and larger percentage of time allocated to the automation were both positively correlated with improved automation failure detection.

3.3 Trust in automation

At the end of each 30-minute session (i.e., three blocks), participants were asked to rate their trust in the automation on a 10-point scale. Participants' trust ratings did not significantly differ between reliability conditions $F(3, 20) = 0.766$, $p > .05$ (Figure 5). However, the difference in trust ratings between the two constant conditions shows a trend towards increasing across blocks, with Constant High participants trusting the automation the most, and Constant Low participants trusting it the least.

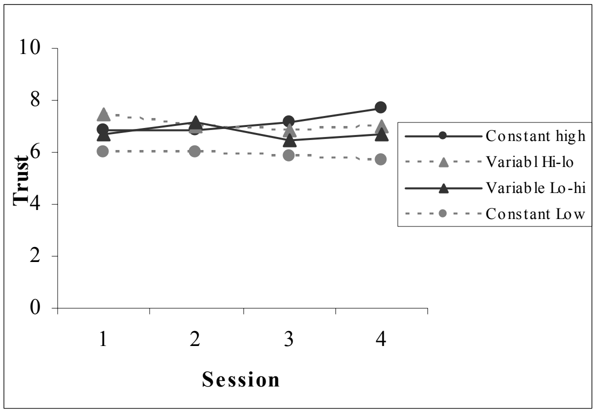

Figure 5. Trust in automation per reliability and session.

In the current study, trust was not significantly correlated to the MTBF of the monitoring lookzone (contrary to what was found in a study conducted by Bagheri and Jamieson [1]), but was significantly correlated with the fixation duration of the monitoring lookzone $r = -0.432$, $p < .0001$, $n = 90$. The more participants trusted the automation, the less they dwelt on the monitoring lookzone when they looked at it. Trust was also significantly correlated with the level of automation reliability $r = 0.342$, $p < .001$, $n = 96$.

4 Discussion

The primary goal of the study presented in this paper was to investigate the effect of providing participants with information about the context-related nature of automation reliability on their detection of automation failures, their sampling behaviour, and their trust in automation. To better understand this effect, participants' performance and behaviour were compared to those of participants in a previous study [1] where no reliability information was provided.

4.1 Effect of context on participants' performance and behaviour

Results revealed that, when participants were provided with some information about the automation behaviour, the automation reliability did not have a significant effect on their detection performance. In particular, detection performance of Constant High participants was significantly better when provided with the reliability context. This is an important result since previous studies, in which no context was provided, have found a significantly poorer performance for Constant High participants (e.g., [1], [9], [10], [13]).

This improvement in detection performance appears to be the result of a more effective attention allocation strategy. Constant High participants, when provided with context, allocated significantly more attention to the monitoring task (i.e., smaller MTBF and longer fixation). This change in attention allocation led to improved detection performance but did not affect performance on the tracking and fuel management tasks. No trade-off in performance seems to have occurred; rather, it appears that a more effective monitoring strategy was adopted.

4.2 A better understanding of the automation performance

4.2.1 The effect of context on attention

The information provided about the automation behaviour during the training session could explain the more effective monitoring behaviour of Constant High participants when provided with context. During this training session, it was first emphasized that the automation reliability would be high but still below 100%. In studies where no context was provided (e.g. [1], [9]), participants were just told orally that the automation would not be 100% reliable. In addition, in the current study, Constant High participants were told that the automation would detect the same percentage of failures in each block, which suggested that a failure could occur within the first block.

The emphasis placed on the fallibility of the automation in the current study could explain the difference in performance and attention strategy. Indeed, it is known that human operators prior to any interaction with automated aids tend to predict near perfect performance from these aids [2]. They tend to have high initial expectations about the automation behaviour. Providing information about the automation reliability might encourage more appropriate expectations of the automation behaviour, and thus lead to more effective monitoring. This is particularly important for Constant High participants, as automation failures were rare, unpredictable, and participants were not told if they missed a failure (i.e., there was no performance feedback). They could thus easily hold to their initial expectations of near-perfect performance of the automation for a long time. We could hypothesize that the effect of providing information about the automation had less consequence for the other participants, as they faced a higher number of automation failures, and could thus realize more quickly through experience, that the automation was fallible, and that the failures were unpredictable.

During the training session, Constant High participants were also told that the automation would detect the same percentage of automation failures in each ten minute block. This information might have allowed them to be vigilant right from the beginning as an automation failure could happen in the first block, and to remain vigilant through time. This might explain the relatively consistent evolution of their MTBF of the monitoring lookzone. It might also have allowed them to better understand the automation behaviour (i.e. to realize the unpredictability of the failures), thereby leading them to adopt a more effective monitoring behaviour.

In sum, providing context might have allowed participants' to better understand the automation behaviour, and this could explain the different attention allocation strategies adopted by participants provided with context. These changes in attention strategy appeared to be positive since they allowed for a more effective interaction with the automation of the MAT Battery.

4.2.2 The effect of context on trust

Analysis of participants' trust in automation also seems to indicate that participants who received information about the automation better understood its behaviour and adopted a more effective monitoring behaviour.

Trust has been conceptualized as a multi-dimensional concept based on the trustee characteristics. The basis for trust is the information that informs the person about the ability of the trustee to achieve the goals of the trustor [4]. Lee and Moray [6] defined '*performance*', '*process*' and '*purpose*' as the general bases of trust. Lee and See [4] considered these three bases of trust as interdependent and evolving depending on information availability and operators' experience with the automation. Any one of these bases might be prominent at any time depending on the information available and the operator's experience. *Performance* refers to the consistency and desirability of past behaviour. Performance information refers to *what* the automation does. *Process* depends on the understanding of the underlying qualities and characteristics that govern behaviour [6]. Process information describes *how* the automation operates. With process, trust is in the agent not in the specific actions of the agent. The process basis of trust relies in part on inferences drawn from performance. *Purpos*e describes *why* the automation was developed. It represents the designer's intent regarding when the automation should and should not be used.

When participants received no information about the automation reliability [1], their detection rate was found to be significantly correlated with their trust in automation. Since participants had no prior experience with the automation, and no information was available about the automation behaviour and competence, their trust might have depended primarily on the observed performance and actions of the automation. Detecting an automation failure is equivalent to observing an undesirable performance of the automation which could explain the significant correlation between participants' trust in automation and their detection rate. In other words, 'performance' [6] might have been the prominent basis of trust in the study conducted by Bagheri and Jamieson [1].

In addition, as participants in the study conducted by Bagheri and Jamieson [1] received no feedback when they missed a failure, their detection rate might have represented in some way their perception of the automation reliability. This could also explain the significant correlation between participants' trust and their detection rate.

Note however that the causal relationship between participants' detection rate and their trust in automation in the study conducted by Bagheri and Jamieson [1] cannot be guaranteed. It could be that trust led to a more skeptical monitoring behaviour and this could explain the relationship between trust and detection rate.

In the current study, trust was not significantly correlated with participants' detection rate. It could be hypothesized that the information provided to participants led to a change in the basis of trust involved. Indeed, Lee and See [4] emphasized that trust in automation would evolve according to the information available about the automation behaviour and competence.

In the current experiment, since participants were informed about the automation reliability and about the factors that may affect it, their focus may have changed from the assessment of the specific behaviour of the automation to 'an assessment of the stable dispositions that guided its behaviour' [7]. That is, 'performance' as a basis of trust might have become less important to the benefit of the 'process' dimension, and this could explain the non-significant correlation between detection rate and trust. The information provided to participants about the automation competence might have allowed them to have a more global apprehension of the automation competence. As a result, trust was less affected by the occurrence of an automation failure, but was influenced by the automation characteristics, which could explain the causal relationship between the level of automation reliability and participants' trust.

Results of trust measures also confirmed that providing information about the automation behaviour led participants to have more appropriate expectations of the automation. For instance, Constant High participants who knew they were facing an automated system that was highly reliable trusted the automation more than Constant Low participants, and this difference tended to increase across sessions. In other words, participants who had been cued to the differential reliability tended to trust the automation more when it was known to be reliable than when it was not. This effect tended to increase across sessions. This appears to confirm Muir's hypothesis [7] that educating operators about the automation competence (e.g., level of reliability) may lead to more accurate expectations of automation performance and thus to a more appropriate level of trust. Results of the present study indeed suggested a better 'resolution' [4] of participants' trust which tended to increase with time. This was supported by the significant correlation that was found between participants' trust and the level of automation reliability.

Finally, the relationship between participants' trust in automation and their monitoring behaviour suggests that participants in the current study adopted a more appropriate behaviour given the dynamic of the automated task than participants in the study conducted by Bagheri and Jamieson [1]. For instance, when participants were provided with context, trust was not significantly related to the sampling frequency of the monitoring lookzone (contrary to the results of Bagheri and Jamieson [1]), but was significantly correlated with the average fixation of the monitoring lookzone. Participants who trusted the automation might not necessarily sample it less frequently, but would dwell on the task less than other participants. Participants knew the automation was fallible and that it could fail anytime. They might thus have realized that even when the reliability of the automation was high, the probability of a failure was low but not null. This might have encouraged them to frequently make quick checks to ensure that no failure was occurring even though they trusted the automation (thereby explaining the non-significant correlation between the MTBF of the monitoring task and participants' trust in automation). This change in the relation between participants' trust and their attention allocation strategy could potentially be the result of participants' better understanding of automation performance. Indeed, the automated task presented no local context that would allow participants to suspect that a failure would happen; the pointers just abruptly went out of range. The unpredictability of automation failures thus requires participants to frequently sample the automation (no matter what the automation reliability was) to detect a failure.

5 Conclusion

We expect well designed automated systems to be both highly reliable and consistent. However, several laboratory studies have shown that human monitoring of such automation tends to be poor compared to automation that is inconsistently reliable. Our observations suggest that providing operators with information about the context-sensitive nature of automation reliability might alleviate this performance decrement. The improvement appears to be associated with changes in attention allocation brought about by the additional information. This observation is important for two reasons. First, it demonstrates that more representative automation reliability models are called for in human-automation interaction research since they lead to different performance and behaviour. Second, it suggests a means by which system designers might foster improved human-automation interaction.

References

[1] Bagheri, N. & Jamieson, G. A. (2004). Considering subjective trust and monitoring behaviour in assessing automation-induced "complacency". In *Proceeding of the Human Performance, Situation Awareness and Automation Technology II Conference* (pp. 54-59), Daytona Beach, FL.

[2] Dzindolet, M., Pierce, L., Beck, H. & Dawe, L. (1999). Misuse and disuse of automated aids. In *Proceedings of the Human Factors Society 43rd Annual Meeting* (pp. 339-343). Santa Monica, CA: Human Factors and Ergonomics Society.

[3] Jamieson, G. A., & Guerlain, S. (2000). Operator Interaction With Model-based Predictive Controllers in Petrochemical Refining. In *Proceedings of the Human Performance, Situation Awareness and Automation Conference* (pp. 172-177). Marietta, GA: SA Technologies.

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

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

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

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

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

[9] Parasuraman, R., Molloy, R. & Singh I. L. (1993). Performance consequences of automation-induced "complacency". *The International Journal of Aviation Psychology, 3*, 1-23.

[10] Prinzel L., DeVries, H., Freeman, F., & Mikulka P. (2001). *Examination of Automation-Induced Complacency and Individual Difference Variates*. Tech. Memorandum No. 211413. Hampton, VA: NASA Langley Research Center.

[11] Rasmussen, J., Rouse, W.B. (Eds) (1981). *Human detection and diagnosis of system failure*. New York: Plenum Press.

[12] Sheridan, T. (1997). Supervisory control. In G. Salvendy (Eds). *Handbook of Human Factors* (2nd Ed., pp.1295-1327). New York: Wiley.

[13] Singh, I., Molloy, R. & Parasuraman, R. (1997). Automation-induced monitoring inefficiency: role of display location. The *International Journal of Human-Computer Studies, 46*, 17-30.

[14] St John, M., Manes, D. & Osga, G. (2002). A "trust-but-verify" design for course of action displays. *Command and Control Technology and Research Symposium*, Monterey, CA: Command and Control Research Program.