In Proceedings of the 22nd European Annual Conference on Human Decision Making and Control (pp. 41-47). Linköping, Sweden: Cognitive Systems Engineering Lab.

Context-Related Reliability and Automation Failure Detection

N. Bagheri, S. Mohite, D. Junta, G. A. Jamieson

Cognitive Engineering Laboratory
Department of Mechanical and Industrial Engineering, University of Toronto
5 King's College Road, Toronto, Ontario, M5S 3G8, Canada
E-mail: jamieson@mie.utoronto.ca

Abstract
This paper addresses how automation failures have been introduced in studies of human reliance on automation. Failures of automation have mostly been treated as random events, neglecting the context-sensitive nature of reliability. We argue that a more realistic model is needed if the interaction between humans and automation is to be studied effectively, with results applicable to the real-world. A pilot study investigates the effects of context-related automation reliability on human performance. Results showed a trend toward superior performance for participants in the quasi-predictable condition. The limits of this exploratory study, and its implications for future research are discussed.

Keywords: human-automation interaction, monitoring behaviour, context-related failure

1 Introduction

Several authors have noted that the role of the human operator has gradually changed from an active manual controller to a passive supervisor (Rasmussen & Rouse, 1981; Sheridan, 1997). As supervisors of automated systems, operators are expected to monitor the automation, and to intervene and perform the task manually if the automation fails. Unfortunately, human operators are not well-suited to be passive monitors (Sheridan, 2002; Parasuraman & Riley, 1997). Indeed, a monitoring task can be tedious while the automation is functioning correctly, and overwhelming when the automation fails. Parasuraman and Riley (1997) noted several factors that directly or indirectly affect automation use, noting the central importance of automation reliability. Several studies have confirmed that automation reliability level does affect joint human-automation failure detection performance (e.g., Parasuraman et al, 1993; Lee & Moray, 1992, 1994).

1.1 Models of Automation Failure

A key difference between existing studies of automation reliability lies in the way that failures have been modeled. This goes beyond the characteristics of the failures themselves (see Bisantz & Seong, 2001) and includes the nature of information given to participants about the failures. A key feature of the failure characteristics would be the cause of automation failure and whether the failure pattern is predictable to some extent. Key characteristics of the information provided to participants would include whether or not they were a) provided with a rationale for automation failure (e.g., Dzindolet et al., 2002), and b) informed about automation reliability (e.g. St. John et al., 2002). Combining the dimensions of predictability (random or predictable), rationale (given or not given), and reliability information (provided or not provided) allows us to differentiate eight experimental models of automation failure (see Table 1).

Table 1: Eight models of automation failure observed in the empirical literature.

	Criterion 1: Random or context-related (predictable) failure?	**Criterion 2**: Operators given rationale for failure?	**Criterion 3**: Operators informed about automation reliability?	Covered by existing literature?
Model A	Random	No Rationale	Uninformed	Yes
Model B	Random	No Rationale	Informed	Yes
Model C	Random	Rationale	Uninformed	Yes
Model D	Random	Rationale	Informed	No
Model E	Context-Related	No Rationale	Uninformed	Yes
Model F	Context-Related	No Rationale	Informed	Yes
Model G	Context-Related	Rationale	Uninformed	No
Model H	Context-Related	Rationale	Informed	No

In the existing literature, automation failure has most frequently been treated as a random event (e.g., Parasuraman et al, 1993; Lee & Moray, 1992, 1994; Wickens et al., 1999; Bisantz & Seong, 2001; Dzindolet et al., 2002; St. John et al., 2002). However, these studies vary along the two other criteria, i.e., whether a rationale was provided and whether feedback was provided. For example, Wickens et al. (1999), in the first of two experiments, had participants perform simulated air-to-ground targeting missions where automated cueing was available. Unbeknownst to the participants, when the cue was present, it had a reliability of 60%. Although the automated cue would fail randomly, no rationale for failure was provided. This experimental design corresponds to model A in Table 1. Dzindolet et al. (2002) provided participants with a rationale for failures of an automated aid that could be used to perform a visual search task. Participants were unaware of the automation reliability. For this study, the rationale provided for automation behaviour was context-related (i.e. linked to the conditions that could affect automation performance), but the actual performance of the automation was not. Thus, the model employed by Dzindolet corresponds to model B. In contrast, St. John & Manes (2002) used an interface design to communicate different levels of automation reliability to participants. The different levels of the automated aid accuracy were not context-sensitive and no rationale was provided to justify why automation performance would be more reliable in some cases. This study thus illustrates model C in Table 1. We have been unable to locate any studies employing model D.

Although numerous studies employ random failures of automation, studies employing context-related failures appear to be rare. Madhani et al. (2002) created four error models for an automated visual inspection system: probabilistic occurrence, periodic occurrence, occurrences predominantly in the start-up phase, and occurrences predominantly in the shutdown phase. Results of this experiment did not show any significant relationship between error pattern and the participants' reliance on and trust in automation. However, participants were not informed of the automation reliability and were offered no rationale for automation failures (model E in Table 1), perhaps explaining the absent effect. In contrast, St. John et al. (2000) explicitly told their participants about the automation reliability, as well as when and how failure would occur. While the automation failure was thus predictable, no rationale was given to explain it (model F in Table 1). Participants tended to use the automation when it was known to be reliable and then switched to manual control when the automatic mode was known to be unreliable. To the knowledge of the authors, this study is unique in employing context-related automation failures where participants are informed about the predictability of these failures. In the literature reviewed, no experiments were found to cover model G or model H. While these two models may be difficult to distinguish in empirical practice, experiments covering both are essential to complete the literature on the impact of automation failure on operator reliance on automation.

1.2 Context-related Failures of Automation

In the real world, automation failures are rarely context-free (St John et al., 2002), and they can often be predicted to some degree. Automation is always designed to work within a defined operating region and tends to be brittle outside of that region (Guerlain et al., 1996). Engineers are often aware of the automation limitations and they may know when automation failures are more likely to occur and why. For example, the expected failure rate of some machines is related to the maintenance cycle. Within this cycle, the reliability of the automation may vary according to a bathtub shaped curve (Jardine, 1973). Knowing where the machine is in the maintenance cycle can give operators an idea of the level of reliability of the machine. It has been observed in actual work settings that operators associate certain contexts with a level of performance of the automation, and adjust their behaviour accordingly (Cohen, 2000; Jamieson & Guerlain, 2001).

To summarize, in order for laboratory experiments to be relevant for industry, the context-sensitive nature of automation reliability must be considered. Models of automation failure that are more realistic may lead to different patterns of performance. Indeed, when automation is unpredictable, operators cannot be warned about when to pay closer attention and must therefore be vigilant at all times. In contrast, when automation is quasi-predictable, operators could more efficiently manage their attention, but only if the nature of reliability and failure contexts were known. The pilot study presented in this paper is the first empirical study of the impact of context-related automation reliability on human performance wherein operators were provided with reliability information and a rationale for failure (i.e., model H in Table 1).

2. Methodology

In an effort to follow Parasuraman et al. (1993), the Multi-Attribute Task (MAT) Battery was used to simulate a multi-task environment. Participants were required to perform three tasks (tracking, monitoring, and fuel management) simultaneously. The monitoring task was automated, and the automation failed occasionally. Participants were placed in one of two groups; the failure model was random for one group, and quasi-predictable (or context-related) for the other group. Participants in the 'random condition' were told that the automation was not perfectly reliable, but they were not given any information about the reliability levels, and no rationale for failure was provided (model A in Table 1). In contrast, participants in the 'quasi-predictable condition' were given a rationale and reliability information that could be put into context to form a general idea about the changes in automation performance over time. They were told that the system undergoes maintenance every 30 days; after the scheduled maintenance, the reliability of the automation decreases with time. For the purpose of this experiment, the 30-day maintenance cycle was condensed into 30 minutes. During a training phase given to all participants, those in the quasi-predictable condition only were also provided with a graphical representation of how the relative reliability would decrease with time. Thus, failures had a somewhat predictable pattern of which participants were aware, and a rationale for failure was given (model H in Table 1). Fourteen subjects (7 male, 7 female) participated in the experiment. A two-factor mixed design was employed, with type of failure (random or context-related failure) as a between subject factor, and the blocks (three block of ten minutes) as the within-subject factor.

3. Results

3.1 Monitoring Task

Analysis of variance of the percentage of automation failures detected in each reliability condition revealed no significant main effect of context on performance. Participants in the 'random' condition detected an average of 76.6% of the total number of automation failures, while the participants in the quasi-predictable group detected 80.2%. The block effect and interaction effect were both non-significant. Figure 1 shows the joint human/automation probability of detecting automation failure in each experimental condition across blocks.

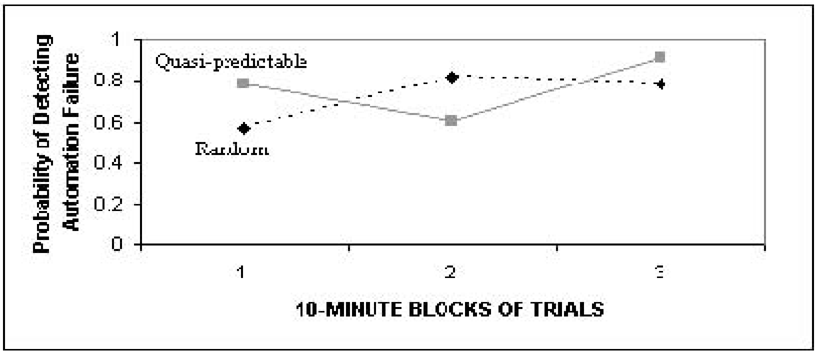

Figure 1: Detection in the 'random' and 'quasi-predictable' conditions across blocks.

The Levene statistic was used to test the homogeneity of variance between the 'quasi-predictable' and 'random' conditions. There was less variation in participants' performance in the 'quasi-predictable' condition (standard deviation = 14.6%) than in the random condition (standard deviation = 20.9%). Although the difference in overall variance between the two conditions was not significant, the variability in performance in the third block was found to be significantly smaller in the quasi-predictable condition, $F(1,12) = 6.611$, $p<0.05$. Moreover, the variability in performance (Figure 2) in the 'quasi-predictable' condition showed a significant decrease across blocks, $F(2, 18)= 3.74$, $p<0.05$. Crossman (1958) indicated that skill acquisition can lead not only to changes in mean performance, but also to changes in performance variability. Thus, the decrease in variability may indicate that in the 'quasi-predictable' condition, participants are progressively able to understand and predict the automation behaviour, and thus learn how to better interact with the imperfect automation.

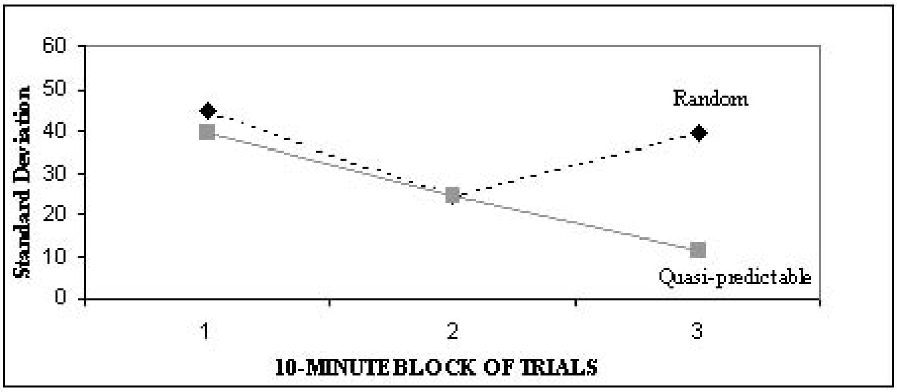

Figure 2: Standard deviation in detection performance (percentage of failures detected) between the two context conditions and across blocks.

The reaction time of participants for the monitoring task was also examined (see Figure 3). No significant difference was found between reaction times in the quasi-predictable and random conditions. However, the mean values indicate that participants' reaction times vary across blocks. In the 'quasi-predictable' condition, participants' reaction times improved across all blocks, which is consistent with the observation made earlier that in the 'quasi-predictable' condition participants are better able to understand how the automation is behaving and interact with it more effectively.

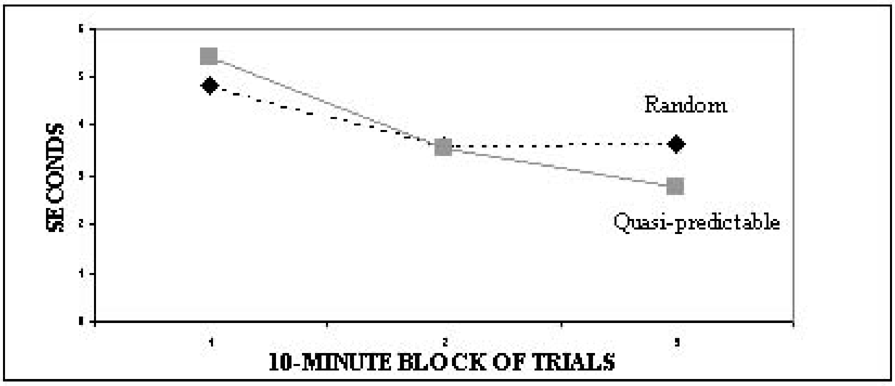

Figure 3: Mean reaction times for both conditions across blocks

3.2 Tracking and Fuel Management Tasks

The participants in the quasi-predictable condition had superior performance in the tracking task over all three blocks, but this difference between the two conditions was found to be non-significant. The main effect of block was also found to be non-significant, and so was the interaction effect. The same results were found in the fuel management task. Participants in the quasi-predictable reliability condition performed consistently better than participants in the random condition. Analysis of variance of the mean RMS error value revealed that neither the context nor blocks main effects on performance in the fuel management task were significant.

4. Discussion

These preliminary results show a modest trend towards superior operator multi-task performance under conditions where automation failure is quasi-predictable. Although there were no significant differences in detection performance between the random-failure and quasi-predictable failure groups, the variability in participants' detection rates and reaction times decreased over time when automation failures were context-related and when participants were made so aware. This pattern in performance was not observed when automation failures were random. These results suggest that, under appropriate conditions, informed participants may be able to cope more effectively with imperfect automation. Moreover, participants in the 'quasi-predictable' condition performed better on the fuel management and tracking tasks. Although these differences were also not statistically significant, they contribute to a trend towards superior task performance. This trend may be an indication that participants are better able to allocate their attention to the different tasks when information is provided about the automation behaviour. However, in order to evaluate this explanation, a means of assessing the participants' locus of attention is required (Moray & Inagaki, 2000). Future studies will thus monitor participants sampling behaviour across the tasks to determine the effect of context on their attention allocation strategies.

Results also revealed that the observed power for the significance test of the main effects and interactions was very low. The power of future studies could be improved by increasing either the sample size or the effect size. The latter might be possible if context is presented in a way that allows participants to better use the information provided concerning automation behaviour. In the experiment presented above, the reliability curve was shown to participants in the 'quasi-predictable' condition on paper before the experiment started. The curve was not provided during the trials and no clear indication of time was given. Debriefings revealed that some participants could not take full advantage of the context provided.

5. Conclusion

The goal of this exploratory study was to introduce context-related automation failures into a multi-task environment. Context-related reliability captures an important aspect of the modeled world that has not been previously considered and this concept will be expanded upon in future studies. This study was the first to a) introduce automation failures as quasi-predictable, b) provide a rationale for the failure, and c) inform the participants of relative reliability. Thus, it covers automation failure model H as presented in Table 1. This shift in perspective is a key development in making human-automation interaction research more relevant to industry.

6. Acknowledgements

Research funds provided by the Natural Science and Engineering Research Council

7. References

Bisantz, A., Seong, Y., J. (2001). Assessment of operator trust in and utilization of automated decision aids under different framing conditions. *International Journal of Industrial Ergonomics*, *28*, p. 85-97.

Cohen, M.S. (2000). A Situation Specific Model of Trust in Decision Aids, Cognitive Technologies, Inc, *Proceedings of the Human Performance, Situation Awareness and Automation: User-centered Design for the New Millennium*, Savannah, Georgia.

Crossman, E.R.F.W. (1958). A theory of the acquisition of speed-skill. *Ergonomics*, *2*, p.153-166.

Dzindolet, M., Pierce, L., Peterson, S., Purcell, L., & Beck, H. (2002). The influence of feedback on automation use, misuse, and disuse. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, Santa Monica, CA: HFES, p.551-555.

Guerlain, S., Smith, P., Obradovich, J., Rudmann, S., Strohm, P., Smith, J.,& Svirbely, J. (1996). Dealing with brittleness in the design of expert systems for immunohematology. *Immunohematology*, *12*(3), p.101-107.

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*, Marietta, GA: SA Technologies, p. 172-177.

Jardine, A.K.S. *Maintenance, Replacement, and Reliability*, Pitman Publishing, 1973.

Lee, J. & Moray, N. (1992): Trust control strategies and allocation of function in human-machine systems. *Ergonomics, 35(10),* p. 1243-1270.

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

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

Madhani, K., Khasawneh, M.T., Kaewkuekool, S., Gramopadhye, A.K., & Melloy, B.J. (2002). Measurement of human trust in a hybrid inspection for varying error patterns. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, Santa Monica, CA: HFES, p.418-422.

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

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

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

Sheridan, T. B. (1997). Supervisory control. In G. Salvendy (Ed). *Handbook of Human Factors*. 2nd Ed., New York: Wiley.

St. John, M., & Manes, D.I. (2000). Designing displays for command and control supervision: Contextualizing alerts and “trust-but-verify” automation. In *Proceedings of the IEA 2000/ HFES 2000. Congress*. Santa Monica, CA: HFES, p. 332-336.

St. John, M., Manes, D.I. & Osga, G.A. (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, no pagination.