In Proceedings of the Human Performance, Situation Awareness and Automation Conference (pp.xx-xx). Marietta, GA: SA Technologies, 2004.

CONSIDERING SUBJECTIVE TRUST AND MONITORING BEHAVIOR IN ASSESSING AUTOMATION-INDUCED "COMPLACENCY"

Nasrine Bagheri

Greg A. Jamieson

Cognitive Engineering Laboratory, University of Toronto, Ontario, Canada

ABSTRACT
This article presents a study of human monitoring of an automated system considering both sampling behavior and subjective reports of trust and self-confidence. It replicates an experiment by Parasuraman et al. (1993) wherein participants were said to be lulled into complacency by unchanging automation. Results confirmed their finding that automation reliability had a significant effect on automation failures detection. In particular, participants using constant highly reliable automation had the poorest failure detection performance. These participants were also observed to have significantly longer intervals between their samples of the monitoring task. Despite this mean difference, the evolution of the attention allocation patterns for participants in the constant-high condition does not support the attribution of their poor performance to complacency. Results of trust and self-confidence ratings are also discussed. These results provide empirical support for Moray's (2000, 2003) assertion that attention allocation and psychological factors should be considered when evaluating monitoring performance and drawing conclusions about complacency.

Keywords : complacency, sampling rate, trust, monitoring of automated systems.

INTRODUCTION
Difficulties with human-automation interaction in complex systems frequently prevent the full benefits of the automation from being realized. This article focuses on one adverse consequence of automation known as *automation-induced complacency* (Parasuraman et al., 1993). Complacency occurs when the role of the human operator is changed from that of an active manual controller to that of a passive monitor of highly reliable automation, and refers to the ensuing decline of that monitoring performance (Farrell & Lewandowsky, 2000). Although researchers generally agree that complacency is a serious problem, little consensus exists to what complacency is and how it can be measured (Prinzel et al., 2001). Previous research has concluded that operators were complacent based primarily on their automation failure detection performance over time (e.g., Parasuraman et al., 1993). Moray (2000, 2003) questioned whether such evidence adequately supports the existence of complacency. He pointed out that (1) complacency is concerned with attention, and (2) that psychological factors such as trust may influence complacency. Missing signals does not necessarily imply complacency as even optimal sampling behaviour can result in missed signals. Rather, complacency may imply under-sampling and defective monitoring (Moray & Inagaki, 2000).

This paper presents a replication of a study conducted by Parasuraman et al. (1993) in which participants who interacted with a consistent and highly reliable automated system were said to show signs of complacency based on detection performance. Participant eye movements, their trust in the automated system, and their self-confidence were evaluated in addition to detection.

METHOD

The experiment was designed to replicate as accurately as possible Parasuraman et al. (1993).

Participants and apparatus

Based on a power analysis of the data obtained by Parasuraman et al. (1993) 24 participants were recruited. Participants had no prior experience with the simulation used in the study.

The Multi-Attribute Task battery (MAT; Comstock and Arnegard, 1992) was used. 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 randomly occurring 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 was also used to track the eye movements of the participants. Gaze location samples were taken 30 times per second.

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 tasks on a 10-point scale similar to the one used by Lee and Moray (1992, 1994).

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.

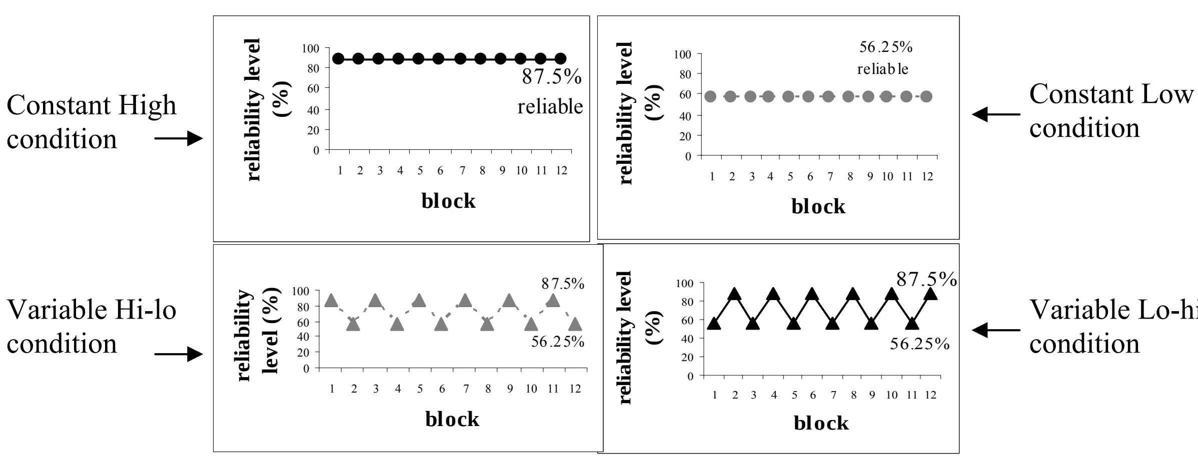

Figure 1. Graphical representation of the automation reliability conditions.

In Proceedings of the Human Performance, Situation Awareness and Automation Conference (pp.xx-xx). Marietta, GA: SA Technologies, 2004.

RESULTS

Detection rate of automation failures.

As in Parasuraman et al. (1993), a 4 (reliability) x 12 (Block) ANOVA of the detection rate indicated a significant effect for automation reliability $F(3, 20) = 11.92$, $p < .001$ (see Figure 2). Post-hoc analysis revealed that the detection performance of the Constant High participants was poorer than that in any other condition. This result differs from that of Parasuraman et al. (1993), who found no significant difference in detection performance between the Constant High and the Constant Low condition. The difference observed in the present study precludes nesting the two reliability groups. As in Parasuraman et al. (1993), the block effect on detection performance was significant $F(11, 220) = 2.23$, $p < .05$. The interaction was not significant $F(33, 220) = 1.32$, $p > .05$.

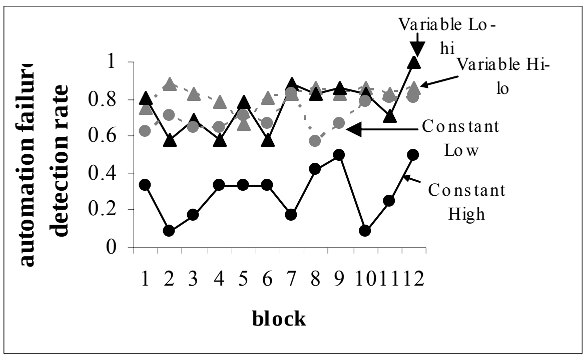

Figure 2A. Effect of automation reliability and block on participants' detection performance in the current study.

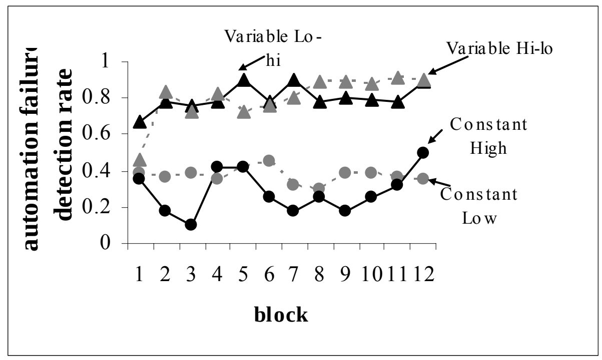

Figure 2B. Effect of automation reliability and blocks on participants' detection performance reconstructed from Parasuraman et al.(1993).

To determine the effect of automation reliability based on equal failure rates, performance of Constant High participants was compared to the performance of those in the variable conditions for the blocks where the reliability was also high. That is, performance in the Constant High condition was compared to the performance in block 1 of the Variable Hi-lo condition and block 2 of the Variable Lo-hi condition, etc. When faced with highly reliable automation, participants in the variable conditions performed significantly better $F(1, 10) = 21.89$, $p < .001$ (Figure 3A). Conversely, for low reliability blocks and conditions, results revealed that whether the reliability was constant or variable did not significantly affect participants' detection rate $F(1,10) = 1.706$, $p > .05$, although Constant Low participants performed poorer in 11 of 12 blocks (see Figure 3B).

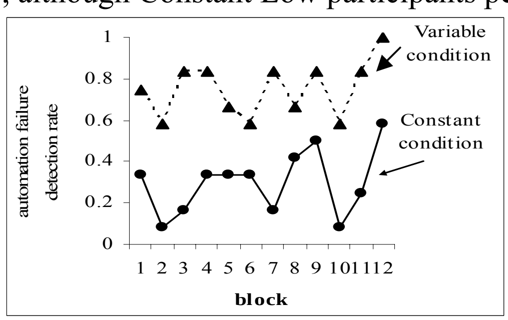

Figure 3A: Comparison of the variable and constant conditions for the high reliability level

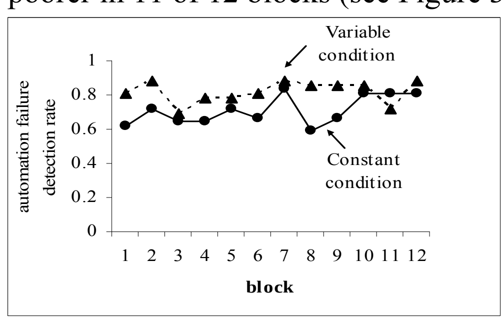

Figure 3B: Comparison of the variable and constant conditions for the low reliability level

There was no significant effect of group difference on either tracking performance $F(3, 20) = 1.27$, $p > .05$ or resource management performance $F(3, 20) = 0.42$, $p > .05$.

Attention and sampling rate.

Parasuraman et al., by informal video observation, did not find any systematic difference in scanning behavior between participants in their constant and variable conditions. In the present study, participants' eye movements were recorded to determine how attention was allocated to the three tasks. The Mean Time Between Fixation (MTBF) for the three lookzones of the MAT battery was measured. The effect of reliability on the log-transformed MTBF of the monitoring task was significant $F(3, 20) = 34.60$, $p <.0001$ (Figure 4), and so was the block effect $F(11, 121) = 2.06$, $p < .05$. The MTBF was transformed to compensate for the skewed variable distribution. The interaction effect was non-significant. Post-hoc analysis further showed that the MTBF of the monitoring lookzone was higher for Constant High participants than for participants in any other condition. Figure 4 shows that the MTBF of Constant High participants gradually increased in the first 3 blocks, but then decreased and converged toward the MTBF of participants in the other three conditions. The detection rate was negatively correlated with the MTBF, $r = -0.57$, n = 189, $p <.01$.

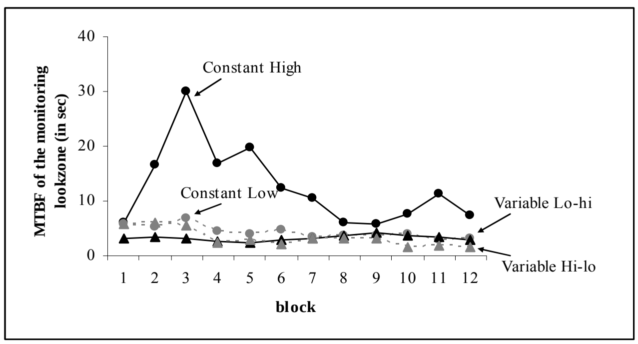

Figure 4: MTBF of the monitoring lookzone

Trust in automation.

Parasuraman et al. (1993) suggested that 'waxing and waning of trust' with the success and failure of the automation could account for part of their detection results. However, the authors did not report any trust measures. In the current study, no significant effect of automation reliability on participants' rating of trust was found $F(3, 20) = 1.19$, $p >.05$ (Figure 5). However, the low power of the test ($1-\beta = 0.3$) should be noted. The block effect was non-significant $F(3,20) = 0.298$, $p>.05$. Correlation analysis revealed that detection rate was inversely correlated with the level of trust (i.e., the more participants trusted the automation, the lower their detection rate), $r =-0.39$, p <.01. Trust was also positively correlated with the MTBF of the monitoring lookzone $r = 0.34$, $p <.01$.

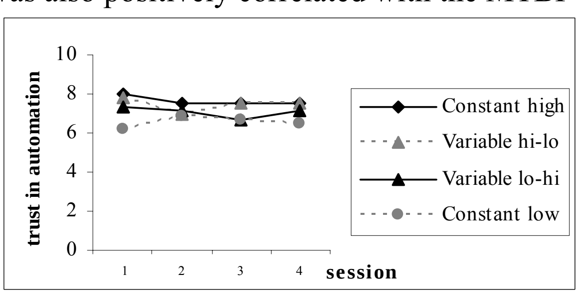

Figure 5. Rating of trust in the automation.

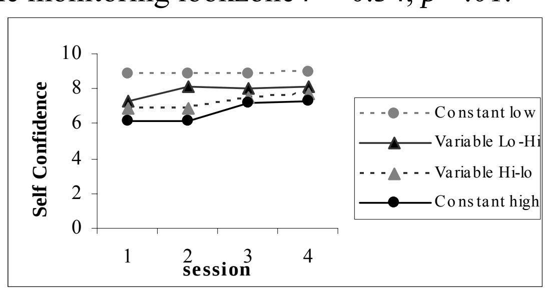

Figure 6. Rating of self-confidence

Self-confidence in performing the monitoring task.

The effect of automation reliability on participants' self-confidence approached significance $F(3, 20) = 2.883$, $p = 0.06$. Post-hoc analysis revealed that the difference between the Constant Low and the Constant High condition approached significance. Constant High participants had the lowest self-confidence in their ability to perform the monitoring task (Figure 6).

DISCUSSION

On the surface, our results replicate much of those found by Parasuraman et al. (1993). The detection of automation failures was significantly worse for participants facing constant, highly reliable automation, which could indicate that these participants showed signs of complacency. Following Parasuraman et al. (1993), several explanations for the observed poor performance should be considered. First, the poor detection performance of Constant High participants could be related to the 'signal rate'. As they faced a low probability of signal occurrence, we might expect their probability of detecting a failure to be lower (Parasuraman, 1986). However, in blocks with equivalent failure rates (i.e., signal rate), we still observed poorer performance for Constant High participants compared to participants in the Variable conditions. Thus, like Parasuraman et al. (1993) we conclude that low signal rates alone do not explain the poor detection performance.

Secondly, Parasuraman et al. (1993) suggested that differences in attention allocation could explain the observed difference. Using informal observations of participants' eye movements, they observed no major differences in scanning behavior between participants in the constant and the variable conditions, although they did not rule out the possibility of small differences. In the present study, eye point of gaze data revealed that Constant High participants had a significantly higher MTBF of the monitoring lookzone. More importantly, the difference in the MTBF between the Constant High condition and the other conditions increased in the first 3 blocks, but then decreased starting in Block 4. This decrease argues against the hypothesis that complacency appeared after a long period in presence of highly reliable automation. This change in attention allocation strategy could not be observed from detection results, which shows the importance of measuring attention in order to accurately evaluate monitoring performance (Moray, 2000, 2003).

Analysis of participants' subjective ratings of trust also forestalls the conclusion that Constant High participants were complacent. Self ratings of trust in the automation revealed no differences between the reliability conditions or across blocks, indicating that poor monitoring performance might not reflect overtrust. This is not to say that trust is not an important factor in monitoring. To the contrary, trust was shown to have a moderate-to-large effect on both monitoring behaviour and detection performance. No trust data were collected by Parasuraman et al. (1993), although the authors cited the 'waxing and waning of trust' as a possible factor in explaining their observations.

Similarly, Constant High participants had the least confidence in their ability to detect failures. Participants with lower self-confidence in their monitoring skills could be expected to be poorer monitors than those with high self-confidence while interacting with a constant highly reliable automation (Prinzel, Pope, and Freeman, 1999). However, it should be noted that Constant High participants were presented with few failures, and did not know if they missed one. Their lower self-confidence might thus be due to a belief that they were missing some signals as they knew that the automation was not 100% reliable. Low self confidence may explain why the sampling patterns of Constant High participants converged towards the level of the other conditions.

The most perplexing observation involves the failure detection performance of Constant Low participants. Parasuraman et al. (1993) observed little difference in detection performance between the Constant Low and the Constant High condition. In contrast, detection performance of Constant Low participants in the present study differed significantly from that of Constant High participants, and was similar to that of participants in the Variable conditions. In the absence of the Parasuraman et al. (1993) results, this observation might indicate that the reliability level was low enough to offset complacency induced by the constant-reliability environment, if complacency there was. Attention data would corroborate this conclusion since the MTBF of the monitoring lookzone of the Constant Low condition was not significantly different from that of the Variable conditions. However, the strong contrast between this observation and that reported by Parasuraman et al. (1993) more likely suggests a discrepancy between the study protocols. All efforts were made to replicate the study as described in the literature, but some details were not readily available.

CONCLUSION

We believe that this study is the first to look at automation-induced complacency based on both sampling behavior and subjective reports of trust and self-confidence. Detection rate results alone might indicate that participants using constant high reliability automation showed signs of complacency. However, assessments of attention allocation, trust, and self-confidence appear to contradict this conclusion. Thus, Moray's (2000, 2003) assertion that investigators must consider allocation of attention and psychological factors when evaluating monitoring performance and drawing conclusions about complacency gains credence from these results. More generalizable conclusions will require that these results be compared against an optimal sampling rate.

ACKNOWLEDGEMENT

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. Thank you to Catherine Burns at the University of Waterloo for the loan of the ERICA system.

REFERENCES

Farrell, S. & Lewandowsky, S. (2000). A connectionnist model of complacency and adaptive recovery under automation. *Journal of Experimental Psychology-Learning Memory and Cognition*, *26*, 395-410.

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

Lee and Moray (1994). Trust, self-confidence, and operators' adaptation to automation. *Human Computer Studies, 40,* 153-184.

Moray, N. (2003). Monitoring, complacency, scepticism and eutectic behaviour. *International Journal of Industrial Ergonomics, 31,* 175-178.

Moray, N., T. Inagaki (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science, 1,* 354-365.

Parasuraman, R. (1986). Vigilance, monitoring and search. In Boff, K., Kaufman, L. & Thomas, J.(Eds), *Handbook of perception and human performance: Vol.2 Cognitive processes and performance* (43.1-43.39).

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

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

Prinzel L., Pope, A. & Freeman, F. (2000). The double-edged sword of self-efficacy: Implications in automation-induced complacency. In *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (124-128). San Diego, 2000.