Level of Automation Effects on Situation Awareness and Functional Specificity in Automation Reliance

Adam G. Smith and Greg A. Jamieson

Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2012 56: 2113
DOI: 10.1177/1071181312561448

Level of Automation Effects on Situation Awareness and Functional Specificity in Automation Reliance

Adam G. Smith, Greg A. Jamieson
University of Toronto
Toronto, ON

This work investigated the relationship between task performance and situation awareness (SA) at different levels of automation (LOA). The conventional wisdom is that routine performance improves with level of automation but that the consequences of automation failure become more severe. This has been characterized as a routine-failure trade-off, However, recent research indicates that the trade-off is subject to unknown contextual factors in addition to the level of automation. Furthermore, it has been suggested that the human operator's SA may impact whether the trade-off between performance under routine and failure conditions is always tenable. The current study therefore aimed to i) provide evidence to support or refute the trade-off and ii) to identify possible extenuating factors. The results generally supported the existence of the routine-failure trade-off, though the strength of this finding was tempered somewhat by operators' apparent selective disuse of higher level automation which limited the effective range of LOA tested. We interpreted that the SA collection method made the goal of SA maintenance explicit and in doing so encouraged operators to preferentially reallocate attention away from other system goals. Thus, the functional structure of the task seems to affect whether the routine-failure trade-off occurs in a given instance.

 DOI 10.1177/1071181312561448

INTRODUCTION

The "out-of-the-loop" (OOTL) effect encapsulates the conventional wisdom surrounding automation, namely that performance is improved under normal circumstances but that operators become progressively dissociated with the tasks and processes they are charged with (Endsley & Kiris, 1995). This latent failure manifests in decreased performance when the automation fails to perform as expected and operators must resume manual control. The OOTL effect is considered to be dependent on level of automation (LOA), with higher levels of automation resulting in progressively more dissociation. In a recent meta-analysis, Wickens, Li, Santamaria, Sebok & Sarter (2010) sought to formalize the conventional interpretation of the trade-off between automation-assisted performance under routine and failure conditions with a description they refer to as the "routine-failure trade-off". The account consists of four components: routine performance, workload, failure performance, and situation awareness (SA). The postulated relationships with LOA are illustrated in Figure 1, which was adapted from Wickens et al. (2010).

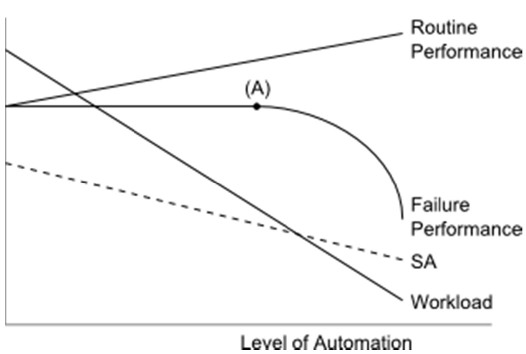

Figure 1: Postulated Routine-Failure Trade-Off

The benefits of automation are represented by an improvement in routine performance and a decrease in workload as LOA increases. The third component, failure performance, refers to performance in the event of automation failure. Failure performance is constant at low LOA because failure performance is *manual* performance and as such is unaffected by automation. Wickens et al. (2010) postulate that the negative impacts of automation may not take effect until a certain threshold level of automation (point (A) in Figure 1). This has implications for system design as the cost/benefit ratio of automation is most desirable at point (A). For this reason, some have argued that intermediate levels or adaptive automation are effective design strategies to combat the negative impacts of automation (Kaber & Endsley, 2004). Beyond the threshold, the operator's ability to control the system manually begins to degrade, ostensibly due to dissociation from the system as described by the OOTL effect. Endsley and Kiris (1995) cite skill decay and loss of situation awareness as key elements of the phenomenon. For knowledge-based work, the latter is likely the more pertinent concern.

The role of the fourth component, situation awareness, is unclear. Wickens et al. (2010) suggest that situation awareness may mediate the trade-off, noting strong correlations between situation awareness and both routine and failure performance. If so, then it may be possible to circumvent the routine-failure trade-off through system design that supports SA. In their meta-analysis, Wickens et al. (2010) noted a deficit of studies in the existing literature that compared SA at different levels of automation. In response to their subsequent call for further data, this study aimed to produce empirical evidence of the role of SA in the routine-failure trade-off.

Wickens et al. (2010) also stipulated that the relationships described by the routine-failure trade-off are likely affected by contextual elements. We therefore also sought to identify possible contextual factors that could impact the applicability of the routine-failure trade-off. Our secondary objective was to introduce and investigate the hypothesis that the functional structure of the task is relevant to the impacts of automation, in addition to characteristics of the automation itself. In particular, we postulated that the effects of automation on task performance and SA are limited to functions where there is a direct mechanistic link to the automated function(s). We broach this topic by first addressing the predominant model of LOA and illustrating how task functional structure is extraneous to the concept of LOA but vital for the routine-failure trade-off.

Characteristics of Automation and Tasks

A function can be defined in terms of the physical and information manipulation as well as the function allocation between agents responsible for performing it. These two elements form the basis of the types and levels taxonomy of automation (Parasuraman, Sheridan, & Wickens, 2000). The "type" dimension of this taxonomy uses a simple four-stage version of the information processing model comprising information acquisition, information analysis, action selection, and action implementation. The "level" dimension refers to the extent to which an automated agent is responsible for carrying out a given function.

However, the taxonomy does not explicitly address the relationships between functions within a task. The hierarchical multi-loop nature of many complex control systems is such that functions exist at different layers of abstraction from system goals (Lorenz, Di Nocera, & Parasuraman, Display integration enchances information sampling and decision making in automated fault management in a simulated spaceflight micro-world, 2002). Clearly, the effects of automation are not expected to manifest only in the function that is directly augmented by automation; the practical purpose of automation is to enhance performance of the end goal. Rationally, performance of the intermediary functions must also be improved. However, in complex tasks, functions may be separate in that they work through independent mechanistic sequences to achieve the same goals, or different goals entirely. This is illustrated in a hierarchical task analysis as parallel branches of means-ends links. Functions in separate parallel branches do not have a causal relationship and as such, performance in one should not be linked to performance in the other.

One exception to the above is that resources which are shared between functions may affect functionally distinct branches of a task. Resources freed by alteration of one function (i.e. through automation) may be reallocated to the other. The resource distributed to functions in the routine-failure trade-off is attention, as indicated by cognitive workload. Thus, this theory is untestable except where workload alleviating effects of automation (task shedding) are eliminated by removing the operator's capacity to reallocate attention, for example by designing a task with both temporally and functionally separate branches.

METHOD

Participants

24 students (11 male, 13 female) participated in the study. None had prior experience with the Cabin Air Management System (CAMS) microworld. Participants received $60 upon successful completion of the study. Ages ranged from 18 to 42 years (M=22.2). Data from two participants were removed for failing to follow task instructions.

Apparatus: AutoCAMS

The Cabin Air Management System (CAMS) microworld simulates a generic life support system (Hockey, Wastell, & Sauer, 1998; Sauer, Hockey, & Wastell, 2000a; Sauer, Wastell, & Hockey, 2000b). Participants fill the role of a system operator whose task is to maintain a livable environment through the management of five cabin parameters (O_2, CO_2, pressure, temperature and humidity). The five parameters are maintained in the normal range automatically by inner control loops which are supervised by an outer control loop. The operator serves as the controller in the outer loop, diagnosing faults in the inner loops as they arise. When any of the system parameters deviate from the normal range, a visual alarm alerts the operator to take action. The operator must then gather information such as tank levels and flow rates to diagnose the system fault. Once the fault is identified, the operator sends a repair order, which takes 60 seconds to complete. During the repair time, the operator must intervene in the inner control loop and manually maintain the affected parameters in the appropriate range. The process for correcting faults is thus divided into temporally separate diagnosis and fault-management phases. The implication of their temporal separation is that task shedding in one phase cannot affect workload in the other.

AutoCAMS augments the outer loop with a decision support system, the Automated Fault Identification and Repair Agent (AFIRA) (Lorenz, Di Nocera, Rottger, & Parasuraman, 2002). AFIRA assists in identifying faults, selecting procedures and implementing control actions. These functions correspond to the information analysis (IA), action selection (AS) and action implementation (AI) stages of automation described by the types and levels model. The automated functions are introduced cumulatively as the information required for later stages is provided by earlier stages. In this way, both the type and level of automation increase simultaneously, a simplification Wickens et al. (2010) refer to as *degree* of automation (DOA).

A diagram of the functional structure of the task is shown in Figure 2, including the aid provided by AFIRA to specific functions.

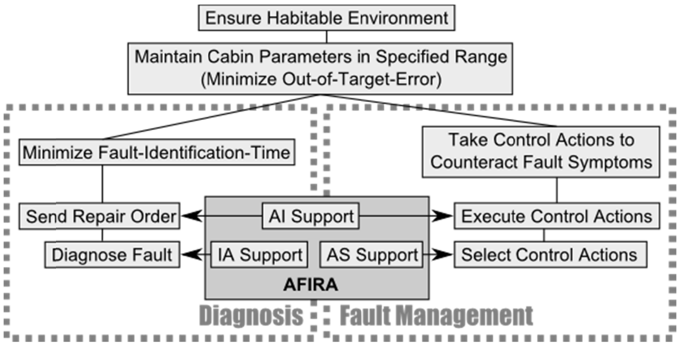

Figure 2: AutoCAMS Task Hierarchy

Procedure

The experiment followed a 4 (DOA) x 5 (Block) mixed design with DOA as the between-subjects factor and blocks of trials as the repeated factor. The approach was based on previous work using the AutoCAMS platform (Manzey, Reichenbach, & Onnasch, 2008).

As in the approach of Manzey et al. (2008), five blocks of six faults were presented to the participants, with each block lasting approximately 40 minutes. The faults were selected to include one sensor fault and one valve stuck-open fault in each block to balance scenario difficulty, and the other four faults randomly selected. The first block of trials was performed manually by all participants as a baseline. In blocks 2-4, participants assigned to one of the automated conditions were aided by AFIRA. In the final block, participants were required to unexpectedly return to manual performance. Using this design, routine performance was taken as the average performance across blocks 2-4 and failure performance as the difference between block 5 and block 1.

Dependent Measures

Fault-identification-time. Fault-identification-time was captured as the duration between initiation of the alarm and sending of the correct repair order.
Out-of-target-error. Out-of-target-error was collected as the total duration that critical system parameters were outside of the normal range.
Situation awareness. Situation awareness was collected using the Quantitative Analysis of Situation Awareness (QASA) freeze and query method (Edgar & Edgar, 2007). The simulation was paused 15 seconds after each alarm and participants were required to identify the symptoms present using a true/false response format. During the pause, the interface was hidden from participants. The timing of the freeze was based on pilot trials, and designed to allow participants enough time to gather some, but not all of the information required to make a full diagnosis. Thus, the SA measure in this study is a snapshot of participants' knowledge as the diagnostic process is underway. Each probe comprised eight query items that as a set were diagnostic of all faults in AutoCAMS.

Hypotheses

We hypothesized that the routine-failure trade-off would be observed, but that effects on performance measures would be null for automation of functions in parallel branches. Following the task structure shown in Figure 2, for routine performance, we predicted that IA automation would improve fault-identification-time and SA as AFIRA's recommendations would directly aid the functions performed during the diagnosis phase. IA was also expected to improve out-of-target-error as any reduction in the duration of the diagnosis phase (fault-identification-time) reduces the overall potential error that may be accumulated. AS automation was expected to only affect out-of-target-error as the procedural information provided is only applicable to the fault-management phase. AI automation was expected to improve both fault-identification-time and out-of-target-error as with IA, but was not expected to affect SA as the automation did not aid so much as supplant the diagnosis process. It was expected that only AI automation would be associated with a decrement in failure performance, as found by Manzey et al. (2008).

RESULTS

The analyses were primarily performed using single-factor ANOVA. Where the assumption of homogeneous variance was not met (as indicated by Levene's test), independent samples t-tests (equal variances not assumed) were performed. Notable results are shown in Figures 3-5. All error bars represent 95% confidence intervals. An alpha criterion of .05 was used in all analyses.

Routine Performance

When aided by AFIRA, a significant improvement from baseline was observed for fault-identification-time, $F(3,18)=24.2$, $p<.001$ (see Figure 3). For out-of-target-error, Levene's test indicated unequal variances ($F=3.76$, $p<.05$), necessitating the use of the independent samples approach. The independent t-test revealed a significant difference between the manual group and the IA group, $t(8.26)=2.33$, $p<.05$ ($\Delta M=140$, $\Delta SD=60.3$) (see Figure 4). However, no further improvement was detected in either measure for LOA beyond the introduction of IA support. A slight improvement in out-of-target-error is apparent in Figure 4 with the introduction of AS support, but it did not reach the level of significance.

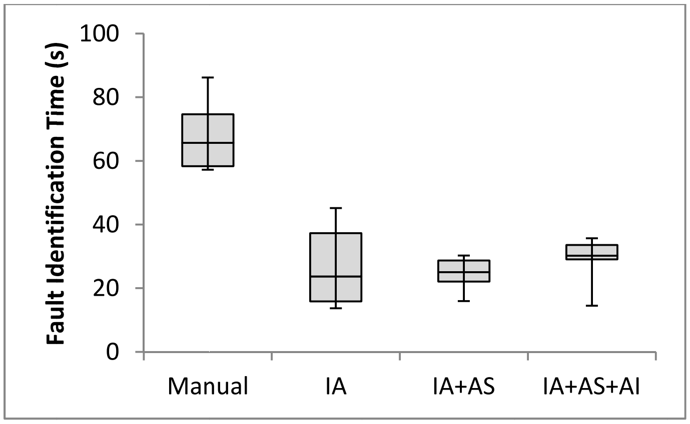

Figure 3: Fault-Identification-Time (Routine)

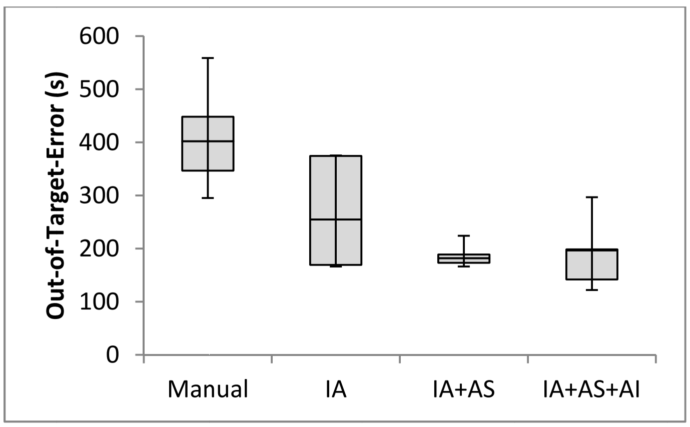

Figure 4: Out-of-Target-Error (Routine)

Failure Performance

Though participants showed a slight improvement in fault-identification-time and out-of-target error from block 1 to block 5, no significant differences were detected across groups for either performance measure. This was an unexpected result as Manzey et al. (2008) found evidence of a failure performance decrement for out-of-target-error at the AI level.

Situation Awareness

Similar to routine-performance, a marked improvement in SA was observed with the introduction of information analysis support. As with out-of-target-error, unequal variances (F=3.64, p<.05), required the use of the independent samples approach. The analysis revealed a significant difference between the manual group and the IA group, $t(5.84)=2.89$, p<.05 (ΔM=36.5, ΔSD=12.6) (see Figure 5). No differences were detected between different LOA groups.

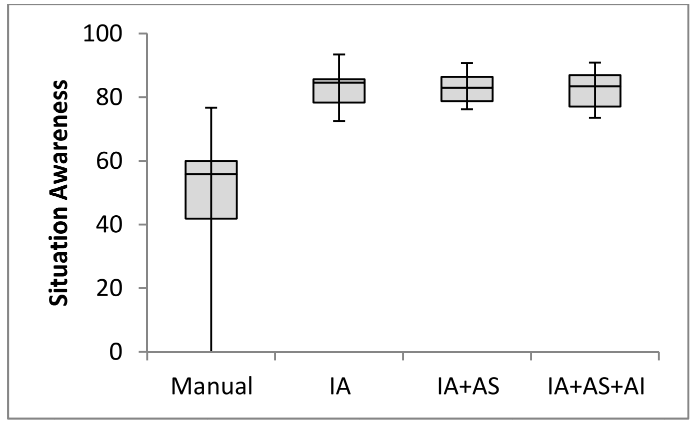

Figure 5: Situation Awareness

DISCUSSION

It was originally hypothesized that automation would only have an effect on goal performance where automation directly augmented the functions that were means to accomplish that goal. This was generally supported by the results. As predicted, all performance measures were improved by IA automation and there was some evidence that AS automation improved out-of-target-error, though not significantly so. However, no change in either routine or failure performance was observed for the AI group. This contrasts with the findings of Manzey et al. (2008), suggesting a key difference between the execution of the two studies. The main methodological change was the addition of the SA method, and the relation of the AI automation to the goal of maintaining SA may be the reason why operators' behaviour was altered.

The mechanism by which AI automation supports performance is rapid acceptance and implementation of AFIRA's recommendations. Because of the short span of time for operators to perform situation assessment in this experiment, operators were faced with the choice to reap the benefit of rapid acceptance at the expense of performance on the SA queries. We therefore interpreted that administering the SA queries made the cost of uncritical reliance explicit and thus caused participants to selectively disuse the AI automation. Note that this trade-off between goals is specific to the modified task and as such is not inherently related to the type or level of the AI automation. The modified task structure is illustrated in Figure 6.

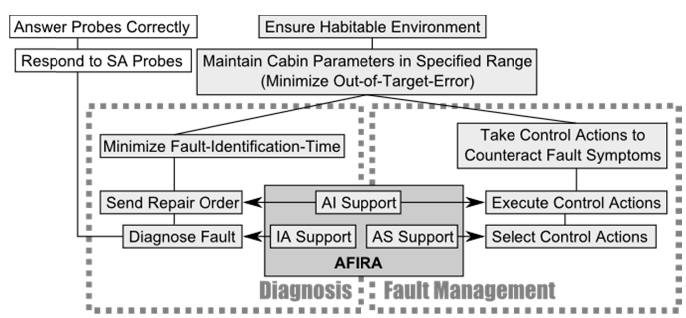

Figure 6: Explicit Goal of SA Maintenance

Unfortunately, this selective reliance masked the effects of higher level automation on SA and both routine and failure performance. Thus, we could not make a strong conclusion regarding the functional specificity of automation effects on performance measures and situation awareness. However, the results are sufficient to suggest that a task's functional structure may be relevant insofar as it affects the operator's strategic fulfillment of goals. Modification of the task, in this case by introducing a new procedure to report system information, may change the operator's behaviour such that the effects of automation are nullified. In terms of the routine-failure trade-off, it does not seem likely that this contextual factor could change the relationships between costs and benefits of automation. However, it does represent a potential pitfall of research in this area wherein otherwise valid effects could be masked.

CONCLUSIONS

In conclusion, the routine-failure trade-off was generally supported, though the strength of this conclusion was tempered by a lack of data for the highest LOA tested due to apparent disuse of that specific automated function. Because of this apparent disuse, the hypothesis that the functional specificity of automation effects is linked to functional structure was not successfully tested. Thus, functionally specific selective reliance seems to be a mechanism by which operators can modulate the effects of automation. Selective reliance was therefore identified as a contextual factor that could impact the presence of routine-failure trade-off effects differentially. This may be particularly relevant to SA, where the goal of SA maintenance may not be explicit but inferred from other tasks and responsibilities. Further work is required to investigate the original hypothesis regarding functional specificity of automation effects.

REFERENCES

Edgar, G. K., & Edgar, H. E. (2007). Using signal detection theory to measure situation awareness: The technique, The tool, the test, the way forward. In J. Noyes, & M. Cook (Eds.), *Decision making in complex environments* (pp. 373-385). Aldershot, UK: Ashgate.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37* (2), 381-394.

Hockey, G. R., Wastell, D., & Sauer, J. (1998). Effects of sleep deprivation and user interface on complex performance: a multilevel analysis of compensatory control. *Human Factors, 40*, 233-253.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*, 113-153.

Lorenz, B., Di Nocera, F., & Parasuraman, R. (2002). Display integration enchances information sampling and decision making in automated fault management in a simulated spaceflight micro-world. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 31-35). Baltimore: Human Factors and Ergonomics Society.

Lorenz, B., Di Nocera, F., Rottger, S., & Parasuraman, R. (2002). Automated fault-management in a simulated spaceflight micro-world. *Aviation, Space, and Environmental Medicine, 73* (9), 886-897.

Manzey, D., Reichenbach, J., & Onnasch, L. (2008). Performance consequences of automated aids in supervisory control: The impact of function allocation. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting. 52*, pp. 297-301. New York: Human Factors and Ergonomics Society.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30* (3), 286-297.

Sauer, J., Hockey, G. R., & Wastell, D. G. (2000a). The effects of training on short- and long-term skill retention in a complex multiple-task environment. *Ergonomics, 43*, 2043-2064.

Sauer, J., Wastell, D. G., & Hockey, G. R. (2000b). A conceptual framework for designing micro-worlds for complex work domains: a case study of the Cabin Air Management System. *Computers in Human Behavior, 16*, 45-58.

Wickens, C. D., Li, H., Santamaria, A., Sebok, A., & Sarter, N. B. (2010). Stages and Levels of Automation: An Integrated Meta-analysis. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (pp. 389-393). San Francisco: Human Factors and Ergonomics Society.