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

by

Adam George Smith

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Department of Mechanical and Industrial Engineering
University of Toronto

© Copyright by Adam George Smith (2012)

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

Adam George Smith

Master of Applied Science

Department of Mechanical and Industrial Engineering
University of Toronto

2012

Abstract

This thesis investigates the relationships between performance, workload, and situation awareness at varying levels of automation. The relationships observed in this study are compared to a description put forth to formalize the conventional interpretation of the trade-off between the benefits of automation during routine operation and the costs under conditions of automation failure. The original work stipulated that this “routine-failure trade-off” is likely a simplification affected by contextual factors. This work therefore aimed to i) provide empirical evidence to support or refute the trade-off and ii) to identify possible extenuating factors. The results generally supported the routine-failure trade-off, and considered in light of the functional structure of the task suggested that the relationships between goals and individual functions specific to a given task seem to affect the overall costs and benefits of automation through the mechanism of selective reliance. Further work is required to validate the findings of this study.

Acknowledgements

This thesis would not have been possible without the support of many friends, family and colleagues.

I appreciate the feedback received from my committee members, Dr. Birsen Donmez and Dr. Mark Chignell, as well as the guidance of my supervisor, Dr. Greg Jamieson. All members of the Cognitive Engineering Laboratory deserve recognition for their help in developing the concepts in this thesis. In particular, I am grateful to Nathan for all of our long discussions which helped me to understand not just the issues at hand but the nuances of Human Factors research more clearly. Additionally, special thanks are due to Nathan and Tony for their help with proofreading this thesis and earlier incarnations of the ideas herein.

I owe a long overdue thank you to Ben for his invaluable help with coding. What seemed like a simple change turned into quite a few nights of head scratching. Nevertheless, it turned out to be essential and worked a charm.

To mum and dad, your unwavering encouragement and patience kept me going back to the drawing board time and again. Yin Ling, your almost peculiar gift for understanding did not go unnoticed during those long months of data collection and writing. Surprise soup and wah mui always brightened my day.

Table of Contents

1	Chapter 1 – Introduction.....	2
1.1	Introduction	2
1.2	Purpose	2
2	Chapter 2 - Foundations	4
2.1	Defining Automation.....	4
2.1.1	Supervisory Control.....	4
2.1.2	Level/Degree of Automation	5
2.2	Costs and Benefits of Automation	7
2.2.1	Descriptions of the Costs of Automation.....	8
2.2.2	Routine-Failure Trade-Off.....	8
2.3	Automation-Induced Complacency and Automation Bias.....	10
2.3.1	Complacency	10
2.3.2	Automation Bias.....	12
2.3.3	Summary	13
2.4	Trust in Automation	13
2.5	Situation Awareness	15
2.5.1	<i>Situation Awareness and Automation</i>	15
2.5.2	Measurement of Situation Awareness	17
2.5.3	Relevant Techniques	19
2.6	Cabin Air Management System (CAMS)	22
2.6.1	(Lorenz, Di Nocera, Rottger, & Parasuraman, 2002)	22
2.6.2	(Manzey, Bahner, & Hueper, 2006)	23
2.6.3	(Manzey, Reichenbach, & Onnasch, 2008)	23
2.6.4	(Reichenbach, Onnasch, & Manzey, 2010)	24

3	Chapter 3 - Methods.....	26
3.1	Hypotheses	26
3.2	Experimental Design	26
3.2.1	Independent Variables.....	26
3.2.2	Dependent Variables	28
3.3	Apparatus.....	30
3.3.1	Workstation	30
3.3.2	CAMS/AutoCAMS 2.0.....	30
3.3.3	Workload.....	37
3.3.4	Trust and Self-Confidence Questionnaire.....	37
3.3.5	Situation Awareness Method	38
3.4	Procedure.....	42
3.4.1	Experiment Overview.....	42
3.4.2	Trial Design.....	44
4	Chapter 4 - Analysis	46
4.1	Expected Effects.....	46
4.1.1	Part 1.....	47
4.1.2	Part 2.....	50
4.2	Statistical Analysis.....	51
4.2.1	Part 1.....	51
4.2.2	Part 2.....	51
4.2.3	Summary	52
4.2.4	Power Analysis	53
5	Chapter 5 - Results.....	56
5.1	Data Corrections	56

5.1.1	Removal of Participants	56
5.1.2	Missing Data.....	56
5.1.3	Outliers.....	56
5.2	Part One	56
5.2.1	Routine Performance.....	59
5.2.2	Workload.....	61
5.2.3	Situation Awareness	62
5.2.4	Failure Performance.....	64
5.2.5	Summary of Part 1 Findings	65
5.3	Part Two.....	66
5.3.1	Trust	67
5.3.2	Verification Sampling	67
5.3.3	SA	68
5.3.4	Summary of Part 2	68
6	Chapter 6 - Discussion.....	71
6.1	Part 1.....	71
6.1.1	Routine Performance.....	72
6.1.2	Situation Awareness	73
6.1.3	Workload.....	73
6.1.4	Failure Performance.....	73
6.1.5	General Findings	74
6.1.6	Part I Summary.....	76
6.2	Part 2.....	77
6.3	Limitations	77
7	Chapter 7 – Conclusions and Future Work	80

Bibliography 82

List of Tables

Table 1: Experimental Design	27
Table 2: Dependent Variables Overview	28
Table 3: Faults and Symptom Patterns	31
Table 4: Experiment Overview.....	43
Table 5: Hypotheses (Part I).....	46
Table 6: Hypotheses (Part II).....	46
Table 7: Predicted Effects (Part I)	47
Table 8: Predicted Effects (Part II)	50
Table 9: Statistical Approach	52
Table 10: Alternate Nonparametric Statistical Approach.....	52
Table 11: Between-Groups Effects – Single Factor ANOVA & Contrasts.....	57
Table 12: Between-Groups Effects - Independent t-tests (Equal Variances Not Assumed).....	57
Table 13: Correlation of Primary Performance Measures.....	60
Table 14: Correlational Analysis of Situation Awareness Measures.....	64
Table 15: Support for Part 1 Sub-Hypotheses	65
Table 16: Block Effects - 3(Block) x 3(DOA) ANOVA.....	66
Table 17: Nonparametric Tests - Friedman’s Test, 3 Levels	66
Table 18: Follow-up on Marginal Situation Awareness Result – 2(Block) x 3(DOA) ANOVA.....	67
Table 19: Summary of Findings in Part 2	69
Table 20: Hypotheses (Part I).....	71
Table 21: Hypotheses (Part II).....	71

List of Figures

Figure 1: Routine-Failure Trade-Off.....	9
Figure 2: Degrees of Automation.....	27
Figure 3: Fault Event Timeline	32
Figure 4: CAMS Task Hierarchy	33
Figure 5: AFIRA Levels	34
Figure 6: AutoCAMS Interface	35
Figure 7: AutoCAMS Task Hierarchy	36
Figure 8: Selection of Query Timing.....	39
Figure 9: Power Analysis (Routine Performance – Fault-Identification-Time)	54
Figure 10: Power Analysis (Failure Performance – Out-of-Target-Error)	54
Figure 11: Routine Performance (Diagnosis Accuracy).....	59
Figure 12: Routine Performance (Fault Identification Time).....	59
Figure 13: Routine Performance (Out-of-Target-Error).....	60
Figure 14: Workload (NASA-TLX)	61
Figure 15: Situation Awareness (SA).....	62
Figure 16: Automation Verification Information Sampling (AVIS).....	63
Figure 17: Situation Awareness (Response Bias)	63
Figure 18: First Failure Response - Trust in AFIRA IA	67
Figure 19: Response to First Failure (AVIS).....	68
Figure 20: Response to First Failure (SA)	68
Figure 21: Wickens’ Postulated (Left), Predicted (Centre) and Observed (Right) Performance Trade-off	72
Figure 22: CAMS Task Hierarchy	76

List of Appendices

Appendix A – Trust/Self-Confidence Questionnaires

Appendix B – Correct Probe Responses

Appendix C – Other Experiment Handouts

Appendix D – Training Manual

Appendix E – Experimental Condition Assignments

Appendix F - Block Scripts

Appendix G – Removed and Missing Data

Appendix H - Normality Tests

Appendix I – Failure Performance Graphs

Chapter 1

Introduction

1 Chapter 1 – Introduction

1.1 Introduction

Automation is pervasive in many work domains, especially where task performance can affect the health and safety of workers and the general public. Automation enables the removal of the human operator from hazardous environments and extends the operator's capabilities. However, there are also known negative impacts attributed to automation, primarily latent failures that manifest when the automation unexpectedly fails to perform its tasks as anticipated. Given the pervasiveness of highly automated processes in safety-critical systems, these deleterious impacts have the potential to cause large scale damage. These latent failures are characterized by the phenomena of skill decay, complacency, and loss of situation awareness (Endsley & Kiris, 1995). The level of automation is thought to directly impact the extent of these latent failures, though the characteristics of specific automated tasks may have some bearing on the trade-off between the costs and benefits of automation (Wickens C. D., Li, Santamaria, Sebok, & Sarter, 2010). In this thesis, I attempt to provide more detail on attributes of tasks and automated agents which influence the introduction of latent failures into the joint human-machine control system.

1.2 Purpose

The purpose of this work is to provide empirical data on the relationship between the benefits and costs of automation. In particular, it has been suggested that situation awareness could moderate the trade-off between these costs and benefits (Wickens C. D., Li, Santamaria, Sebok, & Sarter, 2010). In the account given by Wickens et al., the trade-off stipulates that higher levels of automation are associated with a progressive loss of situation awareness but that performance in the event of an automation failure would not decline so long as situation awareness was maintained. This thesis investigates this position in two stages. First, the effect of level of automation on situation awareness was tested to provide evidence supporting or refuting Wickens' position. The second part investigated whether the relationship held true when operators changed their reliance on the automation. The expected result was that higher levels of automation would not be intrinsically tied to loss of situation awareness and that disuse of the automation would prevent operators from gaining the benefits normally imparted by that automation.

Chapter 2

Foundations

2 Chapter 2 - Foundations

2.1 Defining Automation

In this study, I adopt the definition of automation as “a device or system that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator” (Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000). Automation of functions also enables tasks requiring a level of precision and speed that the unaided human is incapable of and removes the human from physical hazards. Reallocation of a function from a human operator to an automated agent does not simply replace the human but instead changes the nature of the role (Parasuraman & Riley, 1997). As the work environment becomes increasingly automated, the human operator becomes responsible for supervising the automation rather than performing functions personally. This has come to be known as the supervisory control paradigm.

2.1.1 Supervisory Control

In many work domains, the human operator’s role is to supervise a highly automated process. Sheridan lists five roles for the operator in the supervisory control paradigm: planning, teaching, monitoring, intervening and learning (Sheridan, 2006). This thesis deals only with the latter three. Monitoring is the act of sampling relevant system information to determine whether the automation is successfully controlling system processes. If the system is not under stable control, the operator’s role is to then intervene as appropriate. Operators thus must retain the skills required to control the system without the automation. These skills include recognizing the situation, knowing what to do about it and being able to carry out those actions. Finally, by training and/or experience with the automation, operators can learn and adapt to the automation’s capabilities. Therefore, operator’s behaviour with regard to monitoring and intervention can change over time, particularly in response to cues regarding the automation’s capability to handle certain situations.

Supervisory control describes a control configuration with a hierarchy of control loops. In such a configuration, multiple agents act at different levels of abstraction from the overarching system goals. At a low level, system parameters are manipulated to achieve a set-point or acceptable range. At a high level, overall functioning of the system is monitored for abnormal output that may be traced to abnormal operation in the lower level(s) due to equipment malfunction, input disturbances, etc.

In the typical usage of the supervisory control description, the supervisor is a human operator, but either of the high and/or low level control loops may be aided by automation. Automated control at high and low levels of abstraction has been described as inner and outer control loops (e.g. Lorenz, Di Nocera, & Parasuraman, 2002). Inner loop control is mechanistic and controls parameters with sensors and simple controllers. Automation of the outer loop has a greater degree of agency in that the automation is an intelligent system which interacts with the operator on a peer level. An example of such automation is a decision support system (an expert system that assists a human operator with high level decision making). When both control loops are supported by automation such that there are distinct automated entities present, the system might be referred to as multi-agent control architecture (Lorenz, Di Nocera, Rottger, & Parasuraman, 2002). In such a system, an operator must coordinate activities with automated agents, creating the possibility for failures of coordination (Woods, 1996). For example, the autonomy and authority of agents, coupled with low observability of inputs and outputs, may imply that an automated agent possesses capabilities that it does not. The operator's assessment of the automation's capability impacts his/her control strategies. If the operator has inappropriate confidence in the automation's capabilities, then verification of the automation may be biased (Woods, 1996).

The issues inherent in automation interaction are complicated by the fact that automation is not an all-or-none concept (Wiener & Curry, 1980; Billings, 1991); a task may be partially automated. The characterization of level of automation is a major point of discussion in this thesis.

2.1.2 Level/Degree of Automation

An early method for the classification of automation was the Fitts' list (Fitts, 1951). The Fitts' list broadly outlined classes of information transformation and physical manipulation that were more capably performed by a human operator or the technology available at the time. Since that point in history, technology has advanced such that the machine's capability now exceeds that of the human operator in many areas. Subsequently, efforts have been made to classify automation by the extent of automation on a scale representing levels from completely manual performance to fully automated performance (e.g. Sheridan & Verplank, 1978). This type of classification acknowledges that function allocation is a gradient in terms of the extent to which a function may be performed by the human or automated agent. More recently, the level of automation concept was extended into a two-dimensional classification with the inclusion of the type of information or physical manipulation supplemented for a given function (Parasuraman, Sheridan, & Wickens, 2000).

2.1.2.1 Types and Levels of Human Interaction with Automation

A function can be defined in terms of the information and physical manipulation required to perform the function as well as the agent(s) responsible for actually performing the function. These two elements form the basis of the Types and Levels taxonomy of automation put forth by Parasuraman, Sheridan and Wickens (2000). The authors incorporated a simple four-stage version of the information processing model comprising information acquisition, information analysis, action selection, and action implementation with the concept of level of automation to describe automation along two dimensions. Automation of either of the first two stages is referred to as information automation whereas support of the latter two stages comprises decision/action automation. The key differentiator is that information automation aids only processes that occur prior to the point of decision (Parasuraman, Sheridan, & Wickens, 2000). For clarity, we outline the cognitive functions that each stage is understood to represent as well as some examples of what it could mean to automate these functions.

Information acquisition includes the orientation of sensing elements, pre-cognitive processing of perceptual data and selective attention. An example of acquisition automation could be the mechanical movement of sensors to acquire information from the environment, or alarms in a process control application that serve to direct the operator's attention to certain relevant information sources (Parasuraman, Sheridan, & Wickens, 2000). Information analysis includes the conscious/intentional manipulation of acquired and stored/retrieved information in working memory. Automation of the information analysis stage could include trend displays for the prediction of future states, or integration of multiple parameters to specify a higher order concept. Action selection is the output of decision making processes. An example of action selection automation is a decision support system that uses conditional logic to compare the available information to present conditions and offer the operator a filtered selection of possible courses of action or recommendation of one specific choice. Action implementation is the execution of the selected action and comprises mechanical manipulation or communication of information, for example. Implementation automation could be the electromechanical operation of valves, or automatic sending of an order.

As outlined previously, the Types and Levels model classifies automation by the extent to which a function is allocated to an automated agent and the stage of information processing which is supported. However, it has been noted that the taxonomy does not explicitly address abstraction of functions from system goals, which is a pertinent element in many real world systems (Lorenz, Di Nocera, Rottger, & Parasuraman, 2002). While this may not be directly important to the performance of

a single function, it may have ramifications for overall system performance. This will be revisited later in interpreting the results of this study.

2.2 Costs and Benefits of Automation

Modern technology has made it possible to automate essentially any process. However, there are costs and benefits inherent in the implementation of automation that must be weighed. The question is not whether a function can be automated, but whether it should be (Wiener & Curry, 1980). The most fundamental benefits of automation are the enabling of tasks that the human could not perform alone, and the removal of the operator from hazardous conditions. Beyond these basic benefits, there are two main categories of benefits: a decrease in workload and increase in routine performance.

However, it has been found in some cases that implementation of automation did not decrease workload, and in some cases even increased it (Parasuraman, Sheridan, & Wickens, 2000). One example is when the automation is difficult or time consuming to engage. This is referred to as “clumsy” automation (Wiener, 1988). Alternately, automation may increase workload because the operator’s supervisory role is not necessarily less demanding so much as having different demands.

With regard to routine performance, it is generally accepted that routine performance of an automated task exceeds unaided performance (Wickens C. D., Li, Santamaria, Sebok, & Sarter, 2010). Past research has found that automation can also have negative effects on joint human-system performance (Bainbridge, 1983; Wickens, 1994; Woods 1996; Parasuraman & Riley, 1997; Sheridan, 2002). The negative impacts of automation on performance are latent failures as per Reason’s Swiss cheese model (Reason, 1990). A latent failure is a vulnerability in one layer of protection against total system failure. When circumstances unexpected by system designers cause latent failures to align, they manifest as abnormal operation. As Reason suggests, “front-line operators are rarely the principal instigators of system breakdown. Their part is often to provide just those local triggering conditions necessary to manifest systemic weaknesses created by fallible decisions made earlier in the organizational and managerial spheres”(Reason, 1990). In this case, the roles assigned to the operator and automation are such that neither is capable of handling an unusual situation when it arises. The active failure is triggered when the operator is unable, in the face of system limitations and his/her conditioning during long periods of routine performance, to independently overcome the challenge of the unanticipated situation.

2.2.1 Descriptions of the Costs of Automation

There have been various attempts to classify latent failures of automation. One of the better recognized accounts is “out-of-the-loop unfamiliarity” (Wickens C. D., 1994). Whilst the operator is tasked with supervision of the process, the change from active control to passive monitoring decreases the operator’s ability to update and integrate system information in working memory (Endsley & Kiris, 1995). Out-of-the-loop unfamiliarity is manifested by slowed response to abnormal operation and inability to take corrective action during failure mode (Wickens & Hollands, 1999). Two phenomena are cited as integral to out-of-the-loop unfamiliarity: skill decay and loss of situation awareness (Endsley & Kiris, 1995).

Skill decay occurs when the human operator is not required to perform a task over a sufficiently long period of time (Rose, 1989). If the operator must take manual control during an abnormal situation, then his/her skill level is suddenly relevant to maintaining proper system performance. It is particularly pertinent as a failure of automation may be prompted by unusual circumstances, making the situation even more challenging for the operator.

When monitoring highly reliable automation, there is a propensity for operators to allocate less attention than if the automation were unreliable. This may manifest in a decrease in the frequency with which the operator samples system parameters to verify the automation’s control actions. Consequently, the operator may lose situation awareness of the system’s state when that system is the responsibility of another agent, either human or automated (Endsley & Kiris, 1995). Situation awareness is addressed in greater detail in a subsequent section.

In summary, out-of-the-loop unfamiliarity is a phenomena-based description of the latent failures induced by interaction with automation over a long time period. The operator is expected to respond appropriately in the event of an emergency, but is placed in a situation that hampers his/her ability to do so. Out-of-the-loop unfamiliarity is thought to be linked to the level of automation, with higher levels being associated with greater detriment to the operator’s ability to respond in an emergency. This rise in benefits and costs as the extent of automation increases has been referred to as the routine-failure trade-off (Wickens C. D., Li, Santamaria, Sebok, & Sarter, 2010).

2.2.2 Routine-Failure Trade-Off

The routine-failure trade-off has been proposed as a general description of the costs and benefits of automation as level of automation varies (Wickens C. D., Li, Santamaria, Sebok, & Sarter, 2010). The trade-off consists of four components: routine performance, workload, failure performance, and

situation awareness. The postulated relationships are illustrated in Figure 1 (reproduced from Wickens, Li, Santamaria, Sebok, & Sarter, 2010).

Figure 1: Routine-Failure Trade-Off

The benefits of automation are represented by an improvement in routine performance and a decrease in workload as level of automation increases. The third component, failure performance, refers to manual performance in the event of automation failure. Wickens' account posits that as the level of automation increases, failure performance may remain constant up to a point before steeply dropping off (Wickens C. D., Li, Santamaria, Sebok, & Sarter, 2010). Failure performance remains constant for low LOA because failure performance is manual performance, and as such unaffected by the level of automation. Wickens suggests that the negative impacts of automation do not take effect until a certain threshold level of automation (point (A) in Figure 1). Beyond that threshold, the operator's ability to control the system manually begins to degrade and additional levels of automation beyond that point make the effect stronger. For example, some research has indicated that physical implementation of control actions (late stage automation) is associated with failure performance decrement (Endsley & Kaber, 1999; Manzey, Reichenbach, & Onnasch, 2008). This clearly could have implications for system design as the cost/benefit ratio of automation is the most desirable at point (A). For this reason, some have argued that intermediate levels of automation (e.g. Kaber & Endsley, 1997; Parasuraman, Sheridan, & Wickens, 2000) or adaptive automation (Miller & Parasuraman, 2007) are an effective countermeasure against a decrease in failure performance while still accruing benefits.

The impact of level of automation on situation awareness is unclear. While out-of-the-loop unfamiliarity and the routine-failure trade-off both stipulate that increasing levels of automation are associated with decreasing situation awareness, Wickens cites strong correlations between situation awareness and both routine and failure performance as evidence that situation awareness may be a

mediating factor in the trade-off. Conceptually, if operators maintain situation awareness, this may mitigate the onset of failure-performance decrement. Therefore, external factors such as training and interface design may circumvent the routine-failure trade-off, even at high levels of automation (Wickens C. D., Li, Santamaria, Sebok, & Sarter, 2010). Contrary to the suggestions that level of automation must be limited in order to prevent deleterious effects on failure performance, it may be possible to attain the benefits of highly automated processes while avoiding the out-of-the-loop costs. It is therefore important to identify factors that influence the applicability of the routine-failure trade-off. Wickens suggests that situation awareness is an important concept in this respect for which there is a deficit of data in the existing literature (Wickens C. D., Li, Santamaria, Sebok, & Sarter, 2010). This call for data is the primary impetus for this study.

Moving forward, we sought to identify factors that could influence the routine-failure trade-off. Because the latent failures that comprise the costs of automation described earlier are behavioural phenomena, we first require explicit descriptions of the phenomena that arise during automation interaction. The following section addresses the complacency effect and automation bias (section 2.3 Automation-Induced Complacency and Automation Bias). Following this, the situation awareness construct is discussed to clarify what the loss of situation awareness may actually represent (section 2.5 Situation Awareness).

2.3 Automation-Induced Complacency and Automation Bias

The complacency effect and automation bias are behavioural phenomena related to the latent costs of automation.

2.3.1 Complacency

It has been suggested that three elements characterize the phenomenon of complacency (Parasuraman & Manzey, 2010):

1. “human operator monitoring of an automated system”
2. “frequency of such monitoring is lower than some standard or optimal value”
3. “as a result of substandard monitoring, there is some directly observable effect on system performance”

Note that complacency does not imply the colloquial meaning that the operator assumes a passive state but represents an “active reallocation of attention” (Parasuraman & Manzey, 2010). This reallocation is held to be a product of the operator’s management of cognitive resources, wherein tasks

requiring less attention (e.g. those that are reliably automated) are monitored less frequently. Moray argues that this active reallocation can be said to be rational (Moray & Inagaki, 2000; Moray, 2003) even if it prioritizes performance of less critical manual tasks over monitoring of the highly reliable automated tasks, as the allocation is based on the perceived risks and costs. That is, in the operator's assessment, the alleviation of workload and/or benefits to routine performance over a long uneventful period outweigh the costs of a severe but improbable system failure. Moray argues further that complacency can only be cited if monitoring can be shown to be lower than a normatively optimal value (Moray & Inagaki, 2000; Moray, 2003). It is important to note that it is difficult or impossible to quantify the costs of total system failure and that there is almost certainly error in the assessment of risk and outcome, given that even the most experienced operator cannot have perfect knowledge of the consequences and likelihoods of all possible outcomes. In addition to facing insurmountable uncertainty, the operator must also make judgments of the automation's capability across different circumstances. Such judgments are based on information as well as experience, but are influenced by attitudes and beliefs. Therefore, I would describe the operator's attention allocation policy as a rational strategy subject to uncertainty and bias.

Two key factors known to influence operators' sampling are task load and the reliability of automation. Task load is relevant in that the operator must balance resources between all tasks to be performed to an adequate level of performance. Reduced monitoring is typically observed under multiple-task load situations (Parasuraman, Molloy, & Singh, 1993), when the operator must allocate limited resources to multiple tasks. Complacency effects are much weaker or absent when the operator's only responsibility is to monitor the automation. Reliability of automation is pertinent in that it is a determinant of the resources necessary to maintain that level of performance. Reliability has been shown to affect monitoring frequency in two ways: whether reliability is high or low (Prinzel, DeVries, Freeman, & Mikulka, 2001), and whether it is constant or variable (Parasuraman, Molloy, & Singh, 1993; Bagheri & Jamieson, 2004). When automation is highly reliable, sampling is less frequent, as would be expected from the rational perspective. Additionally, when reliability is variable, operators seem to tend towards conservative estimates of reliability and thus sample more frequently. This second point highlights how the operator is forced to base their attention allocation strategy on limited information. Their knowledge of the variability of the automation's reliability is based on past experience and at any given point can only calculate the total reliability based on the total past experiences, unless there is some cue that would suggest to the operator the reliability had changed for contextual reasons. A subset factor of reliability is the operator's perception of reliability, which is closely related to operator trust in

automation. Trust is a more complex construct than an estimate of reliability alone, and will be outlined in more detail subsequently (section 2.4 Trust in Automation).

2.3.2 Automation Bias

Automation bias describes an increased propensity to rely on an automated aid “as a heuristic replacement for vigilant information seeking and processing” (Mosier & Skitka, 1996, p. 205). It is a bias in decision making wherein the automation is credited with more capability than it objectively possesses (Skitka, Mosier, & Burdick, 1999). Though a phenomena of decision making, automation bias has in the past often been indicated by outcomes, specifically errors of omission and errors of commission (Mosier & Skitka, 1996). An error of omission occurs when the aid (incorrectly) recommends against taking an action and the operator complies. An error of commission occurs when the aid (again, incorrectly) recommends to take an action and the operator complies. Examples of errors of omission and commission will be given in context of a hypothetical radiologist working with a decision support system designed to assist in the detection of tumours in radiographs. In cases where the aid fails to find an existing tumour, some percentage of the physicians will also fail to detect the tumour, making a false negative. This is an error of omission. Conversely, if the aid reported a nonexistent tumour and the physician erroneously accepted this conclusion, this would constitute a false positive and an error of commission.

However, it is debatable whether occurrences of errors of omission or commission alone are sufficient evidence of automation bias (Parasuraman & Manzey, 2010). Similar to Moray and Inagaki’s (2000) rational attention allocation interpretation of complacency, the occurrence of single false positives and false negatives do not fully represent the decision maker’s policy governing automation reliance. Instead, automation bias can only be cited if the overall distribution of false positives and false negatives is different from unaided performance. Furthermore, automation bias cannot be classified as a cost of automation unless the weighted outcomes in aggregate are objectively worse. Using the radiologist example, if the occurrence of false negatives and false positives are different from what would be achieved by a radiologist acting alone, it can be concluded that automation bias had an impact on performance. However, if the cost of a missed existing tumour is not equal to the cost of a superfluous treatment, then the costs of the outcomes must be considered when determining the policy that maximizes the expected value. Only if the physician’s performance in aggregate was worse than he/she was independently capable of could automation bias be argued to have had a negative effect on task performance.

The characteristics of automation have been shown to influence whether operators are more likely to rely on an aid. For example, Crocoll and Coury found that operators were more likely to blindly follow automation providing a recommendation (decision automation) than a status update (information automation) (Crocoll & Coury, 1990). Past research has found that reliance on automation is in part determined by trust (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). Higher levels of trust are associated with a greater propensity to accept an aid's recommendation.

2.3.3 Summary

Complacency and automation bias have been studied separately in the past, generally in the separate areas of supervisory control and decision support systems, respectively. More recently, Parasuraman and Manzey categorized them as “different manifestations of overlapping automation-induced phenomena, with attention playing a central role” (Parasuraman & Manzey, 2010). Succinctly, the distinction is that complacency represents suboptimal monitoring/information sampling, while automation bias is a bias in decision making manifested by overreliance on an automated aid. As both phenomena are influenced by trust, monitoring and verification of automated aids, they may be relevant mechanisms of the routine-failure trade-off.

2.4 Trust in Automation

A problem inherent in supervisory control is that the operator's assessment of the automation's performance is based on either the outputs or the process. If the process is unobservable, then the operator must perform the task independently and compare his/her results against the automation's outputs in order to assess whether the task is being performed correctly. Humans are not capable of the same level of rapid, unbiased weighing of sensory evidence as a machine. This is a paradox in that the human has been assigned a task that is impossible to perform proficiently (Bainbridge, 1983). Thus, the operator must commit to a reliance strategy based in part on a judgement of the automation's competence. This assessment of trustworthiness is not simple or static but based on a variety of cues and dynamic in nature (Lee & See, 2004). When automation is highly reliable, trust grows over a long period of time. As monitoring of the automation decreases over this period, the risk that an eventual failure of the automation will go undetected increases. When an automation failure does occur (and is detected), an operator's trust in automation tends to fall (Parasuraman, Molloy, & Singh, 1993). In the following period, trust will be in a recovery phase (Lee & Moray, 1994). Thus, trust continually changes over time based on updates to the operator's perception of the automation's reliability.

In a study using the Pasteurizer microworld, Muir and Moray found that operators' reported trust was mainly based on assessments of the automation's competence (Muir & Moray, 1996). The authors also found that trust was directly correlated to usage of the automation and inversely correlated to monitoring of the automation. Reliance has been shown to also depend on self-confidence (Masalonis & Parasuraman, 1999). Lee and Moray used the Pasteurizer microworld to test operator control strategies in a complex simulated system when automation reliability was not perfect (Lee & Moray, 1992). The authors found that reliance did not correspond to changes in trust alone, but that the balance of trust and self-confidence was a determining factor in the choice between automatic and manual control (Lee & Moray, 1994). That is, low self-confidence and high trust could lead to reliance on the automation, perhaps incurring the performance costs as indicated in the routine-failure trade-off, while high self-confidence and low trust could lead to disuse. This is pertinent to the present study if the disuse of automation leads operators to perform as though working with a lower level of automation.

The predominant model decomposes trust in terms of calibration, resolution and specificity (Lee & See, 2004). Calibration is the correspondence between the trust attributed to an automated agent and the agent's actual capabilities (Lee & Moray, 1994; Muir, 1987). Resolution refers to the precision with which human judgement differentiates between distinct levels of automation capability (Cohen, Parasuraman, & Freeman, 1998). Specificity can be subdivided into functional specificity and temporal specificity. Functional specificity refers to the granularity of trust in specific functions. Low (coarse) functional specificity would imply that trust in the agent is uniform while high (fine) functional specificity implies that the operator maintains independent levels of trust for specific functions of the automation (Lee & See, 2004). Temporal specificity is an indicator of the human trustor's sensitivity to short term fluctuations in automation capability.

Research has shown that human-machine trust is comparable in some ways to human-human trust (Muir, 1987) and similarly, subjective scales can be used to measure human-machine trust (Muir & Moray, 1996). However, previous work has found differences between subjective trust and behavioural measures during trust recovery (Reichenbach, Onnasch, & Manzey, 2010). Thus, subjective reports of trust are useful but should also be interpreted with a measure of caution as reliance behaviour cannot be assumed.

2.5 Situation Awareness

Though owing its origins to military aviation, SA has been studied in air traffic control (e.g. Kaber, Perry, Segall, McClernon, & Prinzl, 2006), nuclear power (e.g. Hogg, Follesø, Volden, & Torralba, 1995), aviation (e.g. Endsley, 1993), driving (e.g. Lee, 1999) and command and control (e.g. Lichacz & Farrell, 2005).

SA is most frequently defined as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (Endsley M. R., 2006, p. 529). Endsley’s definition describes SA as a product of the information gathering processes in three tiers: perception of elements in the current situation, understanding of the current situation and projection of future states (Endsley M. R., 1995b). Though a universally accepted definition of situation awareness (SA) remains elusive, there is a general consensus that accurate knowledge of pertinent, up-to-date system information is a precursor to effective decision making.

Some of the contention over what constitutes SA relates to whether to focus on the *product* or the *process*. As a product, SA is the knowledge of what is happening in the system and what the implications are. Conversely, SA as a process describes the cognitive processes used to build and maintain that knowledge. The product/process debate has also been framed as situation awareness as opposed to situation assessment. Situation assessment refers to “the various perceptual and cognitive activities involved in constructing, updating, and revising the state of awareness”(Adams, Tenney, & Pew, 1995, p. 88). Situation assessment and situation awareness are of course intertwined. As Endsley suggests, one’s current situation awareness influences what further information is required and thus the situation assessment activities that are performed to acquire it (Endsley & Garland, 2000). Endsley notes that acquiring and maintaining SA are influenced by many factors including the operator’s cognitive abilities, training, preconceptions, goals, system design and the task environment (Endsley M. R., 1995b). The significance is that SA is not solely a function of any one of the situation, cognitive ability of the operator, monitoring behaviour, but all of these, as well as features of the task environment that are affected by automation such as availability and format of information as well as task load. By extension, when aided by automation, SA must also be affected by the operator’s expectations and attitudes towards the automation (i.e. trust).

2.5.1 Situation Awareness and Automation

The question of how situation awareness is affected by automation is a complex one. Endsley describes three mechanisms by which automation affects situation awareness: i) complacency in operator

monitoring behaviour, ii) operators assuming a passive role, and iii) changes in the quality or form of feedback provided (Endsley M. R., 1996). Concepts such as out-of-the-loop unfamiliarity imply that allocation of functions to automation by necessity causes operators to lose awareness of the information relevant to that task.

At the same time, it is generally accepted that automation has the potential to improve situation awareness, through either alleviation of workload and/or direct assistance with acquisition and processing of information (Endsley & Kiris, 1995). The former results from allowing the operator to reallocate attention and the latter could comprise filtering or integration of relevant system information such that mental manipulation of the information is performed faster or more completely. The stipulation that often seems to be missing from this discussion is that influences on any portion of the task do not act in isolation. The functions that comprise a task are interrelated and are only performed as a means to achieve system goal(s). For example, the alleviation of workload should only be expected to improve situation awareness if it is known that operators will allocate freed resources to situation assessment activities. Similarly, automation that directly assists information manipulation still carries the requirement of having to support a system goal as maintenance of situation awareness is not necessarily a goal in itself. Thus, task structure is *sine qua non* to issues of SA-aiding automation.

Endsley notes that situation assessment is goal directed (Endsley M. R., 2004). Others have even stated that situation assessment can and should be defined by task analysis techniques such as hierarchical task analysis (Patrick & James, 2004). By extension, theories regarding the effects of automation such as the routine-failure trade-off and out-of-the-loop unfamiliarity should explicitly take task structure information into account. That researchers have not done so in a systematic manner may help to explain why the literature thus far generally supports intermediate levels of automation for optimizing the trade-off between SA and workload (Wickens C. D., 2008) but only partially explains differences in costs and benefits between similar levels of automation in the context of different tasks (e.g. Kaber & Endsley, 1997; Kaber & Endsley, 2003; Endsley & Kaber, 1999)). An alternate explanation is that there are inherent characteristics of the automation-supported functions that independently affect SA or workload. The stage of information processing supplemented, as indicated in the Types and Levels taxonomy may be one such characteristic. For these reasons, this study specifically varies both type of automation and relational task structure across levels of the main independent variable (see Chapter 4 – Experimental Design for details).

2.5.2 Measurement of Situation Awareness

Situation Awareness Global Assessment Technique (SAGAT) (Endsley M. R., 2000b) is likely the most commonly used method and has accrued the most validating evidence (Salmon, Stanton, Walker, & Green, 2006). However, alternate methods are continually being produced as researchers discover ways in which SA measurement could be better conducted in specific domains. Some examples include: SACRI (Hogg, Follesø, Volden, & Torralba, 1995), SALSA (Hauss & Eyferth, 2003), and SA/BARS (Neal, Griffin, Patterson, & Bordia, 2000)). According to Endsley, SA measurement should (Endsley M. R., 1995a):

1. Measure SA and not other processes,
2. Possess a level of sensitivity and diagnosticity appropriate to the technology/manipulation being tested, and
3. Not substantially alter SA itself.

The definition of SA as a product or process plays a role in how it is measured (Salmon, Stanton, Walker, & Green, 2006). The majority of SA measurements techniques treat SA as a product, eliciting knowledge of the current state of dynamic aspects of the work domain rather than the processes used to attain that knowledge. Thus, measurement of SA compares operator knowledge to some objective ground truth. The selection of the objective system information to compare to is of critical importance. To determine what information is relevant to a task, the initial step in most SA measurement techniques is to define what constitutes “good SA” for the specific task. Thus, SA is always measured against a normative standard. The process of determining this standard in SAGAT is called an SA requirements analysis and is developed through goal-directed task analysis (Endsley M. R., 2001). Goal-directed task analysis lists the major goals and sub-goals in a task, along with the decisions that must be made in order to achieve those goals. Then, the system information necessary to make those decisions successfully is elicited from domain experts, literature, observation, etc. Goal-directed task analysis is a form of cognitive task analysis wherein a hierarchy is constructed that links the main goals to the component information needed to successfully complete each goal (Endsley M. R., 2000). In addition to querying only relevant information, SA measurement techniques must also be administered in the relevant timeframe, when that information is needed.

Broad categories of objective methods for measuring SA include real-time probes, freeze probes, performance measures and process indices. An alternative to objective measurements of SA is to collect subjective reports. In subjective methods, the operator evaluates the quality of his/her own SA.

Subjective methods include self-rating and observer-rating. The two categories of method used in this work are freeze probes and self-rating.

2.5.2.1 Freeze & Query Probe

Freeze probes pause (freeze) the task in progress to administer questions to the operator, often automatically by a computer. The displays are usually blanked during the probe and in some methods the timing of the freezes is random. Because the task must be stopped for the duration of the probe, freeze probes are only applicable to simulators or other studies where the task can be paused. Freeze probe techniques have the advantage of being direct but have been criticised for potentially influencing the main task by interruption. Despite some concern that the interruption could affect operators, it has been shown that pauses as long as several minutes had no effect on other performance measures in a combat aviation task performed by experienced fighter pilots (Endsley M. R., 1995a). In the nuclear domain, plant operators similarly did not report any effect from the pauses in the simulation (Hogg, Torralba, & Volden, 1993).

One important consideration with freeze probe methods is the issue of cueing. If the timing or content of the probe itself is a source of information, it has the potential to influence task performance. The timing or content of a query may be a cue in that the appearance of the probe prompts operators to think about or investigate (after the freeze) some aspect of the task that they otherwise would not have. One approach is to hold the query content and timing constant. Thus, cueing may still take place, but will be standardized across participant. Cueing effects can also be nullified by providing the information required by the queries immediately after the probe is administered (Skraaning Jr. G., et al., 2007; Lau, 2012). For example, if a probe queries whether a parameter is out of specification, the answer could be given immediately afterwards with the onset of an alarm on that parameter.

2.5.2.2 Self-Rating

Self-rating methods of SA assessment involve the operator making judgments of their own awareness after the task has been completed. Because they are administered after the trial, they are subject to recall problems and confounds with task performance. Additionally, because they are subjective, they are regarded as less sensitive and potentially more influenced by individual differences than objective measures.

An example of a self-rating method is Situation Awareness Rating Technique (SART) (Taylor, 1990), where the subject rates their SA on a 7-point Likert scale along 10 dimensions (familiarity of the situation, focusing of attention, information quantity, information quality, instability of the situation,

concentration of attention, complexity of the situation, variability of the situation, arousal, and spare mental capacity) or 3 dimensions (demands on attention, supply of attention, and understanding of the situation). It is readily apparent that some of these dimensions are not strictly about SA as a product but as an attentional process. It has been suggested that subjective reports of SA actually measure confidence in one's SA, rather than SA itself (Endsley, Selcon, Hardiman, & Croft, 1998). However, the measurement of regulation of SA may have independent value. Consider that information acquisition for SA maintenance is governed by perceived information requirements. Information requirements can only be known by assessment of one's current knowledge and/or quality of knowledge. In short, self assessment of SA may provide insight into information gathering, much as SA is thought to be a precursor to decision making performance.

In an evaluation of aircraft cockpit displays, Endsley et al. found no correlation between subjective SART scores and objective SAGAT scores (Endsley, Selcon, Hardiman, & Croft, 1998). However, a strong correlation was found between participants' reported confidence and both SART-SA (overall) and SART-understanding. The authors concluded that subjective reports of SA constitute confidence in SA. In a subsequent study of air traffic control displays, Endsley et al. found conflicting evidence in the form of weak correlations between SAGAT and SART-understanding and between SAGAT and SART-supply (Endsley, Sollenberger, Nakata, & Stein, 2000). These weak correlations between objective and subjective SA measures are congruent with a study by Lee (1999), who found low correlations between confidence reports and objective measures of situation awareness in both young and old drivers. Lee refers to this as low meta-SA, alluding to metacognitive regulation of situation awareness. Similarly, Lichacz attributed a weak correlation between confidence and accuracy to low meta-SA in an information gathering and conflict resolution exercise (Lichacz, 2008). Rousseau et al. built on the meta-SA concept by suggesting that subjective measures of SA are information-based judgements, rather than experience based (Rousseau, Tremblay, Banbury, Breton, & Guitouni, 2010). Information-based judgements relate to the background knowledge and expertise in comparable situations, as opposed to the specific evidence at hand.

2.5.3 Relevant Techniques

2.5.3.1 Process Overview

One freeze probe approach of particular relevance to this study is the Process Overview measure developed at the Organisation for Economic Co-operation and Development (OECD) Halden Reactor Project (Lau et al., 2012; Skraaning Jr et al., 2007). Process Overview adapts elements of other methods (SAGAT and SACRI) to evaluate SA in operation of a high-fidelity nuclear power plant simulator. One

notable element of Process Overview is its use of domain experts to select query content and timing. The method specifies that probes be administered at meaningful points in the scenario progression and query knowledge that is relevant in that timeframe. This stems from the need to assess operators' context-sensitive SA at distinct phases of the scenario (e.g., detection, diagnosis, and mitigation of faults). Process Overview was developed for the nuclear domain but is likely applicable to many process control applications as any domain where there are multiple modes of operation will likely have distinct SA requirements for each mode. The parameters queried in Process Overview can be classified as context-sensitive or fault-sensitive. Context-sensitive parameters are relevant during specific modes. Cueing effects are minimal as the need for knowledge of the parameter states is well established by training. Fault-sensitive parameters may not be sampled during normal operation as they pertain to specific abnormal states. The Process Overview approach to query timing and content has implications for this experiment that will be addressed in the methods section (see Chapter 3 - Methods).

Some elements of Process Overview, such as the distinction between context-sensitive and fault-sensitive parameters, collapse at a lower level of complexity. The predominant reason that Process Overview is not directly applicable to this study is that the difference in scope between a microworld simulator and a high-fidelity plant simulator results in fundamental differences in the requirements imposed on operators. The two key differences are that the meaning of parameters may or may not be contextual and there may be different levels of uncertainty in interpreting parameters. First, in a system of greater scope, parameters generally have different meaning depending on the mode of operation. For example, flowrates in pipes when they are being used to fill a vessel in the early stages of a batch process have different implications for correct system functioning than when the same pipes are used later to remove material from that vessel. Second, in a complex system, there may be system parameters that are not directly observable but are inferred from others (e.g. energy balance). Microworld studies generally do not involve these complex elements as the tasks are designed to be easily learned by a naive sample population.

Thus, a simpler SA method was more appropriate for this study, though the guidance offered by Process Overview with regard to query timing and content was taken into consideration (see Chapter 3 – Methods).

2.5.3.2 Quantitative Analysis of Situation Awareness (QASA)

QUASA/QASA (Edgar & Edgar, 2007) is a relatively recent method and one that has seen significant usage (Rousseau, Tremblay, Banbury, Breton, & Guitouni, 2010), particularly in command and control applications, e.g. (McGuinness, 2004; McGuinness, 2007). In the QASA method, probes are

administered during pauses in the task and participants are asked to indicate whether a set of statements regarding system information are true or false. Each response is paired with a confidence assessment where the participant indicates their confidence on a scale ranging from a complete guess to complete certainty. QASA has two notable features: the use of signal detection theory (SDT) to analyze responses, and the collection of both objective and subjective measures of SA.

First, the use of SDT means that participants' responses are analyzed in terms of sensitivity (ability of the operator to distinguish between true and false statements regarding system parameters) as well as bias in reporting (whether participants err towards liberal or conservative reporting of uncertain parameter states) (McGuinness, 2004). Sensitivity and bias scores can be calculated using a variety of measures, both those that make assumptions of the signal distributions and those that do not. The preference across the literature is to calculate the SA score as the distribution-free measure of sensitivity A' and response bias as the distribution-free measure B'' .

Second, the collection of both objective and subjective measures may yield insight into SA and meta-cognitive regulation of SA. Regulation of SA is likely relevant in the context of supervisory control as it may be related to operators' propensity to monitor automation. That is, operators may sample more frequently when they are aware that their level of knowledge is less than ideal. Also, while past research has shown that subjective assessments of self (e.g., self-confidence) and others (e.g., trust in automation) have been shown to influence reliance on automation, those are assessments of agents. Rather, QASA involves assessment of confidence in individual responses, independent of where that response was drawn from. Thus, the assessment may be more directly related to information gathering as a whole than reliance or manual monitoring per se. The output of the subjective portion of QASA may be reported as SA confidence or SA calibration. SA calibration is calculated as the average confidence rating, scaled from confidence of 50 (guess) to 100 (certain), minus the average percentage of query items answered correctly (0-100). At first glance, the difference in scales may seem problematic, but in practice does not present an issue as the effective scale for SA accuracy is actually (50-100). This is because respondents should almost always perform better than a complete guess (50% chance on a 2AFC task) unless there is a fundamental error in the system or query design. The calibration approach has a history of use in psychological research and is generally well accepted (e.g. Lee, 1999). The extension to its use in SA research is a logical step that may provide valuable insight into issues of information gathering, in particular in the context of automation interaction.

2.6 Cabin Air Management System (CAMS)

This study made use of CAMS, a supervisory control microworld, as an experimental platform. CAMS represents a generic life support system for a space capsule. Operators take a supervisory role of managing the system, intervening to diagnose and correct system faults as they occur. More details are given regarding the simulator itself in Chapter 3 – Methods. This section outlines four example studies making use of the CAMS simulator that are particularly relevant to this work.

2.6.1 (Lorenz, Di Nocera, Rottger, & Parasuraman, 2002)

The authors of this study built on the CAMS microworld by implementing automation in the outer control loop in the form of a fault management aid. They referred to this new version as AutoCAMS. The outer loop automation was manipulated as an experimental condition with three levels: a fault identification aid, automated diagnosis and recovery advisory, and automated diagnosis and recovery implementation. The benefits associated with the outer loop automation included more accurate diagnosis of system faults, shorter diagnosis time and reduced subjective workload. In the event of automation failure, the greatest performance decrement was observed for participants who had worked with the intermediate level of automation. This implied that automated diagnosis, which essentially replaced the operator as the decision maker, was most harmful to joint performance, but that this effect was offset in the highest LOA condition by the alleviation of workload from automatic recovery implementation. Subjective reports of trust and self confidence were collected but showed no main effects of LOA. However, an interaction effect was observed wherein self-confidence for participants at low LOA increased when changing from automated to manual performance while participants at high LOA experienced a decrease in self-confidence when returning to manual control.

In essence, AutoCAMS comprises several inner control loops and an outer supervisory loop, each aided by separate automated agents. Thus, AutoCAMS “may be regarded as the simplest variant of more complex multi-agent control architectures” (Lorenz, Di Nocera, & Parasuraman, 2002). Lorenz et al. comment that the hierarchical aspect of these nested control loops is an element of complex systems that may not be adequately addressed in current automation taxonomies. AutoCAMS therefore provides an opportunity to study some of the aspects of automation interaction which may go unaddressed in systems without this element.

The sample population consisted of 24 students (13 male, 11 female), aged 19-28 ($M=21.8$)

2.6.2 (Manzey, Bahner, & Hueper, 2006)

Manzey, Bahner and Hueper tested the effects of training interventions on monitoring. A measure of monitoring was defined as the percentage of a set of parameters required to diagnose each possible system fault that was actually accessed in the event of a fault. Participants were instructed to verify each automated recommendation. As a measure of sampling with a normatively correct response, this was taken to be an index of complacency. The authors found that only very high levels of complacency were associated with an increased risk of commission errors. LOA was not manipulated, so this study did not provide evidence of whether higher LOA were associated with increased levels of complacency.

The training intervention in this study exposed participants to automation failures, which reduced but did not prevent the occurrence of commission errors. Approximately 20% of participants committed a commission error, regardless of training condition. Participants who experienced an automation failure during the training phase continued to sample a higher percentage of the required parameters during the main experiment.

The sample population consisted of 24 engineering students (12 male, 12 female), aged 18-32 ($M=23.1$), without prior CAMS experience.

2.6.3 (Manzey, Reichenbach, & Onnasch, 2008)

Manzey et al. (2008) investigated the effect of level of automation on various task performance measures. The authors found that information analysis automation had a large effect on most of the task performance measures and that higher level action implementation automation had a further effect. Intermediate stage automation, action selection support, showed no tangible effect on performance. From this, the conclusion was drawn that progressively higher LOA led to progressively higher levels of performance. Failure performance was defined as performance in the event of an automation failure that required the operator to resume manual control. Some evidence was found for a failure performance decrement, but only in one performance measure (out-of-target-error), and only for the highest LOA (action implementation). Significant practice effects were found for two other performance measures. The conclusions therefore supported intermediate levels of automation for reducing the risk of operators becoming out-of-the-loop.

No relationship was found between LOA and the number of commission errors committed when the automation provided an incorrect diagnosis. This suggests that automation bias may be a somewhat different type of cost of automation than failure performance. The percentage of participants who

committed a commission error in this study was approximately 50%, notably higher than approximately 20% in the study by Manzey, Bahner and Hueper (2006).

The sample populated consisted of 56 engineering students (40 male, 16 female), aged 20-31 ($M=24.2$), without prior CAMS experience.

2.6.4 (Reichenbach, Onnasch, & Manzey, 2010)

Reichenbach et al. manipulated the amount of time participants spent working with an automated aid before experiencing an automation failure as well as the difference between a first failure event and a second failure. They showed that negative experiences had a stronger effect on subjective reports of trust and verification of the aid than repeated positive experiences. She also showed that a second failure had a weaker effect than the first failure. These results support that the expected relationship between trust in automation and monitoring hold true for CAMS. Furthermore, they highlight that the most pronounced effects tend to be observed in the first instance of automation failure. Overall, this study serves as a strong basis that CAMS can be expected to follow the conventions outlined earlier in this chapter.

The sample populated consisted of 88 engineering students (65 male, 23 female), with a mean age of 24.1.

Chapter 3

Methods

3 Chapter 3 - Methods

3.1 Hypotheses

The objective of this experiment was to show that the effects of automation cannot be considered without explicitly addressing the functional relationships between functions and goals. This was held to be true for both performance measures as well as situation awareness. Thus, the effects postulated by the routine-failure trade-off were expected to be contingent on the automated functions having a direct link to the performance measures in question. Effects due to task shedding were held constant where possible by temporally separating the tasks. Furthermore, the impacts of automation should be contingent on reliance. As this study focused on the role situation awareness to the routine-failure trade-off, reliance effects on situation awareness were also studied. The above research questions were summarized into two hypotheses as follows:

1. An effect on goal performance and situation awareness is expected only where automation directly augments a function that is a means to that end goal, or indirectly by alleviation of workload.
2. Impacts on situation awareness should be mediated by reliance.

These hypotheses were addressed in two separate parts. The specific experimental design and expected effects are explained in detail once the experimental apparatus is introduced.

3.2 Experimental Design

3.2.1 Independent Variables

The experiment followed a 5 x 4 mixed model design varying two factors: Block and Degree of automation (DOA). The first factor (Block) was a repeated measure representing the five blocks of trials participants were required to undergo. Participants completed all blocks in the same order. DOA was a between-subjects variable with four levels representing the different variations of automation provided to participants. The term DOA is taken from Wickens et al. (2010) and indicates both types and levels of automation collapsed to one dimension (see Figure 2).

Figure 2: Degrees of Automation

For each DOA group, the new stage of information processing is added to the previous group(s).

Participants were randomly assigned to DOA groups. Participants worked with their designated aid in blocks 2-4 and controlled CAMS manually in blocks 1 and 5 (see Table 1).

Table 1: Experimental Design

		Block				
		1	2	3	4	5
DOA	1	Manual	Manual	Manual	Manual	Manual
	2	Manual	IA	IA	IA	Manual
	3	Manual	IA + AS	IA + AS	IA + AS	Manual
	4	Manual	IA + AS + AI	IA + AS + AI	IA + AS + AI	Manual

IA: Information Analysis

AS: Action Selection

AI: Action Implementation

In the 1st block, all participants worked without automated assistance to provide a baseline measure of performance. In the 5th block, all participants were returned to manual control unexpectedly to assess failure performance. Routine performance, SA and workload were assessed across blocks 2-4.

At the end of the 3rd block (but not included in the block 3 data), an additional fault was administered wherein the information analysis provided an incorrect integration of information to all participants in automated conditions. This constitutes a first failure event. From the participant's perspective, the first failure event was part of the contiguous block so that the fault did not seem unusual and prompt

suspicion. This approach differs from Manzey et al (2008) where the first failure occurred in block 4, immediately before returning to manual. In this experiment, the automation failure precedes the last automated block to assess how participants' interaction with the aid changed as a result.

3.2.2 Dependent Variables

Table 2 gives an overview of the dependent variables used in the experiment. The variables are organized by the component of the routine-failure trade-off that each corresponds to. Additional dependent variables not part of Wickens' routine-failure trade-off included a measure of monitoring quality and subjective trust in the automation. A brief description of each variable follows the table, and a detailed account of how each fits into the experimental task given in section 3.3 Apparatus.

Table 2: Dependent Variables Overview

Component of Routine-Failure Trade-off	Measure	Description
Routine Performance / Failure Performance	Fault Identification Time [†]	Time taken to diagnose each fault
	Diagnosis Accuracy [†]	Percentage of repair orders sent that were correct
	Out-of-Target-Error [†]	Time that critical parameters were outside the specified range
	Prospective Memory [†]	Proportion of CO ₂ logs entered in the correct interval
	Simple Reaction Time [†]	Time taken to acknowledge the connection check stimulus
Workload	NASA-TLX	Standard workload questionnaire
Situation Awareness	SA [§]	Distribution-free sensitivity measure A'
	Bias [§]	Distribution-free response bias measure B''
	SA Confidence [§]	Reported confidence in query responses
Monitoring	Automation Verification Information Sampling (AVIS) [‡]	Percentage of parameters necessary for diagnosis sampled by the operator
Trust	Trust in IA automation [Ⓢ]	Subjective trust in automation's ability to perform the information analysis function

[†] Adapted from Manzey et al. (2008)

[‡] Adapted from Reichenbach et al. (2009)

[§] Adapted from QASA method (Edgar & Edgar, 2007)

[Ⓢ] Not a previously validated scale

3.2.2.1 Routine Performance/Failure Performance

Routine and failure performance comprised all three primary performance measures and two secondary measures (see below). Routine performance was averaged across blocks 2-4 and failure performance was taken as the difference between the return-to-manual block 5 and baseline block 1.

Primary performance measures were:

1. Fault identification time: The time elapsed between the alarm and the sending of the correct repair order (s).
2. Diagnosis accuracy: The number of correct repair orders divided by the total number of repair orders (%).
3. Out-of-target-error: The duration that critical parameters were outside of the normal range (s).

Secondary performance measures associated with secondary tasks were:

1. Simple reaction time: The interval between the appearance of the connection check stimulus and confirmation by the operator (ms).
2. Prospective memory: The percentage of CO₂ logs recorded in the appropriate time interval (± 5 seconds from the minute) (%).

3.2.2.2 Workload

Workload was measured at the end of each block.

1. NASA-TLX: A computerized version of the two-stage NASA Task Load Index.

3.2.2.3 Situation Awareness

Situation awareness was measured concurrently with routine performance, over blocks 2-4. Situation awareness measures were:

1. SA: The distribution-free measure of sensitivity A' to the probes administered after each fault occurrence (unitless, scaled from 0 to 100).
2. Bias: The distribution-free measure of response bias B'' to the probes administered after each fault occurrence (unitless, scaled from -100 to 100).
3. SA Confidence: The average reported confidence in query responses from 50 to 100, representing the range from a complete guess to complete certainty (unitless).

3.2.2.4 Monitoring

Monitoring was measured throughout blocks 2-4 during the diagnosis phase. The sampling measure used was:

1. AVIS: The percentage of parameters required for manual diagnosis that were accessed by the operator in the period between the fault alarm and sending of the first repair order (%).

3.2.2.5 Trust

Trust was measured throughout blocks 2-4 at the end of each block. The measure used was:

1. Trust in automation: Subjective trust in automation's ability to perform the information analysis function (unitless). Subjective trust in the other automated functions and in participants' self-confidence was also collected, but only for completeness and to avoid cueing participants' expectations of the IA failure in block 3 (see Appendix A).

3.3 Apparatus

3.3.1 Workstation

Participants interacted with a PC using a mouse and keyboard at a desktop workstation with a dual-monitor configuration utilizing two 20" flat panel monitors (viewing distance ~24") in a small cubicle (approximately 6'x4'). The participants were encouraged to adjust the station chair and desk height to a comfortable setting, but viewing distance was not altered. The simulator program was displayed on the left screen and additional information such as training material and questionnaires appeared on the right.

3.3.2 CAMS/AutoCAMS 2.0

The apparatus selected for this experiment was the Cabin Air Management System (CAMS), a microworld simulation designed to test hypotheses relating to supervisory control. CAMS represents a generic life support system for a contained living space such as a space capsule. The system consists of an inner direct control loop and an outer supervisory loop (Lorenz, Di Nocera, Rottger, & Parasuraman, 2002). The inner loop controls 5 system parameters automatically while the outer loop detects and corrects malfunctions of the inner loop. In the original CAMS simulation, the outer loop is solely the responsibility of the operator. The high-level goal of the CAMS simulation is to keep the capsule habitable, which is determined by whether certain critical parameters are kept in their respective safe ranges. This is enabled by two sub-goals: i) to take control of the inner loops in the event of a system fault, and ii) to detect and resolve abnormalities in the system.

In AutoCAMS, the outer-loop control task is aided by a decision support system. The operator's role is to supervise the system, monitoring the activities of the inner-loop automation and working with the outer-loop automation when necessary.

3.3.2.1 System

The core of the CAMS simulation is comprised of coupled oxygen and nitrogen subsystems, which maintain oxygen concentration and pressure levels respectively in the capsule cabin. Bang-bang

controllers open and close on/off valves to maintain parameters in the normal range and the flowrates are moderated by regulator valves. The regulator controller automatically adjusts the flowrates to maintain the proper oxygen/nitrogen mix in the cabin. Inside the pressurized cabin, the oxygen concentration and pressure interact as there is a relationship between the gas inflow (either oxygen or nitrogen) and pressure. The system cannot add oxygen without also increasing pressure. To compensate for this, a drain valve is used to vent air and relieve excess pressure when required. The drain valve is triggered when a pressure sensor in the cabin detects that the threshold level has been reached (the same upper limit as the bang-bang controller).

Additionally, there are 3 subsystems that do not directly impact the control task: temperature, humidity and carbon dioxide. These are automatically managed with a heater/chiller, dehumidifier and CO₂ scrubber respectively. The dynamics of these three subsystems are not pertinent to the main control task (Manzey, et al., 2008).

3.3.2.2 Faults

Periodically, fault states arise due to malfunction of system components. Depending on the location and nature of the malfunctioning component, one or both of the oxygen and nitrogen/pressure subsystems are impaired to varying degrees. Each fault is also associated with a unique symptom pattern that the operator must learn to recognize in order to diagnose the faults. The symptom patterns are made up of cues that are always observable and cues that are revealed through diagnostic procedures using manual control actions. The diagnostic process is entirely rule-based, with no novel faults. Table 3 below outlines each fault and the associated symptom pattern.

Table 3: Faults and Symptom Patterns

Fault Type	Oxygen	Nitrogen/Pressure	Mixer
Valve Leak	<ul style="list-style-type: none"> • Cabin O₂ low • Flow rate < target • Tank level changes > flow rate 	<ul style="list-style-type: none"> • Cabin N₂ low • Flow rate < target • Tank level changes > flow rate 	N/A
Valve Blockage	<ul style="list-style-type: none"> • Cabin O₂ low • Flow rate < target • Tank level changes = flow rate 	<ul style="list-style-type: none"> • Cabin N₂ low • Flow rate < target • Tank level changes = flow rate 	<ul style="list-style-type: none"> • Cabin O₂ & N₂ low • Both flow rates < target
Valve Jammed Open	<ul style="list-style-type: none"> • Cabin O₂ high • Cabin pressure high (at ceiling due to venting) • O₂ flow always on 	<ul style="list-style-type: none"> • Cabin pressure high (at ceiling due to venting) • Cabin O₂ low (due to venting) • N₂ flow always on 	N/A

Fault Type	Oxygen	Nitrogen/Pressure	Mixer
Defective Sensor	(Occurred while O ₂ rising) <ul style="list-style-type: none"> • Cabin O₂ high • Cabin pressure high (at ceiling due to venting) • O₂ flow always on (until shut-off manually, then always off) (Occurred while O ₂ falling) <ul style="list-style-type: none"> • Cabin O₂ low • O₂ flow = always off 	(Occurred while pressure rising) <ul style="list-style-type: none"> • Cabin pressure high (above ceiling, no venting) • Cabin O₂ low (due to venting) • N₂ flow always on (until shut-off manually, then always off) (Occurred while pressure falling) <ul style="list-style-type: none"> • Cabin pressure low • N₂ flow = always off 	N/A

3.3.2.3 Tasks

3.3.2.3.1 Primary Tasks

The operator's main task is to monitor the system to diagnose and manage faults as they occur. Figure 3 illustrates the timeline for each fault event.

Figure 3: Fault Event Timeline

Detection is simplified by a simple two state alarm. The alarm is tripped whenever a fault causes one of the two critical parameters to deviate from the normal range. The alarm is usually triggered within 1-5 seconds of the fault occurring but due to the random elements in CAMS, can take up to one minute. Once detected, the fault must be diagnosed by sampling system parameters to identify symptom patterns and in some cases performing procedures to exclude one of two possibilities. Once the operator decides on a diagnosis, he/she sends a repair order to the crew of the space capsule, who then take 60 seconds to perform the repair. If the diagnosis was correct, the alarm state returns to normal. If it was incorrect, no additional stimulus is given and operators must be aware that 60 seconds have elapsed without the fault being resolved. In that case, operators must then reattempt diagnosis. Once the fault is successfully diagnosed, operators must manually control the affected subsystems during the 60 second repair to prevent the critical parameters from leaving the target range. This is the fault management phase. The control actions taken to mitigate the effects of a known fault are based

on the type of fault, the subsystem affected and current state of the corresponding system parameter. There are two broad classes of control actions that operators must take. For leaks and blocks, the flow rate setting must be increased to replace the lost gas or overcome the back pressure respectively. The adjustment need only be performed once as the bang-bang controller is still operational. Jammed valves and sensor faults impair the controller's ability to open and close the on/off valves and sense parameters respectively. For these types of faults, the operator must take the place of the bang-bang controller, manually opening and closing the on/off valve of the affected subsystem as required to keep the affected parameter in range. A task diagram for CAMS is included here to illustrate the task structure (see Figure 4).

Figure 4: CAMS Task Hierarchy

3.3.2.3.2 Secondary Tasks

The operator is also asked to perform two secondary tasks throughout the experiment. The first is to confirm that the communications link is still connected. This task requires the participant to click an icon as quickly as possible and constitutes a simple reaction time (SRT) task. The second secondary task is to observe and record the cabin CO₂ concentration value $\pm 0.1\%$, on the minute ± 10 seconds without prompting. This is referred to as a time-based prospective memory task, and depends on self-initiated retrieval processes (Sauer, 2000).

3.3.2.4 Automated Fault Identification and Recovery Agent

More recent versions of the CAMS microworld (AutoCAMS) include an automated decision aid (Lorenz, Di Nocera, Rottger, & Parasuraman, 2002). Referred to as Automated Fault Identification and Recovery Agent (AFIRA), the aid assists the operator in performing the outer loop supervisory tasks during fault states by providing the operator with 1 of 3 levels of support. The first level suggests a fault diagnosis to the operator. The second level provides the fault diagnosis and a recommended set of

actions to keep the critical parameters in specification. The third level provides fault diagnoses, recommended actions and the option to have the system execute the repair task and recommended actions. These correspond to the information analysis, action selection, and action implementation types of automation in the types and levels taxonomy (Parasuraman, Sheridan, & Wickens, 2000). AFIRA level is ordinal on both the type and level dimensions of the taxonomy as each higher level of the aid included the lower level features below it. Following the example of Wickens et al, this was termed *degree of automation* (DOA) (Wickens C. D., Li, Santamaria, Sebok, & Sarter, 2010). Examples of the four levels of AFIRA are shown in Figure 5.

Figure 5: AFIRA Levels

3.3.2.5 Interface

The AutoCAMS interface is shown in Figure 6.

Figure 6: AutoCAMS Interface

a) History graph	f) AFIRA panel
b) Mimic display	g) System clock
c) Control panel	h) Connection check
d) Fault alarm	i) CO ₂ log
e) Repair status window	

3.3.2.6 Task Analysis

The task diagram introduced earlier is reproduced here with a representation of AFIRA added for clarity (see Figure 7).

Figure 7: AutoCAMS Task Hierarchy

As mentioned previously, the overall goal of CAMS is to sustain a liveable environment, which is primarily accomplished by maintaining the critical parameters in the specified safe ranges. This is enabled by two main tasks, diagnosis and fault management.

AFIRA provides support to both diagnosis and fault management, but with different functions assigned to different branches of the hierarchy. The diagnosis branch (left side of Figure 7) is aided by information analysis support while the fault management branch (right side) is aided by action selection support. Action implementation automation aids both branches, but through different mechanisms.

AFIRA aids diagnostic performance in two ways. First, information analysis support provides a recommendation immediately following the alarm. Thus, AFIRA provides a diagnosis much faster than the operator can, reducing fault identification time, should the operator choose to rely on AFIRA. Second, action implementation automation provides the operator with the option to have AFIRA execute the order, potentially reducing the fault identification time. Similarly, AFIRA aids fault management performance in two ways. Action selection support reminds the operator of the most effective control actions for mitigating out-of-target error and action implementation performs the selected actions automatically. Note that because the diagnosis branch takes place prior to the fault management branch (see section 3.3.2.3.1 Primary Tasks), automated aid in the latter cannot affect performance in the former. Thus, action selection provides no support to the diagnosis branch.

In addition to diagnosis and fault management, participants must also perform the secondary tasks and respond to SA probes. The secondary tasks are ongoing and serve to keep the operators' task load at full capacity. The SA probes contain information pertinent to the symptom patterns in the diagnosis branch (see section 3.3.5 Situation Awareness Method for details). Participants performing the task unaided must sample parameters in order to answer the queries. Alternately, participants with information analysis support can infer the value of unobserved parameters from AFIRA's recommendations. Because the support provided by action selection automation is only pertinent to fault management, it does not aid in the identification of symptom patterns required for the SA probes. Similarly, there is no mechanism by which action implementation support may affect situation assessment. In the diagnosis phase, action implementation only allows operators to bypass the manually entering of the appropriate repair order.

3.3.3 Workload

Participants were instructed to complete the NASA-TLX (Hart & Staveland, 1988) questionnaire in the context of their experience in the preceding block. The two-stage version of NASA-TLX was used as the nature of the task could potentially change between administrations of the method depending on whether the automated aid was available. Periodically during breaks in the task, participants reported subjective workload on six component scales and provided weights for each component by pairwise comparison. Participants used a computerized version of NASA-TLX where the six component ratings are collected using a visual scale with 20 distinct levels per scale.

3.3.4 Trust and Self-Confidence Questionnaire

Participants' subjective trust in the automation and belief in their own ability was assessed on a 9 point scale. The questionnaire contained 9 items (see Appendix A) querying participants' confidence in specific functions performed by AFIRA and self-confidence in their own ability to perform those functions. The questionnaire items differentiated between the three functions supported by AFIRA (information analysis, action selection and action implementation) and AFIRA as a whole. This distinction corresponds to Lee & See's concept of specificity of trust (Lee & See, 2004). The self-confidence items followed a similar phrasing, except to query participants' confidence in their own ability. Additionally, the operator was also asked to rate overall confidence for both agents jointly. The questionnaire participants received was one of four versions depending on their assigned level-of-automation condition. Participants were queried only on AFIRA functions that applied to their assigned experimental group. The primary item of interest was participants' specific trust in AFIRA's ability to perform the information analysis function. The other items were included to encourage participants to

consciously recognize the three separate functions and to distinguish between AFIRA's ability and their own.

The term "confidence" in the automation's capability was treated as equivalent to "trust" for four reasons. First, to reduce potential misunderstanding, the phrasing of the questions was kept as similar as possible between assessments of the participant's abilities and AFIRA's abilities. Second, use of the word "trust" in the context of an inanimate object/intelligent system (rather than another person) might cause confusion. In a previous study by Lee and Moray (Lee & Moray, 1992), participants had this concept explicitly explained to them, but it was decided that this would add undue burden, given that participants were already required to learn several unfamiliar tasks. Third, it was believed that using the term "trust" might influence participants' baseline expectations of the aid. In particular, explicitly asking whether participants trusted the aid could potentially cause them to become suspicious and anticipate the automation failure. Finally, some previous work has shown that reported trust was based primarily on assessments of capability (Muir & Moray, 1996).

3.3.5 Situation Awareness Method

The situation awareness probes were administered via dialogue boxes that opened at predetermined times. The dialogues appeared on the right-hand display, separate from the CAMS display to avoid accidental responses from clicks that were intended for the CAMS interface. When the probes appeared, the left-hand display with the CAMS interface automatically was blanked and the simulation was frozen. Participants were informed of this procedure and given the opportunity to practice so that they would not assume that there was time pressure in responding to the probes. This section describes the SA requirements analysis conducted to determine the query content, probe timing and response format. Query content and probe timing were related, with certain temporal constraints taken into account in the selection of the goal-relevant information that was elicited by the probes. The response format was defined by the QASA method, and is described fully in this section.

3.3.5.1 SA Requirements Analysis

An SA requirements analysis was conducted to determine the appropriate probe content. The goals for this analysis were to determine a set of queries that met two criteria. The queries were designed to:

1. Elicit information relevant to the goal of fault diagnosis, and
2. Sufficiently describe every symptom pattern uniquely.

To avoid cueing effects, the same set of queries was used in every probe. The probe was administered once for every fault event, yielding a total of 48 queries per data point. Based on pilot trials, it was

determined that this was sufficient to avoid the difficulties in SDT calculations caused by participants not making at least one hit and one false alarm while also avoiding the upper and lower extremes of performance, which are sometimes associated with misestimates of sensitivity and bias.

3.3.5.1.1 Query Timing

Several factors influenced the design of when probes were administered, including the specifics of the system phase, risk of cueing and ceiling effects. Determining in which phase(s) to administer the queries was the overriding factor. The phases that were considered were: pre-alarm, monitoring (normal operation), fault management and diagnosis phases (see Figure 8).

Figure 8: Selection of Query Timing

Pre-alarm phase: It was determined quickly that probes could not be administered during the pre-alarm phase. This is because the phase sometimes lasted as long as one minute but more often was 0-5 seconds and the random variation of system parameters prevented prediction of how long a given pre-alarm phase would last. This practical constraint ruled out the pre-alarm phase.

Monitoring phase: The monitoring phase was considered in order to assess whether participants maintained awareness during normal operation. However, without the possibility of an invisible fault state (because no probes could be administered during the pre-alarm phase), participants would quickly learn that no fault was present so long as the alarm light was not lit. Querying operators during the monitoring phase would therefore only inflate SA scores without providing meaningful data.

Fault-management phase: The fault management phase, after the repair order is sent, was a possibility for administering queries. However, because the operator's goal at this point in the scenario is to mitigate the effects of a known fault, the query content would have to support fault management tasks. The amount of information here is very limited, as the only relevant items are the state of a single parameter and the choice between one of two control strategies (increase regulator valve setting or manually control the parameter with the on/off valve). The operator's knowledge of other system parameters would not constitute goal-relevant information.

Diagnosis phase: Therefore, the probes were limited to the core diagnosis task which took place during the diagnosis phase between the alarm and the repair order. During this phase of the task, operators were required to acquire and integrate information, making situation awareness relevant to task performance.

Next, it was necessary to determine when the probes should be administered within the diagnosis phase. The main concern in this respect was finding a balance between allowing participants time to adequately assess the situation and avoiding ceiling effects by querying too late in the diagnosis process. Through pilot trials, it was found that proficient participants could reliably acquire enough information to diagnose faults within 20 seconds. Therefore, probes were administered 10-15 seconds after the alarm was triggered. Any longer than this and pilot participants were generally able to answer all questions correctly, causing ceiling effects. Thus, the SA probe was used to capture a snapshot of the operator's diagnosis formation as the process evolved. The amount of time elapsed between the alarm and the probe was subject to some random variation (approximately 3 seconds on average) due to the software used.

One valid concern with this approach is that participants could learn to expect the query and alter their control strategy in anticipation. However, because the query was administered during the time-compressed period after the alarm when the operator's task was well-defined, the participants should have been focused on information gathering activities that were identical to those required to answer the SA queries, provided that the query content was representative of the diagnosis task. To nullify the effect of some participants coming to expect the probes more quickly than others, all participants were informed of when the probes would be administered.

3.3.5.1.2 Query Content

The SA requirements analysis was conducted by mapping out the faults and associated symptom patterns (as shown in Table 3), and developing a set of queries that mirrors the diagnosis process participants were trained to follow. As with the diagnosis process, there was some redundancy in the information represented. Eight queries were required to be fully diagnostic of every fault, using the QASA true/false statement approach. This process yielded the following set of queries:

1. Cabin oxygen concentration is outside the normal limits or behaving abnormally within limits
2. Cabin pressure is outside the normal limits or behaving abnormally within limits
3. Oxygen flow rate is less than the preset
4. Nitrogen flow rate is less than the preset
5. Changes in the nitrogen or oxygen tank level are greater than the corresponding flow rate

6. A valve is blocked
7. The automatic controller is not sensing when cabin oxygen or nitrogen crosses the normal limits
8. A valve is jammed open

The set of queries as a whole was sufficient but not necessary to define the fault at hand. There was unavoidable redundancy in that some queries could be answered on the basis of other query responses. Each SA probe consisted of the full set of queries. Note that some of these queries are defined by simple observation of parameters while others describe observable trends and still others require integration of multiple cues. The mapping of correct query responses to faults is available in Appendix B. Each query was paired with a confidence rating from 1-5. Participants' confidence was queried immediately after each query using the same dialogue box format.

3.3.5.2 Calculation of QASA Measures

As described in Chapter 2, the QASA method uses signal detection theory to determine sensitivity and response bias. QASA adopts the distribution-free estimates of sensitivity and bias A' and B'' (Edgar & Edgar, 2007), which were calculated using Equation 1 and Equation 2, (Stanislaw & Todorov, 1999), (Snodgrass & Corwin, 1988). The probabilities of hits (H) and false alarms (FA) were calculated on the basis of 6 probes (48 queries) per data point.

$$SA = 100 \times A' = \begin{cases} 100 \times \left(0.5 + \frac{(H - FA)(1 + H - F)}{4H(1 - FA)} \right), & H \geq FA \\ 100 \times \left(0.5 - \frac{(FA - H)(1 + FA - H)}{4FA(1 - H)} \right), & H < FA \end{cases}$$

where H is the probability of a hit and FA is the probability of a false alarm

Equation 1: Distribution-free sensitivity A'

$$Bias = 100 \times B'' = 100 \times \begin{cases} 100 \times \left(\frac{H(1 - H) - FA(1 - FA)}{H(1 - H) + FA(1 - FA)} \right), & H \geq FA \\ 100 \times \left(\frac{FA(1 - FA) - H(1 - H)}{FA(1 - FA) + H(1 - H)} \right), & H < FA \end{cases}$$

where H is the probability of a hit and FA is the probability of a false alarm

Equation 2: Distribution-free response bias B''

3.3.5.3 Method Validity

In this study, the diagnostic agent provides a diagnosis constituting the integrated whole of the information required by the SA queries. One criticism is that participants need only know the symptom

pattern for a given fault and they will technically be able to answer every SA query perfectly, so long as the aid provides the correct diagnosis. Because of this, one may challenge whether this assessment of knowledge constitutes awareness of the system parameters or simply a learned response to a specific cue. However, there is a theoretical reason why this criticism does not invalidate the study. The rationale is that any interaction with an autonomous system effectively expands the human operator's capabilities, subject to the operator's reliance on the automation. Whether that automation is a decision support system such as AFIRA, a complex display of integrated information or raw sensor data, the operator's reliance on information integrated from sensed data qualifies as inference. Hence, information regarding the actual state of the world inferred from an integrated source is no more or less "awareness" than if the information was acquired by observation because both processes rely on the validity of all intermediate transformations of that information.

3.4 Procedure

3.4.1 Experiment Overview

The experimental protocol underwent ethics review and informed consent was obtained from all participants prior to beginning the study. The consent form is included in Appendix C. The selection criteria were: proficiency in English and enrolment in a postsecondary institution or possession of a diploma/degree.

The experiment was administered in two sessions on two separate days not more than 10 days apart. The first session comprised training with the basic (unaided) CAMS system, familiarization with questionnaires and a test to assess the participants' capability to complete the experiment (see Appendix C). In the second session participants were familiarized with the version of AFIRA associated with their experimental condition, were introduced to the SA method, and completed the main experiment. An overview of the experiment is given in Table 4.

Table 4: Experiment Overview

Session 1 (3 hours)	<ul style="list-style-type: none">•Informed consent•Background information	<ul style="list-style-type: none">•Training (CAMS)•Training (Trust Questionnaire)•Training (NASA-TLX)		<ul style="list-style-type: none">•Proficiency test
Session 2 (6 hours)	<ul style="list-style-type: none">•Refresher exercises	<ul style="list-style-type: none">•Training (SA Method)•Familiarization with experimental condition	<ul style="list-style-type: none">•Main Experiment	<ul style="list-style-type: none">•Debrief

3.4.1.1 Session 1: Preparation

Participants independently worked through a training document explaining the premise of the simulation, interface, tasks and questionnaires (see Appendix D). Participants were encouraged to ask the experimenter for clarification when necessary. This portion of training required approximately 2 - 2.5 hours for each participant.

Participants were then tested to determine whether they were able to complete the main experiment. First, a written quiz (multiple-choice) was administered which tested participants' ability to recognize the symptom patterns of each fault. Participants were required to answer 9 of 11 questions correctly in order to continue. After the test, the experimenter addressed any questions that were answered incorrectly, probing for the reason for misunderstanding. Where required, example trials were run for those specific faults, with the experimenter providing coaching. Up to 30 minutes were required for this process. Second, participants were required to successfully complete 3 practice trials while the experimenter observed passively. This portion required approximately 15 minutes. The total training time was 2.5 - 3 hours for each participant.

3.4.1.2 Session 2: Main Experiment

The second session was allotted 6 hours and comprised several refresher exercises, introduction to the SA method, familiarization with AFIRA, and the main experiment. Preparation for the main experiment required approximately 30 minutes to one hour. The main experiment spanned 4 hours.

The refresher exercises were intended to ensure that participants remembered the fault patterns for each fault. Participants were allowed to review the training material and worked through several practice trials. The experimenter worked with participants at this stage as a secondary check that participants were capable of completing the main experiment.

The introduction to the SA method explained that the simulator would freeze periodically to pose eight true/false questions to the operator, each paired with a confidence rating. The questions were listed in advance so that any necessary clarifications could be made. Participants were informed that the confidence ratings would not affect their performance bonus and asked to respond as truthfully as possible. To standardize the bias measure across participants, the perceived costs and benefits of hits/misses/false-alarms/correct-rejections were explicitly outlined. Participants were informed that an incorrect true has the same value as an incorrect false. Participants practiced responding to one SA probe before beginning the main experiment.

Participants were introduced to AFIRA with a handout specific to their experimental condition (see Appendix E), which explained the basic functions that the aid performed as well as the functions specific to their condition only. Participants were explicitly informed that the diagnosis function was generally reliable, but not perfect, and that they should always verify the aid's recommendations. No numerical value was provided to indicate the aid's reliability. Participants then worked with the aid to resolve a practice fault so that they would be familiar with the format of AFIRA's recommendations. Only one practice trial was used so that participants' strategies for working with the aid did not evolve before data collection began. Following the familiarization stage, participants began the main experiment.

3.4.2 Trial Design

Each 40 minute block of trials contained six scripted faults, spaced randomly but with sufficient time between that operators could be expected to resolve the faults successfully based on the pilot trials (see Appendix F). This increased the monitoring time between faults, but was required to ensure that less proficient operators did not cause an exception fault when the system began a new event before the previous one had been resolved. The scripts were formulated based on the criteria used in a prior study; namely that each block should contain one valve-stuck-open fault, one sensor fault, and the remaining four composed of valve blocks and valve leaks, randomly selected (Manzey, Reichenbach, & Onnasch, 2008). According to Manzey et al., this is sufficient to ensure that fault identification and management demands were equivalent for all blocks. All participants worked with the same fault scripts.

Chapter 4

Analysis

4 Chapter 4 - Analysis

4.1 Expected Effects

As described in section 3.1 Hypotheses, the analysis was split into two parts to investigate separate but related hypotheses. Table 5 and Table 6 reiterate these hypotheses and summarize the sub-hypotheses for each group of dependent variables. The subsequent sections discuss each part of the experiment and detail specific expected effects.

Table 5: Hypotheses (Part I)

Component	Hypothesis 1: An effect on goal performance and situation awareness is expected only where automation directly augments a function that is a means to that end goal, or indirectly by alleviation of workload.
Routine Performance	H1a: Automation is expected to improve routine performance where automation either directly aids performance or alleviates workload
Workload	H1b: Workload is expected to decrease as DOA increases
Situation Awareness	H1c: SA is expected to benefit only from information analysis automation H1d: SA confidence is expected to be affected by all DOA H1e: Response bias is expected to be affected by all DOA
Failure Performance	H1f: Failure performance decrement is expected for one or more performance measures at the implementation level

Table 6: Hypotheses (Part II)

Variable	Hypothesis 2: Impacts on situation awareness should be mediated by reliance.
Trust	H2a: Trust in the automation is expected to decrease following the automation failure
Monitoring	H2b: Participants are expected to increase sampling following the automation failure
Situation Awareness	H2c: SA is expected to decrease following the automation failure due to disuse of automation H2d: SA confidence is expected to decrease following the failure event H2e: Bias is expected to become more liberal following the failure event

4.1.1 Part 1

Table 7 outlines the predicted effects that would support the hypotheses for Part I. Note that predicted changes are stated relative to the previous DOA level, as the support provided by all lower levels is included. Predictions for information analysis are therefore relative to the manual control group baseline. Also note that the blue “increase” indicator refers to an improvement in performance measures rather than an absolute increase as the system goals are to minimize the value of some measures and maximize others. For example, out-of-target-error and diagnosis accuracy should be minimized and maximized respectively.

Table 7: Predicted Effects (Part I)

Component	Measure	Information Analysis	Action Selection	Action Implementation
Routine Performance	Fault identification time	↑	×	↑
	Diagnosis accuracy	↑	×	↑
	Out-of-target-error	↑	↑	↑
	Prospective memory	↑	↑	↑
	Simple reaction time	↑	↑	↑
Workload	NASA-TLX	↓	↓	↓
Situation Awareness	SA	↑	×	×
	SA Confidence	↑ / ↓	↑ / ↓	↑ / ↓
	Response Bias	↑ / ↓	↑ / ↓	↑ / ↓
Failure Performance	Fault identification time	×	×	↓
	Diagnosis accuracy	×	×	↓
	Out-of-target-error	×	×	↓
	Prospective memory	×	×	↓
	Simple reaction time	×	×	↓

- ↑ Predicted increase
- ↓ Predicted decrease
- ↑ / ↓ Non-directional prediction
- × No change predicted

4.1.1.1 Routine Performance

Automation was expected to improve performance measures associated with functions directly aided by the automation, and where the automation alleviated workload. Through one or both of these mechanisms, all levels of automation were expected to impact performance with two exceptions. The action selection level cannot be expected to improve diagnostic performance (fault identification time and diagnosis accuracy) as the automated function does not directly aid diagnosis nor alleviate workload during the diagnosis phase by supporting other functions. In contrast, the information

analysis level directly aids diagnosis through automated recommendations and the action implementation level aids diagnosis through the option to rapidly implement the repair order. All three levels have the potential to aid out-of-target-error, either through support to parameter management or by reducing the total amount of time spent in a fault state by reducing fault identification time. For the secondary performance measures (prospective memory and simple reaction time), which were performed during all phases, all levels of automation had the potential to improve performance by alleviation of workload.

4.1.1.2 Workload

Independent of whether performance was aided, all AFIRA levels had the potential to reduce workload as tasks are offloaded from the operator. The workload effects cannot be localized to the diagnosis or parameter management phases as workload was collected for each block in its entirety. Thus, decreases in subjective workload may not directly correspond to performance improvements.

4.1.1.3 Situation Awareness

Wickens et al. suggest that maintaining awareness of the system state may mitigate degradation of failure performance, citing the meta-analysis findings that SA was directly correlated with routine and failure performance (Wickens, Li, Santamaria, Sebok, & Sarter, 2010). The conventional wisdom is that increasing levels of automation will improve SA if the automation supports information gathering activities or alleviates workload such that the operator can allocate more attention to information gathering activities. Conversely, higher levels of automation might be expected to hinder SA if the automation causes the operator to reallocate attention to other tasks.

In this study, information analysis automation supported fault diagnosis. Because the function does not aid so much as supplant situation assessment, it was possible that participants' SA could increase or decrease. It was expected that participants would be able to use the recommendations provided by AFIRA's information analysis function to infer the correct responses to SA queries.

The action selection and action implementation levels were not expected to provide further benefits to SA either directly or by reducing workload during the relevant timeframe for two reasons. First, the aid provided by the action selection and action implementation automation did not support situation assessment through acquiring, integrating or processing information in any way. Both are decision/action automation, rather than information automation. Second, neither type of automation was expected to affect SA through reduced workload. Action selection automation could not impact workload during diagnosis as the fault management information provided by that automation is only

pertinent during the fault management phase. The potential for reduction of workload afforded by action implementation automation was minimal because operators were not expected to send a repair order in the short time frame before the SA probe. Furthermore, if participants used inference as the primary mechanism for responding to the probes, as previously discussed, then workload would be less pertinent as operators would not need to manually verify the parameters in order to respond correctly.

SA Confidence could change with the introduction of any type of automation. Increases in confidence were expected to coincide with increases in SA but also potentially with changes in the operator's perceptions, independent of actual SA performance. Action selection and action implementation automation were thus predicted to potentially change SA Confidence, though no stipulations were made as to whether confidence would increase or decrease.

The response bias score reflects how operators handle uncertainty when responding to the SA probes. Participants could use a liberal, neutral or conservative strategy in reporting the presence of a symptom when they were unsure whether that symptom was actually present. It was thought that some participants might only report parameters that they observed and none that they did not observe (conservative bias), rather than guessing with even odds (neutral bias). Therefore a slight conservative bias might be expected on average. A change in response bias could be expected from information analysis automation if a) participants inferred system parameters from the aid's recommendation, and b) there was some difference in bias criterion setting associated with inference as opposed to observation. Any differences between higher levels of DOA would only be expected if there was some propensity to change the observation/inference strategy. As workload was factored out by the experimental method, this was not expected. Though workload was factored out by the experimental method and the higher levels of automation did not directly impact situation assessment, it was thus expected that response bias might yet be affected by all DOA.

4.1.1.4 Failure Performance

Manzey et al. (2008) observed a decrement in failure performance only at the action implementation level and only for out-of-target-error. However, it was not known whether participants in the present study would behave similarly or if failure performance in another area might be affected. Thus, all primary and secondary performance measures were analyzed with the prediction that any could decrease at the action implementation level.

4.1.2 Part 2

Table 8 outlines the predicted effects that would support the hypotheses for Part II regarding the effects of an automation failure on SA performance and interaction with AFIRA.

Table 8: Predicted Effects (Part II)

Component	Measure	Predicted Effect
Trust	Trust in IA function	↓
Monitoring	AVIS	↑
Situation Awareness	SA	↓
	SA Confidence	↓
	Bias	↓

Predicted increase

Predicted decrease

4.1.2.1 Trust

Experience of an automation failure tends to decrease the operator's trust in that automation (Lee & See, Trust in Automation: Designing for Appropriate Reliance, 2004). The first step in Part II was to establish that the known phenomenon was occurring. Thus, it was predicted that operators would report lower trust following the automation failure than in the block preceding it.

4.1.2.2 Monitoring

Assuming that a decrease in trust was observed following the automation failure, it was expected that verification of the automation would increase as participants became more wary of the AFIRA's recommendations. This would indicate that operators relied less on the recommendation, opting instead to observe the parameters manually.

4.1.2.3 Situation Awareness

It was expected that participants would be unable to achieve the same level of SA performance when choosing to access parameters manually. Thus, it was predicted that SA performance for the automated groups would decrease in block 4 to a level lower than that observed in the blocks preceding the automation failure. It was not expected that performance would drop to the level of the manual group as participants would likely still use the automated recommendation to some degree.

SA Confidence was expected to decrease following the automation failure as participants would be less sure whether the information inferred from AFIRA's recommendations was reliable.

Following the automation failure, participants were expected to rely on AFIRA less, in particular the information analysis function. Thus, any change in bias observed with the introduction of information

analysis automation in Part I was expected to be reversed following the failure. For simplicity, we have assumed that bias would increase (indicating a conservative shift) in Part I, and thus predicted a decrease in Part II.

4.2 Statistical Analysis

This section outlines the approach to data analysis. The primary tool used was analysis of variance (ANOVA). All analyses were performed with an alpha criterion of .05.

4.2.1 Part 1

In Part 1, single-factor between-subjects ANOVA was used to analyze the components of the trade-off. For the components relevant during routine operation (routine performance, workload and SA), average performance across blocks 2-4 was analyzed. For failure performance, the difference between baseline performance (block 1) and withdrawal of automation (block 5) was analyzed. Repeated contrasts comparing each DOA group to the preceding one were used to further investigate significant effects. Additionally, bivariate correlational analyses were performed for pairs of variables of interest.

In the event that assumptions of ANOVA were violated, the following contingency plans were used. For violations of normality (as indicated by the Shapiro-Wilk test), no action was taken as ANOVA is generally robust to violations of normality. For violations of the assumption of homogeneous variances (as indicated by Levene's test), the data were transformed if possible using a reciprocal, natural log or square root transformation. Where this was not possible, an alternate approach was necessary. Though they are often used as such, many non-parametric analyses including Mann-Whitney, Kruskal-Wallis and Jonckheere-Terpstra tests are unsuitable as they assume identical (if not normal) distributions in all groups (Sheskin, 2003). A suitable compromise was achieved with Welch's independent samples t-test (e.g. Ruxton, 2006), which is robust to differences in variances across groups. Where necessary, three independent sample t-tests were used to compare each DOA group to the preceding one. Though this is not ideal, it was considered better practice to run multiple analyses than to violate test assumptions.

4.2.2 Part 2

In Part 2, a 3x3 mixed model ANOVA was used to evaluate trust, monitoring, situation awareness and confidence over the portion of the experiment when participants worked with AFIRA. For violations of sphericity across the repeated measures variable (block), the Greenhouse-Geisser correction was applied. Violations of normality were dealt with as in Part I. When variances were not equal between groups, the nonparametric Friedman's test was used to analyze all automated groups together. Though Friedman's test is not entirely suitable for use with heterogeneous variances, it is considered more

robust than ANOVA (Maxwell & Delaney, 2003). This alternate approach was required so that dynamics relating to the first failure event could be observed, though the ability to detect interaction effects was sacrificed.

4.2.3 Summary

Table 8 summarizes the statistical approach for each part. Table 10 outlines the statistical tests to be used if the assumption of homogeneous variances is not met.

Table 9: Statistical Approach

Part I		
Measure	Blocks Analyzed	Introduction of Automation
Routine Performance / Workload / Situation Awareness/ Failure Performance	Average of 2-4	Single Factor ANOVA (4 Levels)
Failure Performance	Difference of 1-5	
Part II		
Measure	Blocks Analyzed	Response to Automation Failure
SA / SA Confidence / Sampling	2, 3, 4	3(Block)x3(DOA) Mixed ANOVA

Table 10: Alternate Nonparametric Statistical Approach

Part I			
Measure	Blocks Analyzed	Introduction of Automation	Variation of DOA
Routine Performance / Workload / Situation Awareness/ Failure Performance	Average of 2-4	Welch's t-test (Repeated comparisons of each group to preceding)	
Failure Performance	Difference of 1-5		
Part II			
Measure	Blocks Analyzed	Response to Automation Failure	
Trust / Monitoring / SA / SA Confidence	2, 3, 4	Friedman's test	

4.2.4 Power Analysis

A power analysis was conducted to determine the number of participants required to detect the smallest effects of interest using the findings of Manzey et al. (2008). The inputs to the analysis were partial eta squared (η^2) values observed for differences in routine performance of fault-identification-time and failure performance of out-of-target-error. The output of the analysis for routine performance is shown in Figure 9 and for failure-performance in Figure 10. As the expected correlation between repeated measures of each variable was unknown, three moderate values (0.4, 0.5, 0.6) were selected. Following convention, the minimum sample size was taken as the lowest N necessary to yield power of 0.80. The more conservative estimate was given by the failure performance effect size, which suggested that no fewer than 22 subjects should be used. This is a conservative estimate as i) the failure performance effect was the smallest observed, and ii) correlation between repeated measures of the performance metrics should be fairly high as the trial blocks were standardized as much as possible. In order to have four DOA groups of equal size, the sample needed to be a multiple of four. Therefore, a sample size of 24 was selected.

Figure 9: Power Analysis (Routine Performance – Fault-Identification-Time)

Figure 10: Power Analysis (Failure Performance – Out-of-Target-Error)

Chapter 5

Results

5 Chapter 5 - Results

5.1 Data Corrections

5.1.1 Removal of Participants

28 students participated in the study, 24 of whom reached completion. No participants withdrew but four had to be removed prior to completion for one of the following reasons:

1. The participant(s) were unable to perform the task competently by the end of training (1).
2. The participant(s) caused the simulator to crash by taking too long to resolve a fault (3). In this event, participants could not be asked to repeat the block as they could be expected to quickly realize that the same script was taking place. A replacement block with a different script was not run as the data collected might not be comparable.

Partial data from the 4 participants who were removed were not included in the study as the effects of interest required participants to have experienced all manipulations. The participants who had to be removed completed less than 50% of the main experiment phase.

5.1.2 Missing Data

Separate from the participants who did not complete the experiment, a small quantity of data were missing for various reasons (see Appendix G for a full account).

5.1.3 Outliers

Separate from the participants who did not complete the experiment, two participants' data were removed after the fact (see Appendix G for a full account).

5.2 Part One

The results are organized according to the four components of the routine-failure trade-off: routine performance, workload, situation awareness and failure performance. As outlined in Chapter 4, each category (except workload) comprises more than one variable. Table 11 shows the results of the single-factor ANOVA and contrasts. For the analyses where variances were significantly different between groups (as indicated by Levene's test), independent samples t-tests were performed instead (see Table 12). This was necessary for diagnosis accuracy, out-of-target-error and SA. Results of normality tests are contained in Appendix H.

Table 11: Between-Groups Effects – Single Factor ANOVA & Contrasts

Measure		Levene's		ANOVA				
		<i>F</i>	<i>p</i>	<i>df</i>	<i>df_{error}</i>	<i>F</i>	<i>p</i>	η_p^2
Routine Performance	Diagnosis Accuracy	3.28	.045*					
	Fault Identification Time	2.99	.059	3	18	24.2	.000*	.80
	Out-of-Target-Error	3.76	.029*					
	Prospective Memory	2.05	.14	3	18	0.71	.56	.11
	Simple Reaction Time	1.07	.39	3	16	0.87	.48	.14
Workload	NASA-TLX	1.50	.25	3	18	11.0	.000*	.65
Situation Awareness	SA	3.64	.033*					
	SA Confidence	1.19	.34	3	18	1.64	.22	.21
	Bias	2.37	.11	3	17	3.55	.037*	.39
Failure Performance	Diagnosis Accuracy	0.79	.52	3	17	1.50	.25	.21
	Fault Identification Time	0.17	.91	3	17	1.73	.20	.23
	Out-of-Target-Error	0.45	.72	3	17	1.15	.36	.17
	Prospective Memory	2.70	.078	3	17	0.52	.68	.084
	Simple Reaction Time	0.38	.77	3	15	0.97	.43	.16
Measure		Contrast		<i>p</i>				
Routine Performance	Diagnosis Accuracy	Group 1 – Group 2		.000*				
		Group 2 – Group 3		.68				
		Group 3 – Group 4		.52				
Workload	NASA-TLX	Group 1 – Group 2		.34				
		Group 2 – Group 3		.35				
		Group 3 – Group 4		.000*				
Situation Awareness	Bias	Group 1 – Group 2		.021*				
		Group 2 – Group 3		.67				
		Group 3 – Group 4		.21				

Group 1: Manual

* significant ($p < .05$)

Group 2: IA

Group 3: IA + AS

Group 4: IA + AS + AI

Table 12: Between-Groups Effects - Independent t-tests (Equal Variances Not Assumed)

Measure	Comparison	<i>df</i>	<i>t</i>	<i>p</i>	ΔM	ΔSD
Diagnosis Accuracy	Group 1 – Group 2	7.27	-4.98	.001*	-0.13	0.026
	Group 2 – Group 3	7.31	-0.79	.46	-0.011	0.014
	Group 3 – Group 4	5.93	1.56	.17	0.028	0.018
Out-of-Target-Error	Group 1 – Group 2	8.26	2.33	.047*	140	60.3
	Group 2 – Group 3	4.26	1.74	.15	82.1	47.2
	Group 3 – Group 4	4.62	-0.16	.88	-5.18	31.5
Situation Awareness	Group 1 – Group 2	5.84	-2.89	.028*	-36.5	12.6
	Group 2 – Group 3	7.04	-0.016	.99	-0.068	4.20
	Group 3 – Group 4	7.58	0.15	.88	0.59	3.90

Group 1: Manual

*significant ($p < .05$)

Group 2: IA

Group 3: IA + AS

Group 4: IA + AS + AI

Significant effects were found for routine Diagnosis Accuracy, Fault Identification Time, Out-of-Target-Error, NASA-TLX and SA and response bias. However, only NASA-TLX scores varied between different levels of AFIRA. All other variables showed an effect only with the introduction of information analysis automation. These effects are discussed in greater detail later in this chapter.

5.2.1 Routine Performance

5.2.1.1 Diagnosis Accuracy

Figure 11 shows diagnosis accuracy during routine performance.

Figure 11: Routine Performance (Diagnosis Accuracy)

With the introduction of information analysis automation, diagnosis accuracy increased to nearly 100% for all groups working with the aid (see Figure 11). There were no significant differences between different levels of AFIRA.

5.2.1.2 Fault Identification Time

Figure 12 shows fault-identification-time during routine performance. As with diagnosis accuracy, fault-identification-time improved dramatically when working with the automated aid, but no differences were observable between different levels of AFIRA.

Figure 12: Routine Performance (Fault Identification Time)

5.2.1.3 Out-of-Target-Error

Figure 13 shows out-of-target-error during routine performance. Similar to the measures of diagnostic performance, introduction of the information analysis automation improved out-of-target-error performance. No significant differences were detected between any of the automation-supported conditions, though visual inspection of Figure 13 suggests that the groups with action selection and action implementation support achieved lower mean scores.

Figure 13: Routine Performance (Out-of-Target-Error)

It is notable that the only significant effect on out-of-target-error occurred with the introduction of information analysis support. Information analysis supports only diagnosis and not the use of manual control to mitigate out-of-target-error. Conversely, the automation designed to assist with parameter management (action selection and action implementation) provided minimal (if any) benefits. Therefore, a correlational analysis was used to investigate the influence of diagnosis performance on out-of-target-error (see Table 13).

Table 13: Correlation of Primary Performance Measures

	Fault-Identification-Time	Diagnosis Accuracy
Out-of-Target-Error	$r = .800$ $p < .001^*$ $N = 24$	$r = -.544$ $p = .006^*$ $N = 24$

*significant ($p < .05$)

The correlational analysis revealed that both fault-identification-time and diagnosis accuracy were strongly correlated with out-of-target-error. This suggests that the diagnosis task in AutoCAMS is largely

responsible for all primary performance measures. If so, this represents a limitation of the platform in that each level of automation does not necessarily provide incremental benefits.

5.2.1.4 Summary of Routine Performance

Overall, the predicted effects were observed. However, there were three main caveats:

1. No significant effect of action selection and action implementation on out-of-target-error
2. No significant effect of action implementation on diagnostic performance
3. No effects were found for secondary performance measures

5.2.2 Workload

A significant difference in workload was detected when the manual group was included as well as when it was excluded. However, from Figure 14 it is apparent that the effect was attributable to the action implementation group only. The lower workload reported by the highest DOA group indicates that action implementation automation did yield benefits, despite not providing significant benefits to routine performance. The lack of change in workload from information analysis and action selection automation suggests that any workload alleviated was offset by the requirements of working with the automation to achieve better performance. Therefore, the prediction that higher levels of automation would reduce workload was only partially supported.

Figure 14: Workload (NASA-TLX)

5.2.3 Situation Awareness

5.2.3.1 SA

Figure 15 shows the SA scores during routine performance. Similar to the performance measures, introduction of information analysis automation aided SA. The second single-factor ANOVA did not detect a significant difference between any of the automation supported conditions

Figure 15: Situation Awareness (SA)

To confirm that participants used inference to inform their SA responses, the information sampling behaviour was also analyzed. Figure 16 shows the percentage of necessary parameters sampled during the diagnosis phase. Participants working with AFIRA did not sample all of the parameters required to make a diagnosis manually, whereas participants in the control group consistently did. Therefore, the repair orders sent by the groups working with AFIRA were at least somewhat reliant on the aid as the operators had not made sufficient observations to confirm a single diagnosis. Their higher SA was achieved by reliance on the aid. This illustrates how participants can have high SA under routine conditions without necessarily being robust to failures of automation. The improved SA in this case appears to be attributable to use of the automation, and not independent sampling.

Figure 16: Automation Verification Information Sampling (AVIS)

5.2.3.2 Bias

Figure 17 shows the mean bias scores during routine performance. Introduction of information analysis support caused a conservative shift in bias. Bias scores had more random variation than the SA measure and did not reveal any additional information about participants' behaviour.

Figure 17: Situation Awareness (Response Bias)

A correlational analysis was conducted to gain further insight into the SA results (see Table 14). There were two notable findings. First, SA Confidence was not correlated with either SA or Bias. Second, bias was significantly correlated with SA, though not perfectly, implying that there were some individual differences in participants' responses that did not directly relate to the accuracy of their judgements.

Table 14: Correlational Analysis of Situation Awareness Measures

	Bias	SA Confidence
SA	$r = .525$ $p = .008^*$ $N = 24$	$r = .291$ $p = .168$ $N = 24$
SA Confidence	$r = .277$ $p = .190$ $N = 24$	

* significant ($p < .05$)

5.2.3.3 Summary of Situation Awareness

The prediction that situation awareness would be improved only by information analysis automation was confirmed. There was no evidence that SA decreased for higher levels of automation. Awareness was at least partially based on inference as indicated by participants' decreased sampling. During the entire diagnosis process, participants working with the aid accessed less than 100% of the system parameters required to make the correct diagnoses and answer the SA probes, but still managed to score higher on the probes than the manual group. Participants without the aid continued to access 100% of the necessary parameters.

The prediction that SA confidence might change for all levels of automation was not substantiated. Not only was there no difference between the AFIRA groups, the introduction of AFIRA did not produce a significant effect. This suggests that participants' confidence in their response was not directly related to their actual performance. This was confirmed by a lack of correlation between the two measures.

Similarly, response bias was not affected by different levels of AFIRA, contrary to predictions. However, bias was affected by introduction of the decision aid. This implies that working with the aid influenced how participants dealt with uncertainty in answering the SA probes.

5.2.4 Failure Performance

5.2.4.1 Failure Performance

As shown by the lack of significant findings in Table 11 and Table 12, no significant decrement in failure performance was observed for any group of participants on any performance measure, relative to the manual baseline. For reference, graphs of failure performance are contained in Appendix I.

5.2.5 Summary of Part 1 Findings

Table 15 summarizes the support for the predictions supporting hypotheses **H1a-H1f**. Expected effects are shown by arrows against a green or red background indicating whether the prediction was supported or not.

Table 15: Support for Part 1 Sub-Hypotheses

Component	Measure	Information Analysis	Action Selection	Action Implementation
Routine Performance	Fault identification time	↑	×	↑
	Diagnosis accuracy	↑	×	↑
	Out-of-target-error	↑	↑	↑
	Prospective memory	↑	↑	↑
	Simple reaction time	↑	↑	↑
Workload	NASA-TLX	↓	↓	↓
Situation Awareness	SA	↑	×	×
	SA Confidence	↑ / ↓	↑ / ↓	↑ / ↓
	Response Bias	↑ / ↓	↑ / ↓	↑ / ↓
Failure Performance	Fault identification time	×	×	↓
	Diagnosis accuracy	×	×	↓
	Out-of-target-error	×	×	↓
	Prospective memory	×	×	↓
	Simple reaction time	×	×	↓

5.3 Part Two

In this section, the effects of the first failure event were analyzed. Results were analyzed using a 3(Block) x 3(DOA) ANOVA across blocks 2-4. The main block effects are shown in Table 16. Where the assumption of homogeneous variances was violated, Friedman's test was used to analyze the automated groups together (see Table 17). This was only necessary for subjective trust. Results of normality tests are contained in Appendix H.

Table 16: Block Effects - 3(Block) x 3(DOA) ANOVA

Measure	Block	Levene's		ANOVA				
		F	p	df	df _{error}	F	p	η_p^2
Trust (IA)	2	3.94	.046*					
	3	3.67	.054†					
	4	1.36	.29					
AVIS	2	3.05	.082	2	26	8.63	.001*	.40
	3	1.12	.36					
	4	2.99	.085					
SA	2	0.19	.83	2	26	2.77	.081†	.18
	3	0.83	.46					
	4	0.35	.71					
SA Confidence [§]	2	0.68	.53	1.236	16.074	2.44	.13	.16
	3	0.11	.89					
	4	1.70	.22					
Bias	2	0.42	.67					
	3	5.58	.019*					
	4	1.80	.21					

* significant ($p < .05$)

† marginally significant ($.05 \leq p \leq .10$)

§corrected for sphericity ($X^2(2) = 11.5$, $p = .003$), $\epsilon_{GG} = 0.62$

Table 17: Nonparametric Tests - Friedman's Test, 3 Levels

Measure	df	N	χ^2	p
Trust (IA)	2	16	11.9	.003*
Bias	2	15	1.09	.58

* significant ($p < .05$)

Significant effects were found for trust and AVIS, but not for bias or confidence. Additionally, a marginally significant effect was found for SA. Therefore, a follow-up analysis was performed, comparing block 2 to block 3 and block 3 to block 4 separately. Although the manual group could not be included in the analysis of trust in automation (participants did not work with AFIRA) or AVIS (all manual participants sampling 100% of parameters for every fault), the manual group SA data was

available. Including this data revealed that there was no effect on SA due to Block alone, nor a Block*DOA interaction (see Table 18).

Table 18: Follow-up on Marginal Situation Awareness Result – 2(Block) x 3(DOA) ANOVA

Measure	Contrast	Effect	df	df _{error}	F	p	η_p^2
SA	Block 2 – Block 3	Block	1	18	1.15	.30	.060
	Block 3 – Block 4	Block	1	18	3.01	.10	.14
	Block 2 – Block 3	Block*DOA	3	18	0.50	.69	.076
	Block 3 – Block 4	Block*DOA	3	18	0.47	.71	.072

5.3.1 Trust

Trust in AFIRA's ability to perform the information analysis function significantly decreased following the automation failure (see Figure 18), confirming the predicted effect.

Figure 18: First Failure Response - Trust in AFIRA IA

5.3.2 Verification Sampling

A significant increase in verification was observed in the block following the automation failure (see Figure 19). Together with the observed decrease in trust, this finding indicates that the conventional expectations regarding the effects of operator perceptions of automation reliability were upheld.

Figure 19: Response to First Failure (AVIS)

5.3.3 SA

Contrary to expectations, no significant effects were detected for SA (see Figure 20, left). Though the analysis of the automated groups alone indicated a marginally significant increase in SA, including the manual group revealed no significant main effects or interactions. As the significant difference between the manual group and all AFIRA-aided groups was maintained (see Figure 20, right), it was concluded that participants continued to use the recommendation provided by the information analysis automation. No significant effects were detected for response bias or SA confidence.

Figure 20: Response to First Failure (SA)

5.3.4 Summary of Part 2

Table 19 summarizes the evidence supporting hypotheses **H2a-H2e**. Though the first failure event did impact participants' trust in the automation as well as sampling behaviour, it did not change how participants dealt with the SA probes. This suggests that reliance on automation and the SA benefits realized thereby are not as strongly influenced by operator attitudes as sampling and subjective trust

reports are. Instead, other factors may dictate the operator's reliance strategy. In the following chapter, I postulate that another important factor may be the perceived importance of functions with regard to goal achievement.

Table 19: Summary of Findings in Part 2

Component	Measure	Predicted Effect
Trust	Trust in IA function	↓
Monitoring	AVIS	↑
Situation Awareness	SA	↓
	SA Confidence	↓
	Bias	↓

Chapter 6

Discussion

6 Chapter 6 - Discussion

Table 20 and Table 21 summarize the empirical support for the hypotheses introduced in Chapter 4.

Table 20: Hypotheses (Part I)

Component	Sub-Hypotheses	Findings
Routine Performance	H1a: Automation is expected to improve routine performance where automation either directly aids performance or alleviates workload	Partially supported
Workload	H1b: Workload is expected to decrease as DOA increases	Partially supported
Situation Awareness	H1c: SA is expected to benefit only from information analysis automation	Supported
	H1d: SA confidence is expected to be affected by all DOA	Not supported
	H1e: Response bias is expected to be affected by all DOA	Not supported
Failure Performance	H1f: Failure performance decrement is expected for some performance measures at the AI level	Not supported

Table 21: Hypotheses (Part II)

Variable	Sub-Hypotheses	Findings
Trust	H2a: Trust in the automation is expected to decrease following the automation failure	Supported
Monitoring	H2b: Participants are expected to increase sampling following the automation failure	Supported
Situation Awareness	H2c: SA is expected to decrease following the automation failure due to disuse of automation	Not supported
	H2d: SA confidence is expected to decrease following the failure event	Not supported
	H2e: Bias is expected to decrease following the failure event	Not supported

6.1 Part 1

The routine-failure trade-off postulated by Wickens et al. (2010) suggests that increasing level of automation results in a trade-off of routine performance and failure performance in the event of an automation failure. Increasing level of automation was also posited to decrease workload and cause a loss of situation awareness. These relationships are summarized by the simple illustration in Figure 21 (left). Based on contextual factors of the specific system at hand, somewhat different predictions were

made for AutoCAMS. These are illustrated in Figure 21 (centre). Finally, Figure 21 (right) illustrates the relationships that were observed. The main difference between the Wickens' proposed routine-failure trade-off and the predictions made in this study was that situation awareness was expected to benefit from information analysis automation and otherwise be unaffected by LOA. The predictions were generally supported by the experimental results, except that i) *higher levels of automation had essentially no effect on any measures*, and ii) *a failure performance decrement was not observed*. These two caveats may be interrelated. I first discuss each of the four components of the trade-off before addressing potential explanations for the deviation from the expected results.

Figure 21: Wickens' Postulated (Left), Predicted (Centre) and Observed (Right) Performance Trade-off

6.1.1 Routine Performance

During routine operation with the aid, an improvement was observed for all primary task performance measures when information analysis support was introduced. However, action selection and action implementation automation did not provide any additional benefit. This null effect was unexpected as Manzey et al. (2008) observed a significant improvement in both fault-identification-time and out-of-target-error due to action implementation support. The only substantial difference in the experimental task between that study and this study was the addition of the SA probe method. All other modified aspects took place during the breaks between blocks and thus were unlikely to have affected performance. The most plausible explanation is that the presence of the SA probes somehow negated the benefits of the action implementation automation. It is possible that the need to score well on the SA probes prompted participants to place more emphasis on awareness of system parameters. This requirement may have disallowed participants from relying uncritically on AFIRA. This in turn suggests that uncritical reliance was necessary to realize the benefits of action implementation automation observed by Manzey et al. (2008), meaning that the benefit provided by action implementation automation in AutoCAMS was primarily to reduce fault identification time by bypassing the diagnosis

task. The large impact on out-of-target-error may be attributable to the total amount of time spent in a fault state being reduced.

6.1.2 Situation Awareness

The predicted improvement in SA was confirmed, contrasting with the progressive loss of SA originally stipulated by the routine-failure trade-off. SA was aided only by automation that directly supported situation assessment (information analysis); further automation of other functions did not affect SA. This supports the argument that SA is the product of specific situation assessment activities (Endsley & Garland, 2000) rather than simply a by-product of being directly involved in control tasks.

Reported confidence in the SA queries did not correspond to either SA or bias, and was not affected by DOA. In contrast, bias was strongly correlated with SA and did not seem to reveal any other meaningful information with regards to participants' behaviour.

6.1.3 Workload

Workload only significantly decreased with the introduction of action implementation support. Unexpectedly, participants in the other two DOA groups did not report lower workload than the no-support group. Working with the automation to infer the correct responses to SA queries may have required significant cognitive resources. The decreased workload experienced by the highest DOA group suggests that automatic implementation was used, despite providing no significant performance benefit. The AI participants might have used this function to a lesser degree than participants in the prior study by Manzey et al. (2008), perhaps relying on it for parameter management but not rapid diagnosis. That parameter management performance (out-of-target-error) was unaffected further suggests a problem with the operationalization of automation in AutoCAMS. Specifically, it seems that all of the primary performance measures stem from the speed with which participants performed diagnosis, rather than any other actions. This was supported by the strong correlation found between fault identification time and out-of-target-error.

6.1.4 Failure Performance

Failure-performance decrement was not observed in any DOA condition. In a prior CAMS study, the action implementation group experienced a failure-performance decrement in out-of-target-error (Manzey, Reichenbach, & Onnasch, 2008). It is striking that the prior study also found routine-performance benefits (out-of-target-error and fault identification time) associated with action implementation automation which were not observed here. This further supports both the disuse of

action implementation automation during the diagnosis phase as well as the possible interdependence of all primary performance measures in AutoCAMS.

6.1.5 General Findings

In general, the routine-failure trade-off was observed, but participants effectively only utilized two levels of automation. I posit that this was an artefact of the operationalization of automation in CAMS as well as the influence of the SA method. These interpretations are discussed in sections 6.1.5.1 Operationalization of Automation and 6.1.5.2 Influence of SA Method, respectively.

6.1.5.1 Operationalization of Automation

First, one rationale for the apparent utilization of only two levels of automation relates to the fact that the primary tasks in CAMS are not independent. The speed of diagnosis was known to have a substantial effect on all of the performance measures. This would not be a problem in itself if the automation intended to improve performance during the fault management phase (action selection and action implementation) provided additional benefits. However, this was not the case.

Action selection automation provided a recommended course of action for each fault. While this theoretically should provide a performance benefit, in practice the aid provided was minimal. The action selection support was essentially a reminder, but participants were trained specifically to internalize this simple information. The recommended fault management actions are to manipulate the parameter affected by the fault in one of two ways: If it is a matter of flow rates being low, the operator increases the setting. If the controller cannot manipulate the on/off valve, the operator must do it manually. So long as the correct fault was identified, there is very little room for error in performing this aspect of the task. Action implementation automation executes these actions, but this added benefit is also of limited value. The first control strategy, increasing the flow setting, is a one-time adjustment. The action requires two clicks and only a few seconds to implement. It is difficult to argue that this has a substantial effect on workload. The second control strategy, manipulating the on/off valve manually, requires somewhat more attention. However, this strategy is required a maximum of twice per block, out of six faults, due to the distribution of faults throughout the experiment. Thus the potential for achieving better fault management performance through alleviation of workload or more precise control is small compared to the benefit of reducing the amount of time that the system is in a fault state. Every second longer that the operator spends diagnosing adds to out-of-target-error.

Therefore, it seemed that only the speed of diagnosis had a significant effect on any of the performance measures. Automation such as the action selection level, which affects only the fault management phase, could have no effect. The action implementation level could only have an effect if participants rapidly accepted AFIRA's diagnosis, in direct contradiction to the instructions to manually verify the recommendations.

The automation design thus undercut the usefulness of AutoCAMS substantially. The DOA factor did not produce strong findings because there was little practical distinction between levels. In future work, it is recommended that researchers ensure that the aid provided by automation is well matched to the demands of the experimental task.

6.1.5.2 Influence of SA Method

The second point relates to the performance benefits associated with uncritical reliance. The failure performance decrement observed by Manzey et al. (2008) at the action implementation level suggests that rapid (and ostensibly uncritical) acceptance of AFIRA's recommendation is associated with a decrease in operators' ability to manually manage faults. This experiment sought to find a link between SA and failure performance decrement, but the process of measuring SA apparently altered participants' behaviour such that the action implementation function was disused. This was evidenced by a lack of routine performance benefit for action implementation automation. I postulate that the relative perceived importance and feasibility of the different performance goals was not adequately taken into account. The added requirement for participants to respond correctly to the SA queries seemed to implicitly represent a more relevant task to participants than to reduce diagnosis time. For this reason, participants could not take advantage of the benefits of action implementation automation and as such also did not incur the costs. The task hierarchy presented in Chapter 3 is reproduced here with the addition of the SA maintenance goal to illustrate (see Figure 22).

Figure 22: CAMS Task Hierarchy

As evidenced by comparable verification sampling across all DOA, the SA probes appeared to discourage participants from relying uncritically on AFIRA's recommendations. Instead, participants supported by AFIRA sampled significantly less than the manual group, indicating some level of reliance, but still sampled most of the required parameters. This indicates that the participants did rely on AFIRA to a degree but balanced sampling efficiency against SA performance. Such a strategy suggests that there was a perceived cost to each. I postulate that the perceived benefit of rapid diagnosis was outweighed by the perceived benefit of taking sufficient time to correctly answer the SA queries.

This raises an interesting point, namely that selective use of automation seems to be influenced by the perceived importance of system goals. By extension, the functions that are means to fulfilling system goals must also factor into operator strategies. Furthermore, I suggest that the interrelations between functions, rather than characteristics of functions alone, may be relevant to automation usage. For example, it could be that functions less abstracted from the system goals (that is, those that most directly impact goal performance) are more influenced by the perceived importance of goals. It is thus recommended that future work investigates how the perceived importance of different system goals influences automation usage, as well as the characteristics of functions within a task hierarchy that may influence operators' usage of automation.

6.1.6 Part I Summary

The CAMS platform seems to have a limitation in that the main goal of the system (maintenance of safe cabin environment) is unevenly influenced by different branches of the task hierarchy. Thus, the operationalization of automation reduced the effective levels of the DOA variable. Furthermore, the 3 distinct levels observed by Manzey et al. (2008) were further reduced to 2 in this study due to the

perceived importance of maintaining SA. Therefore, the conclusions that can be drawn regarding the routine-failure trade-off are somewhat limited. The most that can be stated from this study is that the postulated trade-off essentially held true with the key exception of SA, which was shown to increase with automation. Nonetheless, these unexpected effects highlight an important point – that task structure impacts the effects of automation, in addition to characteristics of the automation itself. This is an important finding that is likely not limited to the CAMS platform.

6.2 Part 2

Part 2 investigated the impact of the perceived reliability of automation on subjective trust, verification of automation, and SA performance. The expectation was that the experience of an automation failure would lessen participants' reliance on the automation, allowing them to derive only partial benefit from it. This was only somewhat confirmed, as participants' reliance on the automation did not appear to change. Though trust decreased and independent verification of the automation increased, no resultant change in SA performance was observed. It seems that participants acknowledged the fallibility of the automation and took steps to protect against further automation failures, but continued to use the aid. Interestingly, whereas action implementation automation was disused in Part I in favour of better SA performance, use of information analysis automation could not be discouraged in Part II for the same reason. Part II thus provided supporting evidence that the relative perceived importance of system goals may be a factor in selective use of automation.

6.3 Limitations

This study had several possible sources of error:

1. Differences in sample populations: Parasuraman and Riley note that individual differences can play a large role in how operators use automation (Parasuraman & Riley, 1997). The sample population was known to be different from prior CAMS studies (e.g. Manzey et al., 2008) in several ways:
 - a. Participation in this study was not limited to engineering students. It is possible that prior knowledge of control systems, piping/valves and/or gas law relationships would facilitate understanding of CAMS. This may be a source of unexplained variance between participants.
 - b. Participants in this study were summer students and possibly had different motivations for participation as well as different levels of academic capability.

2. Execution: This was the first study by this research group to use the CAMS microworld, so it is likely that there were small differences in how the study was conducted, relative to prior work. In particular, the training process is critical to preparing participants for the main experiment. The training process was developed by the researcher and informed by an older training document translated from German. This reproduction may have altered the efficacy of the training process somewhat. The extent to which the training process differed is not known.
3. Classification of AFIRA: Early in the history of AutoCAMS, it was acknowledged that there was some difficulty in classifying the different AFIRA functions according to the types and levels model (Lorenz, Di Nocera, Rottger, & Parasuraman, 2002). In conducting this research, we noted that the information analysis level of AFIRA might not actually constitute information automation but decision automation as it synthesized the available fault information into a complete diagnosis. While this does qualify as integration of information, it is not a process that integrates the information and presents it to the operator. Instead, AFIRA operates as an opaque decision maker, only offering the operator the output of a decision making process. In this sense, it is possible that useful DOA information was lost in the gap between complete manual performance and the first level of the DOA factor if AFIRA did not actually contain any instances of information automation.
4. External validity: The CAMS microworld, though more complex than simple lab tasks, lacks certain features of real-world control systems. A detailed account of the task features found in real systems is beyond the scope of this section, but some examples will be discussed. In this and many other microworld studies, the research participants were university students. This has the potential to impact findings, but certain elements are expected to be robust. For example, it is reasonable to assume that participants' ability to allocate attention between tasks is a fundamental constraint of human performance that would be observed in any population. However, the effects of expertise or attitudes developed over time are not necessarily replicated in a laboratory study. These factors may be relevant if complacency in monitoring is dependent on long term attitudes rather than the transient states typically studied in a laboratory setting. Other elements not represented in CAMS include unanticipated failures, false alarms, faults undetected by alarms and control of a greater number of subsystems simultaneously.

Chapter 7

Conclusions and
Future Work

7 Chapter 7 – Conclusions and Future Work

This final chapter outlines the high-level findings of this study. Each conclusion is paired with a recommendation for future work detailed in a subsequent section.

In conclusion, the routine-failure trade-off was observed, but problems with the operationalization of automation temper the strength of this conclusion. It seems that effectively only two points on the DOA axis were tested. Across those two points, the relationship between routine performance and workload followed that postulated by Wickens et al. (2010). Thus, the routine-failure trade-off was supported, at least considering the coarse resolution of DOA in this study. The trade-off should be tested across a greater range of degrees of automation (types and levels), guarding against situations where the experimental platform and method predispose operators to use or disuse automation.

Rather than a loss of SA with increasing automation as stipulated by the routine-failure trade-off, an improvement was observed. Taken as a whole, the situation awareness measures suggest that participants were better informed of the state of the system when working with the aid. Thus, SA did not decrease at higher levels of automation, but increased. Future studies should specifically distinguish between automation designed to aid SA and automation designed to aid other aspects of task performance.

The results here cannot support that higher SA is related to robustness to automation failures as a failure performance decrement did not occur. However, in the context of automation that requires uncritical reliance to yield benefits (as in the rapid acceptance using action implementation), the conventional wisdom seemed to hold true. Participants were faced with a trade-off between maintaining SA and reaping the benefits of automation, though the findings are specific to the experimental platform. In interacting with other aids, it is conceivable that failure performance decrement could occur at a level that also improved SA. In this case, uncritical reliance on the aid would continue to improve SA but not protect against failure performance decrement. Conversely, independent verification of system parameters would help in detecting automation failures but fail to achieve the higher levels of SA performance. I posit that the relationship between SA and robustness to automation failures is strongly dependent on the mechanism by which SA is acquired. Further research should thus address whether different mechanisms for situation assessment have a differential effect on the costs and benefits of automation.

The results of this study strongly indicate that use and disuse of different automated functions was prompted by the introduction of a new system goal. We interpreted that participants were unwilling to sacrifice performance on every trial in order to protect against occasional automation failures. Because the consequences of reliance in the event of occasional automation failures were not severe (similar to a single misdiagnosis) relative to the performance decrement associated with disuse of information analysis automation (significantly longer diagnosis time and lower SA on *every* trial), participants made rational automation usage decisions that amounted to consistent use and disuse of the information analysis and action implementation automation, respectively. This selective use and disuse highlights the relevance of goals to reliance - and by extension the costs and benefits of automation. Further work should investigate how the perceived importance of system goals and the functions that are means to accomplish those goals impact reliance on automation of those functions. Furthermore, the relationship between automated and non-automated functions bears additional inquiry, in particular whether functions support a given goal in series or in parallel.

Finally, the operator behaviours exhibited suggest that only a limited range of strategies were available for gathering information pertinent to the SA probes. In general, participants used IA automation if it was available, and did not if it was unavailable. The only other distinct information sampling behaviour was increased verification in block 4, which generally took place after the SA responses. There was no evidence that participants changed how they dealt with uncertainty in terms of response bias or judged their own performance in terms of SA confidence. Thus, response bias and SA confidence measures did not offer any additional insight in this study, perhaps due to the restrictive nature of the task. The AutoCAMS task was likely not sufficiently complex to allow for different strategies of information gathering. In future, a multivariate approach to SA could be investigated in a setting with greater complexity to determine whether any additional insight could be shed on automation reliance behaviour in terms of information gathering.

In closing, the results of this study show that the routine-failure trade-off describes the relationships of costs and benefits of automation under some circumstances but is a simplification that omits certain relevant elements. Additional work is required to test a broader range of automation and broader range of operator reliance and information sampling behaviours. Further study is required to investigate how goals and task structure affect automation interaction.

Bibliography

- Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex-systems. *Human Factors*, 37 (1), 85-104.
- Bagheri, N., & Jamieson, G. (2004). A sampling model to ascertain automation-induced complacency in multi-task environments. *Human Error, Safety and Systems Development*, 131-145.
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19, 775-779.
- Billings, C. E. (1991). *Human-centred aircraft automation: A concept and guidelines*. Moffet Field, CA: NASA Ames Research Center.
- Cohen, M. S., Parasuraman, R., & Freeman, J. T. (1998). Trust in decision aids: A model and its training implications. (pp. 1-37). Proc. Command and Control Research and Technology Symp.
- Crocoll, W. M., & Coury, B. G. (1990). Status or recommendation: Selecting the type of information for decision aiding. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 34, pp. 1524-1528. Orlando: Human Factors.
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58 (6), 697-718.
- 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. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *International Journal of Aviation Psychology*, 3, 157-168.
- Endsley, M. R. (1996). Automation and Situation Awareness. In R. Parasuraman, & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum.
- Endsley, M. R. (2001). Designing for situation awareness in complex systems. *Proceedings of the Second international workshop on symbiosis of humans, artifacts and environment*. Kyoto.

- Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley, & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 147-173). Mahwah, NJ: Erlbaum.
- Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors*, 37, 65-84.
- Endsley, M. R. (2006). Situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 528-542). New York: Wiley.
- Endsley, M. R. (2004). Situation awareness: Progress and directions. In S. Banbury, & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory, measurement and application* (pp. 317-341). Aldershot, UK: Ashgate Publishing.
- Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32-64.
- Endsley, M. R., & Garland, D. J. (Eds.). (2000). *Situation awareness analysis and measurement*. Mahwah, NJ: Lawrence Erlbaum.
- Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42 (3), 462-492.
- 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.
- Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. *Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 82-86). Chicago: HFES.
- Endsley, M. R., Sollenberger, R. L., Nakata, A., & Stein, E. S. (2000). *Situation awareness in air traffic control: Enhanced displays for advanced operations*. Atlantic City, NJ: Federal Aviation Administration.
- Fitts, P. M. (1951). *Human engineering for an effective air-navigation and traffic-control system*. Washington, DC: National Research Council.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human Mental Workload* (pp. 143-154). New York: Plenum.

- Hauss, Y., & Eyferth, K. (2003). Securing future ATM-concepts' safety. *Aerospace Science and Technology*, 7, 417-427.
- Hogg, D. N., Follesø, K., Volden, F. S., & Torralba, B. (1995). Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. *Ergonomics*, 38 (11), 2394-2413.
- Hogg, D. N., Torralba, B., & Volden, F. S. (1993). A situation awareness methodology for the evaluation of process control systems: Studies of feasibility and the implication of use.
- Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-Loop Performance Problems and the Use of Intermediate Levels of Automation for Improved Control System Functioning and Safety. *Process Safety Progress*, 16 (3), 126-131.
- Kaber, D. B., & Endsley, M. R. (2003). 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*, 3, 1-40.
- Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics*, 36, 447-462.
- Lau, N. (2012). *Situation Awareness in Process Control*. University of Toronto, Department of Mechanical and Industrial Engineering. Toronto, ON: Ph.D dissertation.
- Lau, N., Skraaning Jr, G., Eitrheim, M. H., Karlsson, T., Nihlwing, C., & Jamieson, G. A. (2011). *Situation Awareness in Monitoring Nuclear Power Plants - The Process Overview Concept and Measure*. OECD Halden Reactor Project, Halden.
- Lee, J. D. (1999). Measuring driver adaptation to in-vehicle information systems: disassociation of subjective and objective situation awareness measures. *Proceedings of the Human Factors and Ergonomics Society 43rd annual meeting* (pp. 992-996). Houston: HFES.
- Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35 (10), 1243-1270.
- Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, 40, 153-184.

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46 (1), 50-80.

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

Lichacz, F. M. (2008). Augmenting understanding of the relationship between situation awareness and confidence using calibration analysis. *Ergonomics*, 51 (10), 1489-1502.

Lichacz, F. M., & Farrell, P. S. (2005). The calibration of situation awareness and confidence within a multi-national operational net assessment. *Military Psychology*, 17, 247-268.

Lorenz, B., Di Nocera, F., & Parasuraman, R. (2002). Display integration enhances 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., Bahner, J. E., & Hueper, A. D. (2006). Misuse of automated aids in process control: Complacency, automation bias and possible training interventions. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. 50, pp. 220-224. San Francisco: Human Factors and Ergonomics Society.

Manzey, D., Bleil, M., Bahner-Heyne, J. E., Klostermann, A., Onnasch, L., Reichenbach, J., et al. (2008). AutoCAMS 2.0 manual. Berlin, Germany: Berlin Institute of Technology.

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.

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*, (pp. 297-301).

- Masalonis, A. J., & Parasuraman, R. (1999). Trust as a construct for evaluation of automated aids: Past and future theory and research. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 184-188). Houston: Human Factors.
- Maxwell, S. E., & Delaney, H. D. (2003). *Designing Experiments and Analyzing Data: A Model Comparison Perspective, Second Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- McGuinness, B. (2004). Quantitative analysis of situational awareness (QUASA): Applying signal detection theory to true/false probes and self-ratings. *9th International Command and Control Research and Technology Symposium*. Copenhagen.
- McGuinness, B. (2007). Signal detection theory and the assessment of situation awareness. In M. Cook, J. Noyes, & Y. Masakowski (Eds.), *Decision making in complex environments* (pp. 299-310). Aldershot, UK: Ashgate Publishing Ltd.
- Miller, C. A., & Parasuraman, R. (2007). Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors*, 49 (1), 57-75.
- Moray, N. (2003). Monitoring, complacency, scepticism and eutactic behaviour. *International Journal of Industrial Ergonomics*, 31, 175-178.
- Moray, N., & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science*, 1, 354-365.
- Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman, & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 201-220). Mahwah, NJ: Lawrence Erlbaum Associates.
- Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27, 527-539.
- Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. *Ergonomics*, 39 (3), 429-460.
- Neal, A., Griffin, M., Patterson, J., & Bordia, P. (2000). Development of measures of situation awareness, task performance, and contextual performance in air traffic control. In A. R. Lowe, & B. J. Hayward (Eds.), *Aviation resource management: Proceedings of the Fourth Australian Aviation Psychology Symposium* (Vol. 2, pp. 305-314). Aldershot, UK: Ashgate.

Parasuraman, R., & Manzey, D. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors*, 52 (3), 381-410.

Parasuraman, R., & Riley, V. A. (1997). Humans and automation: use, misuse, disuse, abuse. *Human Factors*, 39 (2), 230-253.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance Consequences of Automation-Induced "Complacency". *The International Journal of Aviation Psychology*, 3 (1), 1-23.

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.

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.

Patrick, J., & James, N. (2004). A task-oriented perspective of situation awareness. In S. Banbury, & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory, measurement and application* (pp. 61-81). Aldershot, UK: Ashgate Publishing.

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 Centre.

Reason, J. (1990). The contribution of latent human failures to the breakdown of complex systems. *Philosophical Transactions of the Royal Society*, 327 (1241), 475-484.

Reichenbach, J., Onnasch, L., & Manzey, D. (2010). Misuse of automation: The impact of system experience on complacency and automation bias in interaction with automated aids. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (pp. 374-378). San Francisco: Human Factors and Ergonomics Society.

Rose, A. M. (1989). Acquisition and retention of skills. In G. McMillan (Ed.), *Human Performance Models to System Design*. New York: Plenum.

- Rousseau, R., Tremblay, S., Banbury, S., Breton, R., & Guitouni, A. (2010). The role of metacognition in the relationship between objective and subjective measures of situation awareness. *Theoretical Issues in Ergonomics Science*, 11 (1), 119-130.
- Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, 17 (4), 688-690.
- Salmon, P., Stanton, N., Walker, G., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics*, 37 (2), 225-238.
- Sauer, J. (2000). Prospective memory: a secondary task with promise. *Applied Ergonomics*, 31 (2), 131-137.
- Sheridan, T. B. (2002). *Humans and automation: Systems design and research issues*. New York: Wiley.
- Sheridan, T. B. (2006). Supervisory Control. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1025-1052). Hoboken, NJ: Wiley.
- Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Cambridge, MA: MIT Man-Machine Systems Laboratory.
- Sheskin, D. J. (2003). *Handbook of Parametric and Nonparametric Statistical Procedures* (3rd ed.). Boca Raton: Chapman & Hall / CRC.
- Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51, 991-1006.
- Skraaning Jr., G., Lau, N., Welch, R., Nihlwing, C., Andresen, G., Brevig, L. H., et al. (2007). *The ecological interface design experiment (2005)*. Halden, Norway: Institutt for energiteknikk.
- Skraaning Jr., G., Lau, N., Welch, R., Nihlwing, C., Andresen, G., Brevig, L. H., et al. (2007). *The Ecological Interface Design Experiment (2005)*. OECD Halden Reactor Project, Halden, Norway.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology*, 117, 34-50.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31 (1), 137-149.

Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. *Situational Awareness in Aerospace Operations (AGARD-CP-478)*, 3/1-3/17.

Wickens, C. D. (1994). Designing for situation awareness and trust in automation. *Proceedings of the IFAC Integrated Systems Engineer* (pp. 77-82). London: Pergamon Press.

Wickens, C. D. (2008). Situation awareness: Review of Mica Endsley's 1995 articles on situation awareness theory and measurement. *Human Factors*, 50 (3), 397-403.

Wickens, C. D., & Hollands, J. G. (1999). *Engineering Psychology and Human Performance* (3rd ed.). Upper Saddle River: Prentice Hall.

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.

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.

Wiener, E. L. (1988). Cockpit automation. In E. L. Wiener, & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 433-461). New York: Academic Press.

Wiener, E. L., & Curry, R. E. (1980). Flight deck automation: promises and problems. *Ergonomics*, 23 (10), 995-1011.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R. Parasuraman, & M. Mouloua, *Automation and human performance: Theory and applications* (pp. 3-17). Mahwah, NJ: Erlbaum.

Appendix A – Trust/Self-Confidence Questionnaires

Confidence Questionnaire

(Block: _____ Participant: _____)

Please rate your confidence in your capability to correctly identify faults (manually).

No Confidence 0 1 2 3 4 5 6 7 8 Complete Confidence

Please rate your confidence in your capability to select the correct fault management actions (manually), given that the fault was correctly identified.

No Confidence 0 1 2 3 4 5 6 7 8 Complete Confidence

Please rate your confidence in your capability to correctly execute the fault management actions (manually), given that the fault was correctly identified and the correct actions selected.

No Confidence 0 1 2 3 4 5 6 7 8 Complete Confidence

Please rate your confidence in your overall capability to handle faults in CAMS (manually).

No Confidence 0 1 2 3 4 5 6 7 8 Complete Confidence

Please rate your confidence in the capability of AFIRA to correctly identify faults. †

No Confidence 0 1 2 3 4 5 6 7 8 Complete Confidence

Please rate your confidence in the capability of AFIRA to select the correct fault management actions, given that the fault is correctly identified. ‡

No Confidence 0 1 2 3 4 5 6 7 8 Complete Confidence

Please rate your confidence in the capability of AFIRA to correctly execute the fault management actions, given that the fault is correctly identified and the correct actions selected. §

No Confidence 0 1 2 3 4 5 6 7 8 Complete Confidence

Please rate your confidence in the overall capability of AFIRA to handle faults in CAMS. †

No Confidence 0 1 2 3 4 5 6 7 8 Complete Confidence

Please rate your confidence in your joint/combined capability to perform all the tasks in CAMS with assistance from AFIRA. †

No Confidence 0 1 2 3 4 5 6 7 8 Complete Confidence

†: Not included for manual group

‡: Not included for manual or information analysis groups

§: Not included for manual, information analysis, or action selection groups

Appendix B – Correct Probe Responses

Query	O ₂ Valve Leak	N ₂ Valve Leak	O ₂ Valve Block	N ₂ Valve Block	Mixer Valve Block	O ₂ Valve Jam	N ₂ Valve Jam	O ₂ Sensor Fault	Pressure Sensor Fault
1	T	F	T	F	T	T	F	T	F
2	T	T	T	T	T	T	T	T	T
3	T	F	T	F	T	F	F	F	F
4	F	T	F	T	T	F	F	F	F
5	T	T	F	F	F	F	F	F	F
6	F	F	T	T	T	F	F	F	F
7	F	F	F	F	F	F	F	T	T
8	F	F	F	F	F	T	T	F	F

Appendix C – Other Experiment Materials

University of Toronto

PARTICIPANT CONSENT FORM

"The Role of Situation Awareness in Supervisory Process Control"

I hereby consent to participate in the research project entitled "The Role of Situation Awareness in Supervisory Process Control". This project looks at the potential benefits of a psychological construct, situation awareness, for evaluating operator performance when working with automated systems.

I understand that my participation in this study involves:

- Using a computer workstation to interact with a simulated process
- Attending two sessions on separate days
- Responding to questions about the simulation content, workload and confidence
- Having my performance on the simulator logged
- Submitting general background information about myself

I understand that the experiment will comprise a 2.5 hour training session and a 4.5 hour main session (approximate times).

I have had the opportunity to discuss this study and all questions have been answered to my satisfaction. Furthermore, I understand that in future I may ask any further questions I may have about the study or the research procedures.

I understand that only the investigators of this study and subsequent related studies will have access to my study data. Furthermore, I understand that results of this study will be published as part of a Masters dissertation and may be presented at conferences or in scientific journals. However, I also understand that my name and any other information that might identify me individually will not appear in publications. Data will be presented in aggregate wherever possible to ensure that all participants remain anonymous.

I understand that participation in this study is strictly voluntary. For my participation, I will receive \$60. I may also earn a bonus \$10 at the end of the experiment if I score among the top 30% of participants. I do, however, have the right to refuse answering any questions asked on the questionnaires, the right to withdraw from the study at any time without penalty, and to request that my data be destroyed. In that case my remuneration will be calculated based on the actual time I shall have spent in the study, at the base rate of \$8/hr.

I also understand that a summary of the research results may be requested and that I may contact the investigators of the experiment to obtain this.

I have been given a copy of this consent form. I understand what this study involves and agree to participate.

Participant's Name (print): _____

Signature: _____

Date: _____

If you have any questions or concerns you may contact the following persons:

Main researcher:

Adam Smith, who may be reached at: RS 317, 164 College St. Tel. (416) 978-0881 or
adam@mie.utoronto.ca.

Supervising professor:

Prof. Greg Jamieson, who may be reached at: RS 306, 164 College St. Tel. (416) 946-8504 or
jamieson@mie.utoronto.ca.

You may also contact the Ethics Review Office at ethics.review@utoronto.ca or 416-946-3273, if you have questions about your rights as a participant.

AutoCAMS Diagnostic Proficiency Test**Score: ____/11**

This quiz will assess whether you have acquired the basic skills required to take part in the experiment. Each question presents a symptom pattern and you must correctly identify the fault which causes that symptom pattern. You must score 80% or greater to pass. If, during the main experiment, you cannot resolve a fault in a reasonable amount of time, the simulation will terminate and the experiment will be aborted. In this case, you will receive partial compensation as per the consent form.

1. O₂ flow rate below preset value. Decreases in O₂ tank match O₂ flow rate.

- ☐ O₂ valve block ☐ N₂ valve block ☐ Mixer valve block ☐ O₂ leak ☐ N₂ leak
☐ O₂ jammed valve ☐ N₂ jammed valve ☐ O₂ sensor fault ☐ Pressure sensor fault

2. Cabin oxygen is falling. Cabin pressure is hovering at the ceiling of the normal range. After manually shutting off N₂ flow then resuming automatic control, cabin pressure continues to hover near the upper limit.

- ☐ O₂ valve block ☐ N₂ valve block ☐ Mixer valve block ☐ O₂ leak ☐ N₂ leak
☐ O₂ jammed valve ☐ N₂ jammed valve ☐ O₂ sensor fault ☐ Pressure sensor fault

3. O₂ flow is zero and cabin oxygen is below the normal range.

- ☐ O₂ valve block ☐ N₂ valve block ☐ Mixer valve block ☐ O₂ leak ☐ N₂ leak
☐ O₂ jammed valve ☐ N₂ jammed valve ☐ O₂ sensor fault ☐ Pressure sensor fault

4. O₂ flow rate below preset value. Decreases in O₂ tank do not match O₂ flow rate.

- ☐ O₂ valve block ☐ N₂ valve block ☐ Mixer valve block ☐ O₂ leak ☐ N₂ leak
☐ O₂ jammed valve ☐ N₂ jammed valve ☐ O₂ sensor fault ☐ Pressure sensor fault

5. O₂ and N₂ flow rates are both below the preset values.

- ☐ O₂ valve block ☐ N₂ valve block ☐ Mixer valve block ☐ O₂ leak ☐ N₂ leak
☐ O₂ jammed valve ☐ N₂ jammed valve ☐ O₂ sensor fault ☐ Pressure sensor fault

6. Cabin oxygen is falling. Cabin pressure exceeds the normal range. After manually shutting off N₂ flow then resuming automatic control, cabin pressure falls.

☐ O₂ valve block ☐ N₂ valve block ☐ Mixer valve block ☐ O₂ leak ☐ N₂ leak

☐ O₂ jammed valve ☐ N₂ jammed valve ☐ O₂ sensor fault ☐ Pressure sensor fault

7. N₂ flow rate below preset value. Decreases in N₂ tank do not match N₂ flow rate.

☐ O₂ valve block ☐ N₂ valve block ☐ Mixer valve block ☐ O₂ leak ☐ N₂ leak

☐ O₂ jammed valve ☐ N₂ jammed valve ☐ O₂ sensor fault ☐ Pressure sensor fault

8. Cabin oxygen is above the normal range and rising. After manually shutting off O₂ flow then resuming automatic control, cabin oxygen continues to rise.

☐ O₂ valve block ☐ N₂ valve block ☐ Mixer valve block ☐ O₂ leak ☐ N₂ leak

☐ O₂ jammed valve ☐ N₂ jammed valve ☐ O₂ sensor fault ☐ Pressure sensor fault

9. N₂ flow is zero and cabin pressure is below the normal range.

☐ O₂ valve block ☐ N₂ valve block ☐ Mixer valve block ☐ O₂ leak ☐ N₂ leak

☐ O₂ jammed valve ☐ N₂ jammed valve ☐ O₂ sensor fault ☐ Pressure sensor fault

10. N₂ flow rate below preset value. Decreases in N₂ tank match N₂ flow rate.

☐ O₂ valve block ☐ N₂ valve block ☐ Mixer valve block ☐ O₂ leak ☐ N₂ leak

☐ O₂ jammed valve ☐ N₂ jammed valve ☐ O₂ sensor fault ☐ Pressure sensor fault

11. Cabin oxygen is above the normal range and rising. After manually shutting off O₂ flow then resuming automatic control, cabin oxygen falls.

☐ O₂ valve block ☐ N₂ valve block ☐ Mixer valve block ☐ O₂ leak ☐ N₂ leak

☐ O₂ jammed valve ☐ N₂ jammed valve ☐ O₂ sensor fault ☐ Pressure sensor fault

Background Information

This questionnaire deals with educational and general background information. It will be used for comparative purposes only.

Participant code: _____

Please indicate your **current** discipline, degree and major area of study.

(e.g. Science, BSc, Biology).

☐ Applied Science/Engineering

(degree)

(area of study)

☐ Science

(degree)

(area of study)

☐ Social Science and Humanities

(degree)

(area of study)

☐ Other

(degree)

(area of study)

☐ Not applicable

Please indicate any **previous** education as above. (Check all that apply)

☐ Applied Science/Engineering

(degree)

(area of study)

☐ Science

(degree)

(area of study)

☐ Social Science and Humanities

(degree)

(area of study)

☐ Other

(degree)

(area of study)

☐ Not applicable

What is your current (or most recently completed) year of study?

- ☐ First
- ☐ Second
- ☐ Third
- ☐ Fourth or higher

Do you have experience with any of the following? (Check all that apply)

- ☐ Piping, valves or P&ID diagrams
- ☐ Control theory or practical applications (e.g. PLCs)
- ☐ Instrumentation
- ☐ Automation interaction
- ☐ Artificial intelligence or decision support systems
- ☐ Basic physics
- ☐ Thermodynamics
- ☐ Process industries (e.g. food, pharmaceutical, etc)
- ☐ Life support systems
- ☐ Space systems

What is your gender?

- ☐ Male ☐ Female ☐ Other

What is your age? ____

Appendix D – Training Manual

Welcome to the CAMS training session!

Today, you will complete a briefing and training session for the Cabin Air Management System (CAMS). The following pages explain the system and your responsibilities. As you progress through the document, you will simultaneously interact with the computer to practice what you are reading. General information is given in the current font while direct instructions are given in bold. Please notify the experimenter if you have questions at any point.

You will now see the desktop in front of you. On the desktop there is a folder titled, "Session 1" Double-click to open this folder.

Open the "CAMS Experiment" folder

Open the "Session 1" folder

You will now see a number of files that you will use later with the CAMS simulation. Each file contains a script of the events that will occur during each scenario block. To start a scenario, double click the file with the name of the scenario. You can terminate a trial in progress by pressing ctrl + q. Please practice starting and stopping the "Introduction" scenario.

Please start the scenario, "Introduction".

Now, terminate the scenario by pressing ctrl + q.

Now that you know how to start and stop a scenario, we will introduce the operation and monitoring of the Cabin Air Management System (CAMS).

Introduction

In this experiment, you play the role of an engineer in the ground control centre of a space agency and your job is to monitor the life support system (LSS) of a space station. This makes you responsible for the well-being and survival of the crew aboard this space station. The LSS consists of five subsystems that work together to ensure suitable atmospheric conditions for the occupants of the station. In order to support life, five parameters must be constantly monitored and regulated:

1. Oxygen (O₂)
2. Carbon dioxide (CO₂)
3. Pressure
4. Temperature
5. Humidity

The LSS monitors these 5 parameters and controls the system autonomously to maintain each in the required range. However, this automation may be unable to control the critical parameters if there is a malfunction in the system. Your task is to monitor the LSS and immediately intervene whenever a fault occurs. When this occurs, you must 1) manually control the system to keep the affected parameters within the safe range, and 2) as soon as possible find the cause of the fault and notify the station crew which component needs to be repaired.

This training session is divided into three parts. First, the five subsystems for controlling the above parameters are explained in more detail. Then, the elements of the interface and their functions are explained. Finally, the possible faults, their symptoms, and the actions necessary for manual control and fault diagnosis are described.

The five subsystems

The following section outlines the five subsystems of the LSS. Each subsystem controls a critical parameter which must be held within certain limits. For each parameter, normal and safe ranges are defined. The safe range indicates the acceptable safe values of each parameter and is broader than the normal range where the parameters are when the system is running as intended. Under normal conditions, the subsystems keep the parameters inside the normal range, so that the values oscillate between the upper and lower bounds of the normal range. Parameter values outside the normal range indicate a fault in the LSS and will trigger the alarm. If the value of a parameter is outside of the safe range, the safety of the crew is threatened. It is not necessary to memorize the exact limits as they are shown on the display.

Oxygen (O₂)

Without oxygen, humans can survive for only a few minutes. Therefore, maintaining a continuous oxygen supply is a central function of the LSS. The proportion of oxygen in the air on board the space station should be about 20%. The safe range for the oxygen concentration is between 19% and 20.5%.

When oxygen levels in the cabin air fall below 19.5%, valve of the oxygen tank is automatically opened and oxygen flows into the cabin. If the O₂ concentration reaches 20%, the oxygen flow is interrupted.

Carbon dioxide (CO₂)

Carbon dioxide is produced by the respiration of the crew members and must be removed from the cabin air. On Earth, plants absorb CO₂; on board the space station, a scrubber filters CO₂ from the air. Breathable air should not contain more than 1% CO₂. The safe range for CO₂ is therefore between 0.1 and 0.8%.

The CO₂ scrubber is active whenever the concentration is above 0.5%.

Pressure

Even abnormal pressure in the cabin of the space station is a direct threat to the lives of the occupants. Normal atmospheric pressure at sea level is about 1000 mbar. On board the space station, the pressure is mainly affected by controlled discharges of nitrogen (N₂) into the cabin, as nitrogen is the largest constituent of air (around 78%). Consequently, pressure is controlled by the nitrogen subsystem. The safe range for pressure is between 970 mbar and 1040 mbar.

If the cabin pressure falls below 970 mbar, the nitrogen tank valve is opened and N₂ flows into the cabin until a pressure of 1020 mbar is reached. If the cabin pressure is greater than 1025 mbar, the pressure controller opens an external drain valve to vent air from the cabin until the pressure drops below 1020 mbar. Note that this function vents air (both O₂ and N₂), not just N₂.

Temperature

The temperature in the cabin must be in the safe range between 18.5°C and 23°C.

If the cabin temperature rises above 22°C, the controller activates a chiller to cool the air and switches it off when the temperature has dropped to 20°C. If the temperature is less than 19.5°C, a heater brings the level back to 21°C. The heating and cooling processes keep the cabin temperature within the normal range of 19.5°C to 22°C.

Humidity

The sweat and breath of the astronauts aboard the station increases the humidity in the cabin. A dehumidifier removes excess moisture from the cabin air which could otherwise damage sensitive equipment. Insufficient humidity, on the other hand, would damage the mucous membranes of the astronauts and result in a higher consumption of drinking water. The humidity must therefore be in the safe range between 36% and 44%.

The dehumidifier turns on whenever the relative humidity exceeds 42% and remains open until 38% is reached.

The Monitoring and Control Screen

Please start the scenario, “Introduction” and leave it running while you read the following section. As you read, locate each display component on the screen.

The LSS schematic and controls are shown in Figure 23.

Figure 23: System Schematic and Controls

A. Oxygen tank and tank level indicator*	L. Control menu (see below)
B. Oxygen on/off valve ¹	M. Cabin (contains occupants and sensors)
C. Oxygen regulator valve ¹	N. CO ₂ scrubber ²
D. Oxygen flow meter	O. Heater ²
E. Nitrogen tank and tank level indicator*	P. Chiller ²
F. Nitrogen on/off valve ¹	Q. Dehumidifier ²
G. Nitrogen regulator valve ¹	R. Drain valve ²
H. Nitrogen flow meter*	S. Alarm indicator ³
I. Mixer valve ¹	T. Repair order status ⁴
J. Air flow meter*	U. Confirm data-link connection
K. Flow rate presets panel*	V. CO ₂ log window

*These components (tank levels, flow meters and flow preset panel) must be clicked on to request the value. This is to minimize the bandwidth required to transfer data from the station.

¹The on/off valves have only two states: open or closed. The regulator valves control the flow rate. ²Indicator illuminates when the component is active (In Figure 23, the CO₂ scrubber and chiller are active)

³Alarm indicator is red if there is a malfunction in the system, otherwise green.

History View

The display in the bottom left region of the screen (Figure 24) shows the history of a selected parameter over the last 240 seconds and is updated in one second intervals. The buttons along the top are used to select a parameter to view. When a parameter is selected, the relevant history graph is shown for 10 seconds before returning to the default (CO₂). The normal range is indicated by the green lines and the red lines indicate the safe range. The specific values of the normal and safe range limits are shown near the top in green and red respectively. The current value is shown next to the name of the selected parameter. Ideally, the measured values oscillate within the normal range; automatic control of each subsystem is activated when an upper or lower limit of the normal range is reached. Please note that sometimes a parameter will briefly cross the upper or lower normal limit without tripping the alarm. However, if the deviation is significant or sustained, a fault is present and the alarm will be triggered. Allowing a parameter to leave the safe range endangers the crew and should be avoided if at all possible.

Figure 24: History View

Control Menu

The buttons in the upper right corner of the screen (see Figure 23) are used to access the control menus for the individual subsystems (see Figure 25). Please note that only one control menu at a time may be active and if left idle the menu will close after 10 seconds.

Figure 25: Control Menu

This menu may be used to take manual control or adjust the setting of each parameter. For example, in the oxygen subsystem, manual/automatic control governs whether the oxygen on/off valve opens and closes automatically while the setting adjusts the oxygen flow rate by controlling the oxygen regulator valve. Note that these two functions are independent. Continuing with the example of the oxygen subsystem, the flow rate setting may be increased while remaining in automatic mode. This will cause the system to oscillate more rapidly and should not be done under normal circumstances as it will lead to increased wear on the valve actuator. To illustrate, we will try this now.

Please change the oxygen flow rate setting to “high” while remaining in automatic control. After observing the difference in the oxygen history graph, return the flow rate setting to “standard”.

Tasks

You must perform a number of tasks. Depending on their importance for the survival of the crew, they are categorized as either primary or secondary.

Your primary responsibilities are 1) to keep the five critical parameters within specification and 2) to diagnose faults in the system. The primary tasks have a much greater impact on your performance bonus. Your secondary tasks are 1) to conserve N_2 and O_2 reserves, 2) to periodically confirm the data connection with the space station and 3) to log the cabin CO_2 level each minute. These tasks are explained in more detail in the follow subsections.

Maintain critical parameters in the safe range (primary task)

Your most important task is to ensure that the values for carbon dioxide concentration, oxygen concentration, pressure, temperature and humidity do not leave their respective safe ranges, using manual control to intervene if necessary. You should only ever take manual control of a subsystem when you have detected a fault in the system. Therefore, when not performing fault management or secondary tasks, you should be scanning through the history graphs to monitor for potential problems.

Manage faults (primary task)

To fulfill the previous objective, malfunctions in the system must be managed. When you discover irregularities in any of the five parameters, you must:

1. Find the cause of the error.
2. Manually maintain the affected parameters within the safe range to ensure the safety of the crew until the error is resolved.
3. Issue a repair order to the crew of the space station.

After successful repair (which requires 60 seconds), the automation will resume regulation of the affected subsystem. The exact procedure for diagnosing and resolving each fault is described later in this document.

Conserve resources (secondary task)

In orbit, both oxygen and nitrogen are limited valuable resources that have to be used as sparingly as possible. Efficient control of the LSS must take this into account as wasteful control strategies can rapidly deplete O₂ and N₂ stocks to the point of endangering the crew.

Confirm data connection to the space station (secondary task)

In order to confirm that the data link between ground control and the station is still connected, a signal is periodically sent from the station which triggers the appearance of an icon on your screen (see Figure 23). When the connection check icon appears, it must be clicked on to confirm.

Please perform the next two connectivity checks

Log CO₂ concentration (secondary)

In order to maintain a backup record of the CO₂ concentration, you must input its value ($\pm 0.1\%$) once per minute, on the minute (± 10 seconds). That is, there should be an entry at 1:00, 2:00, 3:00, and so on. This is performed by clicking on the text box, typing the number, and clicking “Log value”. It is not necessary to type the “%” symbol.

Please perform the next CO₂ logging task

Fault Management

If there are signs of a fault in the system, a series of diagnostic steps needs to be performed in order to give the on-board crew the correct order. It is important not to act prematurely or to guess while the cause of the fault is still unclear. If you are unsure, diagnosis should be continued until the cause for the failure can be identified without ambiguity. All possible repair assignments are listed in the repair menu. Once a repair order is issued through this menu, it takes 60 seconds to perform the repair. During this time, the system may require manual intervention to keep the critical parameters in the safe range. In the following part of the manual, all possible faults and the corresponding fault management procedures are explained.

There are a total of eleven different faults that may occur during this simulation. Faults can be diagnosed on the basis of specific symptom patterns that are unique to each combination of subsystem, component and class of fault. Faults can be divided into four general types affecting either the oxygen or nitrogen subsystem:

1. Valve blockage: If a valve is partially obstructed, the airway is effectively narrower and flow is less than normal. The supply of N₂ or O₂ to the cabin is therefore reduced.
2. Valve leak: A leak in a valve lets gas escape from the piping and leads to a significant loss of O₂ or N₂ so that the supply of these gases to the cabin is also reduced.
3. Valve jammed open: A valve that is stuck open feeds excess gas through the line. N₂ and O₂ are wasted when air must be vented from the cabin to reduce pressure.
4. Faulty sensors: if the O₂ or pressure sensors are defective, deviation of the parameter outside of normal limits is not recognized. This causes the automatic controller to act inappropriately to the actual situation.

These faults will be explained in greater detail in the following sections. Each section addresses a different class of fault and you will be given one or more exercises to practice. During these exercises, do not worry about

performing the secondary tasks. Instead, focus on learning the symptoms and appropriate fault management procedure for each fault.

Blocked Valve

The partial obstruction of a valve results in a reduction of the gas flow through the line. Under normal circumstances, the regulator valves set the flow rates to the preset values shown in the flow rate presets panel. Of course, this only applies while the on/off valve is open to allow gas into the regulator. If the flow rate is less than the preset setting, the valve may be blocked (Figure 26). In addition to the N₂ regulator valve and O₂ regulator valve, the mixer regulator valve is susceptible to blockage. If the mixer valve is blocked, the N₂ and O₂ gas streams are both lower than the presets. Therefore, the gas flows in both systems must be checked in all cases to rule out the possibility of a mixer valve block.

Figure 26: Detection of a blocked valve

Management of blocked valves

For management of a blockage in the nitrogen, oxygen, or mixer valve:

1. Change the flow setting from "standard" to "high" for the relevant gas. This ensures that sufficient oxygen and/or nitrogen is supplied to the cabin despite reduced flow due to the block.
2. Issue the appropriate repair order.

After repair, the default flow rate is restored automatically. Now, you will practice the diagnosis and management of blockages:

1. Please start the scenario, "Block A", find out which valve(s) is/are blocked and manage the fault accordingly. Once the error repair is successfully completed, you can terminate the scenario.
2. Please start the scenario, "Block B", find out which valve(s) is/are blocked and manage the fault accordingly. Once the error repair is successfully completed, you can terminate the scenario.
3. Please start the scenario, "Block C", find out which valve(s) is/are blocked and manage the fault accordingly. Once the error repair is successfully completed, you can terminate the scenario.

Valve Leak

If a valve has a leak, then the amount of gas flowing into the valve is greater than the amount of gas that leaves the valve. The difference is the amount of gas that is lost to the surroundings.

As oxygen flows into the cabin, the oxygen supply is continuously depleted as indicated by the tank level indicator. Under normal conditions the gas flow from the oxygen tank to the oxygen valve is equal to the gas flow from the oxygen valve to the mixer. That is, the difference between two successive values on the tank level indicator (e.g. from 27,510 to 27,500) is equal to the value of the oxygen flow meter indicator (e.g. 10) for one time interval. If the flow meter indicates a lower value, however, the oxygen valve must be leaking (Figure 27). The same pattern of symptoms occurs in the nitrogen system when the nitrogen valve has a leak.

Figure 27: Detection of a leaking valve

Management of valve leaks

The fault management of leaks is the same as the management of blockages. The operator must:

1. Change the flow setting from "standard" to "high" for the relevant gas. This ensures that sufficient oxygen and/or nitrogen is supplied to the cabin despite losses due to the leak.
2. Issue the appropriate repair order.

After repair (one minute) the default flow rate is restored automatically. To practice the procedure just described, follow this exercise:

Start the scenario, “Leak”, find out which valve has a leak and perform the appropriate fault management. Once the error has been repaired, you can terminate the scenario.

It is important to differentiate between blockages and leaks. Both blockages and leaks cause low flow rates. Therefore, before a blockage can be safely diagnosed, the possibility of a leak must be excluded and vice versa. Practice using the following two tasks:

1. **Start the scenario, “Block Leak A”, find out which fault has occurred and perform the appropriate error management.** Once the repair is successfully completed, you can terminate the scenario.
2. **Start the scenario, “Block Leak B”, find out which fault has occurred and perform the appropriate error management.** Once the repair is successfully completed, you can terminate the scenario.

Jammed Open Valve

When an on/off valve is jammed open, gas flows unimpeded into the cabin. In the case of a stuck open oxygen valve, the cabin oxygen concentration increases beyond the upper limit of the normal range and the oxygen gas flow meter shows a constant flow of oxygen rather than the typical on-off behaviour. This continuing stream of oxygen to the cabin also increases the cabin pressure to the upper normal limit, triggering the drain valve to open. This can be detected by observing the drain valve indicator. A jammed open oxygen valve thus causes a characteristic behaviour where pressure values continuously hover at the ceiling of the normal range and oxygen concentration continues to rise if left unchecked (Figure 28).

Figure 28: Detection of a jammed O₂ valve (top: O₂, bottom: Pressure)

If the nitrogen valve becomes jammed open, the symptom pattern is similarly complex. The key symptom is a constant supply of nitrogen to the cabin. The pressure in the cabin rises but remains in the normal range through the intervention of the drain valve. However, the venting of air from the cabin causes the oxygen level to drop (Figure 29).

Figure 29: Detection of a jammed N₂ valve (top: O₂, bottom: Pressure)

Management of jammed open valves

When a jammed open oxygen valve is suspected, it is important to rule out a failure of the oxygen sensor. If the oxygen sensor does not detect that the normal upper limit was exceeded, the consequence will be a continuous inflow of oxygen into the cabin. To determine which of the two possible errors has occurred, use the following diagnostic procedure:

First, manually shut off the oxygen flow from the control menu. After some time, the oxygen concentration will return to the normal range. Once that occurs, the regulation of oxygen can be passed back to the automatic controller. If the oxygen concentration rises again immediately, the oxygen valve is stuck open. However, if the oxygen concentration then falls below the normal lower limit without intervention from the automation, the oxygen sensor is defective (Figure 30).

Figure 30: Diagnostic procedure to differentiate jammed valves from sensor faults

In both cases you maintain the oxygen concentration in the normal range by manually switching on and off the oxygen supply until the station occupants have repaired the fault. This is possible with a jammed open valve because the regulator valve is also closed by the manual off command.

To summarize, management of jammed oxygen valve is as follows:

1. Exclude a defect in the oxygen sensor:
 - Interrupt the oxygen flow in the control menu.
 - After oxygen concentration returns to the normal range, resume automatic control.
 - If oxygen concentration increases once again, the fault is a jammed open valve.
If oxygen concentration decreases, there is a defective sensor.
2. Take manual control of the parameter to keep it in the normal range
3. Submit the repair order

Practice with the following task:

Start the scenario, “Jammed O2 Valve”, and run the corresponding fault management. Once the repair is successfully completed, you can terminate the scenario.

Similarly, when it is suspected that a nitrogen valve is jammed open, it is important to exclude a defect in the nitrogen sensor.

If it the pressure sensor does not detect that the normal upper limit has been exceeded, this results in a continuous flow of nitrogen into the cabin, though the drain valve prevents the pressure from exceeding the normal upper limit. However, if the pressure sensor is defective, the excessive pressure is not detected and the drain valve remains closed, causing pressure to rise above the upper normal limit.

One exception occurs if the sensor fails while the drain valve is open. In this case, the drain valve remains open. Therefore a permanently illuminated drain valve indicator is a sure sign of a pressure sensor fault. To determine which of the two possible errors has occurred in this case, you would use the same diagnostic procedure as before except that the pressure subsystem will be manipulated instead of the oxygen subsystem. First, manually shut off the nitrogen flow from the control menu. After some time, the cabin pressure will return to the normal range. Once that occurs, the regulation of nitrogen flow can be passed back to the automatic controller. If the pressure rises again immediately, the nitrogen valve is stuck open. However, if the pressure decreases further and falls below the minimum boundary of the normal range, then the pressure sensor is defective.

In both cases you must maintain the parameters in the normal range by manually turning on and off the nitrogen supply until the repair order is completed and control can be passed back to the automation.

In brief, the management thus as follows:

1. Exclude a defective nitrogen sensor:
 - Interrupt the nitrogen flow in the control menu.
 - After nitrogen concentration returns to the normal range again, resume automatic control.
 - If nitrogen concentration increases once again, the fault is a jammed open valve.
If nitrogen concentration decreases, there is a defective sensor.
2. Take manual control of the parameter to keep it in the normal range
3. Submit the repair order
4. Turn on automatic control after the repairs is completed

Practice using the following task:

Start the scenario, “Jammed N2 Valve”, and run the corresponding fault management. Once the repair is successfully completed, you can terminate the scenario.

Defective Sensor

As described above, a sensor defect results in the automation not recognizing when a parameter has passed the upper or lower limits of the normal range and the necessary switching on or off of the gas supply is not performed. A sensor fault can therefore be responsible for the parameter deviating above or below the normal range, depending on when the fault occurs.

Sensor failure while pressure/oxygen is falling

If the sensor fault occurs while cabin pressure is decreasing, the pressure will continue decreasing even after leaving the normal range. This could also be the case when the nitrogen valve is blocked or has a leak. The essential feature of a sensor malfunction causing low pressure is that the value on the gas flow meter is zero; there is no nitrogen supply to the cabin. On the other hand, during blockage or leakage a reduced or unusually low gas flow would be observed.

A failure of the oxygen sensor will be similar but affect the oxygen subsystem.

Sensor failure while pressure/oxygen is rising

In the opposite case, the sensor fault occurs while pressure is increasing. This leads to cabin pressure exceeding the upper limit of the normal range. Automatic control does not detect the excess pressure, and therefore the drain valve remains closed. Since this symptom may occur only when a sensor fault in the pressure system has occurred, a pressure above the ceiling in combination with a closed drain valve is a clear indication of a pressure sensor failure.

However, as mentioned previously there is one exception: If the sensor fault occurs while the drain valve is currently open, it remains open. In this case, a defective sensor may be confused with a stuck open nitrogen valve, in which case the pressure increases to the upper normal limit, and then the automatic control opens the drain valve. Therefore, if an increased pressure occurs with an opened drain valve, the test procedure described above must be carried out in order to reliably diagnose the error.

The malfunction of the oxygen sensor while oxygen is increasing in the cabin leads to the normal upper limit being exceeded. Since this symptom occurs similarly in the case of a stuck open oxygen valve, the actual fault must be determined using the previously described test procedure.

Fault management of defective sensors

After determining the present failure (by performing the previously described test procedure), the repair order is issued. In the 60 seconds before the repair is completed, manual control of the parameter is required. Then, control is passed back to the automation.

Even for this last group of failures, there are exercises - this time four!

1. **Start the scenario, "Defective O2 Sensor (O2 falling)", and look for the symptoms of this error. Once the fault is positively identified, run the appropriate fault management. Once the fault is successfully repaired, you may terminate the scenario.**
2. **Start the scenario, "Defective O2 Sensor (O2 rising)", and look for the symptoms of this error. Once the fault is positively identified, run the appropriate fault management. Once the fault is successfully repaired, you may terminate the scenario.**
3. **Start the scenario, "Defective P Sensor (P falling)", and look for the symptoms of this error. Once the fault is positively identified, run the appropriate fault management. Once the fault is successfully repaired, you may terminate the scenario.**
4. **Start the scenario, "Defective P Sensor (P rising)", and look for the symptoms of this error. Once the fault is positively identified, run the appropriate fault management. Once the fault is successfully repaired, you may terminate the scenario.**

Additional Hints & Practice

It is important to keep in mind the 60 second duration between the issue of the repair order and the fault correction. This leads to two immediate consequences for fault management. First, the affected parameters must be regulated with altered settings, and sometimes controlled entirely manually for the duration of the repair before automatic control is resumed. Second, the repair order may have been incorrect. In this case you must go back to fault diagnosis. Therefore, at the time of issuing the repair order, you should immediately look at the clock and after 60 seconds reactivate automatic control. By doing this, you will know if your diagnosis was incorrect as quickly as possible.

To conclude the training, there are now another two exercises where you will not know in advance which faults will occur. Please complete the exercises below:

1. **Start, "Practice A", and perform fault identification and management as practised earlier in this training session for any fault that occurs. Perform all secondary tasks. Once the fault is successfully repaired, you may terminate the scenario.**
2. **Start, "Practice B", and perform fault identification and management as practised earlier in this training session for any fault that occurs. Perform all secondary tasks. Once the fault is successfully repaired, you may terminate the scenario.**

Questionnaires

At the end of each block in the main experiment, you will respond to two questionnaires. The first deals with the level of workload you experienced. The second asks about your confidence in your own ability to perform different tasks as well as your confidence in the capability of the automation.

Workload Questionnaire (NASA-TLX)

NASA-TLX has two parts. First, you will indicate the workload you experienced in terms of mental demand, physical demand, temporal demand, performance, effort and frustration. Second, you will indicate which of each pair of factors contributed more to workload in this task. There are 15 comparisons. These terms are defined below:

Mental demand – how mentally demanding was the task?

Physical demand – how physically demanding was the task?

Temporal demand – how hurried or rushed was the pace of the task?

Performance – how successful were you in accomplishing what you were asked to do?

Effort – how hard did you have to work to accomplish your level of performance?

Frustration – how insecure, discouraged, irritated, stressed and annoyed were you?

We will now fill out a workload questionnaire for your experience during the training session. As you respond, think mainly about the last two practice exercises you completed.

Please run the program, “TLX” and enter your participant code as the Subject ID.

Provide a rating for each factor then indicate which of each pair of factors contributed more to workload in this task - note that performance ranges from good on the left to poor on the right.

If you have any questions about this questionnaire, please ask the experimenter now.

Confidence Questionnaire

This questionnaire assesses how confident you are in yourself and the automation to competently perform the different tasks in CAMS. In particular, there are three essential functions required to successfully handle faults in CAMS: the identification of the fault, the selection of the appropriate fault management procedure, and the execution of that procedure. Questions 1-3 deal with your ability to perform these functions. Question 4 asks how confident you are in performing the task altogether. Questions 5-8 are the same but instead ask about the capability of an automated aid which will be introduced in session 2. Finally, question 9 asks about your ability to manage CAMS as a whole while being assisted by the aid. This includes performance of all primary and secondary tasks. As you look at the sample questionnaire on the next page, think about your experience during the practice exercises and select an answer for each question. You do not have to write this answer down.

If you have any questions about this questionnaire, please ask the experimenter now.

Confidence Questionnaire - SAMPLE

1. Please rate your confidence in your capability to correctly identify faults (manually).

No Confidence

Complete Confidence

0 1 2 3 4 5 6 7 8

2. Please rate your confidence in your capability to select the correct fault management actions (manually), given that the fault was correctly identified.

No Confidence

Complete Confidence

0 1 2 3 4 5 6 7 8

3. Please rate your confidence in your capability to correctly execute the fault management actions (manually), given that the fault was correctly identified and the correct actions selected.

No Confidence

Complete Confidence

0 1 2 3 4 5 6 7 8

4. Please rate your confidence in your overall capability to handle faults in CAMS (manually).

No Confidence

Complete Confidence

0 1 2 3 4 5 6 7 8

5. Please rate your confidence in your joint/combined capability to perform all the tasks in CAMS with assistance from the automated aid.

No Confidence

Complete Confidence

0 1 2 3 4 5 6 7 8

Closing Remarks

In the second session, you will be introduced to an automated decision aid and spend the rest of the session running different scenarios. Please notify the experimenter that you are finished.

Thank you for completing the CAMS training session!

Welcome to session 2 of the CAMS experiment!

Last time, you completed a briefing and training session for the Cabin Air Management System (CAMS). Today, you will be assigned an experimental condition and complete the main experiment.

Review

Open the “CAMS Experiment” folder

Open the “Session 2” folder

The first task for today is to run a brief scenario to refresh your memory. Before we begin, take a moment to review the table of possible faults, their symptoms, and the prescribed fault management procedure. Once you are ready, we will begin the review.

Start the scenario, “Review”, and perform fault identification and management for any faults that occur. Perform all secondary tasks. You may terminate the scenario after repairing 2 faults.

Situation Awareness

This experiment involves assessing your awareness of the situation during each scenario. The simulation will be periodically frozen and a set of questions presented to you. While frozen, the simulation will not advance so you do not need to worry about task performance while answering questions. However, you should try to respond promptly. While frozen, the CAMS display will disappear so that you are answering from your knowledge at the time of the freeze instead of collecting information from the display after seeing the questions. The questions are presented one at a time and are true/false questions paired with a confidence rating. You should answer “true” or “false” to the best of your ability and guess if you are not sure.

There are 8 questions that will be asked. They are:

1. Cabin oxygen concentration is outside the normal limits or behaving abnormally within limits
2. Cabin pressure is outside the normal limits or behaving abnormally within limits
3. Oxygen flow rate is less than the preset
4. Nitrogen flow rate is less than the preset
5. Changes in the nitrogen or oxygen tank level are greater than the corresponding flow rate
6. A valve is blocked
7. The automatic controller is not sensing when cabin oxygen or nitrogen crosses the normal limits
8. A valve is jammed open

A cabin parameter “behaving abnormally within limits” means that the parameter is not exhibiting the typical oscillating behaviour but is still in the normal range. This is the case when cabin pressure hovers near the upper limit due to venting. It can also be more subtle, for example when the pressure changes very gradually due to increased oxygen flow and reduced nitrogen flow.

The confidence rating is on a scale of 1 to 5 as below:

- 5 – Very High
- 4 – High
- 3 – Moderate
- 2 – Low
- 1 – Very Low

The performance bonus does NOT depend on your confidence ratings, so respond as honestly as possible. If you do not know the answer, make an educated guess and indicate your uncertainty using the confidence rating. Note that there is no penalty for wrong answers (an incorrect “true” is the same as an incorrect “false”).

Once the simulation resumes, remember to do any tasks that were not finished before the freeze (fault diagnosis, managing cabin parameters, sending the repair order, etc). If you have any questions about this aspect of the experiment, please ask now. You will have a chance to practice answering the questions in the next section.

Experimental Condition

The next step is to assign you to an experimental condition. This information is contained in the envelope in front of you.

Please open the envelope and read the contents. Take note of the condition number indicated on the assignment sheet.

We will now run an exercise to familiarize you with the specific condition.

Run the program, “Familiarization”, and enter the condition you were assigned to. Perform fault management for any faults that occur. Perform all secondary tasks and respond to situation awareness questions as required. Do not terminate the scenario early – you will be notified when it is complete.

Main Experiment

The main experiment is comprised of five scenarios, each lasting about 40 minutes. In the first block, you will manage CAMS manually. In later blocks, you will work with an automated aid as indicated by your experimental condition. For each block, perform all primary and secondary tasks to the best of your ability. Answer all situation awareness questions. Do not terminate the scenario early – you will be notified when each block is complete.

After each block, you will fill out a workload and confidence questionnaire. The workload questionnaire may be accessed by running the TLX program. Enter your participant code and the block number for the subject ID, for example: **02C Block 1**. The confidence questionnaires are found in the envelope. Remember to write your participant code and block number on this questionnaire as well.

You may also take a break between blocks as required. If you need to get a drink of water or use the washroom, please do so now. To run the main experiment blocks, double click on “Main Experiment” and follow the instructions presented. Once you are ready to begin, run the blocks in order as below:

Run the program, “Main Experiment”

1. Run Block 1
Complete a workload questionnaire
Complete a confidence questionnaire
2. Run Block 2
Complete a workload questionnaire
Complete a confidence questionnaire
3. Run Block 3
Complete a workload questionnaire
Complete a confidence questionnaire
4. Run Block 4
Complete a workload questionnaire
Complete a confidence questionnaire
5. Run Block 5
Complete a workload questionnaire
Complete a confidence questionnaire

Once you have completed all scenarios, notify the experimenter. Thank you for participating!

Appendix E – Experimental Condition Assignment

Condition A

Automatic Fault Identification and Recovery Agent (AFIRA), is an automated aid that can help with management of faults in the system. In the CAMS display, it presents messages in the lower right panel.

The version of AFIRA you will be working with provides a timer that begins counting once a fault is detected.

Condition B

Automatic Fault Identification and Recovery Agent (AFIRA), is an automated aid that can help with management of faults in the system. In the CAMS display, it presents messages in the lower right panel.

The version of AFIRA you will be working with is AFIRA *level 1*. This version is able to identify faults based on their symptom pattern. This function is reliable, but not perfect; AFIRA can make errors. Therefore, you should ALWAYS check the diagnosis offered by AFIRA.

AFIRA *level 1* can also break down. If AFIRA crashes, it will be unavailable for the remainder of that block and you will have to manage CAMS manually. You will know this has happened if AFIRA stops presenting messages when a fault is detected. The basic alarm system will continue to function normally.

Condition C

Automatic Fault Identification and Recovery Agent (AFIRA), is an automated aid that can help with management of faults in the system. In the CAMS display, it presents messages in the lower right panel.

The version of AFIRA you will be working with is AFIRA *level 4*. This version is able to identify faults based on their symptom pattern. This function is reliable, but not perfect; AFIRA can make errors. Therefore, you should ALWAYS check the diagnosis offered by AFIRA. AFIRA *level 4* also suggests fault management actions appropriate to the fault that was identified.

AFIRA *level 4* can also break down. If AFIRA crashes, it will be unavailable for the remainder of that block and you will have to manage CAMS manually. You will know this has happened if AFIRA stops presenting messages when a fault is detected. The basic alarm system will continue to function normally.

Condition D

Automatic Fault Identification and Recovery Agent (AFIRA), is an automated aid that can help with management of faults in the system. In the CAMS display, it presents messages in the lower right panel.

The version of AFIRA you will be working with is AFIRA *level 6*. This version is able to identify faults based on their symptom pattern. This function is reliable, but not perfect; AFIRA can make errors. Therefore, you should ALWAYS check the diagnosis offered by AFIRA. AFIRA *level 6* also selects fault management actions appropriate to the fault that was identified and will execute these actions once you click the "O.K." button.

AFIRA *level 6* can also break down. If AFIRA crashes, it will be unavailable for the remainder of that block and you will have to manage CAMS manually. You will know this has happened if AFIRA stops presenting messages when a fault is detected. The basic alarm system will continue to function normally.

Appendix F - Block Scripts

Block	Trial	Fault	Fault Occurrence
1	1	N ₂ Valve Leak	1:49
	2	O ₂ Valve Block	7:26
	3	O ₂ Valve Jam	14:32
	4	N ₂ Valve Block	20:46
	5	Pressure Sensor Fault (while rising)	27:03
	6	N ₂ Valve Leak	33:21
2	1	N ₂ Valve Jam	1:27
	2	N ₂ Valve Leak	7:32
	3	O ₂ Valve Leak	14:10
	4	N ₂ Valve Leak	20:08
	5	O ₂ Valve Block	26:27
	6	O ₂ Sensor Fault (while falling)	33:09
3	1	O ₂ Valve Leak	1:21
	2	N ₂ Valve Block	7:28
	3	O ₂ Valve Block	13:45
	4	O ₂ Valve Jam	20:36
	5	Mixer Valve Block	26:58
	6	Pressure Sensor Fault (while rising)	33:08
4	1	O ₂ Sensor Fault (while rising)	1:08
	2	Mixer Valve Block	7:48
	3	N ₂ Valve Leak	13:52
	4	N ₂ Valve Block	20:00
	5	N ₂ Valve Block	26:58
	6	O ₂ Valve Jam	32:53
5	1	N ₂ Valve Block	1:09
	2	O ₂ Valve Leak	8:16
	3	O ₂ Valve Jam	14:10
	4	Pressure Sensor Fault (while falling)	20:13
	5	O ₂ Valve Leak	27:03
	6	O ₂ Valve Block	32:45

Appendix G – Removed and Missing Data

Missing Data

Separate from the participants who did not complete the experiment, there were some data that could not be included for various reasons. The list below outlines the gaps in the data and the percentage of the complete set that is absent as a result. Throughout the presentation of results, participants are referred to by number and group, preceded by the pound sign, e.g. #34(no-support), #12(IA).

1. Block 1 data for participant #14(AI) were incomplete. The participant encountered a system error during the block and failed to notify the experimenter. The missing data were detected later when all files were checked for completeness and validity. Participant #14(AI) completed all other blocks normally, so only block 1 was removed. **(4.2% of data for failure performance only)**
2. 2 participants' reaction time data were missing due to failing to perform the secondary task (participants #22(IA) and #11(AS)). **(8.3% of reaction time data)**
3. Prospective memory (CO₂ log) accuracy data were not usable as two participants (#5(no-support) and #10(no-support)) performed the task incorrectly. It was apparent that the participants misunderstood the training manual ("±0.1%" in the instructions was taken to mean that the CO₂ value should always be entered as 0.1%). However, both of these participants' CO₂ log frequency (percentage of logs entered) was valid. Because accuracy was high for all other participants, the frequency of CO₂ response was used in all analyses rather than frequency*accuracy. **(0% loss)**

The amount of missing data was minimal with the exception of simple reaction time. The missing block 1 data affects only the analysis of failure performance. For all instances of missing data, participants were excluded from analyses casewise.

Outliers

Additionally, two areas were identified where participants' data was substantially different from the sample population and it was deemed appropriate to remove data.

1. Response bias data for participant #11(AS) in block 4 dropped to -100, more than 3 standard deviations below the sample mean (see below). This was thought to be due to a problem with the software used, rather than a systematic difference in how the participant interacted with the system. Thus, only the response bias data was removed for that participant. **(4.2% of response bias data)**

Outliers in Response Bias Measure

2. AVIS data indicated that two participants did not follow the instructions given (see below). One participant was highly complacent throughout the study (#12IA), and one who drastically reduced sampling following the first failure (#24AI). These were the only participants who sampled less than 50% of the parameters at any point during the experiment. Such low sampling could be due to accessing very few parameters for each fault, or by checking one or two faults fully and relying completely on the aid for the others. Detailed examination of the log file revealed that participant #24AI did check all necessary parameters for the first two faults in block 4. In fact, the participant accessed all parameters in the system, regardless of their applicability to the fault at hand. After fault 2, the participant ceased to check parameters at all and immediately sent a repair orders corresponding to AFIRA's recommendations (mean fault-identification-time of 5 seconds). Removing this data was deemed appropriate as participants were explicitly instructed that verification of the aid was integral to correct task performance. While some decrease in sampling was expected and relevant to the topics of discussion in this thesis, the performance observed here (less than 50% sampling) constitutes a different class of behaviour that the experimental design sought to exclude. Thus, these two participants were removed from the study entirely **(8.3% of all data)**

Outliers in AVIS Measure

Appendix H - Normality Tests

Normality Tests for Routine Measures (Part 1)

Measure	DOA Group	Shapiro-Wilk		
		Statistic	df	sig.
Fault Identification Time	No Support	.872	6	.236
	IA Support	.941	6	.665
	AS Support	.946	6	.705
	AI Support	.808	6	.070
Diagnosis Accuracy	No Support	.956	6	.785
	IA Support	.640	6	.001*
	AS Support	.496	6	.000*
	AI Support	.821	6	.090
Out-of-Target-Error	No Support	.969	6	.888
	IA Support	.818	6	.085
	AS Support	.860	6	.188
	AI Support	.906	6	.409
Prospective Memory	No Support	.945	6	.702
	IA Support	.780	6	.038*
	AS Support	.854	6	.169
	AI Support	.887	6	.303
Simple Reaction Time	No Support	.947	6	.714
	IA Support	.980	5	.936
	AS Support	.939	5	.661
	AI Support	.835	6	.118
SA	No Support	.861	6	.194
	IA Support	.946	6	.707
	AS Support	.960	6	.822
	AI Support	.932	6	.596
Bias	No Support	.996	6	.999
	IA Support	.938	6	.646
	AS Support	.955	5	.771

	AI Support	.957	6	.794
SA Confidence	No Support	.948	6	.724
	IA Support	.931	6	.585
	AS Support	.858	6	.183
	AI Support	.781	6	.039*
NASA-TLX	No Support	.926	6	.550
	IA Support	.853	6	.167
	AS Support	.945	6	.698
	AI Support	.930	6	.582

Normality Tests for Failure Performance (Part 1)

Measure	DOA Group	Shapiro-Wilk		
		Statistic	df	sig.
Fault Identification Time	No Support	.933	6	.607
	IA Support	.938	6	.642
	AS Support	.953	6	.768
	AI Support	.962	5	.818
Diagnosis Accuracy	No Support	.866	6	.212
	IA Support	.906	6	.408
	AS Support	.821	6	.090
	AI Support	.710	5	.012*
Out-of-Target-Error	No Support	.678	6	.004*
	IA Support	.818	6	.086
	AS Support	.964	6	.851
	AI Support	.918	5	.517
Prospective Memory	No Support	.928	6	.568
	IA Support	.912	6	.450
	AS Support	.897	6	.355
	AI Support	.938	5	.652

Simple Reaction Time	No Support	.935	6	.622
	IA Support	.975	5	.909
	AS Support	.922	5	.544
	AI Support	.943	5	.686

Normality Tests (Part 2)

Measure	DOA Group	Block	Shapiro-Wilk		
			Statistic	df	sig.
Trust (IA)	IA Support	2	.822	6	.091
		3	.902	6	.389
		4	.857	6	.178
	AS Support	2	.866	6	.212
		3	.827	6	.101
		4	.683	6	.004*
	AI Support	2	.958	6	.804
		3	.805	6	.065
		4	.496	6	.000*
SA	IA Support	2	.982	6	.961
		3	.912	6	.447
		4	.890	6	.320
	AS Support	2	.829	6	.105
		3	.910	6	.435
		4	.916	6	.474
	AI Support	2	.894	6	.342
		3	.951	6	.749
		4	.781	6	.040*
Bias	IA Support	2	.935	6	.619
		3	.932	6	.595
		4	.877	6	.255
	AS Support	2	.884	5	.327
		3	.857	5	.219
		4	.895	5	.383
	AI Support	2	.922	6	.521
		3	.941	6	.668
		4	.981	6	.958
SA Confidence	IA Support	2	.938	6	.642
		3	.952	6	.756
		4	.977	6	.933
	AS Support	2	.946	6	.704
		3	.837	6	.123
		4	.858	6	.182
	AI Support	2	.953	6	.766
		3	.950	6	.737
		4	.951	6	.745
AVIS	IA Support	2	.894	5	.377
		3	.881	5	.314
		4	.687	5	.007*
	AS Support	2	.905	6	.407
		3	.799	6	.057
		4	.805	6	.065
	AI Support	2	.942	5	.680
		3	.939	5	.656
		4	.552	5	.000*

Appendix I – Failure Performance Graphs

The figures below show the change in all performance measures from block 1 to block 5. In addition to showing each separate DOA group (left), the aggregate of AFIRA-supported groups is shown in comparison to the manual control group (right) for clarity. All error bars represent 95% confidence intervals.

Failure Performance (Diagnosis Accuracy)

Failure Performance (Fault Identification Time)

Failure Performance (Out-of-Target-Error)

Failure Performance (Prospective Memory)

Failure Performance (Simple Reaction Time)