2024

An empirical study of the relationship between automation and aircrew accident performance in high performance aircraft operating in the U.S. naval shipboard

Richard Matthew Gensley

AN EMPIRICAL STUDY OF THE RELATIONSHIP BETWEEN AUTOMATION AND AIRCREW ACCIDENT PERFORMANCE IN HIGH PERFORMANCE AIRCRAFT OPERATING IN THE U.S. NAVAL SHIPBOARD ENVIRONMENT

Richard Matthew Gensley

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
Industrial Engineering
to
The Graduate School
of
The University of Alabama in Huntsville
May 2024

Approved by:

Dr. James Swain, Research Advisor
Dr. Sampson Gholston, Committee Chair
Dr. Phillip Farrington, Committee Member
Dr. Sherri Messimer, Committee Member
Dr. L. Dale Thomas, Committee Member
Dr. Sampson Gholston, Department Chair
Dr. Shankar Mahalingam, College Dean
Dr. Jon Hakkila, Graduate Dean

Abstract

AN EMPIRICAL STUDY OF THE RELATIONSHIP BETWEEN AUTOMATION AND AIRCREW ACCIDENT PERFORMANCE IN HIGH PERFORMANCE AIRCRAFT OPERATING IN THE U.S. NAVAL SHIPBOARD ENVIRONMENT

Richard Matthew Gensley

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Industrial Engineering

The University of Alabama in Huntsville
May 2024

The research described in this dissertation was an empirical study of cockpit automation and aircrew accident performance in high performance aircraft. The data set consisted of 3,249 accident records released by the U.S. Naval Safety Center and publicly available information for high performance aircraft based aboard U.S. aircraft carriers between the years of 1980 and 2013. Five conclusions resulted from this study. The first result was a demonstration that through statistical analysis, it is possible to assess if different aircraft over a prolonged period of time have been exposed to a common operating environment. The second result was that while accident rate is the traditional method of measuring accident performance, the costs and/or fatalities associated with accidents may be more useful measurements. The third result was the use of current taxonomies of category, type and level of automation present in systems was sufficient for correlation of automation attributes to measures of human accident performance. Additionally, it was discovered and recommended that the list of automation categories be expanded to include one for human life support systems. The fourth conclusion from this study was that correlation did exist between certain configurations of cockpit automation and accident

performance. The fifth conclusion is the observation of a potential connection between group identity fusion and fatality rate for accidents involving automated cockpit systems.

Acknowledgements

First, I would like to acknowledge the University of Alabama in Huntsville for providing the course of study and opportunity for the pursuit of the degree of Doctor of Philosophy in Industrial and Systems Engineering. Specifically, I would like to thank Dr. Swain, Dr. Gholston, Dr. Farrington, and Dr. Componation for their guidance and council during my time at UAH.

I would also like to express my appreciation to my family for their support during this effort. To my parents, I am thankful for the guiding principles you instilled in me and the council you continue to provide. To my wife, I would like to express my deep thanks and gratitude for your encouragement, patience, and well-timed inputs as my progression at times was not as straight-forward as hoped. To my children, I would like to say “thank you very much” for your patience and encouragement over these past few years. It is my hope that you forge a path of contribution and excitement as you continue to grow into spectacular people.

It is my sincerest hope that the information contained in this summary of research is of value to the community and contributes to the body of knowledge in a way that will support additional discovery and application in a practical setting.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Summary

Due to the increasing complexity of aircraft, automation has become a key design attribute. Whether it fulfills the role of keeping an aircraft in stable flight or simply administering to nuisance tasks, such as modulating mixture of fuel for the engines, automated systems have become present in the vast majority of aircraft flying today. Since its introduction in aviation, the footprint of automation in aircraft cockpits has steadily increased and is expected to expand for the foreseeable future.

The common expectation associated with integrating automated systems into aircraft cockpits is that their presence will result in an increase in human-machine performance. To date, research has indicated that assumption may not be accurate for the human portion of the human-machine system. Based on surveys and simulator-based experiments, there is evidence indicating that the increase of automated systems in aircraft cockpits may be leading to an associated increase in errors due to aircrew complacency and an atrophy of basic flight skills.

Previous studies of the effects of cockpit automation on aircrew performance have focused on surveys and simulator-based laboratory events. Both methods have been successful in gathering useful data to identify and characterize the benefits and drawbacks of automated systems. However, research into the changes of human performance affected by cockpit automation while in an operational environment is limited. As a result, the purpose of the research summarized in this document was to assess if attributes of cockpit automation correlated to differences in aircrew performance, specifically accident performance, in the operational environment.

1.2 Hypothesis

Systems can be automated in many ways through the use of different types and degrees of automation. An example would be the spectrum of automation available for operating windshield wipers on a vehicle. Some vehicles require the driver to sense and recognize the presence of rain, determine that the condition of rain on the windshield requires removal, and select one of a limited number of modes of operation of the windshield wipers to remove the rain. Conversely, other vehicles have systems that automatically recognize the presence of rain and modulate the rate at which the wipers move across the windshield to effectively remove the rain drops from the driver's field of view. In both cases, automation is present; however, the attributes of the automation are different. Continuing with the example of windshield wipers in an automobile, the same concept of different attributes of automation could be applied to other systems in the car such as speed control (*i.e.*, cruise control), cabin temperature regulation (*i.e.*, climate control), and tuning of the radio to listen to different stations. When the attributes of these systems are characterized together through a system-of-systems view, a unique characterization of automation for the car emerges specific to that make and model of vehicle. This uniqueness in automation attributes applies to aircraft as well. As a result, the first premise of this research was that each aircraft cockpit used in this study had a unique configuration of automation.

In aviation, the human-machine system is comprised of the aircraft and aircrew. While both components are required to execute a flight event, the aircrew (*i.e.*, pilots and other crewmembers) have been and continue to bear the ultimate responsibility for safety. When a flight is not completed successfully due to an accident, a safety investigation, led by an independent accident investigation board, occurs and the result is a report containing a list of factors causal to the accident. Those factors can be categorized as either human performance related (*i.e.*, errors in following procedures, etc.) or non-human performance related (*i.e.*, material failure, etc.). By comparing the prevalence of human performance related causal factors between aircrew of different aircraft, an assessment of aircrew accident performance could be obtained. As a result, the second premise of this research was that aircrew accident performance could be measured in the operational environment.

Based on the two premises above, it was hypothesized that correlation between different attributes of cockpit automation and aircrew accident performance would be possible at a statistically significant level of $\alpha \leq 0.05$.

RESEARCH HYPOTHESIS: There is no statistically significant relationship ($\alpha \leq 0.05$) between attributes of cockpit automation and aircrew accident performance in the operational environment.

1.3 Importance of Topic

In 1980, the Naval Safety Center recorded a total of 180 accidents (with an average rate of 50.6 accidents per 100,000 flight hours), totaling $512M (adjusted to calendar year 2000 values) in damage, and the loss of 19 lives for the tactical aviation community. In 2000, the number of accidents recorded was 46 (with an average rate of 11.1 accidents per 100,000 flight hours), damages totaling approximately $530M, and the loss of 7 lives. While there was a significant reduction in the number of accidents (and accident rate), the average financial cost per accident in the year 2000 was greater than 4 times the average cost per accident in 1980 and the probability that a life was lost in an accident increased by 45%.

Interestingly, between the years of 1980 and 2000, the profile of carrier-based tactical naval aircraft in operation changed as well. Most notably, four of the older models of aircraft were retired from active use due to reaching the end of service life. Additionally, the quantity of the newest aircraft, when compared to quantities in 1980, increased significantly. With the retirement of the older aircraft and increase in quantity of newer aircraft, the average level of cockpit automation increased.

With the assumption that changes to automation were in part to improve aircrew accident performance, it would be expected that decreases in accident rate, cost, and fatalities would occur. While all three did decrease according to the "bottom line", analysis shows that reductions in cost and fatalities were a byproduct of the reduction in rate. When average costs of accidents (in dollars and lost lives) are compared, it is apparent that human accident performance declined between 1980 and 2000 as the average cost (financial and number of fatalities) per accident increased. If the average cost per accident remained constant between the years 1980 and 2000, the reduction in accident rate would have resulted in an approximate $399M avoidance of loss. Additionally, if the probability of loss of life per accident had remained constant, two deaths could have been avoided.

Understanding what changes in cockpit automation correlated to improvements and degradations in aircrew accident performance to a level of statistical significance ($\alpha \leq 0.05$) was the primary objective of this research. The results have the potential to refine the development of automated systems and provide significant benefit in the aviation and other fields.

Chapter 2. Prior Research

Prior research of the effects of automation on human operator performance can be grouped into the three main areas of characterization of automated systems, observations of human-automation performance, and theories of application. The review of prior work will be binned into those three groups.

2.1 Characterization of Automated Systems

Automation and automated systems are complex and tailored for their specific application. As a result, the methods used to characterize automated systems is equally complex. Four attributes have been proposed in the literature to describe automated systems (Kaber & Endsley, 2004), (Parasuraman, Sheridan, & Wickens, 2000), (Billings, 1997), (Calhoun, 2022). The first is a description of the magnitude of automation present, expressed on a scale of ten discrete levels. The second description is of the type of function the automation performs, expressed on a scale of four different types of automation. The third description is of the category of the function the automation performs, expressed as three distinct categories. The fourth description addresses if the automation is adaptable or adaptive. A summary of each of these characterizations of automation are summarized in the following paragraphs.

The most intuitive characteristic of automation is the magnitude, or level, of automation. A taxonomy consisting of ten discrete levels of automation was proposed by Endsley and Kaber as shown in Table 2.1 (Kaber & Endsley, 2004). A basic description of each level and whether the computer (system) or human (operator) is performing a function is provided. Of the ten levels, it can be observed that there are natural breaks between levels three and four as well as levels seven and eight. For levels one to three, the human is generating and selecting solutions while at levels four through seven, the human and computer share the role of generating solutions, but the human still retains the authority to select the best one for the given situation. For levels eight to ten, the computer assumes the role of selecting and implementing the solution. These breaks (between levels three and four as well as seven and eight) allow the levels of automation to be generalized into three groups of levels 1-3, 4-7, and 8-10.

Table 2.1: Endsley and Kaber's LOA Taxonomy.

The ten discrete levels of automation proposed by Endsley and Kaber describing each level with allocation of functions between the computer and human operator (Kaber & Endsley, 2004).

TABLE 1. Endsley and Kaber's (*15*) LOA Taxonomy

	FUNCTIONS			
LEVEL OF AUTOMATION	MONITORING	GENERATING	SELECTING	IMPLEMENTING
1. Manual Control	Human	Human	Human	Human
2. Action Support	Human/Computer	Human	Human	Human/Computer
3. Batch Processing	Human/Computer	Human	Human	Computer
4. Shared Control	Human/Computer	Human/Computer	Human	Human/Computer
5. Decision Support	Human/Computer	Human/Computer	Human	Computer
6. Blended Decision Making	Human/Computer	Human/Computer	Human/Computer	Computer
7. Rigid System	Human/Computer	Computer	Human	Computer
8. Automated Decision Making	Human/Computer	Human/Computer	Computer	Computer
9. Supervisory Control	Human/Computer	Computer	Computer	Computer
10. Full Automation	Computer	Computer	Computer	Computer

Another characteristic of automated systems is the function the system performs. As proposed by Parasuraman and Sheridan, systems can be characterized by using a parallel between the four-stage model of human information processing and the automated functions of a system as shown in Figure 2.1 (Parasuraman, Sheridan, & Wickens, 2000) . Of the four types shown, action implementation is the most commonly thought of form of automation. However, the acquisition and analysis of information as well as appropriate selection of the appropriate solution (*i.e.,* decision selection) have become more prevalent in everyday systems. An example is the lane departure warning system on many modern cars. It has subsystems that acquire information regarding the car's relative position in the lane, analyze information to determine if the car will stay in the lane, and decision selection through a cue to the driver as to which way to move the car. For that system, it still requires the driver of the car to implement the action of turning the steering wheel in the recommended direction. In this example, it is also important to note that "a particular system can involve automation of all four types at different levels" (Parasuraman, Sheridan, & Wickens, 2000) as demonstrated with the lane departure warning system.

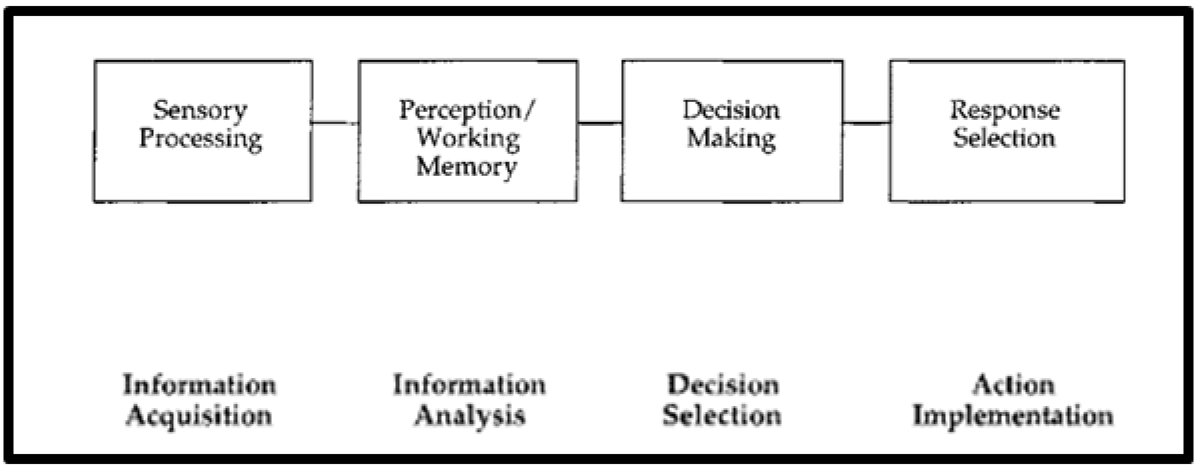


Figure 2.1: Model for Types of Human Interaction with Automation.

Parallel model of automated functions to the four-stage model of human information processing proposed by Parasuraman and Sheridan (Parasuraman, Sheridan, & Wickens, 2000).

A third attribute is characterizing what the automated system is controlling or managing from a larger system of systems perspective. Recently, Dudley, *et al.* developed a framework to distinguish information automation from control and management automation as shown in Figure 2.2 (Dudley, *et al.*, 2014). While the focus of this framework is to show the role of information automation in other types of automation present in an aircraft, it provides a matrix connecting the automation type (*i.e.,* processing step) to the three automation categories of aircraft systems as suggested by Billings (Billings, 1997). The incorporation of automation categories provides a better framework to evaluate the role of the automated system and potential interactions between the operator and the typical system of systems being used. While the particular framework suggested by Billings applies to aircraft operation, it is proposed that the framework could apply to any other system such as a vehicle, train, or other piece of equipment.

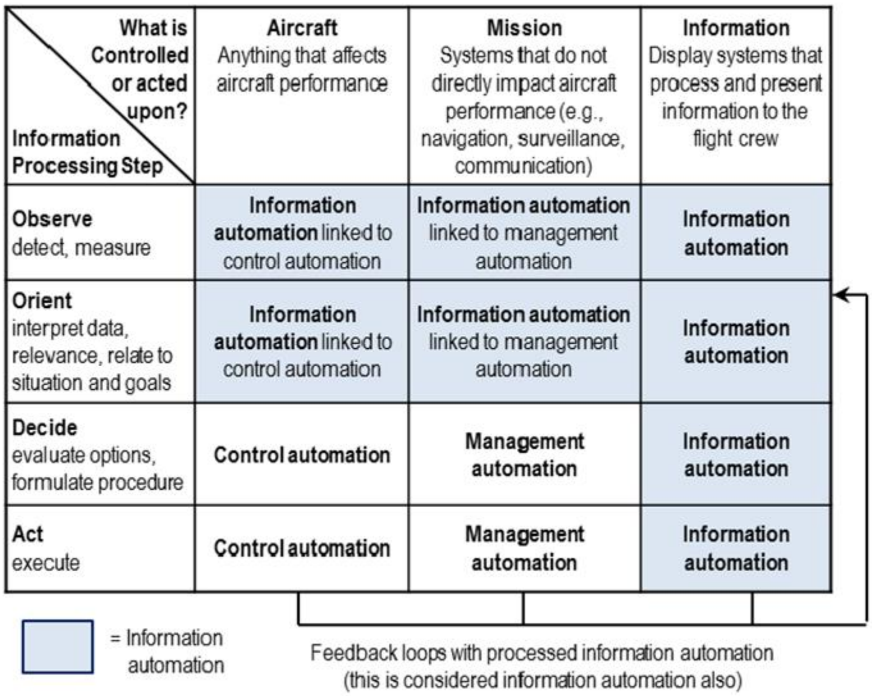


Figure 2.2: Automation Categories and Types.

Framework proposed by Dudley *et al.* to distinguish information automation from control and management automation (Dudley, *et al.*, 2014).

Typically, the attributes previously described (automation level, type, and category) are fixed for a system and do not change. However, there is an application of automation where the level of support provided to the operator varies. Systems with this flexible attribute are characterized as either being adaptable or adaptive. Adaptable automation is defined as allowing the human operator to assign the level of automation. Adaptive automation is defined as the system automatically assigns the appropriate level of automation based on feedback of operator performance or the environment (Calhoun, 2022).

The attributes of level, type, category, and adaptable or adaptive automation have been a common method to describe automated systems in the literature. These attributes have been observed to correlate or have a degree of influence on human performance. Connections between the attributes previously described and human performance will be summarized in the following paragraphs.

2.2 Observations of Human Performance and Automated Systems

A significant portion of the research conducted on human performance with automated systems has focused on observations in the laboratory as well as empirical studies. Correlation to differences in performance and attributes listed previously have been highlighted in several research publications. Additional factors external to the automated systems have been observed to have an impact on human-machine performance as well. In the subsequent paragraphs, a summary of the previous observations is presented in three groups. The first is the observations associated with the characterizations of automated systems described previously. The second group is observations associated with the human performance attributes of automation induced complacency, automation bias, skill atrophy, and operator fatigue. The third group will be a summary of observations regarding internal operator motivations such as perception, trust, and intra-group dynamics when multiple users are operating a common system.

2.2.1 Observations of Automation Attributes and Performance

The configuration of automated systems, as described by the attributes of level, type, category, and adaptiveness, has correlated to differences in human-machine performance. Research has observed that levels of automation impact multiple skills required by operators and does not appear to have a linear correlation to performance. Research into differences in performance for different automation types has yielded the observation that there appears to be different applications of each type to achieve an optimal level of performance. Automation categories have been observed to not have an equal level of impact on operator engagement or level of monitoring providing a nuisance in their role in performance. Recent research into adaptive automation has produced results indicating that the impact of adaptive systems may be more applicable to performance on secondary tasks. A summary of the observations associated with the four attributes of automation is provided in the following paragraphs.

Differences in levels of automation have been shown to impact human machine performance although not in a linear manner as many may assume. A positive correlation between performance and levels of automation was observed in a lab experiment where airline pilots using flight planning tools of increasing levels of automation resulted in increased performance with increased automation (Gil G.-H. , Kaber, Kaufmann, & Kim, 2012). Positive correlation was also observed between situational awareness and increased levels of automation when subjects operated a nuclear power plant simulator (Jamieson & Skraaning, 2020). However, in the event of a system failure, a negative correlation was observed where increased automation resulted in worse performance when subjects performed a simulated robotic arm task. In this experiment it was assessed that medium levels of automation were preferable as the low and high levels of automation resulted in suboptimal results (Li, Wickens, Sarter, & Sebok, 2014). Research conducted through a meta-analysis of 18 experiments came to a similar conclusion where it was observed that the benefit of automation was clear for routine system performance, but there was a negative impact of higher levels of automation with system failures and operator situational awareness (Onnasch, Wickens, Li, & Manzey, 2014). While a number of studies show a positive correlation between increased levels of automation and performance, an experiment to assess impact of levels of automation on unmanned air vehicle (UAV) operators indicated very little variance between operator performance when different levels of automation were used (Ruff, Calhoun, Draper, Fontejon, & Guilfoos, 2004). A similar counter observation occurred when a survey of 420 Aviation Safety Reporting System (ASRS) incident reports indicated that task prioritization errors were more prevalent in cockpits with greater automation (Wilson, 1998). However, it was from an experiment comparing the effects of three different levels of flight planning automation on aircrew performance that produced results indicating that human performance may not be linearly correlated to levels of automation as the best aircrew performance was achieved with use of the mid-level automated system (Gil G.-H. , Kaber, Kaufmann, & Kim, 2012). Ultimately, it is shown through previous research that levels of automation have an impact on human-machine performance, but it is not a linear relationship and is dependent on additional factors such as system reliability.

Just as differences in levels of automation have shown to impact human machine performance, the same has been observed with different types of automation. In a series of four experiments requiring an operator to conduct a visual search to locate a target, it was observed that information automation improved human performance much greater than decision aiding automation (Galster, 2003). In a team setting, an experiment consisting of forty teams performing a simulated Theater Defense Task to assess the impact of automation of information acquisition, information analysis and decision selection resulted in the finding that of the three types of automation, “decision-making automation may provide benefits in more limited contexts” compared to great benefits with the other two types (Wright & Kaber, 2005). However, this is not to say that decision aiding does not provide a benefit as it was observed during an experiment involving a decision task, the presence of an automated aid did improve human-machine performance (Yamani & McCarley, 2016). Additionally, the observation that “a human being augmented with a diagnostic aid does more poorly than the automation itself” (Rice, Trafimow, & Hunt, 2010) indicates that optimal application of each type of automation may be different based on the type of automation and expected operating conditions.

Research into the different categories of automated systems has resulted in observed differences in human-machine performance as well. While the categories listed in Figure 2.2 are aligned to aircraft, the concept can be applied to different systems. A theme in the research has been that operators are more aware of the status of a system when consequences of automation error may result in physical injury. For example, in an experiment measuring the difference in automobile brake reaction times associated with use of traditional or adaptive cruise control, it was observed that reaction times were much shorter when an “an increase in kinematic criticality” was present when either type of cruise control was used (Piccinini, *et al.*, 2020). In another study involving adaptive cruise control, it was found that driving with adaptive cruise control “tended to be associated with increased physiological arousal and improved driving behavior” when compared to driving manually (Weaver, Roldan, Gonzalez, Balk, & Philips, 2022). A similar observation was made during an experiment involving twenty B-747-400 pilots where it was determined that pilots “monitor basic flight parameters to a much greater extent” than the configuration of automated systems and in some cases “do not process mode annunciations in sufficient depth *to understand their implications for aircraft behavior* (emphasis added)” (Sarter, Mumaw, & Wickens, 2007)**.** The previous research indicates that operator performance is different between the three categories of automation with improved awareness occurring with the automation that falls into the “aircraft” type category (or similar category depending on the system).

Adaptive and adaptable automation has been an area of research as differences in performance between adaptive and fixed automation systems have been observed. In an experiment involving thirty university students performing a primary control task and a secondary monitoring task, it was observed that the overall system level of automation had significant impact on the performance of the primary task while the presence of adaptive automation significantly affected performance on the secondary (Kaber & Endsley, 1997). In a subsequent study, an experiment comprised of another dual-task scenario and dynamic allocation of levels of automation indicated that the effects of level of automation and adaptive automation "did not appear to be 'additive' in nature" with each providing different benefits to performance (Kaber & Endsley, 2004). An experiment consisting of forty participants operating a simulated air traffic control task illustrated that humans appear to better adapt to adaptive automation when applied to action implementation tasks than to information analysis and decision-making tasks (Kaber, Wright, Prinzel III, & Clamann, 2005). The research indicates that the use of adaptive automation is best suited for action implementation tasks to support improvement in secondary monitoring tasks.

The literature has documented performance differences associated with different configurations of automated systems. While one study used top-level empirical data to propose correlation between types of errors and presence of automation (Wilson, 1998), most of the studies used experiments and simulators to presumably increase precision in exploring relationships between human performance and attributes of automated systems.

2.2.2 Characterizations of Operator Performance

Operator performance associated with automation has been characterized into four common attributes of automation complacency, automation bias, skill atrophy, and operator fatigue. Automation complacency is a phenomenon where the operator does not adequately monitor automated systems. This type of complacency has been observed to occur in several environments with negative impacts to performance. Automation bias is phenomenon where the operator overly trust automation and does not effectively challenge automated answers. Similarly, to automation complacency, this phenomenon has been observed in several environments with negative impacts to performance. A concern for a few decades has been with the potential for skill atrophy by operators when using automated systems. Research has confirmed this concern as well as observed that while it does occur in some situations, is does not occur in all. Operator fatigue is another situation that research has observed occurring due to automation as well as being a factor for consideration in the design and implementation of automated systems. All four attributes contribute to the performance of automated systems and a summary of applicable research is summarized in the following paragraphs.

Automation complacency is "reflected in an inappropriate checking and monitoring of automated functions" (Bahner, Hüper, & Manzey, 2008) by the operator and has been identified as a significant risk as automation becomes more advanced and prevalent. Conditions in which it occurs include those of "multiple-task load, when manual tasks compete with the automated task for the operator's attention" (Parasuraman & Manzey, 2010). During an experiment where eleven participants were asked to drive a simulated vehicle using either an autopilot (automatic driving unless the human actively takes over), Active Safety mode (the human drives and the system takes over if a dangerous situation develops), or Haptic Shared Control mode (driving responsibilities shift based on human muscle activation); it was observed that in instances where the automation failed, the ability for the operator to recover was lowest when using the autopilot mode (Bahardwaj, *et al.*, 2020). A similar observation was made during an experiment where airline pilots were required to complete three challenging automation-related tasks. Over the course of the experiment, it was observed that detection of an error by pilots were often delayed, the recovery strategies were inefficient, and they "relied on high levels of automation to manage the consequences" (Nikolic & Sarter, 2007). In both of these examples, the automation was at a high level. In an experiment where subjects were asked to control a simulated robotic arm, it was observed that "a medium stage of automation" was preferable as it reduced the occurrence of complacency (Li, Wickens, Sarter, & Sebok, 2014). A similar conclusion was made when an experiment varied the levels of automation reliability for operators. It was observed that "operator detection of automation failures was substantially worse" for systems that had constant reliability compared to those that had variable reliability (Parasuraman, Molloy, & Singh, 1993). It should be noted that operator biases to monitoring strategies can predispose them to periods of complacency as it was observed that system "monitoring failures are one major contributor to breakdowns in pilot-automation interaction" (Sarter, Mumaw, & Wickens, 2007). To find a way to prevent or reduce rates of complacency, it was observed in a laboratory experiment that "exposing operators to automation failures during training significantly decreased complacency" (Bahner, Hüper, & Manzey, 2008). Automation complacency is a phenomenon that can produce reduced performance, but strategies of selecting mid-level automation and managing the operator work load have been observed to help mitigate its occurrence.

Automation bias is "the tendency to over-rely on automation" (Goddard, Roudsari, & Wyatt, 2012). It occurs when human operators ignore or not look for contradictory information due to assuming that a solution provided by an automated system is correct (Cummings, 2004). In research to determine the conditions that lead to automation bias, specifically in the evaluation of a clinical healthcare decision support systems, several factors associated with the user(s), environment, and implementation of automated systems were identified to include task complexity and time constraints (Goddard, Roudsari, & Wyatt, 2012). It has also been identified that automation bias occurs for "both naïve and expert participants, cannot be prevented by training or instructions, and can affect decision making in individuals as well as in teams" (Parasuraman & Manzey, 2010). In an experiment to assess automation complacency and bias through use of an environmental process control simulation, it was concluded that automation errors or "'automation wrong' had a much greater effect on accuracy, reflection the automation bias, than did 'automation gone,' reflecting the impact of complacency" (Wickens, Sebok, Li, Sarter, & Gacy, 2015). Previous research efforts have observed that over-reliance on automation, or automation bias, tends to occur in environments with high mental workload for both expert and novice users and tends to result in negative performance when automation errors occur.

Atrophy of operator skills due to operators opting to use automated systems in place of maintaining proficiency of manual skills has been a subject of concern and research for several decades. Supporting the concern, it was observed in an experiment where 126 randomly selected airline pilots were asked to perform a manual precision approach that "manual flying skills are subject to erosion due to a lack of practice" (Haslbeck & Hoermann, 2016). The consequence of the erosion of this particular skill was highlighted when research into aircraft accidents that occurred during the instrument approach phase observed that the leading cause was pilots failing to control the aircraft (Keller, 2013). In addition to the erosion of manual skills, the skill of maintaining situational awareness can atrophy as well as has been a focus area of research. Endsley identified an "automation conundrum" where "as more autonomy is added…the lower the situation awareness of human operators and the less likely that they will be able to take over manual control" (Endsley, 2017). In a similar theme, it has been observed that with advances in commercial aircraft automation, the "reduced workload seems to create a trend toward lack of vigilance and even boredom" (Archer, Keno, & Kwon, 2012). In an experiment involving 18 airline pilots, it was observed that while "cockpit automation may provide pilots more time to think, it may encourage pilots to reinvest only some of this mental free time in thinking flight-related thoughts" (Casner & Schooler, 2014). In another experiment involving medical students using a decision support system, it was observed that errors of omission correlated to lower cognitive load of the operator when compared to operators that had fewer errors, suggesting that "errors may stem from an insufficient allocation of cognitive resources" (Lyell, Magrabi, & Coiera, 2018). While the research summarized above indicates the presence and consequences of skill atrophy (manual skill and vigilance), it was observed in an experiment assessing operator engagement with vehicle automation that "drivers new to the technology remained engaged" with the system indicating that reductions in vigilance may develop over time. While automation aims to reduce operator workload, a theme in previous research is that a balance is needed to maintain operator proficiency and avoid atrophy of basic skills.

The impact of operator fatigue has been observed to affect human-machine performance when automation was present. In a survey of safety occurrences involving aircraft maintenance, a link was observed between memory lapses and fatigue as well as rule violations and time pressure (Hobbs & Williamson, 2003). In a study to determine if providing automation could help alleviate operator fatigue and stress states, it was observed that when implemented in a driving simulator, "automation use increased distress, especially in fatigue-prone drivers' (Neubauer, Matthews, Langheim, & Saxby, 2012). In another experiment where subjects were asked to perform a simulated supervisory process control task with a decision aid, those with a sleep deficit were more careful in using the decision aid but did experience a decline in performance of a secondary task (Reichenbach, Onnasch, & Manzey, 2011). To demonstrate the mutual impact of human errors and machine degradation, Haiyang, Shecgkui, and Jianbin developed a model that illustrated the phenomenon where "human errors usually…. Accelerate its (machine) degradation" which in turn increases operator fatigue and subsequently more human errors (Haiyang, Shengkui, & Jianbin, 2019). While fatigue impacts human performance, the impact to human-machine performance is significant as well.

Previous research has placed focus on four characteristics of operator performance when using automated systems. In exploring these characteristics, significant findings have been made primarily through experiments utilizing simulators. The use of empirical data has been limited.

2.2.3 Operator Mindset

The mindset of the operator and its impact on performance has been an area of research. The first area of focus to be summarized is the research in perceptions of automation and automated systems by various operators. The second area of focus is operator trust in automation. While initial views of automation tend to be positive, it has been observed that if the performance of an automated system either meets or fails to meet the expectations of the operator, it will impact overall performance. The third focus area addresses observations on group dynamics when multiple operators are part of a team that is utilizing a system. The influence of peer review and team interdependencies has been observed as potentially impacting performance. The fourth and fifth focus areas extend the concept of group dynamics to the phenomenon of tightly fused groups and the extreme sacrifice members may make for the benefit of the group. As systems increase in size and require more operators, the impacts of the ideas and mindsets of individual operators as well as groups of operators has been observed to impact performance. Research into these five areas is summarized in the following paragraphs.

Research into the general perceptions of operators has indicated favorable views of automation with a few areas of concern. In a survey of 132 pilots of advanced automation aircraft, it was noted that they appreciated automation especially when it was simple, reliable, and produced predictable results. On the contrary, the same group of pilots indicated that automation that required extensive interaction (*i.e.*, systems requiring excessive data input or constant monitoring) would be "perceived as obtrusive and pilots' attention will be focused on the automation instead of the underlying function" (Tenney, Rogers, & Pew, 1995). In another survey of pilots operating 13 different types of commercial aircraft, key themes observed were a general positive attitude towards automation with specific concerns towards what is now labeled as automation bias and complacency and the subsequent result of decreased safety (Rudisill, 1995). When experts with broad experience and knowledge of human factors and flight deck automation were surveyed regarding human factors issues, the top concerns included inadequate understanding of automation by the pilots, automation induced surprises, and complacency. The bottom concerns included reduction in pilot job satisfaction, increase in pilot workload, reduced inter-pilot communication, and lack of use of automation when it should be used (Funk & Lyall, 2000). During a study on the relationship between a pilot's flight hours and their performance, where performance between airline Captains and First Officers were compared, it was observed that First Officers with less than 1,500 flight hours kept the autopilot engaged until a significantly lower altitude with no statistically significant differences in technical measures of performance (Todd & Thomas, 2012). In another experiment involving human-automation collaboration with a simulated flight routing system for multiple unmanned vehicles, it was observed that “poor performance was associated with … negative attitudes towards unmanned aerial vehicles in general” (Cummings, Clare, & Hart, 2010). While perception is a qualitative attribute, it does appear to have an impact on human-machine performance.

Operator trust in automation has been another area of focus for research and has aligned with some previous findings on operator perceptions. In an experiment where 225 subjects were asked to rate trust and make automation use decisions during a visual search task, it was observed that with constant machine performance, operator “perceptions account for 52% of trust variance” indicating that trust is more subjective than previously thought (Merritt & Ilgen, 2008). However, operators may not be aware of their feelings of trust as shown during an experiment to explore the influence of implicit and explicit operator attitudes towards trust in automation. It was observed that the explicit attitude stated by the operators did not correlate with the observed implicit attitudes with the implicit attitudes having a significant effect on automation trust (Merritt, Heimbaugh, LaChapell, & Lee, 2013). It was also observed that operator trust in automation may be affected by operator emotions. In an experiment where operators first observed video clips designed to induce positive or negative moods and then interact with a simulated automated system, it was observed that “happiness significantly increased trust” (Merritt, 2011). In another experiment, participants were asked to self-report their expectations regarding the performance of automated systems. They were subsequently subjected to imperfect automation and then asked to report on their level of trust in the imperfect system. It was observed that operators who possessed an “all-or-none thinking had significant associations with decreases in trust following aid errors” (Merritt, Unnerstall, Lee, & Huber, 2015). When 40 participants were tasked to select a simulated convoy route between inputs provided by human aid or an automated map, it was observed that “their reliance on the human aid decreased during high-risk decisions" indicating an increase in reliance on the automation (Lyons & Stokes, 2012). In an experiment where passenger's trust and preferences were measured while being driven by a human or automation, it was observed that "passengers most preferred and trusted the human-defensive driver" and their individual preferences influenced their "trust and subjective driving characterizations" (Mühl, *et al.*, 2020). Additionally, the results of focus group discussions and online surveys on factors that influence user acceptance of full automobile driving automation included the finding that "perceived safety is strongly influenced by trust" (Motamedi, Wang, Zhang, & Chan, 2020). An experiment to explore how operators of different age groups relied on automation to provide alerts and how they complied with automation provided directions when workload varied. The results indicated that older operators maintained the same level of reliance and compliance as the workload increased whereas the younger operators increased compliance as workload increased (McBride, Rogers, & Fisk, 2011). In a series of experiments, it was observed that when workload capacity measurements are used, there is evidence that an automated aid "speeded human participant's responses" (Yamani & McCarley, 2016) even though in complex situations, "assistance from an automation decision aid may cause operators to delay their own responses" (Yamani & McCarley, 2018) indicating a reliance or trust in the automation. In an experiment where participants performed a target detection task where the automated aid made apparently "easy" errors, it was observed that "automation errors on task that appeared 'easy' to the operator severely degrade trust and reliance" (Madhavan, Wiegmann, & Lacson, 2006). An experiment to further explore the impact of automation false alarms versus automation misses resulted in indications that "a multiple process theory of operator trust" was needed to explain the effects of automation errors on operator behaviors (Rice, 2009). It was also observed in an experiment where subjects were exposed to two levels of automation reliability that while trust scores where greater for decision automation than information automation, “they did not vary with overall automation reliability as expected” indicating that operator trust involves multiple factors (Rovira, McGarry, & Parasuraman, 2007). Operator trust in automated systems has been observed to be affected by multiple factors to include operator attitudes, emotions, experience, workload, perceived risk, and observed performance of the automated system.

As automation has spread, the number of operators has increased. Locations or systems where there may have been only one operator managing a small system in the past now require many operators to work in concert to manage a much larger system. This has introduced a new dynamic of how the teams of operators work and behave because of their close proximity and interaction with a common system. An experiment to evaluate if redundant automation monitoring by multiple operators resulted in “social loafing” observed a correlation between performance and expectation of individual performance feedback. Specifically, the operators who worked collectively but did not expect individual performance feedback found 25% fewer automation failures than operators working alone. However, when they expected individual performance feedback, their performance was similar to those working alone and “a team advantage became apparent” (Cymek, 2018). However, in another study of accident rates of single-piloted and multi-crewed U.S. Navy and Marine Corps tactical jets between 1997 and 2007, it was observed that accident rates were similar except for the least severe category where mishap rates were “significantly higher for the multiple-operator system(s)” (Davis, 2010). A study of 95 severe Naval aviation mishaps between 2011 and 2016 observed that “teamwork failures were seen to be considerably damaging to both aviator skill and judgement” (Miranda, 2018) providing some insight as to potential pitfalls of multiple-operator systems. The interactions and performance of multiple operators has been observed to have an impact on performance.

Systems that have multiple operators can be viewed as having a single "group" of operators. When viewing the behaviors of groups, researchers have made several observations regarding loyalty of individuals to the group with resulting impacts to behavior. It was observed through interviews with soldiers on loyalty, the themes of "loyalty as reciprocity", "importance of emotional connection for cohesion", and that loyalty enabled them to prioritize competing demands were present (Connor, Andrews, Noack-Lundberg, & Wadham, 2021). In another study, it was identified that when members of a group share "core characteristics, they are more likely to project familial ties" resulting in becoming strongly fused or having a "powerful, visceral feeling of oneness with" a group (Swann Jr. W. B., *et al.*, 2014). For the military, the standard model of group cohesion identifies a primary group of cohesion between peers and leaders that are based on trust and teamwork (Siebold, 2007). When individuals are strongly fused to a group (either between peers or others), they have been observed to be "especially inclined to endorse pro-group action" when they perceive that the members of the group share core values among other reasons (Swann Jr. & Buhrmester, 2015). While a default view of pro-group actions may be benevolent, it is important to note that the observations have only identified behaviors in support of the group and its objectives. It has been identified that pro-group behavior can have negative consequences when groups are focused on extreme, anti-social behaviors (Fredman, *et al.*, 2015) or illegal activities (Connor J. M., 2010). It is the dynamic of group cohesion and identity fusion that has been observed to correlate with pro-group behaviors.

Previous research has also observed that for tightly fused groups, the pro-group behaviors exhibited by group members has included instances of extreme self-sacrifice. In an experiment to assess whether, when faced with a dilemma of sacrificing themselves or another person in order to save others, people tended to bias towards a self-serving solution and select the option to sacrifice another person (Bahnik, Efendic, & Vranka, 2021). In an experiment to assess the relationship between an individual's concern for the greater good of others and their decision in a sacrificial dilemma, it was observed that "'utilitarian' judgment was associated with greater endorsement of rational egoism…and less identification with the whole of humanity" (Kahane, Everett, Earp, Farias, & Savulescu, 2015). However, in a similar study it was observed that when subjects were presented with similar dilemma that included illustrations of sacrificing another or themselves, the subjects approved "of self-sacrifice more than directly harming another person." Additionally, it was noted that "the difference between self-sacrifice and murder" appeared to be an important one for the participants (Sachdeva, Iliev, Ekhtiari, & Dehghani, 2015). During a study of the testimony of Carnegie Hero Medal Recipients who risked their lives to save others, it was observed that "highstakes extreme altruism may be largely motivated by automatic, intuitive processes" (Rand & Epstein, 2014). It was observed over a series of seven experiments that "only those who were strongly fused with the group… endorsed self-sacrifice" (Swann Jr., *et al.*, 2014). At an extreme level, it has been observed that "extreme self-sacrifice is motivated by identity fusion, a visceral sense of oneness with the group" (Whitehouse, 2018). For tightly fused groups, it has been observed that individuals are much more prone to making extreme sacrifices for the good of others.

The mindset of human operators has been observed to play a role in performance. Factors affecting an operator's perception and trust in an automated system have been proposed based on surveys, laboratory experiments, and limited empirical data. Additionally, the role of group cohesion, fusion, and potential for extreme self-sacrifice by individuals for the benefit of the group have been explored. However, the literature was limited in how tightly fused groups interact and behave with automated systems requiring multiple operators.

2.3 Theories of Automation

Research into the design and implementation of automation has resulted in the development of a number of tools and design approaches with the objective of improved performance. The findings of these efforts are summarized below in three categories. The first is a summary of approaches and frameworks for implementation during the design phase of a system. Primary themes in this area of research include the human centered design philosophy, implementation of adaptive automation, and the role that operator trust plays in system performance. The second is a summary of methods to predict where a design may have attributes that will result in operator error and includes tools typically used in laboratory settings with systems nearing the end or complete with the design phase. The third category is a summary of models used to assess what went wrong after an accident for systems that have been placed into an operational environment.

2.3.1 Design Approaches and Frameworks

Approaches and frameworks for incorporating automation and automated systems into larger system designs has been an area of research that includes three themes. The first is the philosophy of human centered design where an emphasis is placed on optimizing the operator's role and performance. The second theme consists of the opportunities and potential approaches for implementing adaptive automation. The third theme is the role of operator trust and methods to optimize it and overall human-machine performance.

A very prevalent design approach addressed in existing literature is human centered design due to observed shortfalls with previous design philosophies. Multiple studies have identified performance shortfalls with previous design approaches where the full performance capability of the system was not achieved with a human operator present. In an experiment where human operators conducted an automation-aided search task, it was observed that operator performance was "far short of optimal" meaning that the full potential of improved performance was not achieved (Boskemper, Bartlett, & McCarley, 2022). Similarly, it has been acknowledged that when automated systems are implemented to improve specific safety related tasks, "it is highly questionable whether total system safety is always enhanced by allocating functions to automatic devices rather than human operators" (Wiener & Curry, 1980). During an experiment consisting of a simulated medical visual search task, it was observed that the negative impact of time pressure on human performance can be reduced when support is received by an automated decision support system (DSS), but "joint overall performance remains below DSS-alone performance" (Rieger & Manzey, 2022). To explain why total performance has been observed to be less than optimal, it has been assessed that human-machine performance predictions have been difficult to reliably acquire due to the influence of trust, mental workload, and risk which has an impact on an operator's use, misuse (overreliance on automation), and disuse (neglect or underutilization) of the system (Parasuraman & Riley, 1997). Another aspect associated with the difficulty of predicting performance is the presence of individual psychological precursors of the operators. The impact of which was highlighted in a survey of 743 Dutch drivers where it was found that "'unsafe' attitudes, physical precursors, and psychological precursors" explained up to 9% of performance variance in relation to driving related errors, violations, and accident involvement over the preceding three years (Verschuur & Hurts, 2008). Additionally, it has been observed that the introduction of automation has subjected operators to new stressors associated with "technical problems, poor usability, low situational awareness, and increased requirements on employees' qualification" (Körner, *et al.*, 2019). Ultimately, it has been proposed that to improve safety performance, numerous drawbacks in person-centered safety theories (where accident causation tends to ultimately fall on the human operator) have been identified with the finding that focusing on whole-system design is a more optimal approach (Holden, 2009). It has been identified that total system performance can be improved with increased focus and consideration of human variability in the design of automated systems.

The human centered design philosophy has resulted in a number of proposed approaches to development and integration of automated systems. When viewing human centered design, it has been recommended that a "socio-technical systems approach" be taken to account for the multiple human-human interactions and associated impacts to system operators (Harris, 2006). On the topic of automated vehicles, it has been suggested that performance of the driver and vehicle should be viewed as a joint cognitive system when examining the impacts of different systems (Horrey & Lee, 2020). This is supported by the observation that "intuitive cognition dominates human reasoning and decision making" based on literature review of 120 articles and books published within the last 50 years (Patterson, 2017). Additionally, it was discovered through a study to understand how to maintain driver engagement while using highly reliable automation that "a key component of driver engagement is cognitive (understanding the need for action), rather than purely visual (looking at the threat), or having hand on wheel" (Victor, *et al.*, 2018). To assist with assessing and designing towards an optimal level of cognitive engagement by the operator, research into the application of all five phases of the cognitive work analysis (CWA) framework was conducted within the context of a home to demonstrate benefits and potential shortfalls (Naikar, 2006). Research that included the application of CWA analysis in the design of automated systems resulted in the proposal that a layered approach towards implementing degrees of automation with cognitive work analysis may provide insight into design options for improved human-automation performance (Li & Burns, 2017). Application of cognitive systems engineering and human centered design continues to be a focus of research as evidenced by the successful use of the Usability Engineering Repository (UsER) in the development of supervisory controls systems to address a lack of integrated tooling for user-centered software development (Herczeg, Kammler, Mentler, & Roenspieß, 2013). At a more general level, it has been recommended that for near-perfect systems, human operators should be understood at the individual level and the system should be actively designed for the given individual (Forough, *et al.*, 2021). It is through the application of the human centered approach that a number of benefits are expected.

The benefits of human centered design have been another focus area of research with many benefits having been identified that are expected to improve the interface between human operators and automated systems. It has been proposed that if human factors were used as a design driver, a single crew commercial aircraft using largely existing technology would be possible (Harris, 2007). Additionally, it has been proposed that adopting a "context-aware automation design philosophy that promotes a more communicative and collaborative human-machine interface" could assist in countering potential skill atrophy of airline pilots due to overreliance on automation (Geiselman, Johnson, & Buck, 2013). Similarly, research in the Human-Centered Automation (HCA) concept has shown that out-of-the-loop and skill degradation problems can be avoided. Research in the application of HCA in fighter aircraft design has resulted in multiple suggestions for changes in development processes (Helldin & Falkman, 2011). Using a traffic control example, the benefit of having systems operate in a human-machine cooperation approach was observed as superior to "*Robot First* or *Human First* decision approaches" (Klumpp, Hesenius, Meyer, Ruiner, & Gruhn, 2019). In an experiment designed to test changes in human-machine performance between four levels of touch steering guidance, the mode of continuous guidance yielded "improved performance and satisfaction" at the cost of delayed response when system was unexpectedly deactivated when compared to bandwidth activated guidance modes (Petermeijer, Abbink, & de Winter, 2015). The impact of time pressure has been shown to reduce the quality of human decision making, but an automated decision support system can mitigate this effect if the "automation's advice follows the assessment of the human" (Rieger & Manzey, 2022). The field of research in human-centered design approach has demonstrated many benefits as well as numerous methods of implementation.

Implementation of adaptive automation is another approach that has been a focus of research as it provides a potential solution to multiple areas of concern with human-machine integration. In a driving experiment using either a haptic feedback force through the steering wheel or an automated steering control system to avoid collisions, it was observed that the "necessity of an adaptive automation that can strike balance between the processing ability of the human and the system, and hazardous context encountered" was needed for optimal results (Muslim & Itoh, 2019). In a review of empirical studies on human-automation interaction and the implications for design, it was found that adaptive automation can help balance operator workload and maintain their situational awareness with the caveat that more research is needed in determining when "adaptation should be user controlled or system driven" (Parasuraman & Wickens, 2008). For adaptable automation, it has been shown that operator satisfaction and overall performance is improved when the operator defines what tasks will be performed by the operator or automated systems versus when it is defined during the system design phase (Tausch & Kluge, 2022). However, in two experiments conducted on adaptive and adaptable methods to invoke automation, results indicated that brain based adaptive automation systems improve perceptual level situational awareness and reduce mental workload when compared to systems requiring user-initiated control (*i.e.*, adaptable automation) (Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006). Additionally, an experiment using adaptive automation and a "Vigilance and Attention Controller" (a system based on electroencephalography (EEG) and eye-tracking (ET) techniques) was conducted using air traffic controllers. The results indicated that the pairing of the two systems was able to decrease the presence of the out-of-the-loop phenomena associated with long duration and highly automated tasks thereby keeping the operators more involved within operative tasking (Di Flumeri, *et al.*, 2019). For future designs, a Coloured-Petri-Net simulation environment based approach to measure operator task performance has been proposed as an input for adaptive automation (Hasselberg & Söffker, 2013). Research into adaptive automation has shown a number of potential benefits associated with changing the human-machine interface to keep the total system operating closer to an optimal level.

Research has shown a correlation between an operator's level of trust in a system and total performance. When initial descriptions are provided to operators that state an automated system is less reliable than it actually is, long-lasting effects of those statements on operator perceptions have been observed (Barg-Walkow & Rogers, 2016). Similarly, in an experiment involving a simulated X-ray screening task, it was observed that when operators were informed that a system had a high false alarm rate and low positive predictive value (probability that a target is present when automation alarms), operators ignored "about one-half of the true automation alarms on difficult targets-a strong cry-wolf effect" (Huegli, Merks, & Schwaninger, 2020). To assess the accuracy of human estimates of automation reliability, an experiment was conducted where it was observed that participants' initial assessments were below the true setting and remained lower than true reliability after the system reliability was modified during the test (Hutchinson, Strickland, Farrell, & Loft, 2022). In exploring if trust affected operator compliance to automation alerts and operator reliance on alerts differently, an experiment was conducted that demonstrated false alarms "clearly affected both operator compliance and reliance, whereas mis-prone automation appeared to affect only operator reliance" (Dixon, Wickens, & McCarley, 2007). Additionally, a separate experiment observed that false alarms correlated with reduced performance in a primary task while misses correlated to reduced performance in secondary or concurrent task (Dixon & Wickens, 2006). When operators were tasked with performing a simulated surveillance task with the help of an imperfect automated detector, it was observed that "operators informed of the predictive values or the overall likelihood value, rather than the hit and correct rejection rates, relied on the decision aid more appropriately and obtained higher task scores" (Du, Huang, & Yang, 2020). The observations of previous research present a trend of conservative trust levels by human operators that may change when presented with additional information.

It has been documented that the level of trust a system operation has in the automation will affect the performance of the human-machine system resulting in significant research into the principles behind operator trust in automation. In study involving experienced rail operators in four signaling centers, it was observed that an understanding of the automation was the strongest aspect of trust development indicating that "development and maintenance of trust in real-world, safety-critical automation system may be distinct from artificial laboratory automation" (Balfe, Sharples, & Wilson, 2018). In a study of factors that influence trust in automation, understanding the capabilities and limitations of both the system (*i.e.,* error, feedback, etc.) and the human operator (*i.e.,* emotive factors, cognitive factors, etc.) were identified as important to system performance (Schaefer, Chen, Szalma, & Hancock, 2016). Similar findings occurred in an experiment to explore the impact of expected automation reliability on trust where it was observed that both expectations on automation reliability and task experience played a role in determining trust (Bowden, Griffiths, Strickland, & Loft, 2021). To provide quantitative models of trust in automation, it has been proposed that the existing models of signal detection, statistical parameter estimation calibration, and internal model-based control can be effectively applied to identify quantitative trust measures in system designs (Sheridan, 2019). Additionally, a model has been developed that identified three layers of variability in human-automation trust as dispositional trust, situational trust, and learned trust (Hoff & Bashir, 2015). In a field study to empirically compare two models of automation surprises, it was observed that human interaction with automation aligns more with the concept that automation surprises mark "the cognitive realization that what is observed does not fit the current frame of thinking" than the concept that surprises are due to complacency (De Boer & Dekker, 2017). It has been proposed that when addressing inappropriate human automation interaction (*i.e.*, use, misuse, disuse), a focus on the psychophysiological factors of decision making will be critical to mitigating inappropriate behavior (Dmec, Marathe, Likos, & Metcalfe, 2016). Similarly, a set of guidelines have been developed for designing automated systems that encourage operator trust with the three key concepts being to "Make Automation Trustable", "Relate Context to Capability of the Automation", and "Consider Cultural, Organizational, and Team Interactions" (Lee & See, 2004). It has been proposed that a deeper understanding of system limitations and performance could be beneficial to maintaining appropriate levels of trust and greater system performance.

Research has been conducted in how alignment between operator mental models to system attributes and limitations affect trust and therefore performance. To reduce the potential for automation-related operator error, it has been proposed that improvements may arise from integrating automation based on "expected operator expertise levels" and planned necessary training to develop operator automation knowledge (Strauch, 2017). The impact of training on automated systems was observed in an experiment that assessed the impact of prior training for operators on how to execute takeover requests from an automated driving system. It was observed that prior training correlated to improved takeover performance and lower initial automation trust (Hergeth, Lorenz, & Krems, 2017). Similarly, the importance of pilot training on the "design and function theories of automation system" and the subsequent ability to "recover from automation failure and …be more willing to take over in case of malfunctions of automation systems" has been highlighted and discussed (Liu, 1997). An alternative approach to preventing accidents has been proposed where the focus is on what "goes right and identify how to replicate that process" to further the resiliency of a system where it "can sustain required operations under expected and unexpected conditions by adjusting its functioning" (Null, *et al.*, 2019). In an experiment to assess the impact of performance feedback and scenario training on operator misuse and disuse rates, it was observed that while there was little misuse, "a combination of feedback and scenario training was more effective in mitigating disuse than either intervention used in isolation" (Beck, Dzindolet, & Pierce, 2007). While training typically focuses on operating a particular system, it has been proposed that training as part of behavior modification efforts may mitigate some of the negative effects of psychological precursors resident in operators, as previously discussed. Research has proposed that three classes of behavioral factors consisting of those that reduce operator capability on a long-term basis, short-term basis, and those that promote risk taking behavior with long-term impact may assist in understanding the causes of road traffic accidents and contribute to behavior modification efforts (Petridou & Moustaki, 2000). The impact of training on trust and performance clearly supports its consideration as an important tool for improved performance of automated systems.

In addition to training, research has been conducted to explore the impact of real-time system feedback regarding levels of uncertainty to the operator. In an experiment to evaluate whether presenting automation uncertainty improves driver-automation interaction, it was observed that the presentation of a symbol indicating the automation uncertainty improved cooperation between the drive and automated system (Beller, Heesen, & Vollrath, 2013). It has been proposed that inadequate feedback by automated systems to the operator is the source of many difficulties and an “appropriate design should assume the existence of error, it should continually provide feedback” (Norman D. A., 1990). This is supported by two simulator studies involving automated vehicles where it was discovered that trust is “calibrated along provided information prior to and during the initial drive with an automated vehicle” and temporary decreases in trust due to automation malfunctions can be avoided through providing transparent information of potential limitations of the system (Kraus, Scholz, Stiegemeier, & Baumann, 2020). To address part of that concern, the automation transparency design principle was developed and proposed that the activities, effects, capabilities, and responsibilities of automation should be “directly observable in the human-system interface” and has shown benefit at the system component level, but not at the procedural level (Skraaning & Jamieson, 2021).

Accurate and timely feedback has been shown to facilitate operator trust and improved performance.

The potential benefits associated with use of human centered design, adaptive automation, and improvement of operator trust have been focus areas in the literature. Methods of implementation and proposed design philosophies have been provided and have shown promise in laboratory environments.

2.3.2 Prediction of Operator Performance

The ability to assess designs of automated systems and predict potential strengths and pitfalls has included two areas of focus. The first is the development of tools to assist with evaluating designs through analysis. The second is the use of physiological measurements of human operators to provide real-time feedback of their mental states and performance. The use of physiological measurements can be used in design test phase with surrogate operators, or it can be implemented into the design of automated systems to provide performance predictions to the system. Both of these focus areas are summarized in the succeeding paragraphs.

There has been significant research into tools or models that may assist in predicting areas of a design that may result in suboptimal human-machine performance when automation is present. The lumberjack analogy applied to automation has been used to describe the general observation that while increased levels of automation improve performance in normal or routine operations, when a system failure occurs, the greater levels of automation result “in more significantly impaired performance” (Sebok & Wickens, 2017). This general observation has been widely accepted, yet there are some conditions that present exceptions, such as complex work situations, where observed experimental results indicated “situational awareness increased with DOA (degree of automation), which contradicts the lumberjack model” (Jamieson &

Skraaning, 2020). Whether the lumberjack model applies to a design or not, a number of model-based tools have been developed. For example, the model-based tools of Automation Design Advisor Tool (ADAT), Machine Integration Design Analysis System-Function Allocation Support Tool (MIDAS-FAST), and Space Performance Research Integration Tool (S-PRINT); have been assessed to provide “useful ways to predict operator performance in complex systems” when evaluated with a focus on black swan events and the lumberjack analogy (Sebok & Wickens, 2017). To improve precision in application of automation design concepts, a more methodical approach of analyzing user interaction with automated systems has been proposed through highlighting the design flaw in an altitude hold system of a commercial airliner as an example (Degani & Heymann, 2002). Research into development of an object-oriented Bayesian network model has shown significant promise in identifying emerging causal factors requiring mitigations that do not currently exist (Ancel, *et al.*, 2015). Similarly, research into modeling operator cognitive behavior resulted in proposed modifications to the GOMS (Goals, Operators, Methods and Selection rules) method which showed promise in assessing the “potential for automation-induced pilot performance problems” (Gil & Kaber, 2012). Additionally, the N-SEEV (noticing-salience, expectancy, effort, and value) model was developed to predict operator failure to notice unexpected subtle changes in the system (*i.e.*, change blindness) and has demonstrated success in a simulated environment (Wickens, Hooey, Gore, Sebok, & Koenicke, 2009). A control-theoretic framework has been developed to investigate automation and potential human-factors problems. This framework has effectively been used to develop a set of design principles to improve human-automation performance (Jamieson & Vicente, 2005). In an experiment involving the inject of failures during operation of a robotic arm, a computational model of automation complacency was validated (Wickens, Sebok, Li, Sarter, & Gacy, 2015).

Conversely, in a series of experiments researching operator vigilance when the operator is required to switch attention between multiple subtasks while supported by automation (*i.e.*, operation of multiple Unmanned Aerial Systems), it was discovered that Warm's resource theory of vigilance may require modification (Wohleber, *et al.*, 2019). The looming prediction and lower gain models successfully predicted human performance when exposed to situations with "an increase in kinematic criticality" and adaptive cruise control systems (Piccinini, *et al.*, 2020). Most notably, the models adequately modeled when drivers were exposed to scenarios with increased kinematic risk, response times were reduced and when the drivers were exposed to adaptive cruise control (ACC), reaction times "were significantly delayed" when compared to normal cruise control. To assess usability of new systems, a hybrid approach of involving many levels of automation and degrees of user participation has been proposed to address eight dimensions in usability testing (Norman & Panizzi, 2006).

Another area of research has been in the associations between physiological changes in operators and performance have been researched to gain deeper insight into human-machine performance. Operator trust in a system has been a key focus area due to its impact on performance. In a study involving sixteen participants, it was discovered that trust and distrust "can be two distinctive neural processes in human-automation interaction" with "distrust being a more complex network than trust, possibly due to the increased cognitive load" (Huang, Choo, Pugh, & Nam, 2022). In an experiment where passengers were driven around a given course either by a human driver or an automated system, it was observed that there was "a close relation between subjective trust ratings and skin conductance" was observed (Mühl, *et al.*, 2020). Additionally, in another experiment involving 35 drivers using an automated driving system, it was observed that gaze behavior correlated to their level of automation trust resulting in the proposal that "automation trust during highly automated driving might be inferred from gaze behavior" (Hergeth, Lorenz, Vilimek, & Krems, 2016). Leveraging research in eye-tracking and performance, a framework of "four stages of step-by-step integration of eye-tracking systems" has been developed in an effort to gain insights into pilot focus and underlying decision processes, (Peysakhovich, Lefraçois, Dehais, & Causse, 2018). The stages progressed from monitoring eye-tracking and performance during training on the ground with the fourth stage implemented in the flight environment with the inclusion of "authority taking by the aircraft". Research in the monitoring of pilot mental states has produced promising test results with the use of Functional Near-InfraRed Spectroscopy (fNIRS) where it was part of a passive Brain Computer Interface to monitor pilot mental states (Verdière, Roy, & Dehais, 2018). Experiments involving the use of a "Vigilance and Attention Controller" system where electroencephalography (EEG) and eye-tracking (ET) techniques provided inputs to an adaptive automation system demonstrated success in keeping operators more involved with their assigned tasks, countering the Out-Of-The-Loop phenomenon (Di Flumeri, *et al.*, 2019). It has been observed that physiological changes correlate to changes in operator performance and provide potentially useful inputs for emerging concepts such as adaptive automation.

Predictive methods to assess performance have been focused on design analysis and monitoring of operator mental states through physiological measurements. These methods have been proposed to provide feedback during the design phase as well as during operation of the system if implemented. Analysis and experimentation have indicated that both methods show promise.

2.3.3 Diagnosis of Operator Performance (Accident Investigation Tools)

Performance of automated systems and methods to measure it has been another topic of research. For human supervisory control of automated systems, the generalized metric classes of: mission effectiveness, autonomous platform behavior efficiency, human behavior efficiency, human behavior precursors, and collaborative matrices have been proposed (Pina, Donmez, & Cummings, 2008). Follow on research proposed evaluation criteria for the metrics that included experimental constraints, comprehensive understanding of metrics and relationships to other measurements, construct validity, statistical efficiency, and measurement technique efficiency (Donmez, Pina, & Cummings, 2008). While use of all the metric classes may be desired, the evaluation criteria does acknowledge that limits of the measurement conditions may restrict data acquisition. As a result, an approach that has been pursued is investigation into human factor related errors and their role in accidents.

In the case of accident investigation, the Taxonomy of Unsafe Operations was proposed to capture the causal factors associated with the human operator (Shappell & Wiegmann, 1997). Another proposed matrix model for accident causation analysis and classification is the Accident Causation Analysis and Taxonomy (ACAT) model which considers combinations of system factors and control functions (Li, Zhang, & Liang, 2017). A comprehensible system adopted by the commercial and general aviation sectors known as the Human Factors Analysis and Classification System (HFACS) has been used with the development and theoretical foundation summarized by Shappell and Wiegmann (Shappell & Wiegmann, 2000). While not the only accident causation model, when compared to the accident causation "2-4" model, it was assessed that the HFACS cause classification is more practical and the accident analysis is more convenient at the cost of being less comprehensive than the "2-4" model (Fu, Cao, & Xiang, 2017). The HFACS system has been successfully used to conduct research into general aviation accidents resulting in proposed interventions to address human causal factors in accidents (Weigmann, *et al.*, 2005). A tailored version of the HFACS system has been successfully developed to assess the human and organizational factors specific to coal mine accidents resulting in the identification of four key risk paths originating from three risk areas (Fa, Li, Liu, Qui, & Zhai, 2021). The Department of Defense adopted a tailored Human Factors Analysis and Classification System as a taxonomy for accident investigation (Department of Defense, 2005). The use of HFACS by the Department of Defense has been successfully used to recognize some trends in accident causal factors such as a "steady increase in…. skill-based errors" beginning in 1991 for U.S. Navy and U.S. Marine Corps major aircraft accidents (Shappell & Wiegmann, 2000). The HFACS system has been used successfully in several fields to identify trends in causal factors of accidents.

When discussing the systems use of the HFACS system, it is important to note a theory that is part of the system's foundation. Reason proposed the dynamic of accident causation includes active and latent human failures and summarized this theory in the Swiss Cheese Model (Reason J. , 1990). The Swiss Cheese Model recognizes that an accident is a product of human error at multiple levels of management with some errors classified as latent and some as active. The key point is that if any level of management does not act in a deficient or erroneous manner, the accident will not occur as it requires all levels to fail for the accident to occur. Additionally, it proposes that if there are more 'layers of cheese' or 'smaller holes in the cheese', the occurrence and severity of accidents would be reduced. This model is illustrated in Figure 2.3. Subsequent research based on this model has identified that approaches to preventing human error tend to focus on the human or the system. The same study observed that organizations that have a much lower accident rate recognize that “human variability is a force to harness in averting errors, but they work hard to focus that variability and are constantly preoccupied with the possibility of failure” (Reason J. , 2000). This model has been successfully used and recommended as a framework for adverse event analysis in health care (Elliott, Page, & Worrall-Carter, 2012). It was also successfully used to diagnose that active failures, rather than latent failures, contributed to the most medication incidents for direct oral anticoagulants in a study of a 48-month period in hospital setting (Haque, *et al.*, 2021). While the model has demonstrated usefulness in a number of settings, research has identified that it does not provide detail into “how the multitude of functions and entities in a complex socio-technical system interact and depend on each other” (Reason, Hollnagel, & Paries, 2006). However, the Swiss Cheese Model has been successfully used to diagnose causal factors of accident in several fields.

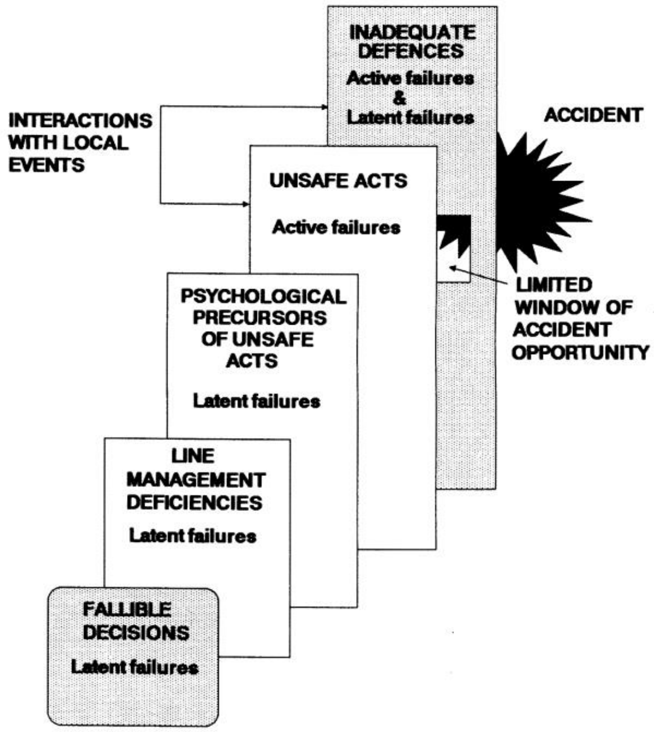


Figure 2.3: Reason's Swiss Cheese Model.

Reason proposed the dynamic of accident causation includes active and latent human failures and summarized this theory in the Swiss Cheese Model where failures must be present at many levels for an accident to occur (Reason J., 1990).

The complexity of automation design and integration has been a topic of discussion and research with identification of the interdependent nature of advances in automation and human-machine interface design. An observation within the field is that in a rush to take advantage of the benefits of automation, "it was integrated into existing work flows without fully appreciating how such a shift would change the work itself", resulting in unexpected consequences (Marquez & Gore, 2017). For example, aircraft cockpit automation has been identified as delivering many benefits, but with costs such as "mode confusion, errors of omission, and automation surprises" resulting in the proposal that a "system of defenses in depth is required" that includes improved training, procedures, and design (Olson, 2000). Additionally, it has been identified that the coevolution of automation and humans, “in which both adapt to the responses of the other” has highlighted the fact that although automation seems to relieve operators of work, “automation requires more, not less, attention to training, interface design, and interaction design” (Lee J. D., 2008). This aligns with the aviation safety community’s position that “preventing human error in aviation disasters is now the principal challenge” resulting in many proposed changes to pilot training to increase safety awareness (Tetteh, 2006). For vehicle automation, a multidisciplinary team of experts identified that automation “is not a technical innovation alone but is a social as much as a technological revolution” with five critical challenges ahead of it (Hancock, *et al.*, 2020). As a result, it has been proposed that addressing how automation influences operators, how operators can influence automation and “how interdependent interactions affect trusting automation” be a focus area of future work (Chiou & Lee, 2021).

The design and implementation of automation continues to be an area of research. The summary above was organized into the three areas of: approaches and frameworks, methods to predict operator error, and models to assess what went wrong after an accident. The field of study clearly indicates that data and discovery is possible at all phases of the system lifecycle as well as opportunities to improve human-machine performance.

To summarize, previous research into automation and human performance has resulted in frameworks to characterize automated systems, several observations of human performance with automated systems, and multiple theories of system design and performance measurement. The literature heavily referenced studies and experimental data from the use of simulators in reaching proposed findings. The use of empirical data was limited but did align at a general level with the experimental findings. Correlation between specific attributes of automation and characterizations of operator performance based on empirical data was not found in the literature.

The purpose of this research was to test the hypothesis that there is no statistically significant relationship ($\alpha \leq 0.05$) between attributes of cockpit automation and aircrew accident performance in the operational environment. Specifically, aircrew accident performance as measured using the Department of Defense tailored HFACS model (Department of Defense, 2005) was assessed for statistically significant ($\alpha \leq 0.05$) correlation to the attributes of automation level, type, and category as previously defined (Kaber & Endsley, 2004), (Parasuraman, Sheridan, & Wickens, 2000), (Dudley, *et al.*, 2014). Previous research has observed that operator performance experienced a linear correlation between automation types and categories (Wright & Kaber, 2005), (Sarter, Mumaw, & Wickens, 2007). It has also observed a nonlinear correlation between operator performance and levels of automation (Li, Wickens, Sarter, & Sebok, 2014). Results from previous experiments have suggested that this is due in part to automation induced complacency, bias, and skill atrophy (Bahardwaj, *et al.*, 2020), (Wickens, Sebok, Li, Sarter, & Gacy, 2015), (Keller, 2013).

Chapter 3. Research Methodology

3.1 Theory Base of Research

Previous research has shown that a correlation between human performance and automation exists. Operator performance has been observed to vary depending on the level, type, and category of automation present in laboratory experiments as discussed in the Chapter 2 section on Observations of Human Performance and Automated Systems. The variations of operator performance have included complacency, bias, skill atrophy, and fatigue. Limited research using empirical data from aircraft safety and accident reports has aligned at a general level with the laboratory findings. However, correlation between specific attributes of automation and operator performance in the aviation domain has not been accomplished. The theory proposed in this research is that the general alignment between empirical aircraft data and laboratory results as discussed in the Chapter 2 section on Observations of Human Performance and Automated Systems extends to a deeper level of specificity when accident data for high performance aircraft operating in the U.S. naval shipboard environment are used.

3.2 Research Approach

This research was conducted following the four steps listed below:

1) Standardize the test environment.
2) Measure the test conditions.
3) Measure the test results.
4) Analyze the test results.

3.2.1 Standardize the Test Environment

To accurately measure performance, a common test environment was required. The two primary attributes of concern were differences in the operating environments for the aircraft included in the study and any pre-existing conditions (outside the area of interest) that would cause an inequality in aircrew performance. The primary method used to address the two concerns and establish a common test environment was data selection.

Operating environments are as varied as the aircraft that fly in them. The differences between commercial, private, cargo, and international aircraft operations are numerous and require the aircrew to become proficient and maintain skillsets that are as unique as the environments themselves. To accurately compare aircrew performance between different aircraft, it was assessed that normalizing the effects of different operating environments exceeded the scope of this research. As a result, a requirement for the data to be from aircraft from a common environment with a similar exposure rate was levied.

Pre-existing conditions outside of the focus of this research that were of concern included initial skill development (initial training) and follow-on proficiency training. To operate an aircraft, aircrew are required to meet a common level of performance in multiple skills to obtain a license. While the licensing requirements are common, the training paths to earn one are numerous and can result in differences in performance of skills not evaluated during the exam. A similar effect was of concern regarding post licensing proficiency training. Proficiency training conducted post licensing exam is as varied as the number of companies and schools that comprise the aviation community. The result is that some skills demonstrated during the licensing exam will continue to be developed while others will atrophy to a certain extent. It was the impact of this variance in skill development post licensing on aircrew performance that was an additional point of concern for this research. The variances in initial skill development and post licensing proficiency training resulted in the requirement that data for this research must be from a community of aircrew subjected to a common initial and follow-on training program.

The source of data selected for this study was that of aircraft that operated from U.S. Navy ships. Accident data were acquired from the U.S. Naval Safety Center encompassing records for aircraft that operate in a sea-based environment between the years of 1980 to 2013.

To statistically evaluate if the aircraft in this study were subjected to a common operating environment, accident rates were compared over time to assess if changes for the individual aircraft occurred at the same time thus indicating exposure to a common environmental stimulus. The premise was that while changes in the environment may not produce a common response across the subject aircraft, a response would nonetheless occur. If changes in accident rates (response) correlated between aircraft, it could be surmised the operating environment was common.

Correlation was evaluated using a Spearman's Rank Correlation analysis with α=0.05 to assess the monotonic relationship between aircraft. The null hypothesis was that correlation was not present. The Spearman's Rank Coefficient was expressed between +1 to -1 with the further away the coefficient was from zero indicating the strength of the monotonic relationship. The assumptions for Spearman's Rank Correlation were met as the data were in the interval/ratio format. The selection of Spearman's Rank Correlation for analysis was based on the monotonic assessment compared to the linear evaluation of Pearson's. The assessment of a monotonic relationship was selected as changes in the environment may not produce a linear response across the subject aircraft. While the Spearman's Rank Correlation Test is robust to deviations from the normality assumption, due to the small size of some samples as a result of being removed from service during the years assessed (F-4, A-6, and F-14), a Johnson's Transformation was performed on the data for all aircraft to address any potential skewness or outliers.

3.2.2 Measure the Test Conditions

For this research, four factors to characterize composition of cockpit automation, crew size, and workload were considered for potential correlation to human performance:

1) Quantity of automation and associated characteristics (*i.e.,* level, type, category);
2) Quantity of interfaces between the aircrew and aircraft;
3) Quantity of checklists assigned to the aircrew;
4) Quantity of aircrew for each aircraft.

The quantity of automation and associated characteristics in each cockpit was assessed through completing an inventory of the automation present in each cockpit in accordance with the taxonomies provided in previous research (Kaber & Endsley, 2004), (Parasuraman, Sheridan, & Wickens, 2000), (Dudley, *et al.*, 2014). Descriptions of each cockpit present in publicly available manuals for each aircraft were reviewed and an inventory (or accounting) of the descriptions that matched the definitions in the previous literature was conducted. Each system was assessed for the category, type, and level of automation present per the descriptions provided in Figure 2.1, Table 2.1, and Figure 2.2. An example of the automation inventory data for a single aircraft is provided in Table 3.1. An example of this process with an example system is provided in Appendix A.

The quantity of interfaces between the aircrew and aircraft were also measured via completing an inventory of cockpit diagrams from data present in publicly available manuals for each aircraft. Each interface was classified as either a “control” or a “display”. Controls were defined as interfaces that allowed the aircrew to manipulate systems within the aircraft or transfer information from the human to the aircraft. Displays, in contrast, were defined as interfaces that transferred information from the aircraft to the aircrew. For example, an airspeed gauge would be classified as a “display” while a button or switch would be classified as a “control”. An example of this process is provided in Appendix B.

Table 3.1: Example of Automation Survey Data for Single Aircraft.

An example of the automation inventory data for each aircraft. The level, type, and category of automation was assessed per the descriptions provided in Figure 2.1 and Table 2.1.

Aircraft: XX	**Level (1-10)**	**Type** • Information Acquisition • Information Analysis • Decision Selection • Action Implementation	**Category** • Aircraft Performance • Mission Performance • Information Management • Life Support
System A	4	Information Acquisition	Aircraft Performance
System B	10	Action Implementation	Life Support
System C	1	Information Acquisition	Information Management
…	…	…	…

The quantity of checklists assigned to the aircrew for each aircraft was measured through referencing data present in publicly available manuals for each aircraft. Each checklist was placed into one of eight categories: Normal Procedures, Emergency Procedures, Emergency Memory Actions, All-Weather Procedures, Limitations, Warnings, Cautions, and Notes. The quantity of checklists in each category and the total number of steps in the checklists were recorded for each aircraft. An example of this process is provided in Appendix C.

Crew size was acquired through research of publicly available information. In the situation where an aircraft was designed to operate with different quantities of crew on board, the largest number was assumed as the typical quantity for flight operations.

3.2.3 Measure the Test Results

For this research, human performance was measured by assessing the impact of six bins failure modes in alignment with Department of Defense Human Factors Model (Department of Defense, 2005) as implemented by the Naval Safety Center. The bins of failure modes are characterized by a 3x2 matrix. The horizontal axis of the matrix identifies if the error resulting in an accident was attributed to a human (*i.e.*, aircraft or facility maintenance personnel, management, aircrew, etc.) or if it specifically identified aircrew error as the causal factor. The vertical axis of the matrix identifies if the error (human or aircrew) occurred in combination with a material failure of the aircraft or if the only cause of the accident was due to human error. An additional row is present to identify all accidents attributed to human or aircrew error and is a sum of the first two rows. The bins are shown in Table 3.2.

Table 3.2: Human Error Categories.

The categories of human error are divided into six subsets support analysis of different accident conditions. Aircrew Error (AE) is a subset of Human Error (HE). Similarly, the presence of material failure (MF) or lack of material failure (O) are both categories that are a subset of all errors (ALL).

	Human Error (HE)	**Aircrew Error (AE)**
Material Failure Present (MF)	HE-MF	AE-MF
Material Failure Not Present (O)	HE-O	AE-O
Total Error (ALL)	HE-ALL	AE-ALL

The impact of each bin of the failure modes was measured by calculating the cost of the accident in financial cost (expressed as a percentage of aircraft cost per accident), lives lost (expressed as a percentage of crew size per accident), and accident rate (expressed as average number of accidents per 1000 flight hours per year). The impact measurements are shown in Table 3.3.

Table 3.3: Impact Measurements.

Financial Cost, Lives Lost (Fatalities), and Accident Rate are the three values used to assess accident performance. Financial Cost was normalized across the different aircraft through expression as a percentage of total aircraft cost. Lives Lost (Fatalities) was normalized across the different aircraft through expression as a percentage of crew size. Accident Rate was calculated as the number of accidents per 1000 flight hours accumulated by the specific model of aircraft in a given year.

Impact	**Measurement Calculation**
Financial Cost	**% of Aircraft Cost** = Cost ($) / Total Aircraft Cost ($)
Lives Lost (Fatalities)	**% Crew Size** = Lives Lost (#) / Aircraft Crew Size (#)
Accident Rate	**Rate** = Number of Accidents (#) / 1000 Flight Hours

For accidents where multiple factors were assessed to be causal, the impact was divided equally to each of the causal factors identified. This is in alignment with Reason's Swiss Cheese Model (Reason J. , 2000) and the DoD HFACS accident investigation model (source of accident data) (Department of Defense, 2005) as all causal failures were required to allow the accident to occur. If any of the causal factors had not be present, the 'hole in the layer of cheese' would not have been present and the accident would have been avoided. Thus, each causal factor is equally responsible for the accident as if any factor was not present, the accident would have been avoided. An example of this process is included in Appendix D. The result of this approach was a total of 18 measurements of performance were calculated annually for each aircraft. An example is shown in Table 3.4.

Statistical analysis of human performance was conducted through the use of a Randomized Complete Block Design (RCBD) with Fisher's Pairwise Comparison (α=0.05) and blocking performed on the individual years. The assumptions for this method were met as the data consisted of yearly averages, thus invoking the Central Limit Theorem, and were inclusive of all events over the given period. Other statistical methods were considered such as Tukey, Bonferroni, Sidak, and Dunnet. Due to concerns with Type II error and the lack of a control group for comparison, Fisher's Pairwise Comparison was selected. The null hypothesis for the Fisher's Pairwise Comparison was that the performance of each aircraft (measured as percentage cost, percentage crew fatalities, and accident rate) was the same.

Table 3.4: Performance Measurement Example.

This is an example of the 1-year performance data calculated for each aircraft. The three impact measurements (Financial, Fatalities, and Accident Rate) are calculated for the six human factor categories (Human Error–Material Failure, Human Error–Only, Human Error-All, Aircrew Error-Material Failure, Aircrew Error-Only, and Aircrew Error-All).

Aircraft: XX-XX **Year: XXXX**	**Financial** **(% Aircraft)**	**Fatalities** **(% Crew Size)**	**Accident Rate** **(Events / 100K Hours**
Human Error – Material Failure (HE-MF)	XX %	XX %	XX / 100K
Human Error – Only (HE-O)	XX %	XX %	XX / 100K
Human Error – All (HE-ALL)	XX %	XX %	XX / 100K
Aircrew Error – Material Failure (AE-MF)	XX %	XX %	XX / 100K
Aircrew Error – Only (HE-O)	XX %	XX %	XX / 100K
Aircrew Error – All (HE-ALL)	XX %	XX %	XX / 100K

3.2.4 Analyze the Test Results

For this research, analysis focused on correlation between the attributes of automation described in the Test Conditions section and the 18 performance measurements described in the Test Results section expressed as deviations from a Generalized Linear Model. Correlation was evaluated using a Pearson's Correlation with α=0.05 to assess the linear relationship between automation and performance with the null set to no correlation was present. The Pearson's Rank was expressed between +1 to -1 with the further away the coefficient was from zero indicating the strength of the linear relationship. The assumptions for Pearson's Correlation were met as the data were interval. Spearman's Correlation was considered for the analysis, but ultimately not selected as the sensitivity associated with linear analysis was preferred.

3.3 Data Preparation

Preparation of data for this research consisted of selecting data that met the requirements for the research, standardizing the performance metrics for comparison, and appropriate allocation of performance measurements to the failure modes present for each accident.

3.3.1 Data Selection

Data selection for this research was limited to information available in the public forum or acquired through the Freedom of Information Act (FOIA) process. To meet the criteria that the subjects of this research be exposed to a common environment, it was decided to request accident data from the U.S. Naval Safety center for aircraft that operate in the ship-based environment. It was also assessed that the initial and follow-on training for the subject aircrew was standardized and common.

The accident data consisted of 3,249 accident records that covered the period from 1980 to 2013. Each record identified the year, cost, number of fatalities and causal factors for each accident. The cost was listed in U.S. dollars and the causal factors of each accident were listed as one or more of the items shown in Figure 3.1.

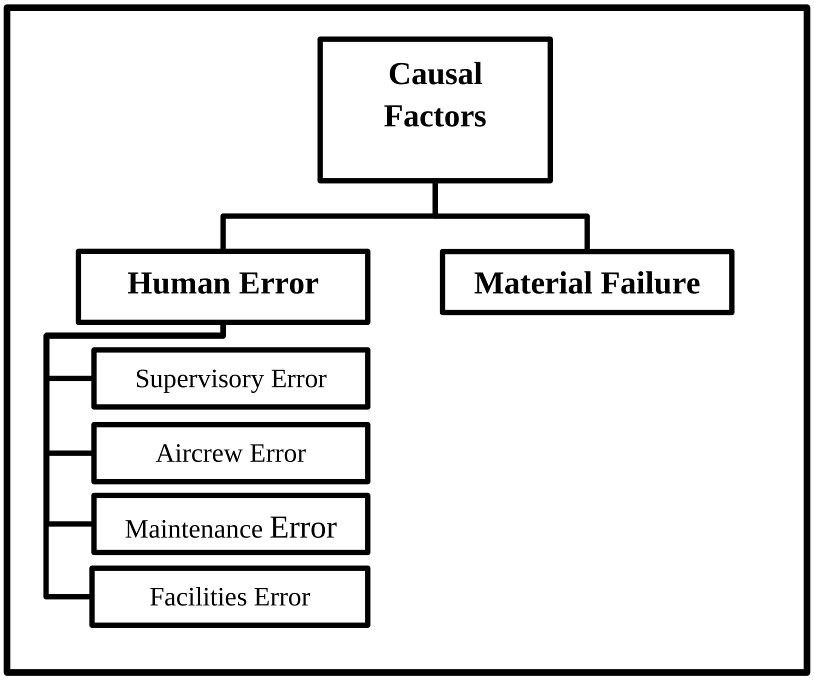


Figure 3.1: Causal Factor Hierarchy.

Accident data provided by the U.S. Navy Safety Center listed causal factors for each of the events. Of the five causal factors, four were listed in the human error category and one was material failure. This figure illustrates the hierarchy of potential causal factors for each accident.

Detailed specifics of the causes of each accident were not available, but a one-line narrative was provided for 1,746 (53.7%) of the records. Additionally, the total number of hours each aircraft operated annually was provided and indicated that four of the eight aircraft used in this study retired from operational service between the years of 1980 to 2013. For the purpose of this study, retirement from operational service was defined as the year when total annual flight time for a model of aircraft dropped below 2,000 hours. The aircraft, associated years of operation, and values that were used in this study are shown in Table 3.5.

Table 3.5: Aircraft Summary.

* "In Operation" was assessed as of 2013 based on data provided by the U.S. Naval Safety Center. Aircraft may have left since that time.
** FA-18A/B/C/D had a calculated value of $26M. The FA-18E/F had a calculated value of $80.3M.
*** FA-18 is a single seat aircraft for the FA-18A/C/E models. The FA-18B/D/F models are two seat aircraft.

Aircraft	Entry Into Operational Service	Last Year of Operational Service*	Aircraft Value	Aircraft Crew Size
C-2	1966	In Operation	$39,000,000	2
E-2	1964	In Operation	$40,717,000	5
AV-8	1985	In Operation	$18,746,000	1
A-6	1963	1996	$12,221,000	2
EA-6B	1971	In Operation	$19,662,796	4
F-4	1960	1991	$2,847,835	2
F-14	1974	2006	$19,256,805	2
FA-18	1983	In Operation	$26,378,904**	1 or 2***

3.3.2 Performance Metrics

Human performance was measured using three parameters. The first was the annual accident rate per aircraft measured as the average number of accidents that occurred per 1,000 hours of flight time for each year the aircraft was in operation. The second was the cost of each accident measured as the percentage of aircraft damaged. The third parameter was the number of fatalities associated with each accident expressed as the percentage of aircraft crew size.

In assessing cost and fatalities for each accident, percentages were used to account for differences between the aircraft. Without taking into consideration differences in aircraft value, the data would have erroneously indicated that accidents involving older aircraft were much less severe than those involving newer aircraft. For example, an accident resulting in the total loss of an F-4 would cost the same as an accident resulting in relatively minor damage to an FA-18E.

While the loss of an aircraft is a much larger impact, the differences in aircraft cost would not indicate as much. The same reasoning was used in accounting for the number of fatalities associated with each accident. The total value and crew size used in this study for each aircraft is listed in Table 3.5.

3.3.3 Allocation of Performance Metrics

For the accidents that listed more than one causal factor, the performance metrics of cost and fatalities per accident were distributed as an equal ratio among the multiple factors. An example of the method is shown in 0. This approach was used based on the theory underlying Reasons Swiss Cheese Model (Reason J. , 2000), specifically that if either one of the factors was not present, the accident would not have occurred. With the factors being equally culpable, each factor was assigned an equal burden of the consequence.

3.4 Data Analysis

Analysis of the data was conducted in three steps. The first was to assess if the test environment was common between the subjects. The second was to compare performance between the subjects. The third was to assess if there was correlation between attributes of cockpit automation and performance.

To assess if the test environment was common, correlation between accident rates was conducted. The accident rate for each aircraft was calculated per year. A Spearman's Correlation Test with α=0.05 was conducted to determine in changes in accident rates correlated between the aircraft. If correlation was present, it would suggest that the subject aircraft were exposed to a common event that caused the change in performance. If the aircraft experienced changes in performance that did not correlate to each other, it would suggest the subjects were not exposed to a common environment. For this correlation, the direction or magnitude of change was not of concern as each subject could react differently to changes in the environment.

Comparison of the three performance measurands (accident rate, accident cost, and accident fatalities) were analyzed by using a randomized complete block design (RCBD) with each year being treated as a block. The difference in annual performance for each aircraft when compared to the average performance of all aircraft was calculated. Using Fisher's Pairwise Comparison with $\alpha=0.05$, the differences in annual performance for each aircraft were analyzed. Correlation between the three aircrew accident performance measurands (assessed as statistically different through the method described above) and attributes of automation systems were assessed using Pearson's Correlation with $\alpha=0.05$. The factors tested for correlation are listed in Table 3.6.

For this research, α was set to 0.05 with an associated potential of Type I errors. To mitigate against the potential of false positives, the focus of the analysis was on pairings that had previously demonstrated correlation in previous research as described in Chapter 2 (Li, Wickens, Sarter, & Sebok, 2014), (Wright & Kaber, 2005), (Sarter, Mumaw, & Wickens, 2007). With α set to 0.05, a potential of 8 positive findings could be the result of Type I error. Since the purpose of this research was to assess if statistically significant ($\alpha \leq 0.05$) correlation between the documented characteristics of operator performance and specific attributes of automation exist using empirical data outside of a scripted environment, the potential of Type I error is acknowledged.

Table 3.6: Correlation Factors.

* Category: Aircraft, Mission, Information (Dudley, *et al.*, 2014).
** Types: Information Acquisition, Information Analysis, Decision Selection, Action Implementation (Parasuraman, Sheridan, & Wickens, A Model for Types and Levels of Human Interaction with Automation, 2000).
*** Levels: The 10 levels recommended by Endsley and Kaber (Kaber & Endsley, The Effects of Level of Automation and Adaptive Automation on Human Performance, Situational Awareness and Workload in a Dynamic Control Task, 2004).

Attributes			Performance	
Automated Systems	**Checklists**	**Human-Vehicle Interface**	**Causal Factor**	**Performance Measurement**
Category*	Normal Procedures	# of Displays	Human Error-All (HE-ALL)	Accident Rate
Types**	Emergency Procedures	# of Switches	Human Error-Only (HE-O)	Accident Cost (% Aircraft Price)
Levels***	Emergency Memory Action Item	# of Crew	Human Error w/ Material Failure (HE-MF)	Accident Fatalities (% Crew Size)
	All-Weather Procedures		Aircrew Error-All (AE-ALL)	
	Limitations		Aircrew Error-Only (AE-O)	
	Warnings		Aircrew Error w/ Material Failure (AE-MF)	
	Cautions			
	Notes			

3.5 Limitations and Key Assumptions

Limitations to this research were primarily due to data availability. Accident data were limited to what was available from the Naval Safety Center through the Freedom of Information Act (FOIA) process. The research was limited in depth and granularity of accident specifics due to the privacy rights of those involved in the accidents and limitations of access to the reports used in this study.

This research was also limited to the cockpit automation data available for the subject aircraft. The data used for this research were acquired from manuals published on the internet. The manuals provide a list and description of systems that are either in the cockpit or related to systems in the cockpit. A general observation was that the descriptions of systems were limited to a depth of what was assessed as useful to the aircrew. As a result, it is understood that the descriptions are not all inclusive of the attributes of automation that would be expected in engineering level documents.

A third limitation of this research was that assessment of the configuration of automated systems in the subject aircraft was conducted by one person, the author. The assessments were conducted in alignment with the taxonomies proposed by previous research. Descriptions of the categories and types of automation were very clear, and it is expected that similar results would be achieved by other assessors. Descriptions of the levels of automation did provide more potential for differences in assessment between the 10 discrete levels. To mitigate the impact of a difference in assessment of a single level (*i.e.*, a system assessed at a level 4 instead of a level 5), the correlation between levels of automation and performance used the quantity of systems in three groups of automation (*i.e.*, Levels 1-3, 4-7, and 8-10).

The first assumption of this research was that human error events recorded in accident reports were proportionally representative of all human error events that occur during flight events. The data for this research were limited to human errors that were assessed to have a causal role in an accident. Since there are human error events that do not result in an accident, it is assumed that the number of human errors reported in the accident reports were proportionate to the total number of human errors that occurred during the years 1980-2013. This assumption was the basis behind using accident reports as a safety measurement of performance.

The second assumption was that although the aircraft studied in this research each have a different specific mission within the aircraft carrier environment, the basis of aircraft operation comes from a standard training program that provided a common level of performance between each of the groups (*i.e.*, aircrew for each aircraft). Follow-on safety training is prescribed by a common organization and each aircrew must pass standardized exams annually that are managed and proctored by the same common organization. While mission specific skills are required for each aircraft, safe operation of the aircraft has been standardized across the different aircraft and aircrew. As a result, the differences in assigned mission and aircrew training for the subject aircraft were assumed to be negligible.

Chapter 4. Research Results and Analysis

4.1 Overview

The results of this research are divided into four areas. The first area is the assessment of a standardized test environment for the aircraft in this study. The second is the measurement of the test conditions (*i.e.*, differences in cockpit automation present in the different aircraft). The third is the measurement of differences in human accident performance between the different aircraft. The fourth and final area is an assessment of correlation between the presence of cockpit automation and human accident performance.

4.2 Standard Test Environment Results and Analysis

To evaluate if the aircraft in this study were subjected to a common operating environment, accident rates were compared over time to assess if changes in rates for the individual aircraft occurred at the same time indicating exposure to a common environmental stimulus.

A Spearman's Rank Correlation analysis with α set to 0.05 was conducted for all combinations of aircraft in each year group with the null hypothesis set as no monotonic relationship exists. The selection of $\alpha=0.05$ for analysis was to identify correlation that met the commonly accepted value of $\alpha=0.05$. Pairings with alpha values just outside of 0.05 (but less than 0.10) were noted due to proximity to the commonly accepted criteria to reject the null and potential to reject the null if more data is available in the future. A summary of the results is shown in Table 4.1.

Table 4.1: Rate of Human Error (RoHE) Correlation (1980-2013).

To evaluate if changes in accident rates correlated between aircraft, a Spearman's Rank Correlation analysis with α set to 0.05 and 0.10 was conducted for all combinations of aircraft in each year group with the null hypothesis set as no monotonic relationship exists. With α set to 0.05, the null was rejected for all pairs except those outlined with a solid or dashed box. With α set to 0.10, the null was rejected for all pairs except those outlined with a solid box.

	C-2 RoHE	F-4 RoHE	A-6 RoHE	AV-8 RoHE	E-2 RoHE	EA-6 RoHE	F-14 RoHE
F-4 RoHE	0.147 (0.648)						
A-6 RoHE	0.076 (0.771)	0.781 (0.003)					
AV-8 RoHE	0.048 (0.787)	0.662 (0.019)	0.623 (0.008)				
E-2 RoHE	0.209 (0.235)	0.661 (0.019)	0.552 (0.022)	0.396 (0.020)			
EA-6 RoHE	0.081 (0.649)	0.960 (0.000)	0.777 (0.000)	0.246 (0.162)	0.355 (0.040)		
F-14 RoHE	-0.045 (0.825)	0.886 (0.000)	0.846 (0.000)	0.406 (0.036)	0.372 (0.056)	0.798 (0.000)	
F-18 RoHE	0.368 (0.032)	0.764 (0.004)	0.843 (0.000)	0.255 (0.145)	0.535 (0.001)	0.570 (0.000)	0.579 (0.002)
Legend	Spearman's Rank Correlation Coefficient (P-Value)						

A total of 19 of the 28 pairings rejected the null with α set to 0.05 indicating a monotonic relationship between annual accident rates. The expected number of false positives was 1.4 (with α set to 0.05). If the risk of false positives was realized and the quantity of pairs that rejected the null was 17 of the 28, the results would still indicate that the majority of pairs exhibited a monotonic relationship. The nine pairs that did not reject the null with α set to 0.05 are highlighted in Table 4.1. The two aircraft that are most prevalent in the pairings that did not reject the null are the C-2 and AV-8 (accounting for a total of 8 of the 9 pairs). Of note, the E-2 and F-14 pairing did not reject the null with α set to 0.05 but did produce a p-value of 0.056.

This is noted due to the proximity to the commonly accepted threshold for rejecting the null and potential to meet that criterion if a greater scope of data were available.

Looking at the accident rate summary shown in Figure 4.1, the C-2 experienced a smoother trend between different years than the other aircraft and appeared to be less susceptible to fluctuations in operational tempo during the time period evaluated. Upon evaluating the annual flight hours flown and number of accidents per year, the C-2 experienced the lowest variance in both for the aircraft included in the study. This indicates the operational tempo, or pace, of the C-2 community as well as the number of human errors (*i.e.*, accidents) was relatively stable in contrast to the changes observed for the rest of the aircraft in the study.

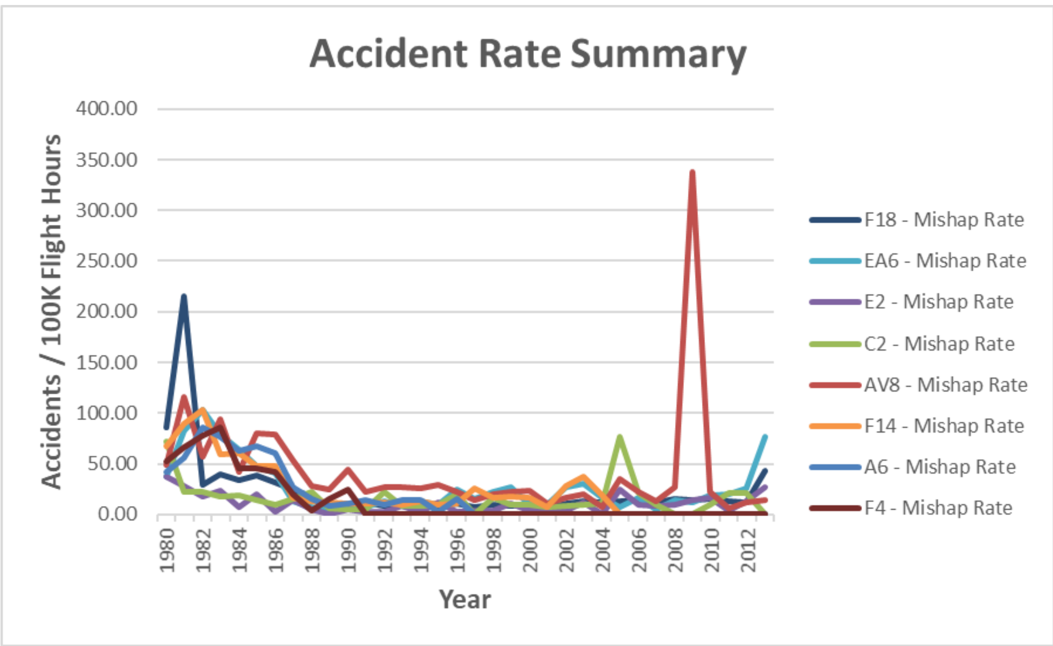


Figure 4.1: Accident Rate Summary.

Annual accident rates for each model of aircraft are displayed. Of note, the C-2 accident rate was more consistent than the other aircraft and appeared to be less susceptible to fluctuations in operational tempo during the time period evaluated. The AV-8 experienced a significant increase in accident rate in 2009 which corresponded to a significantly reduced number of flight hours flown that year resulting in increased sensitivity any accident event. The F-18 entered operational service in 1983 resulting in the uptick in accident rate observed in 1981 not included in the analysis of this report. It is included here for data completeness.

Of the nine pairs that did not reject the null, two had the AV-8 as one of the two samples in the pair. This indicates that a monotonic relationship between the annual accident rates of the AV-8 when paired with the EA-6 or F-18 did not exist. However, when the Spearman's Rank Correlation Test was run with the 2009 accident rate removed for the AV-8, the AV-8/F-18 Correlation Coefficient was 0.351 with a p-value of 0.045 and the AV-8/EA-6 Correlation Coefficient was 0.333 with p-value of 0.058. This indicates that when the data from a year with exceptionally low flight hours (and significantly increased sensitivity to impact of an accident to the accident rate measurement) are removed, monotonic relationships appear. As a result, it is assessed that the AV-8 did have a monotonic relationship with the F-18 when and EA-6 during the period of study.

Of the nine pairs that did not reject the null, one was the pairing of the E-2 and F-14. For this pairing, α=0.056 which was slightly greater than the community accepted value of 0.05. In this case, the sample size was smaller than others due to the F-14 being retired from service at the end of 2006. While the Spearman Rank Correlation Test accounts for the unequal sample sizes, the deviation from the accepted α value is small. Data prior to 1980 for this pairing of aircraft were not available, however the data indicate that a monotonic relationship between the E-2 and F-14 accident rates may exist.

In evaluating if the subject aircraft were exposed to a common environment, a Spearman's Rank Correlation Test was conducted for each combination of pairing of subject aircraft. A monotonic relationship was observed between the annual accident rates for each pair of aircraft except when the C-2 was paired with any aircraft other than the F-18. This indicates that the aircraft were exposed to common environmental changes during the period of study (except for the C-2). Based on this finding and the research of the organizational relationship between the subject aircraft, it is assessed that the subject aircraft were exposed to a common operational environment.

4.3 Measurement of Test Conditions and Analysis

Four factors were surveyed for potential correlation to human performance:

1) quantity of automation and associated characteristics;
2) quantity of interfaces between the aircrew and aircraft;
3) quantity of checklists assigned to the aircrew;
4) quantity of aircrew for each aircraft.

The quantity of automation and associated characteristics are summarized in Table 4.2 and the quantity of interfaces and checklists are summarized in Table 4.3. It is acknowledged that there is some subjectivity associated with classification of automation characteristics. To mitigate impacts of differences in how a system would be ranked or classified, groupings were used for automation levels at clear conceptual break points and classification criteria for automation types and categories were applied consistently by the author.

4.4 Measurement of Human Performance

To assess performance, the three impact measurements (financial cost, lives lost, and accident rate) were calculated for the six human error categories (HE-MF, HE-O, HE-ALL, AE-MF, AE-O, AE-ALL) summarized in the example shown in Table 3.4. A RCBD with Fisher's Pairwise Comparison (α=0.05) was conducted with blocking performed on the individual years for each combination of human error category and impact. The result was groupings of aircraft with similar performance that failed to reject the null. The results and analysis of each impact measurement are described below.

Table 4.2: Survey of Cockpit Automation.

Survey of Cockpit Automation is summarized by aircraft. To mitigate impacts of subjective criteria, groupings were used for automation levels at clear conceptual break points and classification criteria for automation types and categories were applied consistently.

Automation Characteristics	F-4	A-6	E-2	C-2	EA-6B	F-14	FA-18	AV-8
Total Systems	412	59	160	155	154	184	131	178
Level 8-10	154	29	104	72	92	100	83	119
Level 4-7	57	10	17	13	19	34	28	6
Level 1-3	201	20	39	70	43	50	30	53
Information Acquisition	112	14	62	55	69	55	55	95
Information Analysis	1	0	0	0	0	0	2	0
Decision Selection	2	0	0	0	0	5	0	0
Action Implementation	297	45	98	100	85	124	74	83
Aircraft Performance	144	17	70	79	54	88	46	60
Mission Performance	83	15	29	24	26	40	11	17
Information Management	131	9	35	32	38	42	55	54
Life Support	54	18	26	20	36	14	19	47

Table 4.3: Survey of Interfaces and Checklists.

Survey of human-machine interfaces and checklists items is summarized by aircraft.

Aircrew-Aircraft Interfaces	F-4	A-6	E-2	C-2	EA-6B	F-14	FA-18	AV-8
Displays	151	93	109	168	193	223	125	133
Switches	226	348	202	306	491	386	443	247
# Checklists	214	175	185	166	145	232	391	302
# Checklist Steps	937	1635	1611	1474	1180	1574	2301	1463
# Memory Items	82	172	215	227	241	325	213	442

4.4.1 Accident Rates - Results

Accident rates were calculated annually for each aircraft using the metric of number of accidents per 100,000 flight hours and are summarized in Figure 4.1. Of note, the AV-8 experienced a spike in accident rate during 2009 due to an extremely low number of hours flown as discussed in the analysis of a standard test environment. While the number of accidents involving an AV-8 was near the average of all the years evaluated, the exceptionally low number of flight hours resulted in the rate being exceedingly high for that year. A similar situation occurred involving the F-18 in 1982, however it did not have the same effect as the AV-8. The analysis was conducted for the six human error categories and are described below.

4.4.1.1 Accident Rate – HE-ALL

The first accident rate evaluated was that of accidents caused by any human error (HE-ALL). The results of the Fisher's Pairwise Comparison are listed in Table E.1. Three statistically significant performance groups were identified with 5 of the 8 aircraft having performance that corresponded to more than one group. Group A experienced the greatest accident rate while Group C experienced the lowest. Of note, Group C does not include the F-18 despite it having a relative mean that was less than the F-4 (a member of Group C). This is due to the small sample size of the F-4 performance relative to the F-18. It is expected that if the F-4 had a larger sample size with consistent performance, it would have been removed from Group C. Of the three groups, there is a significant overlap between the aircraft with the exception of three. The AV-8, F18, and E-2 each only belong to one group while all other aircraft in the study belong to two or all three groups. The AV-8 human error rate rejected the null when compared to that of the F-18, C-2, and E-2. Key attributes that differ between the F-18 and the other three aircraft are the increased number of checklists and associated steps for the F-18. When the E-2 human error rate was evaluated against all aircraft other than the C-2 and F-4, the null was rejected as well. Key attributes of the E-2 that separate it from the other aircraft are the low number of systems in the A-6, greater quantity of checklists and associated steps in the F-18, and the large quantity of memory items in the AV-8.

While there were differences in the attributes associated with the automation characteristics associated with the aircraft, ultimately, they did not demonstrate signification correlation. The observations are annotated here for the potential of future research.

4.4.1.2 Accident Rate – HE-MF

The second accident rate evaluated was that of accidents caused by human error and material failure (HE-MF) with the results of the Fisher's Pairwise Comparison listed in Table E.2. The results indicated that the AV-8 experienced a greater accident rate than the other aircraft and the E-2 experienced the lowest rate of accidents. Four statistically significant performance groups were identified with 6 of the 8 aircraft having performance that corresponded to more than one group. Group A experienced the greatest accident rate while

Group D experienced the lowest. Of note, Group A does not include the F-14 despite it having a relative mean that was less than the F-4 (a member of Group A). This is due to the small sample size of the F-4 performance relative to the F-14. It is expected that if the F-4 had a larger sample size, its performance would remove it from Group A.

The rate of accidents caused by human error with material failure accounted for errors due to aircrew and non-aircrew personnel. The results of the analysis indicated that performance among the aircraft statistically fell into four groups. While the groups were not mutually exclusive, the AV-8 and E-2 were the only two aircraft with performance that only fell into one group. Key attributes that differentiate the AV-8 from the C-2 and F-4 are the increased quantity of memory items the AV-8 aircrew are required to recall in event of an accident, the relatively low total number of checklist steps for the operation of the F-4, and the large number of total systems in the F-4. The main attributes of the E-2 that differed from the F-14, C-2, and AV-8 were the crew size of the E-2 and aforementioned memory items of the AV-8. While none of these three attributes ultimately correlated to a difference in performance (as will be covered in a later section), they are items of note for potential future research.

4.4.1.3 Accident Rate – HE-O

Rate of accidents caused only by human error was the third measure of performance evaluated and the results of the Fisher's Pairwise Comparison are listed in Table E.3. The results indicated that the AV-8 experienced a greater accident rate than the other aircraft and the E-2 experienced the lowest accident rate. Three statistically significant performance groups were identified with six of the eight aircraft having performance that corresponded to more than one group. Group A experienced the greatest accident rate while Group D experienced the lowest.

Of the eight aircraft evaluated, six belonged to two or more groups while the AV-8 and E-2 both only belonged to one. The accident rate associated only with human error for the AV-8 rejected the null when compared to the C-2 and E-2 with the primary difference between the AV-8 and the other aircraft being the AV-8's large number of memory items. Additionally, the error rate of the E-2 rejected the null as well when compared to all aircraft except the F-4 and C-2. The primary difference between the E-2 and the other aircraft was its larger crew size. When correlation analysis was conducted for all attributes included in this study, the null was not rejected for any of them when compared with rates of accidents caused only by human error.

4.4.1.4 Accident Rate – AE-ALL

The fourth accident rate evaluated was that of accidents caused by aircrew error with or without material failure and the results of the Fisher's Pairwise Comparison are listed in Table E.4. The results indicated that the AV-8 experienced a greater accident rate than the other aircraft and the E-2 experienced the lowest rate of accidents. Three statistically significant performance groups were identified with 4 of the 8 aircraft having performance that corresponded to more than one group. Group A experienced the greatest accident rate while Group C experienced the lowest.

Of the eight aircraft evaluated, four belonged to two groups while the AV-8, F-14, C-2, and E-2 each only belonged to one. The accident rate associated with aircrew error for the AV-8 rejected the null when compared to all other aircraft. The accident rates of the F-14 and C-2 aircraft rejected the null when compared with the other six aircraft. Additionally, the accident rates associated with aircrew error for the E-2 rejected the null when compared to the C-2, F-14, and AV-8. Correlation analysis between attributes of automation and the rate of accidents caused all or in part by aircrew error failed to reject the null for all attributes.

4.4.1.5 Accident Rate – AE-MF

Accidents caused by aircrew error and material failure was the fifth accident rate evaluated with the results of the Fisher's Pairwise Comparison listed in Table E.5 and the associated. In a change from the trend of performance in previous rankings, the results indicated that the F-14 experienced a greater accident rate than the other aircraft and the E-2 experiencing the lowest rate of accidents. Three statistically significant performance groups were identified with 5 of the 8 aircraft having performance that corresponded to more than one group. Group A experienced the greatest accident rate while Group C experienced the lowest.

Of the eight aircraft evaluated, three belonged to only one group. The F-14, AV-8, and E-2 each belonged to one group while the other five were associated with two or more performance groups. The F-14 and AV-8 rejected the null when the accident rates associated with aircrew error and material failure were compared with the F-18 and E-2. The E-2 accident rate associated with aircrew error and material failure rejected the null when compared to the C-2 as well as the F-14 and AV-8. Correlation analysis between attributes of automation and rates of accidents caused by aircrew error and material failure did not indicate the presence of correlation for any attribute as they all failed to reject the null.

4.4.1.6 Accident Rate – AE-O

The final accident rate evaluated was that of accidents caused by only by aircrew error with the results of the Fisher's Pairwise Comparison listed in Table E.6. The results indicated that the AV-8 experienced a greater accident rate than the other aircraft and the E-2 experienced the lowest rate of accidents. Two statistically significant performance groups were identified with the AV-8 being the only one in Group A. Group A experienced the greatest accident rate while Group B experienced the lowest.

The rate of accidents caused only by aircrew error indicated that performance among the aircraft statistically fell into two groups. The groups were mutually exclusive with the AV-8 being the only aircraft in Group A and all other aircraft populating Group B as shown in Table E.6. Most notably is the difference between the relative mean rate of only aircrew error events of the AV-8 compared to the other aircraft. While the range of mean rates of aircrew only error events is significant, over half of the range is accounted for by the performance of the AV-8. This indicates that an attribute of the AV-8 design may have a significant influence on aircrew error when absent any material failure.

When correlation analysis was performed, crew size correlated negatively to accident rate only due to aircrew error. A negative correlation was observed as shown in Figure 4.2 with a Pearson's Correlation Coefficient of -0.716. Of note, subsets of aircrew error rate (*i.e.*, rate of accidents due only to aircrew error and rate of accidents due to aircrew error and material failure) showed a similar correlation.

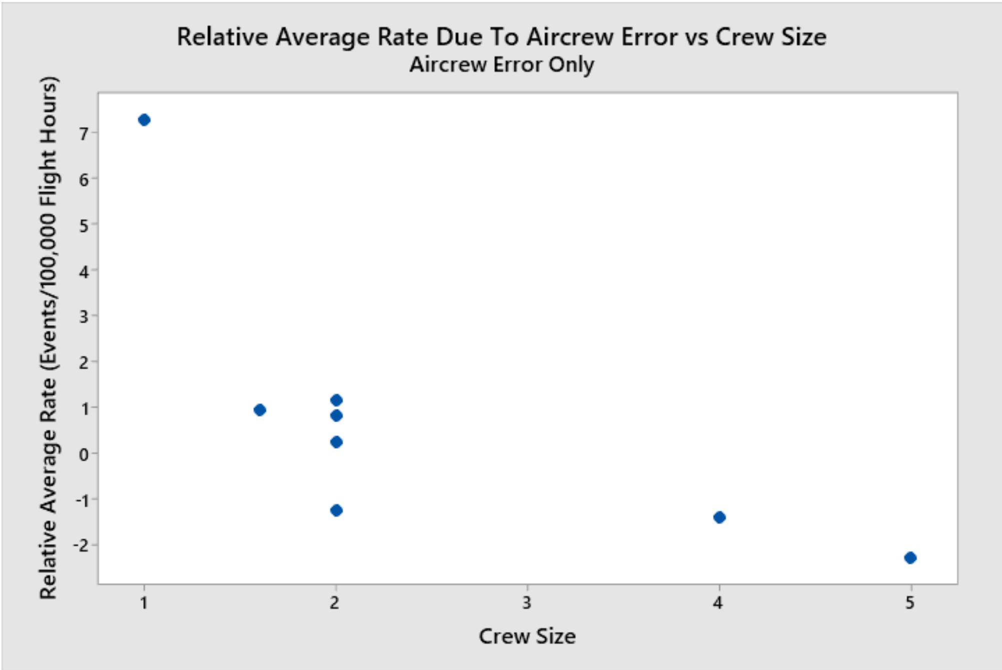


Figure 4.2: Crew Size vs Rate of Aircrew Error Without Material Failure. Pearson's Correlation Coefficient of -0.716 was observed between accident rates due to any aircrew error and crew size. This observation is in line with Reason's Swiss Cheese Model (Reason J., 2000) as Reason states that in systems with multiple levels of redundancy, a number of failures must align for an accident to occur. Larger crew sizes inherently have additional layers of redundancy to prevent aircrew error resulting in an accident.

This result is in line with Reason's Swiss Cheese Model (Reason J. , 2000) as Reason states that in systems with multiple levels of redundancy, several failures must align for an accident to occur. Larger crew sizes inherently have additional layers of redundancy to prevent aircrew error resulting in an accident. However, the lack of statistically significant correlation between crew size and rate of events caused by aircrew error and material failure indicates that a similar relationship may not be present between crew size and aircrew performance when forced to react to an unplanned event such as the failure of an aircraft component.

4.4.1.7 Accident Rates - Summary

Accident rates differed between the six combinations of human error, but relative ranking of the aircraft remained relatively consistent as shown in Table E.7. Of note, the AV-8 demonstrated the highest rate of accidents for five of the six combinations of human error and the E-2 had the lowest rate for all six. The A-6 and EA-6 experienced a relative higher accident rate for accidents caused by human error when compared to the other aircraft than when only aircrew error was measured. This indicates that errors due to supervisory, maintenance, or facility personnel had an impact on accident rates that resulted in an increase. Conversely, accident rates involving material failure indicated a lower ranking than those solely attributed to human and aircrew error for the F-18. It should be noted that the A-6 entered operation in 1963 and the F-18 was introduced in 1983.

The only attribute that correlated to accident rates was crew size. The size of the aircrew present in each aircraft negatively correlated to the rate of accidents caused only by aircrew error. This is in line with Reason's Swiss Cheese Model (Reason J. , 2000). All other attributes failed to reject the null when assessed for correlation to any of the six combinations of error.

4.4.2 Accident Financial Costs - Results

The financial costs of accidents due to the causal factors listed in Table 3.6 were calculated annually for each aircraft using the metric of damage due to the accident expressed as a percentage of aircraft value.

4.4.2.1 Accident Cost – HE-ALL

The first causal factor evaluated was human error. The results of the Fisher's Pairwise Comparison are listed in Table E.8. The results indicated that the F-4 experienced a greater level of damage on average than the other aircraft and the C-2 experienced the lowest. Five statistically significant performance groups were identified with 6 of the 8 aircraft having performance that corresponded to more than one group. Group A experienced the greatest average level of damage while Group E experienced the lowest. Of note, Group B does not include the AV-8 despite it having a relative mean that was less than the F-4 (a member of Group B). This is due to the small sample size of the F-4 performance relative to the AV-8. It is expected that if the F-4 had a larger sample size, its performance would remove it from Group B.

Of the eight aircraft evaluated, two belonged to only one group. The AV-8 and C-2 each belonged to one group while the other six were associated with two performance groups. The F-4 rejected the null when the accident costs (expressed as a percentage of aircraft value) associated with human error were compared with all other aircraft except the AV-8 and F-14. The C-2 rejected the null when accident costs associated with human error were compared with all other aircraft except the E-2 and A-6. Correlation analysis between attributes of automation and costs of accidents caused by human error did not indicate the presence of correlation for any attribute as they all failed to reject the null.

4.4.2.2 Accident Cost – HE-MF

The second causal factor evaluated was human error with material failure. The results of the Fisher's Pairwise Comparison are listed in Table E.9. The results indicated that the F-4 experienced a greater level of damage on average than the other aircraft and the C-2 experienced the lowest. Three statistically significant performance groups were identified with only 2 of the 8 aircraft having performance that corresponded to more than one group. Group A experienced the greatest average level of damage while Group C experienced the lowest.

Of the eight aircraft evaluated, six belonged to only one group and only two belonged to multiple groups. The costs (expressed as a percentage of aircraft value) associated with accidents due to human error and material failure for the A-6, F-14, and AV-8 rejected the null when compared to the other five aircraft. The costs of accidents due to human error and material failure for the F-18 rejected the null when compared to all aircraft except the A-6 and EA-6.

Additionally, the costs associated with accidents for the E-2 and C-2, rejected the null when compared to all aircraft except the A-6 and EA-6. Correlation analysis between attributes of automation and costs of accidents caused by human error with material failure did not reject the null, indicating no presence of statistically significant correlation.

4.4.2.3 Accident Cost – HE-O

The third causal factor evaluated was human error only events. The results of the Fisher's Pairwise Comparison are listed in Table E.10. The results indicated that the F-4 experienced a greater level of damage on average than the other aircraft and the E-2 experiencing the lowest. Three statistically significant performance groups were identified with the two aircraft in Group B aircraft having performance that corresponded to more than one group. Group A experienced the greatest average level of damage while Group C experienced the lowest. Of note, Group B does not include the EA-6 despite it having a relative mean that was greater than the A-6 (a member of Group B). This is due to the small sample size of the A-6 performance relative to the EA-6. It is expected that if the A-6 had a larger sample size, its performance would remove it from Group B and result in the removal of the entire group.

Of the eight aircraft evaluated, six belonged to only one group and only two belonged to multiple groups. The costs (expressed as a percentage of aircraft value) associated with accidents due only to human error for the F-4, AV-8, F-18 and F-14 rejected the null when compared to the other four aircraft. The costs of accidents due only to human error for the E-2, C-2, and EA-6 rejected the null when compared to all aircraft except the A-6. Correlation analysis between attributes of automation and costs of accidents caused only by human error did not indicate the presence of statistically significant correlation as the null was not rejected.

4.4.2.4 Accident Cost – AE-ALL

The fourth causal factor evaluated was aircrew error. The results of the Fisher's Pairwise Comparison are listed in Table E.11. The results indicated that the F-4 experienced a greater level of damage on average than the other aircraft and the C-2 experienced the lowest. Three statistically significant performance groups were identified with none of the 8 aircraft having performance that corresponded to more than one group. Group A experienced the greatest average level of damage while Group C experienced the lowest. Correlation analysis indicated that three characteristics of automation correlated to the costs associated with accidents due to aircrew error. The total number of systems present in the cockpit displayed positive correlation as did the quantity of systems present with a level of automation between four and seven. Further investigation of the correlation between accident costs and level four to seven systems revealed that the quantity of level four systems displayed a negative correlation to the costs of accidents due to aircrew error. All three characteristics and correlation are discussed below.

The total number of systems present correlated positively to the costs of accidents due to aircrew error with or without material failure as shown in Figure 4.3. The Pearson's Correlation Coefficient was 0.831. This indicates that the number of systems that aircrew must manage may directly relate to the magnitude of damage resulting from an error. Additionally, it indicates that as more systems required monitoring by the aircrew, the changes in automation did not mitigate the impacts of aircrew error with and without material failure. Of note, when the data were further reduced to measure the costs of accidents only due to aircrew error or aircrew error and material failure, correlation to the total number of systems present remained. The scatter plots for both were similar to Figure 4.3 and the Pearson's Correlation Coefficients were similar as well.

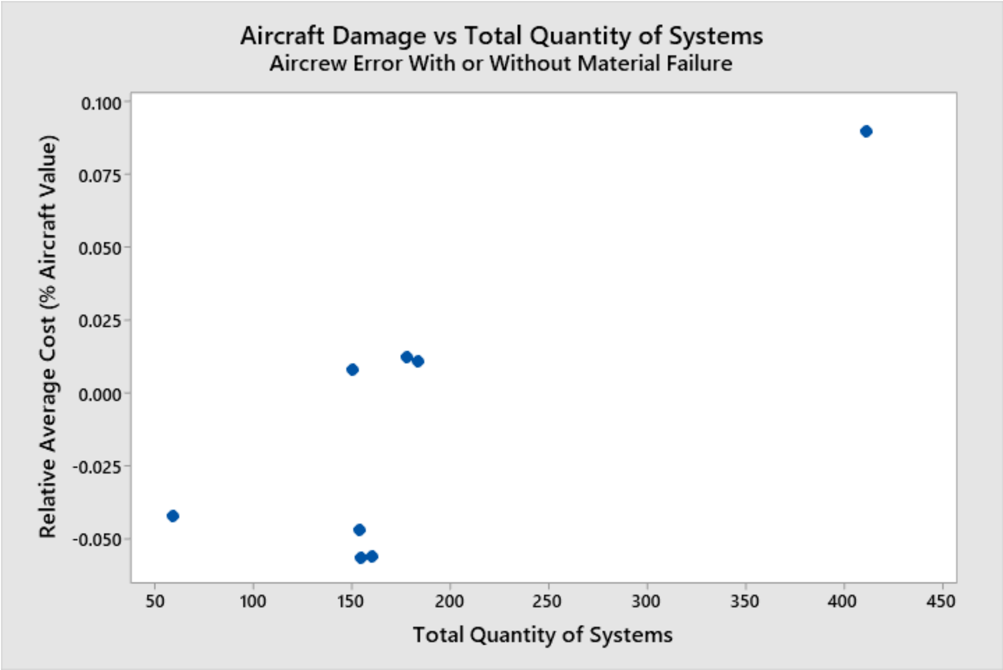


Figure 4.3: Total Quantity of Systems vs Accident Cost (AE-ALL).

Pearson's Correlation Coefficient of 0.831 was observed between aircraft damage due to any aircrew error and the total quantity of systems. This indicates that the number of systems aircrew must manage may directly relate to magnitude of damage result from an error.

The second grouping of automation levels that showed correlation was the quantity of systems ranked between levels 4-7. The quantity of level 4-7 automated systems correlated to the costs of accidents due to aircrew error with or without material failure as shown in Figure 4.4 with a Pearson's Correlation Coefficient of 0.792. When the data were further reduced to measure the costs of accidents only due to aircrew error or aircrew error and material failure, correlation to the number of systems ranked between levels 4-7 remained. The scatter plots for both were similar to Figure 4.4 and the Pearson's Correlation Coefficients were similar as well.

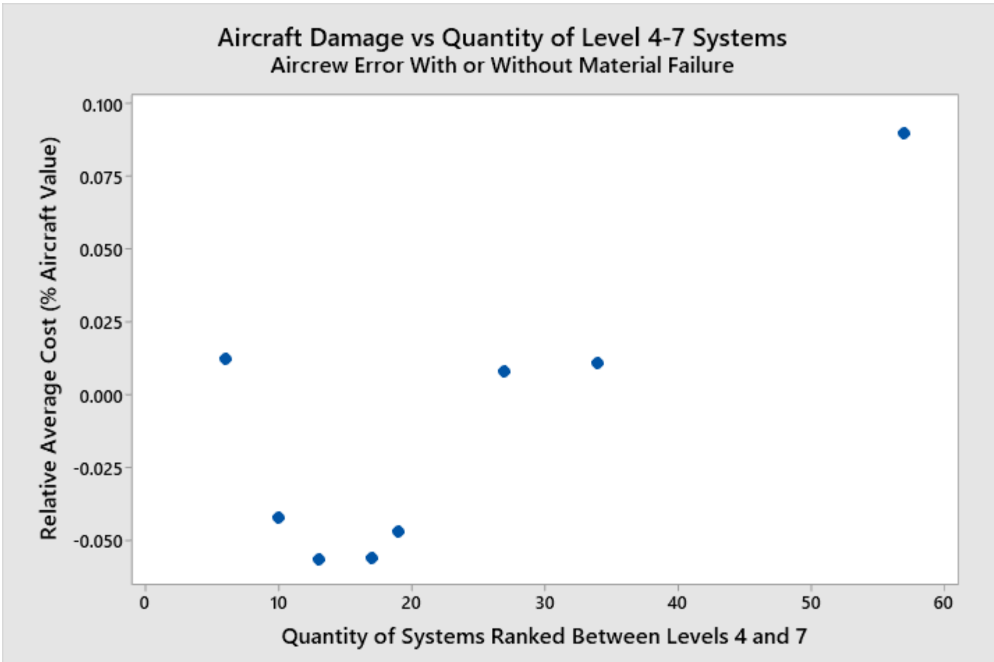


Figure 4.4: Level 4-7 Systems vs Accident Cost (AE-ALL).

Pearson's Correlation Coefficient of 0.792 was observed between aircraft damage due to any aircrew error and quantity of level 4-7 systems.

While the grouping of level 4-7 automated systems had a strong positive correlation coefficient with aircrew error with or without material failure, when the quantity of systems ranked only at a level four of automation were evaluated, the correlation to aircrew error with or without material failure exhibited negative correlation as shown in Figure 4.5 with a Pearson's Correlation Coefficient of -0.775. When the data were further reduced to measure the costs of accidents only due to aircrew error or aircrew error and material failure, correlation to the number of systems ranked at level 4 remained. The scatter plots for both were similar to Figure 4.5 and the Pearson's Correlation Coefficients were similar as well. This finding indicates that the laboratory results of Gil, Kabe, Kaufmann, and Kim indicating that human performance may not be linearly related to levels of automation (Gil G.-H. , Kaber, Kaufmann, & Kim, 2012) may have been observed in an operational environment.

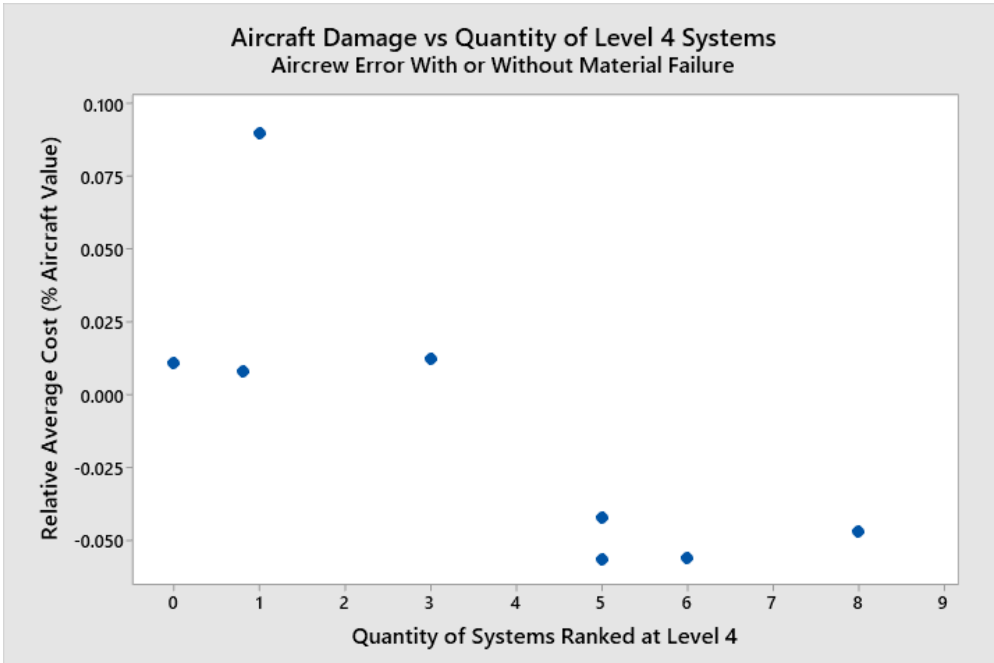


Figure 4.5: Level 4 Systems vs Accident Cost (AE-ALL).

Pearson's Correlation Coefficient of -0.775 was observed between aircraft damage due to any aircrew error and quantity of level 4 systems. When compared to the results in Figure 4.4, it indicates that the laboratory results of Gil, Kabe, Kaufmann, and Kim indicating that human performance may not be linearly related to levels of automation (Gil G.-H. , Kaber, Kaufmann, & Kim, 2012) may have been observed in an operational environment.

The third category of automation that correlated to human performance was the quantity of systems associated with information management. The quantity of information management systems correlated to accident costs due to aircrew error with or without material failure as shown in Figure 4.6 with a Pearson's Correlation Coefficient of 0.895. Of note, when the data were further reduced to measure the costs of accidents only due to aircrew error or aircrew error and material failure, correlation to the number of systems associated with information management remained. The scatter plots for both were similar to Figure 4.6 and the Pearson's

Correlation Coefficients were similar as well. This indicates that a relationship may exist between the quantity of systems managing information and the magnitude of damage resulting from aircrew error either with or without an unexpected material failure. This finding lends support to Norman's suggestion that a problem with automation may be insufficient feedback to the operator (Norman D. A., 1990) since aircraft status or feedback is fundamentally based on management of information.

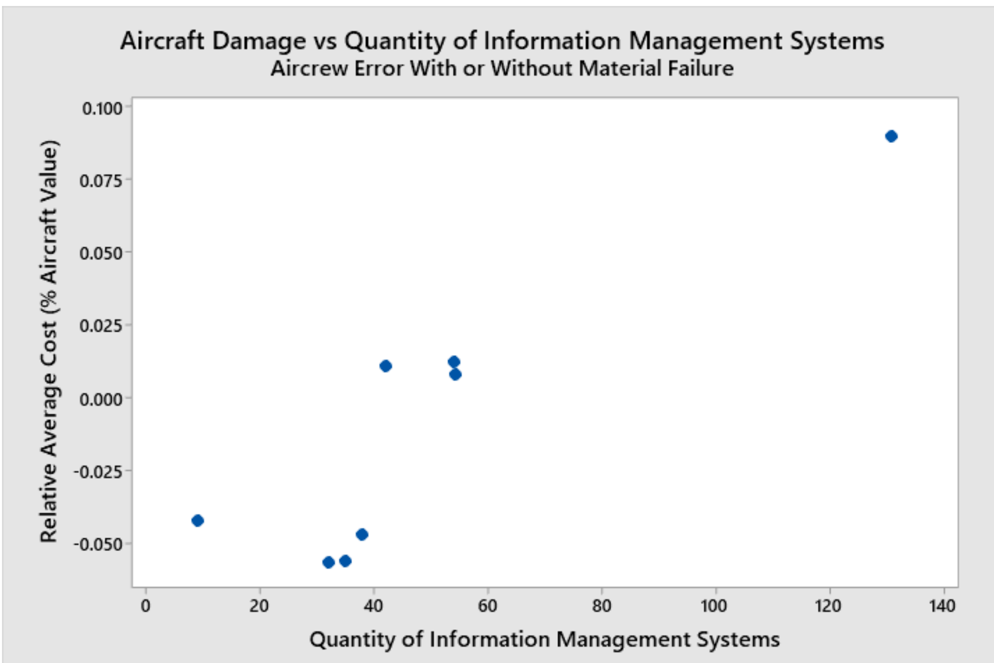


Figure 4.6: Information Management Systems vs Accident Cost (AE-ALL).

Pearson's Correlation Coefficient of 0.895 was observed between aircraft damage due to any aircrew error and quantity of information management systems. This indicates that a relationship may exist between the quantity of systems managing information and the magnitude of damage resulting from aircrew error either with or without an unexpected material failure and may lend support to Norman's suggestion that a problem with automation may be insufficient feedback to the operator (Norman D. A., 1990) since aircraft status or feedback is fundamentally based on management of information.

4.4.2.5 Accident Cost – AE-MF

The fifth causal factor evaluated was aircrew error with material failure. The results of the Fisher's Pairwise Comparison are listed in Table E.12. The results indicated that the F-4 experienced a greater level of damage on average than the other aircraft and the C-2 experienced the lowest. Four statistically significant performance groups were identified with 4 of the 8 aircraft having performance that corresponded to more than one group. Group A experienced the greatest average level of damage while Group D experienced the lowest.

Correlation analysis indicated that in addition to the three characteristics discussed in the previous section (quantities of total systems, systems with automation between levels four and seven, and systems with level four automation), four additional characteristics correlated to the costs of accidents due to aircrew error and material failure. The quantity of systems possessing automation between levels one and three displayed positive correlation as did the quantities of systems designed for action implementation, aircraft performance, or mission performance. All four characteristics and correlation are discussed below.

The quantity of level 1-3 automated systems correlated to the costs of accidents due to aircrew error with material failure as shown in Figure 4.7 with a Pearson's Correlation Coefficient of 0.738. Automation levels one to three are defined as systems that require the operator to use manual control, action support, or batch process activities for successful task execution. The strong positive correlation indicates that a relationship may exist between the quantities of low automation systems present to the magnitude of damage resulting from aircrew error with a material failure event requiring a reaction. This finding indicates that the presence of low-level automation may reduce the performance of aircrew to recover from an emergency due to a material failure and minimize damage to the aircraft.

The second type that showed correlation was the quantity of systems that are designed for action implementation. The quantity of action implementation systems present displayed a positive correlation to costs associated with accidents due to aircrew error and material failure with a Pearson's Correlation Coefficient of 0.815 as shown in Figure 4.8. This indicates that a relationship may exist between the quantity of systems present to implement a desired action and the magnitude of damage resulting from aircrew error with an unexpected material failure. This finding indicates that if fewer systems are required to operate an aircraft, the damage from an emergency event due to material failure may be lower.

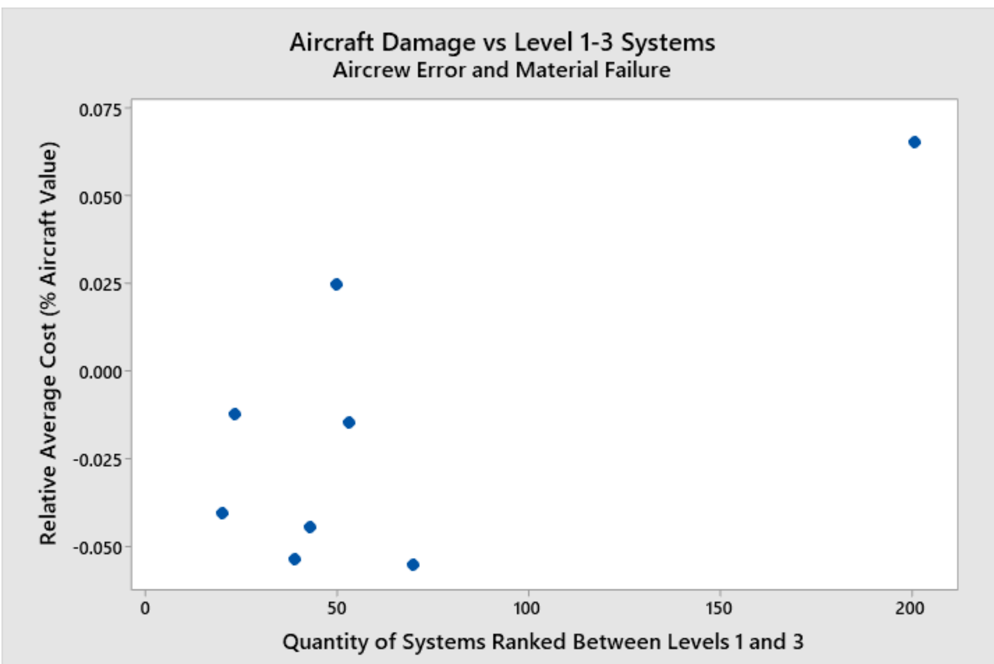


Figure 4.7: Level 1-3 Systems vs Accident Cost (AE-M).

Pearson's Correlation Coefficient of 0.738 was observed between aircraft damage due to aircrew error associated with material failure and quantity of level 1-3 systems. The strong positive correlation indicates that a relationship may exist between the quantities of low automation systems present to the magnitude of damage resulting from aircrew error with a material failure event requiring a reaction. This finding indicates that the presence of low-level automation may reduce the performance of aircrew to recover from an emergency due to a material failure and minimize damage to the aircraft.

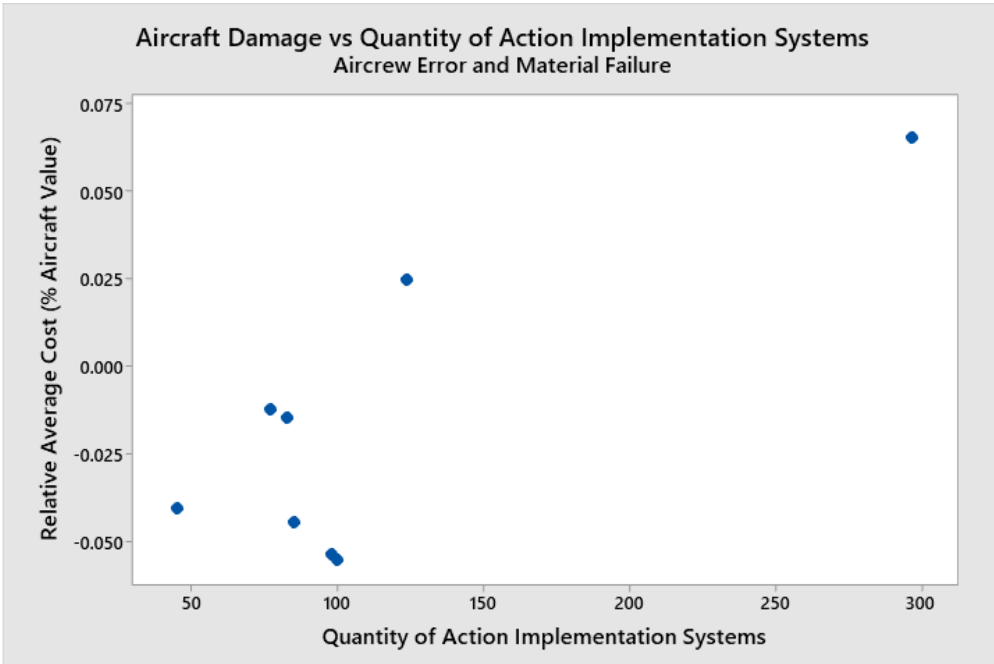


Figure 4.8: Action Implementation Systems vs Accident Cost (AE-M).

Pearson's Correlation Coefficient of 0.815 was observed between aircraft damage due to aircrew error associated with material failure and quantity of action implementation systems. This indicates that a relationship may exist between the quantity of systems present to implement a desired action and the magnitude of damage resulting from aircrew error with an unexpected material failure and that if fewer systems are required to operate an aircraft, the damage from an emergency event due to material failure may be lower.

The third automation characteristic that correlated was the quantity of systems associated with aircraft performance. The amount of aircraft performance systems in the cockpit showed a positive correlation to the costs of accidents due to aircrew error and material failure with a Pearson's Correlation Coefficient of 0.75 as shown in Figure 4.9. Of note, this correlation was not present when evaluated for accidents without material failure as a cause indicating that the presence of these types of systems may have a negative effect on aircrew performance when an aircraft component fails unexpectedly. This supports the results from the two experiments conducted by Parasuraman, Molloy, and Singh that showed reduced human performance in constant-reliability systems due to automation-induced complacency (Parasuraman, Molloy, & Singh, Performance Consequences of Automation-Induced "Complacency", 1993).

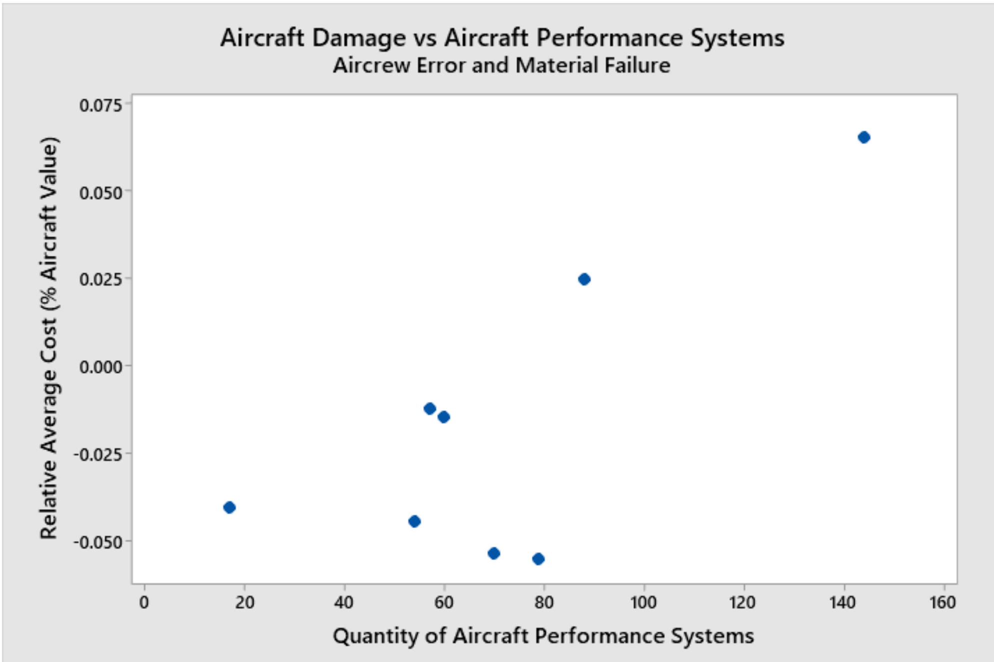


Figure 4.9: Aircraft Performance Systems vs Accident Cost (AE-M).

Pearson's Correlation Coefficient of 0.75 was observed between aircraft damage due to aircrew error associated with material failure and quantity of aircraft performance systems. This correlation was not present when evaluated for accidents without material failure as a cause indicating that the presence of these types of systems may have a negative effect on aircrew performance when an aircraft component fails unexpectedly.

The fourth characteristic of automation that correlated accident costs due to aircrew error and material failure was the quantity of mission performance systems. The quantity of mission performance systems present displayed a positive correlation to the costs of accidents due to aircrew error and material failure with a Pearson's Correlation Coefficient of 0.814 as shown in Figure 4.10. Of note, this correlation was not present for accident costs due only to aircrew error indicating that a relationship may exists between the quantity of non-flight critical automation (*i.e.,* mission performance) and the magnitude of damage resulting from aircrew error when reacting to an unexpected material failure. Aligning these results with Wilson's finding that task prioritization errors occurred more frequently in advanced technology aircraft (Wilson, 1998), it appears that the impact of task prioritization errors may be greater in aircraft with more automation.

4.4.2.6 Accident Cost – AE-O

The final causal factor evaluated was aircrew error only events. The results of the Fisher's Pairwise Comparison are listed in Table E.13. The results indicated that the F-4 experienced a greater level of damage on average than the other aircraft and the C-2 experienced the lowest. Three statistically significant performance groups were identified with only 1 of the 8 aircraft having performance that corresponded to more than one group. Group A experienced the greatest average level of damage while Group C experienced the lowest.

Correlation analysis indicated that in addition to the three characteristics discussed in the section regarding all accidents due to aircrew error (AE-ALL) (*i.e.,* quantities of total systems, systems with automation between levels four and seven, and systems with level four automation), two additional characteristics correlated to the costs of accidents due only to aircrew error. The quantity of systems possessing automation between levels eight and ten displayed positive correlation as did the quantities of systems designed for information acquisition. Both characteristics and correlations are discussed below.

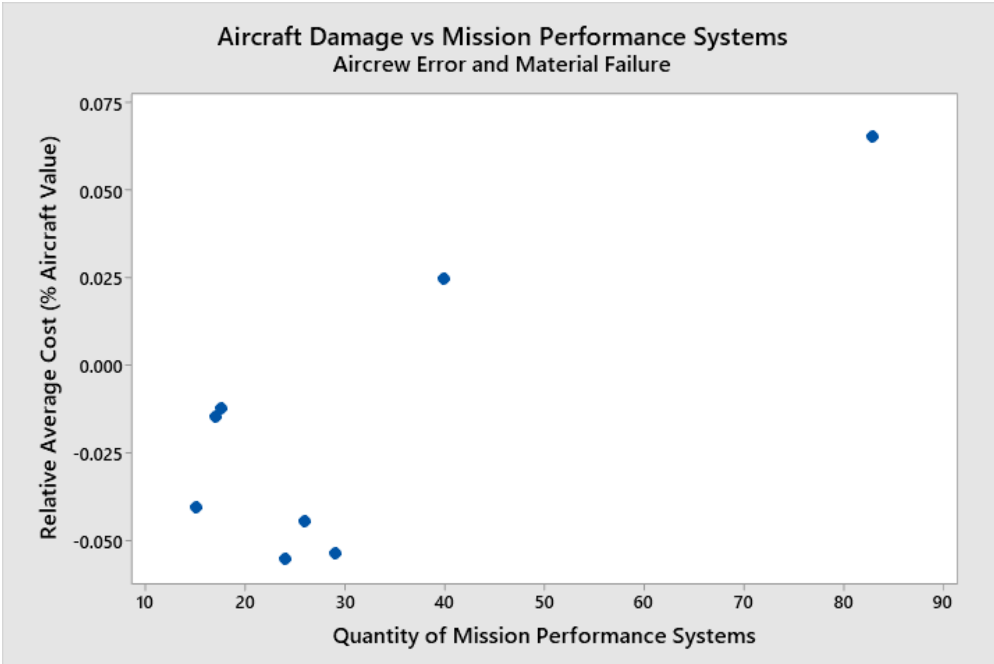


Figure 4.10: Mission Performance Systems vs Accident Cost (AE-M).

Pearson's Correlation Coefficient of 0.814 was observed between aircraft damage due to aircrew error associated with material failure and quantity of mission performance systems. This correlation was not present for damage due only to aircrew error indicating that a relationship may exists between the quantity of non-flight critical automation (*i.e.*, mission performance) and the magnitude of damage resulting from aircrew error when reacting to an unexpected material failure. Aligning these results with Wilson's finding that task prioritization errors occurred more frequently in advanced technology aircraft (Wilson, 1998), it appears that the impact of task prioritization errors may be greater in aircraft with more automation.

The quantity of level 8-10 automated systems correlated to the costs of accidents only due to aircrew error as shown in Figure 4.11 with a Pearson's Correlation Coefficient of 0.735. The total number of systems present with an automation level between 8-10 showed a strong positive correlation with statistical significance to accident costs associated with accidents due only to aircrew error. This indicates that a relationship may exist between the quantity of highly automated systems present to the magnitude of damage resulting from aircrew error without a material failure event requiring a reaction. This analysis is in line with the results from Endsley and Kaber's experiment and supports their proposal that automation at high levels will leave operators "out-of-the-loop" (Kaber & Endsley, 1997). Additionally, the results indicate that the reduction in performance is only for errors that could be considered self-induced (*i.e.*, no external event occurred requiring a reaction) since correlation did not exist between level 8-10 systems and aircrew errors with material failure.

The quantity of information acquisition systems correlated to the costs of accidents only due to aircrew error as shown in Figure 4.12 with a Pearson's Correlation Coefficient of 0.712. This strong positive correlation indicates that a relationship may exist between the quantity of systems acquiring information and the magnitude of damage resulting from aircrew error with no external event requiring a reaction. Since the function of information acquisition systems is to

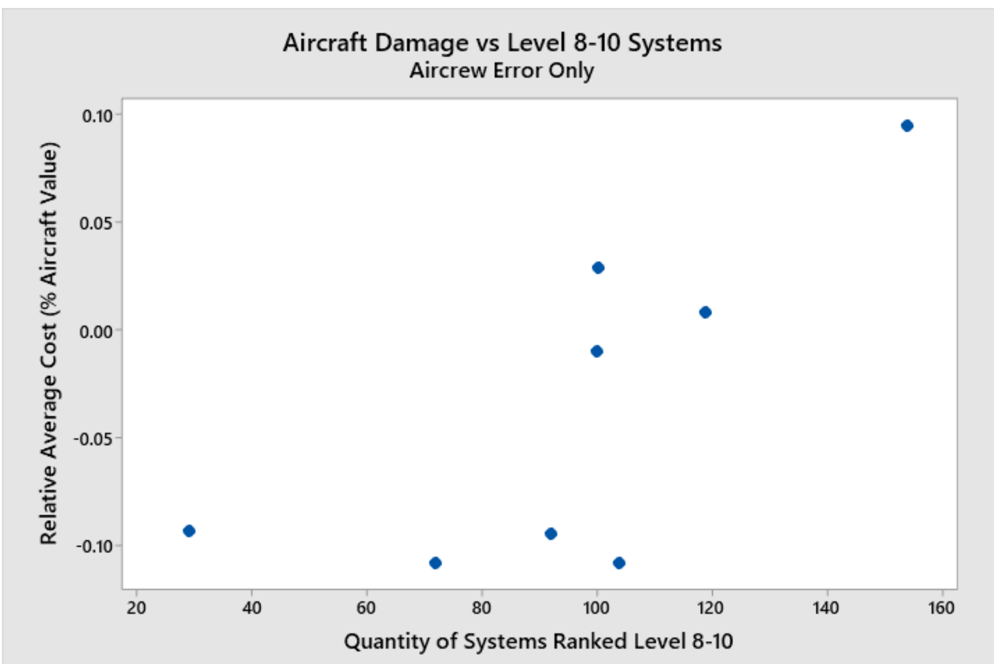


Figure 4.11: Level 8-10 Systems vs Accident Cost (AE-O).

Pearson's Correlation Coefficient of 0.735 was observed between aircraft damage due only to aircrew error and quantity of level 8-10 systems. The results indicate that the reduction in performance is only for errors considered self-induced (*i.e.*, no external event required a reaction). This indicates that a relationship may exist between the quantities of highly automated systems present and magnitude of damage resulting from aircrew error without a material failure event providing support aligning with the results from Endsley and Kaber's experiment and supports their proposal that automation at high levels will leave operators "out-of-the-loop" (Kaber & Endsley, 1997).

acquire and present information to the aircrew without analysis (information analysis is a different type of automation), the result of this analysis indicates that there may be a limit to the number of inputs aircrew can monitor and process.

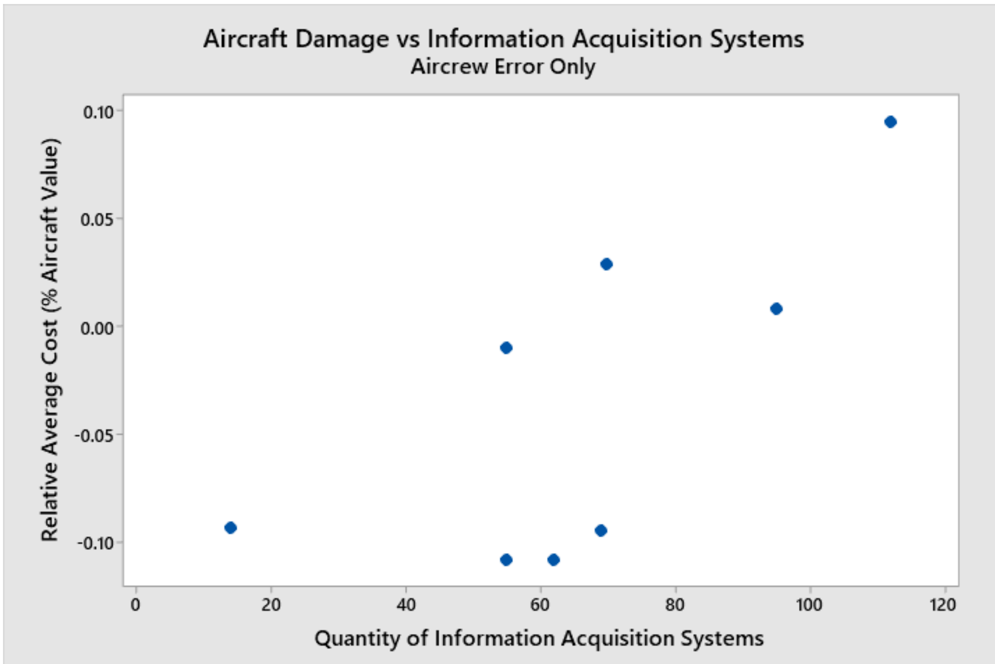


Figure 4.12: Information Acquisition Systems vs Accident Cost (AE-O).

Pearson's Correlation Coefficient of 0.712 was observed between aircraft damage due only to aircrew error and quantity of information acquisition systems. The results indicates that a relationship may exist between the quantity of systems acquiring information and the magnitude of damage resulting from aircrew error with no external event. Additionally, the results indicate that there may be a limit to the number of inputs aircrew can monitor and process.

4.4.2.7 Accident Financial Costs - Analysis

Accident costs differed between the six combinations of human error, but relative ranking of the aircraft remained relatively consistent as shown in Table E.14. Of note, the F-4 demonstrated the highest rate of accidents for all six combinations of human error and the E-2 and C-2 had the lowest with the A-6 and EA-6 maintained the ranking positions of five and six for all combinations. Overall, no fluctuations over the two ranks were observed by any aircraft indicating that the costs (expressed as a percentage of aircraft value) associated with the different combinations of human error were relatively consistent.

Correlation of attributes of automation were observed only for accidents involving aircrew error. Total number of systems present in the cockpit, systems with an automation level between four and seven, and information management systems each demonstrated positive correlation with accidents involving aircrew error (AE-ALL) as well as when analysis was conducted with the subsets of accidents caused only by aircrew error (AE-O) or accidents caused by aircrew error and material failure (AE-MF). In addition to those three configurations, the costs of accidents due to aircrew error and material failure displayed positive correlation to the quantity of systems with an automation level between one and three and systems designed for action implementation, aircraft performance, or mission performance. The costs associated with accidents due only to aircrew error, positive correlation was observed for systems with automation levels between eight and ten, and information acquisition.

4.4.3 Accident Fatalities

The median number of fatalities associated with accidents due to the causal factors listed in Table 3.6 was calculated annually for each aircraft using the metric of percentage of crew size for each aircraft. per recorded accident.

4.4.3.1 Accident Fatalities – HE-ALL

The first accident fatality rate evaluated was that of accidents caused by any human error. The results of the Fisher’s Pairwise Comparison are listed in Table E.15. The results indicated that the EA-6 experienced a greater fatality rate than the other aircraft and the C-2 experienced the lowest. Two statistically significant performance groups were identified. While the groups were not mutually exclusive, the EA-6, E-2, and C-2 each had performance that was placed in only one group. The other five aircraft had performance that placed them in both groups as shown in Table E.15.

The only attribute that showed correlation was crew size. Crew size showed a strong positive correlation between accident fatalities (expressed as percentage of crew size) and accidents due to human error with or without material failure. The Pearson's Correlation Coefficient between crew size and accidents due to human error was 0.844 and is shown in Figure 4.13. This indicates that a relationship may exist between crew size and the number of fatalities resulting from human error either with or without an unexpected material failure. As mentioned earlier, Reason's Swiss Cheese Model (Reason J. , 2000) implies that more aircrew should relate to greater performance. However, the data do not appear to be in line with that proposal. As a result, this finding indicates potential presence of behaviors aligned with identity fusion and associated acceptance of extreme self-sacrifice previously discussed in the Chapter 2 section on Operator Mindset (Swann Jr. W. B., *et al.*, 2014) (Whitehouse, 2018) (Rand & Epstein, 2014). It is suspected that a group dynamic within the crew may be a factor that results in a larger percentage of the crew experiencing fatal injuries when in an aircraft with a larger crew size.

4.4.3.2 Accident Fatalities – HE-MF

The second accident fatality rate evaluated was that of accidents caused by human error and material failure. The results of the Fisher's Pairwise Comparison are listed in Table E.16. Two statistically significant performance groups were identified with the E-2 experiencing a greater fatality rate than the other aircraft and the EA-6 experiencing the lowest rate of fatalities. While the two groups were not mutually exclusive, the E-2, C-2, and EA-6 each had performance that was placed in only one group. The other five aircraft had performance that placed them in both groups as shown in Table E.16. Correlation analysis between attributes of automation and fatalities associated with accidents caused by human error and material failure did not indicate the presence of a statistically significant correlation as the null failed to be rejected.

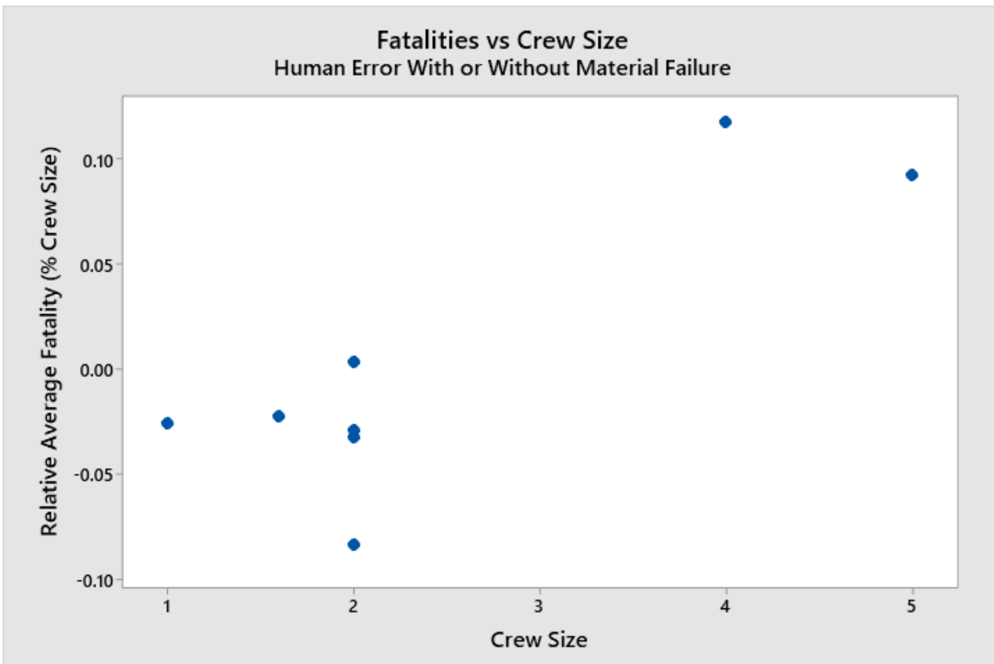


Figure 4.13: Crew Size vs Fatalities (HE-ALL).

Pearson's Correlation Coefficient of 0.844 was observed between fatalities due to any human error and crew size. This indicates that a relationship may exist between crew size and the number of fatalities resulting from human error either with or without an unexpected material failure. This observation indicates potential presence of behaviors aligned with identity fusion and associated acceptance of extreme self-sacrifice previously discussed (Swann Jr. W. B., *et al.*, 2014) (Whitehouse, 2018) (Rand & Epstein, 2014).

4.4.3.3 Accident Fatalities – HE-O

The third accident fatality rate evaluated was that of accidents caused only by human error. The results of the Fisher's Pairwise Comparison are listed in Table E.17. The number of fatalities associated with accidents caused only by human error statistically fell into two groups with the EA-6 experiencing the greatest rate of fatalities and the C-2 experiencing the lowest. While the groups were not mutually exclusive, the EA-6, AV-8, E-2, and C-2 each had performance that was placed in only one group while the other four aircraft had performance that placed them in both. As was observed in earlier performance evaluations, the small sample size of the F-4 resulted in its inclusion in Group A. However, if the F-4 sample size was larger, it is expected that it would not be included in Group A. These results indicate that a difference in human performance when no material failure occurred was observed. Correlation analysis between attributes of automation and the rate of fatalities associated with accidents caused only by human error failed to reject the null.

4.4.3.4 Accident Fatalities – AE-ALL

The fourth accident fatality rate evaluated was that of accidents caused by any aircrew error with or without material failure. The results of the Fisher's Pairwise Comparison are listed in Table E.18. The rate of fatalities associated with accidents caused all or in part by aircrew error statistically fell into two groups. While the groups were not mutually exclusive, only the E-2 and C-2 had performance that was placed in a single group. These results indicate that statistically significant differences in aircrew performance with or without material failure did exist. Correlation analysis between attributes of automation and the rate of fatalities associated with accidents caused by aircrew error failed to reject the null.

4.4.3.5 Accident Fatalities – AE-MF and AE-O

The fifth and sixth accident fatality rates evaluated were of accidents caused by aircrew error with material failure and accidents only caused by aircrew error. Both rates did not show performance that could be separated into multiple groups with statistical significance. However, correlation analysis did show that a strong positive correlation between accident fatalities due to aircrew error and material failure and the quantity of systems with automation between levels four and seven. Additionally, while analysis for correlation with quantity of systems between levels one and three failed to reject the null, further analysis showed strong negative correlation between fatalities due to aircrew error and material failure and quantities of level three and two systems. The correlation between the three characteristics of automation is discussed below.

The total number of systems present with an automation level between 4-7 showed a strong positive correlation with statistical significance to quantity of accident fatalities associated with accidents due to aircrew error with material failure. The quantity of systems with

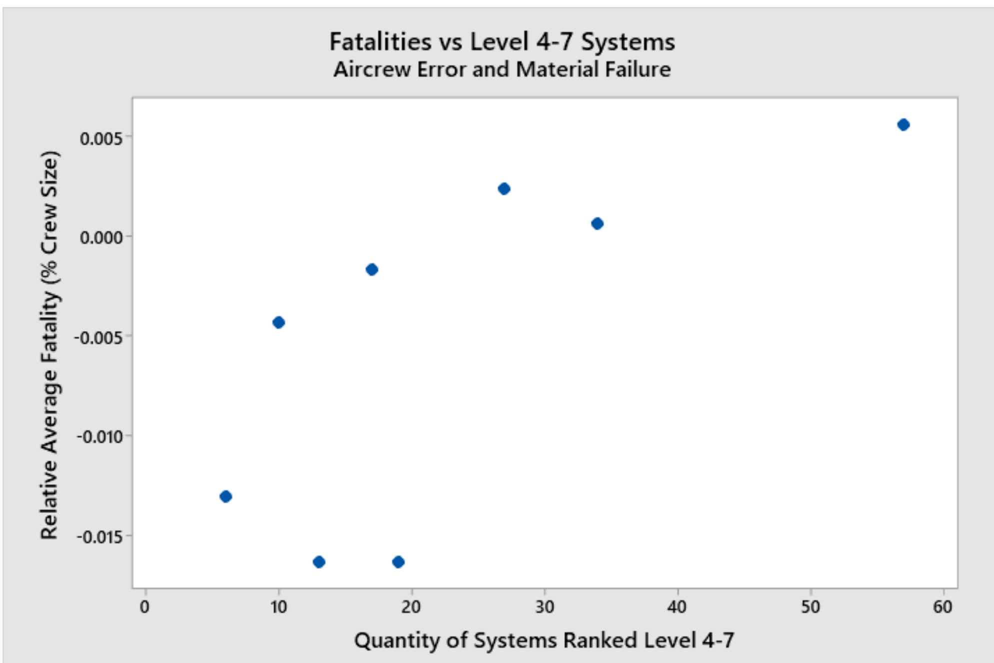


Figure 4.14: Level 4-7 Systems vs Fatalities (AE-M).

Pearson's Correlation Coefficient of 0.714 was observed between fatalities due to aircrew error associated with material failure and quantity of level 4-7 systems. When viewed with the findings in Figure 4.15 and Figure 4.16, the results support the finding of Gil, Kabe, Kaufmann, and Kim indicating that human performance may not be linearly related to levels of automation and may have been observed in an operational environment (Gil G.-H. , Kaber, Kaufmann, & Kim, 2012). Additionally, it indicates that lower levels of automation may help aircrew avoid being "out-of-the-loop" and improve the probably of surviving an accident.

automation levels between 4-7 experienced a Pearson's Correlation Coefficient of 0.714 as shown in Figure 4.14.

Analysis of potential correlation between fatalities due to aircrew error with material failure and quantity of automated systems with levels between one and three failed to reject the null. However, when analysis was conducted only on the systems categorized with level 3 and then with level 2 automation, the total number of systems with level 3 or level 2 automation both showed strong negative correlation. The quantity of level 3 automated systems correlated to the quantity of fatalities in accidents due to aircrew error with material failure as shown in Figure 4.15 with a Pearson's Correlation Coefficient of -0.845 and the quantity of level 2 automated systems correlated to the quantity of fatalities in accidents due to aircrew error with material failure as shown in Figure 4.16 with a Pearson's Correlation Coefficient of -0.864. This finding indicates that the laboratory results of Gil, Kabe, Kaufmann, and Kim indicating that human performance may not be linearly related to levels of automation and may have been observed in an operational environment (Gil G.-H. , Kaber, Kaufmann, & Kim, 2012). Additionally, it indicates that lower levels of automation may help aircrew avoid being "out-of-the-loop" and improve the probably of surviving an accident.

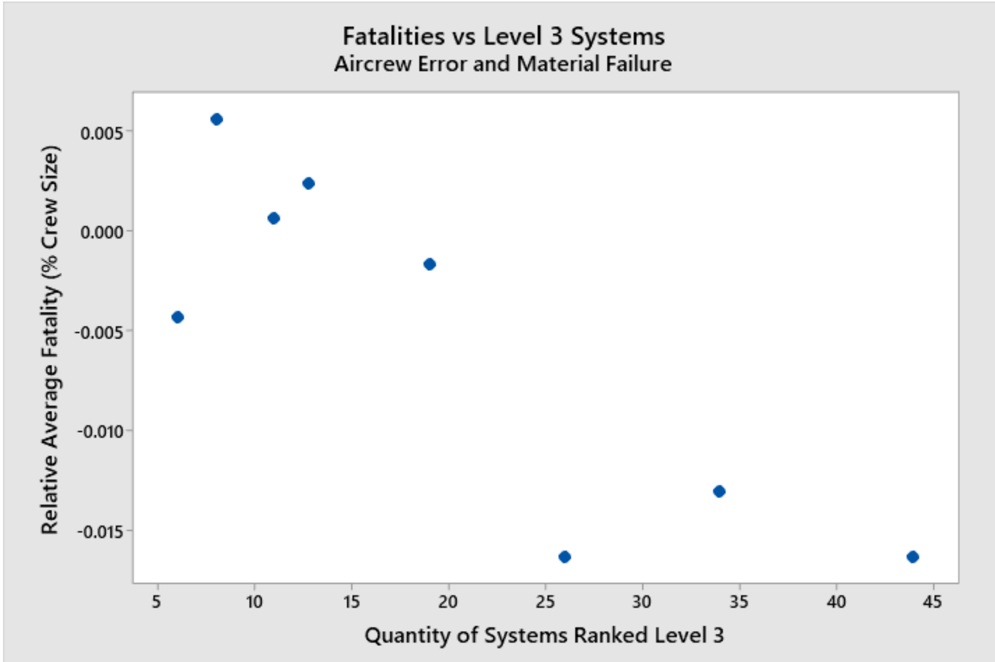


Figure 4.15: Level 3 Systems vs Fatalities (AE-M).

Pearson's Correlation Coefficient of -0.845 was observed between fatalities due to aircrew error associated with material failure and quantity of level 3 systems. When viewed with the findings in Figure 4.14 and Figure 4.16, the results support the finding of Gil, Kabe, Kaufmann, and Kim indicating that human performance may not be linearly related to levels of automation and may have been observed in an operational environment (Gil G.-H. , Kaber, Kaufmann, & Kim, 2012). Additionally, it indicates that lower levels of automation may help aircrew avoid being "out-of-the-loop" and improve the probably of surviving an accident.

4.4.3.6 Accident Fatalities - Analysis

Accident fatalities differed between the six combinations of human error with significant shifts in relative ranking of the aircraft as shown in Table E.19. Of note, the EA-6 shifted from having the highest occurrence of fatalities in accidents due to any human error, but the lowest occurrence when assessed only with accidents due to human error and material failure. Conversely, the C-2 remained in the bottom two rankings for the least quantity of fatalities.

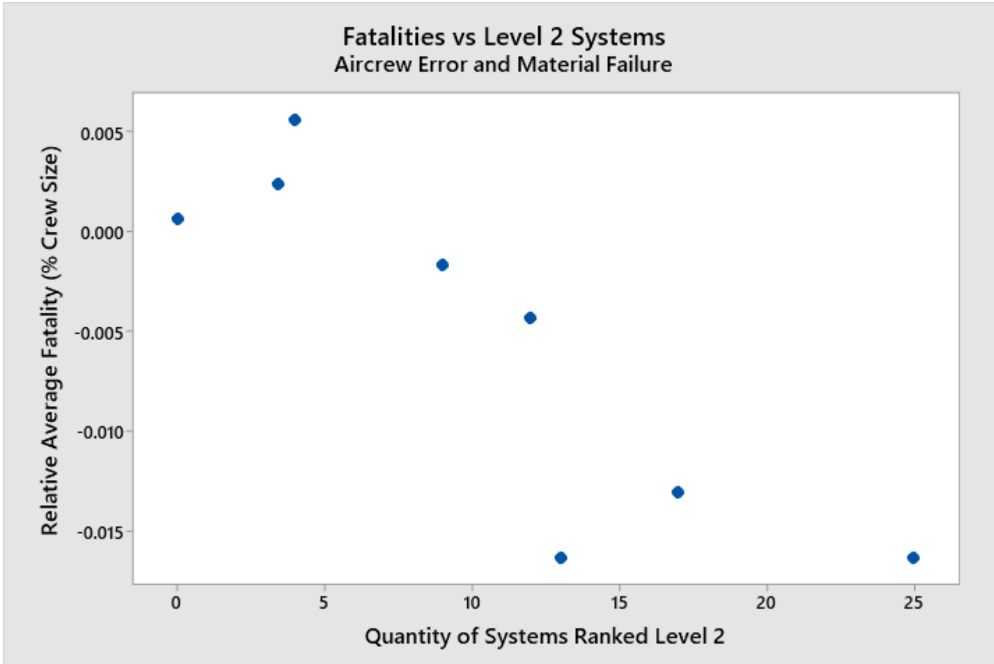


Figure 4.16: Level 2 Systems vs Fatalities (AE-M).

Pearson's Correlation Coefficient of -0.864 was observed between fatalities due to aircrew error associated with material failure and quantity of level 3 systems. When viewed with the findings in Figure 4.14 and Figure 4.15, the results support the finding of Gil, Kabe, Kaufmann, and Kim indicating that human performance may not be linearly related to levels of automation and may have been observed in an operational environment (Gil G.-H. , Kaber, Kaufmann, & Kim, 2012). Additionally, it indicates that lower levels of automation may help aircrew avoid being "out-of-the-loop" and improve the probably of surviving an accident.

Correlation of attributes of automation were observed only for accidents involving any human error and aircrew error involving material failure. Crew size showed a strong positive correlation between accident fatalities and accidents due to human error with or without material failure indicating another factor at play in addition to Reason's Swiss Cheese Model (Reason J. , 2000). The total number of systems present with an automation level between 4-7 showed a strong positive correlation with statistical significance to quantity of accident fatalities associated with accidents due to aircrew error with material failure while the total number of systems with level 3 or level 2 automation both showed strong negative correlation lending support to the laboratory results of Gil, Kabe, Kaufmann, and Kim indicating that human performance may not be linearly related to levels of automation (Gil G.-H. , Kaber, Kaufmann, & Kim, 2012).

4.5 Correlation Results

Correlation between human performance and cockpit automation was evaluated using Pearson's Correlation with $\alpha=0.05$. A total of 12 of the pairings rejected the null. The analysis indicated that crew size, quantity of systems, levels of automation, types of automation, and categories of automation correlated to differences in aircrew accident performance. A summary of the results is provided in Table 4.4.

To address the potential for multiplicity, previous research has discussed a number of concerns with either reducing the potential for Type I error due to random chance or preventing large Type II errors due to conservative approaches of analysis. Ultimately, it has been "strongly recommended that both the corrected and uncorrected P-values be reported and that all finding be reported as tentative and hypothesis generating, rather than hypothesis testing" (Streiner, 2015). As a result, the Bonferroni, Holm, and Hochberg methods were considered as corrections for multiplicity with the Hochberg method selected as it was the least conservative of the three. The Hochberg method was applied to the pairings in Table 4.4 with a False Discovery Rate set to 50%. Only one pairing rejected the null (information management systems and costs associated with all subsets of aircrew error). While the results of the Hochberg method significantly reduce the number of significant results, it is important to reference the American Statistical Association's statement that "a p-value without context or other evidence provides limited information" (Wasserstein & Lazar, 2016).

Table 4.4: Summary of Automation and Performance Correlation.

A summary of the results of a Pearson's Correlation ($\alpha \leq 0.05$) analysis between attributes of automation and human performance. The characteristic is listed in the left most column, the causal factor for the accident and the performance measurement are listed under Dependent Variables. Correlation is identified by an "X" between the Characteristic and Independent Variable and Pearson's Coefficient/P-Value between the Characteristic and Dependent Variable. Example: Crew Size Correlated to Accident Rates due to errors only caused by aircrew (*i.e.*, no material failure). The Pearson's Coefficient = -0.716 and p-value = 0.046.

Characteristic (Independent Variables)	Dependent Variables						
	Aircrew Error Only	Aircrew Error - All	Aircrew Error with Material Failure	Human Error - All	Accident Rate	Accident Cost	Accident Fatalities
Crew Size	X[1]			X[2]	-0.716[1] (P=0.046)		0.844[2] (P=0.008)
Total # of Systems		X				0.831 (P=0.011)	
Levels of Automation							
Levels 8-10	X					0.735 (P=0.038)	
Levels 4-7		X[3]	X[4]			0.792[3] (P=0.019)	0.714[4] (P=0.047)
Levels 1-3			X			0.738 (P=0.037)	
Types of Automation							
Information Acquisition	X					0.712 (P=0.047)	
Action Implementation			X			0.815 (P=0.014)	
Categories of Automation							
Aircraft Performance			X			0.750 (P=0.032)	
Mission Performance			X			0.814 (P=0.014)	
Information Management		X				0.895 (P=0.003)	

4.5.1 Crew Size

The size of the crew correlated negatively to accident rate and positively to rate of fatalities as summarized in Table 4.4. Accident rate and crew size displayed a Pearson's Correlation Coefficient of -0.716 ($p = 0.046$) indicating that a larger number of crewmembers in an aircraft was related to a lower accident rate. This corresponds to Reason's Swiss Cheese Model (Reason J. , 2000) as each person adds a layer of protection. Additionally, this finding supports the idea that "overall, the performance of operators when overseeing and intervening in automation tasking is dependent on their level of SA (Situational Awareness) and workload" (Endsley, 2017) as a larger crew provides greater ability to optimize workload.

Conversely, it was interesting to discover that accident fatalities, when measured as a percentage of total crew size, showed positive correlation to crew size with a Pearson's Correlation Coefficient of 0.844 ($p = 0.008$) (Table 4.4). The observed findings of this study have similarities to the observations from interviews with soldiers on loyalty. Themes of "loyalty as reciprocity" and "importance of emotional connection for cohesion" (Connor, Andrews, Noack-Lundberg, & Wadham, 2021) encouraged identity fusion where members of a group feel a deep sense of oneness and are "compelled to make extreme sacrifices." (Fredman, *et al.*, 2015). That sense of oneness has been also been characterized as a group of strongly fused persons in which the emotional engagement between them "overrides the desire for self-preservation and compels them to translate their moral beliefs into self-sacrificial behavior." (Swann Jr., *et al.*, 2014) While the study of interpersonal relationships is not the focus of this research, it does provide an explanation for this particular finding. Considering that "high stakes extreme altruism may be largely motivated by automatic, intuitive processes" (Rand & Epstein, 2014), the results suggest that interpersonal relationship between operators may play a role in system performance.

4.5.2 Total Quantity of Systems

The total quantity of systems present in the cockpit showed positive correlation to the cost of aircrew error with a Pearson's Correlation Coefficient of 0.831 ($p = 0.011$) as summarized in Table 4.4. Analysis of the subsets of aircrew error events (*i.e.*, aircrew error with material failure and errors only due to aircrew error) displayed similar results. While the specific reasons for the aircrew errors were not identified and correlation between the quantities of systems did not exist with accident rate, the fact that the financial impact did increase with the quantity of systems indicates a relationship. This aligns to the expectation from previous studies and experiments that "automation complacency occurs under conditions of multiple-task load" (Parasuraman & Manzey, 2010) where operators do not monitor system states adequately resulting in development a dangerous condition and accident. Given each system in the cockpit is designed to address a task, the increased number of systems indicates an increased task load for the aircrew and supports the findings of Parasurman and Manzey.

4.5.3 Levels of Automation

Levels of automation displayed positive correlation to aircrew performance when measured against fatalities due to aircrew error and material failure as well as the cost of aircrew error. Of note, it did not show correlation to errors attributed to non-aircrew personnel which is not unexpected as the systems included in this study were ones located in aircraft cockpits with the primary users being aircrew. In general, these findings were expected and support previous research in which nonlinear performance improvements were observed in an experiment utilizing a simulated robotic arm with varying levels of automation (Li, Wickens, Sarter, & Sebok, 2014).

The quantity of systems with high levels of automation (levels 8-10) correlated with a Pearson's Correlation Coefficient of 0.735 ($p = 0.038$) to increased costs associated with accidents caused only by aircrew error (AE-O). Given that the causal factor excluded anything other than aircrew error, the result matches the definition of automation bias where operators "over-rely on automation" (Goddard, Roudsari, & Wyatt, 2012). This phenomenon is characterized by human operators assuming a solution provided by an automated system is correct (Cummings, 2004). This supports the observation during a simulator-based experiment that automation bias "appears to be a more problematic automation error" than the consequences associated with the pure failure of automated systems (Wickens, Clegg, Vieane, & Sebok, 2015).

The presence of systems with mid-level automation (levels 4-7) displayed strong correlation to increased costs of accidents due to all categories of aircrew error with a Pearson's Correlation Coefficient of 0.792 ($p = 0.019$). Correlation was also displayed between the systems with mid-level automation and fatalities attributed to aircrew error and material failure with a Pearson's Correlation Coefficient of 0.714 ($p = 0.047$). This finding matches the expectation based on Kaber and Endsley's findings from an experiment involving a dual-task scenario that "intermediate LOAs (levels of automation) facilitated higher SA (situational awareness), but…not associated with improved performance" (Kaber & Endsley, 2004)" as the higher levels of automation displayed better performance (*i.e.*, failed to reject the null for positive correlation) with aircrew errors and material failure.

Systems with low-level automation (levels 1-3) displayed correlation to cost of accidents attributed to aircrew error and material failure with a Pearson's Correlation Coefficient of 0.738 ($p = 0.037$). Of note, correlation was not sufficient to reject the null for accidents due only to aircrew error. When viewed in comparison to the correlation results of high levels of automation, it appears that systems automated at a low-level correlated to reduced performance for events with a component of time pressure (*i.e.*, accidents due in part to material failure). In contrast, systems automated at a high-level correlated to reduced performance where there was no time pressure (*i.e.*, accidents due only to aircrew error). This observation matches expectations based on laboratory results indicating that "high automated tasks induce vigilance decreasing and OOTL-related phenomena" (Di Flumeri, *et al.*, 2019) and that "low-level automation produced superior performance" (Kaber & Endsley, 2004). However, it should be noted that task loading appears to have exceeded the performance gains of low-level automation as the consequence of accidents due to aircrew error and material failure were greater. The observations regarding accident costs and levels of automation indicate that the benefits of automation are not homogenous across human errors included in the HFACS model.

It is noted that in Table 4.4, the data indicate that different types of error were associated with the different levels of automation. In general, higher levels of automation correlated with accidents solely due to human error, whereas lower levels of automation correlated with accidents due to material failure playing a role. These results support previous findings that as automation levels shift from high to low, human performance shortfalls shift from bias to workload saturation.

4.5.4 Types of Automation

Types of automation displayed positive correlation to the cost of accidents due to aircrew error and the cost of accidents attributed to aircrew error and material failure as summarized in Table 4.4. This finding was expected as previous experiments and research have observed correlation between operator performance and automation type as well as a dependency on implementation of the system based on different operating environments (Galster, 2003), (Wright & Kaber, 2005), (Rice, Trafimow, & Hunt, 2010). It should be noted that the survey results of the systems present in the subject cockpits yielded very low quantities of systems attributed to the other two types of automation. As a result, correlation to accident performance for systems designed for information analysis or decision selection should be considered as not evaluated in this study.

For systems with the purpose of information acquisition, there was positive correlation to the cost of accidents due only to aircrew error (Pearson's Correlation Coefficient of 0.712, $p = 0.047$). It is noteworthy that this type of automation is not directly tied to machine performance, but rather the acquisition of information necessary for the aircrew to make decisions and act. The fact that increased presence of these types of systems correlate to the costs of accidents solely due to aircrew error indicates the potential atrophy of critical thought or discipline in acquiring all pertinent information for appropriate decision making. This aligns with the expectation associated with automation bias as discussed in the results of systems with high levels of automation.

Systems with the purpose of implementing an action showed strong positive correlation to the cost of accidents due to aircrew error and material failure (Pearson's Correlation Coefficient of 0.815, $p = 0.014$). This observed correlation matches the expectation from a simulator based experiment that "manual flying skills are subject to erosion due to lack of practice" due to intensive use of automation (Haslbeck & Hoermann, 2016) as the situation associated with these types of accidents is one where aircrew were required to respond in an adequate manner to an unexpected material failure. It is the sudden requirement to compensate for a condition that the automated systems were unable or not designed to handle which resulted in the reliance on aircrew skill.

4.5.5 Categories of Automation

Categories of automation displayed positive correlation to aircrew performance when measured against the cost of aircrew errors and material failure as well as the cost of all aircrew error as summarized in Table 4.4. This result was expected as previous research from simulator based experiments have observed correlation between categories of automation and aircrew monitoring strategies as well as skill atrophy (Sarter, Mumaw, & Wickens, 2007), (Haslbeck & Hoermann, 2016). Three of the four categories displayed positive correlation with only life support systems failing to reject the null when Pearson's correlation analysis was conducted. While the number of life support systems was significant across the subject aircraft, it is suspected that the inherent necessity of resilience and redundancy of systems associated with supporting human life places them at a greater level of design maturity.

Systems designed to support aircraft performance demonstrated a positive correlation with the cost of accidents due to aircrew error and material failure (Pearson's Correlation Coefficient of 0.750, $p = 0.032$). As discussed earlier, accidents resulting from material failure and aircrew error present a time limited opportunity for aircrew to react and recover from the failure of an aircraft component. The correlation between increased quantities of systems designed for aircraft performance and costs associated to these errors matches expectations based on the findings from simulator based experiments that the erosion of manual flying skills appears to be associated with the intensive use of automation (Haslbeck & Hoermann, 2016) and that although "cockpit automation may provide pilots with more time to think, it may encourage pilots to reinvest only some of this mental free time in thinking flight-related thoughts" (Casner & Schooler, 2014) resulting in the out-of-the-loop (OOTL) phenomenon.

The systems designed to support mission performance also positively correlated to the costs associated with aircrew error and material failure with a Pearson's Correlation Coefficient of 0.814 ($p = 0.014$). The systems in this category are designed to support and facilitate the execution of a mission, but not the performance of the aircraft. The displayed correlation between the quantity of these systems and accidents caused by material failure and aircrew error indicates that in addition to the task of flying an aircraft, the crew were significantly tasked with aspects of mission execution. Given the expectation that "automation complacency occurs under conditions of multiple-task load, when manual tasks compete with automated task for the operator's attention" (Parasuraman & Manzey, 2010) and experimental observation that "time pressure compromises the quality of decision-making" (Rieger & Manzey, 2022), it can be concluded that increased presence of mission performance systems may compete for aircrew attention and potentially contribute to an operator-out-of-the-loop (OOTL) situation for other categories.

Information management systems correlated positively to the costs associated with all subsets of aircrew error (Pearson's Correlation Coefficient of 0.895, $p = 0.003$). Systems designed for information management are primarily focused on transfer of information to the aircrew. The results indicate that an increase in systems to manage information flow in the cockpit correspond to increased costs or consequences of aircrew error. This supports the theory of the presence of automation bias where aircrew are not sufficiently acquiring the information needed to make the correct decisions and instead are relying on solutions provided by the system. The observations in this study appear to add to the anecdotal evidence to laboratory observations that when "given an unreliable system, humans are still likely to approve computer-generated recommendations" (Cummings, 2004).

4.5.6 Correlation Summary

Correlation analysis resulted in connections between some attributes of automation and human performance. It was interesting to observe that generally accident rate did not correlate to attributes of automation, except for crew size, as well as accident fatalities. However, costs associated with accidents due to aircrew error did show positive correlation with several attributes which is somewhat expected as this research focused on automation present in the cockpit with the primary user being aircrew.

The measure of performance that correlated to the most attributes of automation was aircrew error associated with material failure. Since the seven attributes displayed positive correlation, it appears that room for improvement in design or approach exists. The results of this study for the select population correspond to the observations and concerns addressed in previous research. Five of the attributes displayed positive correlation to accidents due only to aircrew error indicating that the operator-out-of-the-loop and automation bias may be present in the subject population as well. Based on the correlation analysis, further research into how high and low-level automation affects operator performance as well as systems designed for information acquisition, action implementation, aircraft performance, and mission performance would potentially yield improvements.

Chapter 5. Summary and Conclusion

The research summarized in this document was able to show relationships between attributes of automation and human accident performance in the subject setting. The empirical evidence contained in this report highlighted the successful use of existing taxonomies to characterize automation, the effective measurement of operator performance in the subject environment, and the observation of correlation between attributes of automation and human accident performance.

First, it was demonstrated (through statistical analysis) that it is possible to assess if different aircraft were exposed to a common operating environment over a prolonged period of time. Though analysis, the data indicated the presence of a monotonic relationship of the subject aircraft over time with relation to accident rate. This correlation in accident rates over time suggests the aircraft were exposed to a common environmental stimulus.

Second, it was demonstrated that while accident rate is the traditional method of measuring accident performance, use of accident cost and number of fatalities may be more useful. Traditional methods to measure accident performance have relied on dollar values adjusted for inflation and accident rates to show improvement or regression. The method of using a percentage of aircraft or crew lost due to an accident provided a comparative scale between the aircraft to illustrate differences in performance with statistical significance.

Third, it was demonstrated that the use of existing taxonomies for categories, types, and levels of automation are of sufficient granularity for correlation to performance. While costs associated with accidents caused by aircrew error correlated to the most attributes of automation, accident rate, and fatalities displayed correlation as well. The addition of the life support category to the list proposed by Dudley, *et al.* (Dudley, *et al.*, 2014) addressed systems that did not fit into the other categories during the survey portion. However, life support systems did not correlate to any aspect of human performance.

The fourth conclusion from this study was that correlation did exist between certain configurations of cockpit automation and accident performance. Correlation analysis showed a consistent relationship between increased quantities of automation and aircrew error. Of note, empirical data indicated that as levels of automation increase, aircrew error modalities change from situations where automation complacency would be expected to situations where automation bias is more expected. Additionally, it was observed that increased quantities of systems assigned to the categories of aircraft performance and mission performance both correlate only to increases of costs associated with accidents caused by aircrew error and material failure. However, increased quantities of systems assigned to the category of information management correlated to increased costs associated with accidents caused by all combinations of aircrew error indicating that those systems have a broader relationship to aircrew performance. Finally, it should be pointed out that for types of automation, action implementation correlated to increased costs of accidents caused by aircrew error and material failure which indicates the presence of time critical situation. Conversely, information acquisition types of systems correlated to the cost of accidents caused only by aircrew error without an obvious indicator of a time critical scenario.

It is important to note that while the potential for multiplicity was identified through the use of the Hochberg method, it was determined that the results of the study did find the existence of correlation between certain configurations of cockpit automation and accident performance. While the Hochberg method reduced the family-wise error rate through reduction of individual test alpha, there was a corresponding increase in potential for Type II error. Considering the fact that the source data for the study was empirical, all environmental factors external to the focus of this study could not be removed, the findings associated with crew size aligned to well documented observations in previous research (section 4.5.1), and the remaining findings aligned to previous laboratory and simulator-based experiments (sections 4.5.2-4.5.5); it was assessed that while the potential of Type I error is present, it is unlikely that the results of this study are a product of chance.

The fifth conclusion of this study is the observation of a potential connection between group identity fusion and fatality rates for accidents involving automated cockpit systems. The impact and interaction of crew size is something that presented an interesting observation as a reduction in accident rate correlating to increased crew size would be expected, but an increase in fatalities (and a percentage of crew size) correlating to larger crew sizes was not expected. The observations from this research indicate that the concept of tightly fused groups and potential for extreme self-sacrifice may add another dimension to performance of automated systems and multiple operators when in a crew construct. It is recommended that further research be conducted to investigate the relationship between crew size or dynamics and effectiveness of attributes of automation.

Recommendations from this study regarding future designs of automated systems include consideration of adaptive automation, changes to operator training, and consideration of group dynamics when implementing an automated system with multiple operators. The data aligned with previous research regarding automation induced bias and complacency. It is recommended that future designs take these phenomena into consideration and explore options to mitigate their presence, to include potential application of adaptive automation. The data also aligned with previous research regarding operator skill atrophy when exposed to automated systems. It is recommended that future consideration is given to mitigating this through design and training requirements. Symptoms of identity fusion and associated connections to extreme self-sacrifice were observed in the data used in this study. It is recommended that consideration is given in future designs to account for these social behaviors amongst groups of operators. Through application of these recommendations, it is expected that improvements to human-machine performance may result.

Proposed areas for future study include application of contemporary automation technologies, impact of group and social interactions, and replication of this study to mitigate any impacts of the limitations previously stated. The results of this study corresponded to a number of observations from previous research. The preponderance of positive correlation to undesired performance (*i.e.*, increased accident cost and fatalities) indicates that further application and study in the fields of adaptive automation, operator feedback systems, and training would be beneficial. Specifically, methods to keep operators mentally stimulated and engaged in the system. This study also indicated that findings from previous research regarding identity fusion and the associated symptoms were present in the data. While the presence of identity fusion and the associated connection to extreme self-sacrificial behavior has been observed in other settings, the literature was very limited regarding observations in the aviation field. It is recommended that future study explore the presence of identity fusion in the aviation community. The third recommendation of this study is to replicate the process and determine if similar findings are observed with other aircraft and systems. The research and data processing of this study was conducted by one individual (author). Replication would help mitigate the impact of any unintentional bias or skewing of the results.

The research described in this dissertation was an empirical study of cockpit automation and aircrew accident performance in high performance aircraft operating in the U.S. naval shipboard environment. Previous research has proposed a number of impacts that automation places on human performance. It was the purpose of this study to assess if a correlation between attributes of cockpit automation and aircrew accident performance in the operational environment could be measured to a statistically significant level of $\alpha \leq 0.05$. The results indicate that the objective was achieved. It is the hope of the author that the work contained in this document contributes meaningfully to the body of knowledge and is a steppingstone for more discovery.

References

Ajmi, A. A., Mahmood, N. S., Jamaludin, K. R., Talib, H. H., Sarip, S., & Kaidi, H. M. (2022). Intelligent Integrated Model for Improving Performance in Power Plants. *Computers, Materials & Continua, 70*(3), 5783-5801. doi:10.32604/cmc.2022.021885

Ancel, E., Shih, A. T., Jones, S. M., Reveley, M. S., Luxhoj, J. T., & Evans, J. K. (2015). Predictive Safety Analytics: Inferring Aviation Accident Shaping Factors and Causation. *Journal of Risk Research, 18*(4), 428-451.

Antunes, R., Coito, F. V., & Duarte-Ramos, H. (2013, September 03). Skill Evaluation in Point-to-Point Human-Machine Operation. *Applied Mechanics and Materials, 394*, pp. 463-469. doi:10.4028/www.scientific.net/AMM.394.463

Archer, J., Keno, H., & Kwon, Y. (2012). *Effects of Automation in the Aircraft Cockpit Environment: Skill Degradation, Situation Awareness, Workload.* Purdue University, School of Industrial Engineering, West Lafayette.

Bahardwaj, A., Ghasemi, A. H., Zheng, Y., Febbo, H., Jayakumar, P., Ersal, T., . . . Gillespie, R. B. (2020, January 7). Who's the Boss? Arbitrating Control Authority Between a Human Driver and Automation System. *Transportation Research Part F, 68*, 144-160. doi:10.1016/j.trf.2019.12.005

Bahner, J. E., Hüper, A.-D., & Manzey, D. (2008, June 2). Misuse of Automated Decision Aids: Complacency, Automation Bias and the Impact of Training Experience. *International Journal of Human-Computer Studies*, *66*, 688-699. doi:10.1016/j.ijhcs.2008.06.001

Bahnik, S., Efendic, E., & Vranka, M. A. (2021). Sacraficing Oneself or Another: The Difference Between Prescriptive and Normative Judgments in Moral Evaluations. *Psychological Reports, 124*(1), 108-130. doi:10.1177/0033294119896061

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006, Winter). Comparison of a Brain-Based Adaptive System and a Manual Adaptable System for Invoking Automation. *Human Factors, 48*(4), 693-709.

Balfe, N., Sharples, S., & Wilson, J. R. (2018, June). Understanding Is Key: An Analysis of Factors Pertaining to Trust in a Real-World Automation System. *Human Factors, 60*(4), 477-495. doi:10.1177/0018720818761256

Barg-Walkow, L. H., & Rogers, W. A. (2016, March). The Effect of Incorrect Reliability Information on Expectations, Perceptions, and Use of Automation. *Human Factors, 58*(2), 242-260. doi:10.1177/0018720815610271

Bartlett, M. L., & McCarley, J. S. (2017, March). Benchmarking Aided Decision Making in a Signal Detection Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(6), pp. 881-900.

Beck, H. P., Dzindolet, M. T., & Pierce, L. G. (2007, June). Automation Usage Decisions: Controlling Intent and Appraisal Errors in a Target Detection Task. *Human Factors, 49*(3), pp. 429-437. doi:10.1518/001872007X200076

Beller, J., Heesen, M., & Vollrath, M. (2013, December). Improving the Driver-Automation Interaction: An Approach Using Automation Uncertainty. *Human Factors*, *55*(6), 1130-1141.

Billings, C. (1997). The Search For A Human-Centered Approach. *Aviation automation*.

Boskemper, M. M., Bartlett, M. L., & McCarley, J. S. (2022). Measuring the Efficiency of Automation-Aided Perforance in a Simulated Baggage Screening Task. *Human Factors, 64*(6), 945-961. doi: 10.1177/ 0018 7208 20983632

Bowden, V. K., Griffiths, N., Strickland, L., & Loft, S. (2021). Detecting a Single Automation Failure: The Impact of Expected (But Not Experienced) Automation Reliability. *Human Factors*, 1-13. doi:10.1177/00187208211037188

Brandenburger, N., & Jipp, M. (2017, September 13). Effects of Expertise for Automatic Train Operations. *Cogn Tech Work*, *19*, 699-709. doi:10.1007/s10111-017-0434-2

Broccia, G., Milazzo, P., & Ölveczky, P. C. (2019, April 8). Formal Modeling and Analysis of Safety-Critical Human Multitasking. *Innovations in Systems and Software Engineering*, 169-190. doi:10.1007/s11334-019-00333-7

Cabrall, C. D., Janssen, N. M., & de Winter, J. C. (2018, October 1). Adaptive Automation: Automatically (Dis)engaging Automation During Visually Distracted Driving. *PeerJ Computer Science*, 1-27. doi:10.7717/peerj-cs.166

Cabrall, C. D., Stapel, J. C., Happee, R., & de Winter, J. C. (2020, March). Redesigning Today's Driving Automation Toward Adaptive Backup Control With Context-Based and Invisible Interfaces. *Human Factors*, *62*(2), 211-228. doi:10.1177/0018720819894757

Calhoun, G. (2022, March). Adaptable (Not Adaptive) Automation: Forefront of Human-Automation Teaming. *Human Factors, 64*(2), 269-277. doi:10.1177/00187208211037457

Casey, T., Griffin, M. A., & Harrison, H. F. (2017). Safety Climate and Culture: Integrating Psychological and Systems Perspectives. *Journal of Occupational Health Psychology*, *22*(3), 341-353. doi:10.1037/ocp0000072

Casner, S. M., & Schooler, J. W. (2014, May). Thoughts in Flight: Automation Use and Pilots' Task-Related and Task-Unrelated Thought. *Human Factors*, *56*(3), 433-442. doi:10.1177/0018720813501550

Chiou, E. K., & Lee, J. D. (2021). Trusting Automation: Designing for Responsivity and Resilience. *Human Factors*, 1-29. doi:10.1177/00187208211009995

Chok, N. S. (2010). *Pearson's Versus Spearman's and Kendall's Correlation Coefficients for Continuous Data.* University of Pittsburgh, Graduate School of Public Health.

Connor, J. M. (2010). Military Loyalty; A Functional Vice? *Criminal Justics Ethics, 29*(3), 278-290.

Connor, J., Andrews, D. J., Noack-Lundberg, K., & Wadham, B. (2021). Military Loyalty as a Moral Emotion. *Armed Forces & Society*, 530-550. doi:10.1177/0095327X19880248

Cummings, M. L. (2004). Automation Bias in Intelliegent Time Critical Decision Support Systems. *AIAA 1st Intelliegent Systems Technical Conference Vol. 2*, (pp. 557-562).

Cummings, M. L., Clare, A., & Hart, C. (2010, February). The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling. *Human Factors, 52*(1), 17-27. doi:10.1177/0018720810368674

Cymek, D. H. (2018, November). Redundant Automation Monitoring: Four Eyes Don't See More Than Two, if Everyone Turns a Blind Eye. *Human Factors, 60*(7), 902-921. doi:10.1177/0018720818781192

Davis, T. A. (2010). *A Study of the Contribution of Human Factors to Human-Machine System Failures in Dynamic Mission Operations.* Dissertation, The George Washington University, The School of Engineering and Applied Science.

De Boer, R., & Dekker, S. (2017, September 11). Models of Automation Surprise: Results of a Field Survey in Aviation. *Safety, 3*(20), 1-11. doi:10.3390/safety3030020

Degani, A., & Heymann, M. (2002, Spring). Formal Verification of Human-Automation Interaction. *Human Factors, 44*(1), 28-43.

Department of Defense. (2005, January 11). *DOD HFACS.* Retrieved January 02, 2022, from North American Electric Reliability Corporation: https://www.nerc.com/pa/rrm/ea/CA_Reference_Materials_DL/Department%20of%20Defense%20Human%20Factors%20Analysis%20and%20Classification%20System.pdf

Di Flumeri, G., De Crescenzio, F., Berberian, B., Ohnelser, O., Kramer, J., Arico, P., . . . Piastra, S. (2019, September 06). Brain-Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems. *Frontiers in Human Neuroscience, 13*.

Dixon, S. R., & Wickens, C. D. (2006, Fall). Automation Reliability in Unmanned Aerial Vehicle Control: A Reliance-Compliance Model of Automation Dependence in High Workload. *Human Factors, 48*(3), 474-486.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007, August). On the Independence of Compliance and Reliance: Are Automation False Alarms Worse Than Misses? *Human Factors, 49*(4), pp. 564-572.

Dmec, K., Marathe, A. R., Likos, J. R., & Metcalfe, J. S. (2016, June 30). From Trust in Automation to Decision Neuroscience: Applying Cognitive Neuroscience Methods to Understand and Improve Interaction Decisions Involoved in Human Automation Interaction. *Frontiers in Human Neuroscience, 10*, 1-14. doi:10.3389/fnhum.2016.00290

Dmec, K., Marathe, A. R., Lukos, J. R., & Metcalfe, J. S. (2016, June 30). From Trust in Automation to Decision Neuroscience: Applying Cognitive Neuroscience Methods to Understand and Improve Interaction Decisions Involved in Human Automation Interaction. *Frontiers in Human Neuroscience, 10*, 1-15. doi:10.3389/fnhum.2016.00290

Donmez, B., Pina, P. E., & Cummings, M. L. (2008, August 19-21). Evaluation Criteria for Human-Automation Performance Metrics. *PerMIS'08*, 77-82.

Du, N., Huang, K. Y., & Yang, X. J. (2020, September). Not All Information Is Equal: Effects of Disclosing Different Types of Likelihood Information on Trust, Compliance and

Reliance, and Task Performance in Human-Automation Teaming. *Human Factors, 62*(6), 987-1001. doi:10.1177/0018720819862916

Dudley, R., Dorneich, M. C., Letsu-Dake, E., Rogers, W., Whitlow, S. D., Dillard, M., & and Nelson, E. (2014). Characterization of Information Automation on the Flight Deck. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (p. Paper 9.). Chicago.

Elliott, M., Page, K., & Worrall-Carter, L. (2012, December). Reason's Accident Causation Model: Application to Adverse Events in Acute Care. *Comtemporary Nurse, 43*(1), 22-28.

Endsley, M. R. (2017, February). From Here to Autonomy: Lessons Learned From Human-Automation Research. *Human Factors, 59*(1), 5-27. doi:10.1177/0018720816681350

Endsley, M. R. (2017, February). From Here to Autonomy: Lessons Learned From Human-Automation Research. *Human Factors, 59*(1), 5-27. doi:10.1177/0018720816681350

Endsley, M. R. (n.d.). Automation and Situation Awareness. In R. Parasuraman, & M. Mouloua, *Automation and Human Performance: Theory and Applications* (pp. 163-181). Mahwah, N.J.: Lawrence Erlbaum.

Fa, Z., Li, X., Liu, Q., Qui, Z., & Zhai, Z. (2021). Correlation in Causality: A Progressive Study of Hierarchical Relations within Human and Organizational Factors in Coal Mine Accidents. *International Journal of Environmental Research and Public Health, 18*(5020), 1-16.

Forough, C. K., Devlin, S., Pak, R., Brown, N. L., Sibley, C., & Coyne, J. T. (2021). Near-Perfect Automation: Investigating Performance, Trust, and Visual Attention Allocation. *Human Factors*, 1-16. doi:10.1177/00187208211032889

Fredman, L. A., Buhrmester, M. D., Gomez, A., Fraser, W. T., Talaifar, S., Brannon, S. M., & Swann, Jr., W. B. (2015). Identity Fusion, Extreme Pro-Group Behavior, and the Path to Defusion. *Social and Personality Psychology Compas, 9*(9), 468-480. doi:10.1111/spc3.12193

Fu, G., Cao, J.-L., & Xiang, Y.-C. (2017). Comparative Study of HFACS and the 24Model Accident Causation Models. (Y.-H. Sun, Ed.) *Petroleum Science, 14*, 570-578. doi: 10.1007/s12182-017-0171-4

Funk, K., & Lyall, B. (2000). *A Comparative Analysis of Flightdecks With Varying Levels of Automation.*

Galster, S. M. (2003). An Examination of Complex Human-Machine System Performance Under Multiple Levels and Stages of Automation. Washington, D.C, USA.

Geiselman, E. E., Johnson, C. M., & Buck, D. R. (2013, July). Flight Deck Automation: Invaluable Collaborator or Insidious Enabler? *Ergonomics In Design*, 22-26. doi:10.1177/1064804613491268

Gil, G.-H., & Kaber, D. B. (2012). An Accessible Cognitive Modeling Tool for Evaluation of Pilot-Automation Interaction. *The International Journal of Aviation Psychology, 22*(4), 319-342. doi:10.1080/10508414.2012.718236

Gil, G.-H., Kaber, D., Kaufmann, K., & Kim, S.-H. (2012). Effects of Modes of Cockpit Automation on Pilot Performance and Workload in a Next Generation Flight Concept of Operation. *Human Factors and Ergonomics in Manufacturing & Service Industries, 22*(5), 395-406. doi:10.1002/hfm.20377

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators. *Journal of the American Medical Informatics Association, 19*, 121-127. doi:10.1136/amiajnl-2011-000089

Grimell, J. (2019). Suffering for Others While Making Others Suffer: Military Narratives of Sacrifice. *Journal of Pastoral Care & Counseling, 73*(1), 30-40. doi:10.1177/1542305019828658

Haass, M. J., Warrender, C. E., Burnham, L., Jeffers, R. F., Stevens-Adams, S. M., Cole, K. S., & Forsythe, C. (2015). Toward an Objective Measure of Automation for the Electric Grid. *6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015* (pp. 5285-5292). Elsevier B.V. doi:10.1016/j.promfg.2015.07.611

Haiyang, C., Shengkui, Z., & Jianbin, G. (2019). Reliability Assessment of Man-Machine Systems Subject to Mutually Dependent Machine Degradation and Human Errors. *Reliability Engineering and System Safety, 190*, 1-11.

Hancock, P. A., Jagacinski, R. J., Parasuraman, R., Wickens, C. D., Wilson, G. F., & Kaber, D. B. (2013, April). Human-Automation Interaction Research: Past, Present, and Future. *Ergonomics In Design: The Quarterly of Human Factors Applications, 21*(9), 9-14. doi:10.1177/1064804613477099

Hancock, P. A., Kajaks, T., Caird, J. K., Chignell, M. H., Mizobuchi, S., Burns, P. C., . . . Vrkljan, B. H. (2020, March). Challenges to Human Drivers in Increasingly Automated Vehicles. *Human Factors, 62*(2), 310-328. doi:10.1177/0018720819900402

Haque, H., Alrowily, A., Jalal, Z., Tailor, B., Efue, V., Sarwar, A., & Paudyal, V. (2021). Direct Oral Anticoagulant-Related Medication Incidents and Pharmacists' Intervention in Hospital In-Patients: Evaluation Using Reason's Accident Causation Theory. *International Journal of Clinical Pharmacy*(43), 1693-1704.

Harris, D. (2006). The Influence of Human Factors on Operational Efficiency. *Aircraft Engineering and Aerospace Technology, 78*(1), 20-25.

Harris, D. (2007). A Human-Centered Design Agenda For The Development Of Single Crew Operated Commercial Aircraft. *Aircraft Engineering and Aerospace Technology: An International Journal, 79*(5), 518-526. doi: 10.1108/00022660710780650

Haslbeck, A., & Hoermann, H.-J. (2016, June). Flying the Needles: Flight Deck Automation Erodes Fine-Motor Flying Skills Among Airline Pilots. *Human Factors*, *58*(4), pp. 533-545.

Hasselberg, A., & Söffker, D. (2013). A Human Cognitive Performance Measure Based on Available Options for Adaptive Aiding. *12th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems*, (pp. 442-449). Las Vegas, NV, USA.

Hauke, J., & Kossowski, T. (2011). Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quaestiones Geographicae, 30*(2), 87-93. doi:10.2478/v10117-011-0021-1

Helldin, T., & Falkman, G. (2011). Human-Centered Automation and the Development of Fighter Aircraft Support Systems. *Proceedings of the Swedish Human Factors Network (HFN) Conference*, (p. 21). Linköping, Sweeden. Retrieved from http://www.humanfactorsnetwork.se/

Herczeg, M., Kammler, M., Mentler, T., & Roenspieß, A. (2013). The Usability Engineering Repository UsER for the Development of Task- and Event-based Human-Machine-Interfaces. *12th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems* (pp. 483-490). Las Vegas, NV: IFAC. doi:10.3182/20130811-5-US-2037.00011

Hergeth, S., Lorenz, L., & Krems, J. F. (2017, May). Prior Familiarization With Takeover Requests Affects Drivers' Takeover Perforance and Automation Trust. *Human Factors, 59*(3), 457-470. doi:10.1177/0018720816678714

Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2016, May). Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust During Highly Automated Driving. *Human Factors, 58*(3), 509-519. doi:10.1177/0018720815625744

Hobbs, A., & Williamson, A. (2003). Associations Between Errors and Contributing Factors in Aircraft Maintenance. *Human Factors, 45*(2), 186-201.

Hoff, K. A., & Bashir, M. (2015, May). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors, 57*(3), 407-434. doi:10.1177/0018720814547570

Holden, R. J. (2009, December). People or Systems? *Professional Safety, 54*(12), 34-41.

Horrey, W. J., & Lee, J. D. (2020, March). Preface to the Special Issue on Human Factors and Advanced Vehicle Automation: Of Benefits, Barriers, and Bridges to Safe and Effective Implementation. *Human Factors, 62*(2), 189-193. doi:10.1177/0018720820901542

Howard, A. M. (2007, July). A Systematic Approach to Predict Performance of Human-Automation Systems. *IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews, 37*(4), 594-601.

Huang, J., Choo, S., Pugh, Z. H., & Nam, C. S. (2022, September). Evaluating Effective Connectivity of Trust in Human-Automation Interaction: A Dynamic Causal Modeling (DCM) Study. *Human Factors, 64*(6), 1051-1069. doi: 10. 1177/ 0018 7208 20987443

Huegli, D., Merks, S., & Schwaninger, A. (2020). Automation Reliability, Human-Machine System Performance, and Operator Compliance: A Study With Airport Security Screeners Supported by Automated Explosives Detection Systems for Cabin Baggage Screening. *Applied Ergonomics, 86*, 1-12.

Hutchinson, J., Strickland, L., Farrell, S., & Loft, S. (2022). The Perception of Automation Reliability and Acceptance of Automation Advice. *Human Factors*, 1-17. doi:10.1177/00187208211062985

Jamieson, G. A., & Skraaning, G. (2020, June). The Absence of Degree of Automation Trade-Offs in Complex Work Settings. *Human Factors, 62*(4), pp. 516-529.

Jamieson, G. A., & Vicente, K. J. (2005, Spring). Designing Effective Human-Automation-Plant Interfaces: A Control-Theoretic Perspective. *Human Factors, 47*(1), 12-34.

Jaussi, J. A., & Hoffmann, H. O. (2018). Manned Versus Unmanned Aircraft Accidents, Including Causation and Rates. *International Journal of Aviation, Aeronautics, and Aerospace, 5*(4), 1-40.

Jipp, M. (2016, February). Expertise Development With Different Types of Automation: A Function of Different Cognitive Abilities. *Human Factors, 58*(1), 92-106. doi:10.1177/0018720815604441

Kaber, D. B. (1996). *The Effect Of Level Of Automation And Adaptive Automation On Performance In Dynamic Control Environments.* Dissertation, Texas Tech University, Industrial Engineering.

Kaber, D. B. (2000). *Situation Awareness & Levels Of Automation: Empirical Assessment Of Levels Of Automation In The commercial Cockpit.* National Aeronautics and Space Administration, Langley Research Center, Langley, VA.

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, no. 3*, 126-131.

Kaber, D. B., & Endsley, M. R. (1997). The Combined Effect of Level of Automation And Adaptive Automation On Human Perforamnce With complex, Dynamic Control Systems. *Proceedings of the Human Factors and Ergonomics Socienty 41st Annual Meeting*, (pp. 205-209).

Kaber, D. B., & Endsley, M. R. (2004). The Effects of Level of Automation and Adaptive Automation on Human Performance, Situational Awareness and Workload in a Dynamic Control Task. *Theoretical Issues in Ergonomics Science 5, no. 2*, 113-153.

Kaber, D. B., Wright, M. C., Prinzel III, L. J., & Clamann, M. P. (2005). Adaptive Automation of Human-Machine System Information-Process Functions. *Human Factors*, *47*(4), 730-741.

Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015, January). 'Utilitarian' Judgments In Sacrificial Moral Dilemmas Do Not Reflect Impartial Concern For The Greater Good. *Cognition, 134,* 193-209. doi:10.1016/j.cognition.2014.10.005

Keller, J. C. (2013). Flight Skill Proficiency Issues In Instrument Approach Accidents. *Aviation Technology Graduate Student Publications*, p. 21. Retrieved from http://docs.lib.purdue.edu/atgrads/26

Klinect, J. (2005). *Line Operations Safety Audit: A Cockpit Observation Methodology for Monitoring Commercial Airline Safety Performance.* PhD Dissertation, The University of Texas at Austin.

Klumpp, M., Hesenius, M., Meyer, O., Ruiner, C., & Gruhn, V. (2019, May 12). Production Logistics and Human-Computer Interaction -- State-Of-The-Art, Challenges and

Requirements for the Future. *The International Journal of Advanced Manufacturing Technology*, *105*, 3691-3709. doi:10.1007/s00170-019-03785-0

Körner, U., Müler-Thur, K., Lunau, T., Dragano, N., Angerer, P., & Buchner, A. (2019). Perceived Stress In Human-Machine Interaction In Modern Manufacturing Environments - Results of a Qualitative Interview Study. *Stress and Health, 35*, 187-199. doi:10.1002/smi.2853

Kraus, J., Scholz, D., Stiegemeier, D., & Baumann, M. (2020, August). The More You Know: Trust Dynamics and Calibration in Highly Automated Driving and the Effects of Take-Overs, System Malfunction, and System Transparency. *Human Factors, 62*(5), pp. 718-736.

Lee, B. C., Park, J., Jeong, H., & Park, J. (2020, February 14). Validation of Trade-Off in Human-Automation Interaction: An Empirical Study of Contrasting Office Automation Effects on Task Performance and Workload. *Applied Sciences*, *10*, 1-14. doi:10.3390/app10041288

Lee, J. D. (2008, June). Review of a Pivotal Human Factors Article: "Humans and Automation: Use, Misuse, Disuse, Abuse". *Human Factors, 50*(3), 404-410. doi:10.1518/001872008X288547

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

Lee, R. L. (2004). *The Impact of Cognitive Task Analysis on Performance: A Meta-Analysis of Comparative Studies.* University of Southern California, The Rossier School of Education. Author.

Lesley Univeristy. (n.d.). *An Introduction to Flipped Learning*. Retrieved September 11, 2022, from Lesley University: https://lesley.edu/article/an-introduction-to-flipped-learning

Lesley University. (n.d.). *An Introduction To Flipped Learning*. Retrieved September 11, 2022, from Lesley University: https://lesley.edu/article/an-introduction-to-flipped-learning

Li, H., Wickens, C. D., Sarter, N., & Sebok, A. (2014, September). Stages and Levels of Automation in Support of Space Teleoperations. *Human Factors, 56*(6), 1050-1061. doi:10.1177/0018720814522830

Li, W., Zhang, L., & Liang, W. (2017). An Accident Causation Analysis and Taxonomy (ACAT) Model of Complex Industrial System From Both System Safety and Control Theory Perspectives. (E. Ltd., Ed.) *Safety Science, 92,* 94-103.

Li, Y., & Burns, C. M. (2017, December). Modeling Automation With Cognitive Work Analysis to Support Human-Automation Coordination. *Journal of Cognitive Engineering and Decision Making, 11*(4), 299-322. doi:10.1177/1555343417709669

Liu, K. K. (1997). *The Highly-Automated Airplane: Its Impact On Aviation Safety And An Analysis Of Training Philosophy.* Thesis, Air Force Institute of Technology, Graduate School of Logistics and Acquisition Management.

Loft, S., Bhaskara, A., Lock, B. A., Skinner, M., Brooks, J., Li, R., & Bell, J. (2021). The Impact of Transparency and Decision Risk on Human-Automation Teaming Outcomes. *Human Factors*, 1-16. doi:10.1177/00187208211033445

Lyell, D., Magrabi, F., & Coiera, E. (2018, November). The Effect of Cognitive Load and Task Complexity on Automation Bias in Electronic Prescribing. *Human Factors, 60*(7), 1008-1021. doi:10.1177/0018720818781224

Lyons, J. B., & Stokes, C. K. (2012, February). Human-Human Reliance in the Context of Automation. *Human Factors, 54*(1), 112-121. doi:10.1177/0018720811427034

Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006, Summer). Automation Failures on Tasks Easily Performed by Operators Undermine Trust in Automated Aids. *Human Factors, 48*(2), 241-256.

Marquez, J. J., & Gore, B. F. (2017, March). Measuring Safety and Performance in Human-Automation Systems: Special Issue Commentary. *Human Factors*, *59*(2), pp. 169-171.

McBride, S. E., Rogers, W. A., & Fisk, A. D. (2011, December). Understanding the Effect of Workload on Automation Use for Younger and Older Adults. *Human Factors, 53*(6), 672-686. doi:10.1177/0018720811421909

McDonnell, A. S., Simmons, T. G., Erickson, G. G., Lohani, M., Cooper, J. M., & Strayer, D. L. (2021). This Is Your Brain on Autopilot: Neural Indices of Driver Workload and Engagement During Partial Vehicle Automation. *Human Factors*, 1-16. doi:10.1177/00187208211039091

Merritt, S. M. (2011, August). Affective Processes in Human-Automation Interactions. *Human Factors, 53*(4), 356-370. doi:10.1177/0018720811411912

Merritt, S. M., & Ilgen, D. R. (2008, April). Not All Trust Is Created Equal: Dispositional and History-Based Trust In Human-Automation Interactions. *Human Factors, 50*(2), 194-210. doi:10.1518/001872008X288574

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013, June). I Trust It, But I Don't Know Why: Effects of Implicit Attitudes Toward Automation on Trust in an Automated System. *Human Factors*, *55*(3), 520-534. doi:10.1177/0018720812465081

Merritt, S. M., Unnerstall, J. L., Lee, D., & Huber, K. (2015, August). Measuring Individual Differences in the Perfect Automation Schema. *Human Factors, 57*(5), 740-753. doi:10.1177/0018720815581247

Miller, C. A., & Parasuraman, R. (2007, February). Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors, 49*(1), 57-75.

Miranda, A. T. (2018, September). Understanding Human Error in Naval Aviation Mishaps. *Human Factors*, *60*(6), 763-777. doi:10.1177/0018720818771904

Motamedi, S., Wang, P., Zhang, T., & Chan, C.-Y. (2020, March). Acceptance of Full Driving Automation: Personally Owned and Shared-Use Concepts. *Human Factors, 62*(2), 288-309. doi:10.1177/0018720819870658

Mühl, K., Strauch, C., Grabmaier, C., Reithinger, S., Huckauf, A., & Baumann, M. (2020, December). Get Ready for Being Chauffeured: Passenger's Preferences and Trust While Being Driven by Human and Automation. *Human Factors, 62*(8), 1322-1338. doi:10.1177/0018720819872893

Muslim, H., & Itoh, M. (2019, June 6). Effects of Human Understanding of Automation Abilities on Driver Performance and Acceptance of Lane Change Collision Avoidance Systems. *IEEE Transactions on Intelligent Transportation Systems, 20*(6), 2014-2024.

Naikar, N. (2006). An Examination of the Key Concepts of the Five Phases of Cognitive Work Analysis with Examples From a Familiar System. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, 447-451.

Neubauer, C., Matthews, G., Langheim, L., & Saxby, D. (2012, October). Fatigue and Voluntary Utilization of Automation in Simulated Driving. *Human Factors, 54*(5), 734-746. doi:10.1177/0018720811423261

Nikolic, M. L., & Sarter, N. B. (2007, August). Flight Deck Disturbance Management: A Simulator Study of Diagnosis and Recovery From Breakdowns in Pilot-Automation Coordination. *Human Factors, 49*(4), 553-563. doi:10.1518/001872007X215647

Norman, D. A. (1990). The 'Problem' With Automation: Inappropriate Feedback and Interaction, Not 'Over-Automation'. *Philosophical Transactions of the Royal Society of London, Biological Sciences 327, no. 1241*, 585-593.

Norman, K. L., & Panizzi, E. (2006, March). Levels of Automation and User Participation In Usability Testing. *Interacting With computers, 18*(2), 246-264. doi:10.1016/j.intcom.2005.06.002

Null, C. H., Adduru, V., Ammann, O. C., Cardoza, C. T., Stewart, M. J., Avrekh, I., . . . Smith, B. E. (2019, February). *Human Performance Contributions to Safety in Commercial Aviation.* Langely Research Center, NASA Engineering and Safety Center. Hampton, Virginia: National Aeronautics and Space Administration.

Olson, W. A. (2000). *Identifying And Mitigating The Risks Of Cockpit Automation.* Air Command and Staff College, Maxwell Air Force Base, Alabama.

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014, May). Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis. *Human Factors, 56*(3), 476-488. doi:10.1177/0018720813501549

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

Parasuraman, R., & Riley, V. (1997, June). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, 39*(2), 230-253.

Parasuraman, R., & Wickens, C. D. (2008, June 3). Humans: Still Vital After All These Years of Automation. *Human Factors, 50*(3), 511-520. doi: 10.1518/001872008X312198

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance Consequences of Automation-Induced "Complacency". *The International Journal of Aviation Psychology 3, no. 1*, 1-23.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A Model for Types and Levels of Human Interaction with Automation. *Systems, Man and Cybernetics- Part A: Systems and Humans, IEEE Transactions on 30, no. 3*, 286-297.

Patterson, R. E. (2017, February). Intuitive Cognition and Models of Human-Automation Interaction. *Human Factors, 59*(1), 101-115. doi:10.1177/0018720816659796

Petermeijer, S. M., Abbink, D. A., & de Winter, J. C. (2015, February). Should Drivers Be Operating Within an Automation-Free Bandwidth? Evaluating Haptic Steering Support Systems With Different Levels of Authority. *Human Factors, 57*(1), 5-20. doi:10.1177/0018720814563602

Petridou, E., & Moustaki, M. (2000). Human Factors in the Causation of Road Traffic Crashes. *European Journal of Epidemiology, 16*(9), 819-826.

Peysakhovich, V., Lefrançois, O., Dehais, F., & Causse, M. (2018, February 27). The Neuroergonomics of Aircraft Cockpits: The Four Stages of Eye-Tracking Integration to Enhance Flight Safety. *Safety, 4*(8), 1-15. doi:10.3390/safety4010008

Piccinini, G. B., Lehtonen, E., Forcolin, F., Engström, J., Albers, D., Markkula, G., . . . Sandin, J. (2020, November). How Do Drivers Respond to Silent Automation Failures? Driving Simulator Study and Comparison of Computational Driver Braking Models. *Human Factors, 62*(7), 1212-1229. doi:10.1177/0018720819875347

Pickup, S., Paton, K., Hayes, C., & Morrison, B. (2020, September 13). A Day In The Life of Frontline Manufacturing Personnel: A Diary-Based Safety Study. *Safety Science, 132*, 1-12.

Pina, P. E., Donmez, B., & Cummings, M. L. (2008). *Selecting Metrics to Evaluate Human Supervisory Control Applications.* Massachusetts Institute of Technology, Humans and Automation Laboratory. Cambridge, MA: MIT Department of Aeronautics and Astronautics.

Rand, D. G., & Epstein, Z. G. (2014, October). Risking Your Life without a Second Thought; Intuitive Decision-Making and Extreme Altruism. *PLOS One, 9*(10), 1-6. doi:10.1371/journal.pone.0109687

Reason, J. (1990). The Contribution Of Latent Human Failures To The Breakdown Of Complex Systems. *Philosophical Transactions of the Roypal Society of London. B, Biological Sciences, 327*, 475-484.

Reason, J. (2000). Human Error: Models and Management. *EMJ Volume 320*, 768-770.

Reason, J., Hollnagel, E., & Paries, J. (2006). *Revisiting The << Swiss Cheese >> Model of Accidents.* EUROCONTROL, EUROCONTROL Experimental Centre.

Reichenbach, J., Onnasch, L., & Manzey, D. (2011, December). Human Performance Consequences of Automated Decision Aids in States of Sleep Loss. *Human Factors, 53*(6), 717-728. doi:10.1177/0018720811418222

Rice, S. (2009). Examining Single- and Multiple-Process Theories of Trust in Automation. *The Journal of General Psychology, 136*(3), 303-319.

Rice, S., Trafimow, D., & Hunt, G. (2010). Using PPT to Analyze Suboptimal Human-Automation Performance. *The Journal of General Psychology, 137*(3), 310-329.

Rieger, T., & Manzey, D. (2022). Human Performance Consequences of Automated Decision Aids: The Impact of Time Pressure. *Human Factors, 64*(4), pp. 617-634. doi:10. 1177/ 0018 7208 20965019

Rieger, T., & Manzey, D. (2022). Understanding the Impact of Time Pressure and Automation Support in a Visual Search Task. *Human Factors*, 1-17. doi:10.1177/00187208221111236

Roberts, N., Gerow, J. E., Jeyaraj, A., & Roberts, S. (2017, November). A Meta-Analysis of Organizational Learning and IT Assimilation. *The DATA BASE for Advances in Information Systems, 48*(4), 51-68.

Rovira, E., McGarry, K., & Parasuraman, R. (2007, February). Effects of Imperfect Automation on Decision Making in a Simulated Command and Control Task. *Human Factors, 49*(1), 76-87.

Rudisill, M. (1995). Line Pilots' Attitudes About And Experience With Flight Deck Automation: Results Of An International Survey And Proposed Guidelines. *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 1-6). Columbus, OH: The Ohio State University Press.

Ruff, H. A., Calhoun, G. L., Draper, M. H., Fontejon, J. V., & Guilfoos, B. J. (2004). Exploring Automation Issues In Supervisor Control of Multiple UAVS. *Proceedings of the Human Performance, Situation Awareness, and Automation Technology Conference*, (pp. 218-222).

Sachdeva, S., Iliev, R., Ekhtiari, H., & Dehghani, M. (2015, June 15). The Role of Self-Sacrifice in Moral Dilemmas. *PLOS One, 10*(6), 1-12. doi:10.1371/journal.pone.0127409

Sarter, N. B. (1994). *"Strong, Silent, And 'Out-Of-The-Loop'": Properties Of Advanced (Cockpit) Automation And Their Impact On Human-Automation Interaction.* The Ohio State University, Department of Industrial and Systems Engineering. Columbus, OH: The Ohio State University.

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' Monitoring Strategies and Performance on Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data.". *Human Factors: The Journal of the Human Factors and Ergonomics Society 49, no. 3*, 347-357.

Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016, May). A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems. *Human Factors, 58*(3), 377-400. doi:10.1177/0018720816634228

Schultz, T. P. (n.d.). The Problem with Pilots: How Physicians, Engineers, and Airpower Enthusiasts Redefined Flight.

Sebok, A., & Wickens, C. D. (2017, March). Implementing Lumberjacks and Black Swans Into Model-Based Tools to Support Human-Automation Interaction. *Human Factors, 59*(2), 189-203. doi:10.1177/0018720816665201

Shappell, S. A., & Wiegmann, D. A. (1997). A Human Error Approach to Accident Investigation: The Taxonomy of Unsafe Operations. *The International Journal of Aviation Psychology, 7*(4), 269-291.

Shappell, S. A., & Wiegmann, D. A. (2000). *The Human Factors Analysis and Classification System - HFACS.* Federal Aviation Administration, Office of Aviation Medicine. Washington D.C.: U.S. Department of Transportation.

Shappell, S., & Wiegmann, D. A. (2000). Is Proficiency Eroding Among U.S. Naval Aircrews? A Quantitative Analysis Using The Human Factors Analysis And Classification System. *Human Factors And Ergonomics Society Annual Meeting Proceedings*, (pp. 345-348).

Sheridan, T. B. (2019, November). Extending Three Existing Models to Analysis of Trust in Automation: Signal Detection, Statistical Parameter Estimation, and Model-Based Control. *Human Factors, 61*(7), 1162-1170. doi:10.1177/0018720819829951

Siebold, G. L. (2007). The Essence of Military Group Cohesion. *Armed Forces & Society, 33*(2), 286-295. doi:10.1177/0095327X06294173

Skraaning, G., & Jamieson, G. A. (2021, May). Human Performance Benefits of The Automation Transparency Design Principle: Validation and Variation. *Human Factors, 63*(3), 379-401. doi:10.1177/0018720819887252

Solis-Marcos, I., Ahlström, C., & Kircher, K. (2018, September). Performance of an Additional Task During Level 2 Automated Driving: An On-Road Study Comparing Drivers With and Without Experience With Partial Automation. *Human Factors, 60*(6), 778-792.

Strauch, B. (2017, March). The Automation-by-Expertise-by-Training Interaction: Why Automation-Related Accidents Continue to Occur in Sociotechnical Systes. *Human Factors, 59*(2), 204-228. doi:10.1177/0018720816665459

Streiner, D. L. (2015, October). Best (but oft-forgotten) practices: the multiple problems of multiplicity-whether and how to correct for many statistical tests. *The American Journal of Clinical Nutrition, 102*(4), 721-728.

Swann Jr., W. B., Buhrmester, M. D., Gómez, A., Jetten, J., Bastian, B., Vázquez, A., . . . Zhang, A. (2014). What Makes a Group Worth Dying For? Identity Fusion Fosters Perception of Familial Ties, Promoting Self-Sacrifice. *Journal of Personality and Social Psychology, 106*(6), 912-926. doi:10.1037/a0036089

Swann Jr., W. B., Gómez, Á., Buhrmester, M. D., López-Rodríguez, L., Jímenez, J., & Vázquez, A. (2014). Contemplating the Ultimate Sacrifice: Identity Fusion Channels Pro-Group Affect, Cognition, and Moral Decision Making. *Journal of Personality and Social Psychology, 106*(5), 713-727. doi:10.1037/a0035809

Swann Jr., W., & Buhrmester, M. D. (2015). Identity Fusion. *Current Directions in Psychological Science, 24*(1), 52-57. doi:10.1177/0963721414551363

Tausch, A., & Kluge, A. (2022). The Best Task Allocation Process Is to Decide On One's Own: Effects of the Allocation Agent in Human-Robot Interaction on Perceived Work Characteristics and Satisfaction. *Cognition, Technology & Work, 24*, 39-55. doi:10.1007/s10111-020-00656-7

Tenney, Y. J., Rogers, W. H., & Pew, R. W. (1995). *Pilot Opinions on High Level Flight Deck Automation Issues: Toward the Development of a Design Philosophy.* National Aeronautics and Space Administration, Langley Research Center, Hampton, VA.

Tetteh, E. G. (2006). *Human Factors Analysis of Commercial Aircraft Accidents in the United States: 1960-2000.* Purdue University, Department of Industrial Technology, West Lafayette, IN.

Todd, M. A., & Thomas, M. J. (2012, August). Flight Hours and Flight Crew Performance in Commercial Aviation. *Aviation, Space, and Environmental Medicine, 83*(8), 776-782. doi:10.3357/AESM.3271.2012

U.S. Navy. (1965, December 15). *NATOPS-Flight-Manual-Navy-Model-RF-4B-Aircraft.pdf.* Retrieved December 26, 2015, from f4phantom.com: http://www.f4phantom.com/docs/NATOPS-Flight-Manual-Navy-Model-RF-4B-Aircraft.pdf

U.S. Navy. (1973, May 1). *Queue | Navair 01-85adf-1 - Natops Flight Manual - A-6e - PDFCOFFEE.COM.* Retrieved January 16, 2022, from pdfcoffee.com: https://pdfcoffee.com/qdownload/navair-01-85adf-1-natops-flight-manual-a-6e-pdf-free.html

U.S. Navy. (1991, September 1). *Avialogs: Aviation Library - A1 - C2AHA-NFM-000 Natops Flight Manual C-2A Aircraft.* Retrieved December 29, 2015, from avialogs.com: https://www.avialogs.com/aircraft-g/grumman/item/4225-a1-c2aha-nfm-000-natops-flight-manual-c-2a-aircraft

U.S. Navy. (1999, September 30). *NAVAIR 01-E2AAB-1 Natops Flight Manual Navy Model E-2C Plus Aircraft.* Retrieved 12 30, 2015, from avialogs.com: https://www.avialogs.com/reader.php?jid=4234#p=1

U.S. Navy. (2000, February 15). *NATOPS-Flight-Manual-US-Navy-FA-18ABCD-McDonnell-Douglas-fighter-A1-F18AC-NFM-000-Chg-6-2000-BBS.pdf.* Retrieved December 29, 2015, from jasonblair.net: https://jasonblair.net/wp-content/uploads/2015/06/NATOPS-Flight-Manual-US-Navy-FA-18ABCD-McDonnell-Douglas-fighter-A1-F18AC-NFM-000-Chg-6-2000-BBS.pdf

U.S. Navy. (2000, August 15). *NAVAIR 01-85ADC-1 Natops Flight Manual Navy Model EA-6B Aircraft.* Retrieved December 29, 2015, from avialogs.com: https://www.avialogs.com/reader.php?jid=4240#p=1

U.S. Navy. (2001, August 1). *NAVAIR 01-F14AAP-1.* Retrieved December 29, 2015, from 3rd-wing.net: https://server.3rd-wing.net/public/Ked/natops%20F14B.pdf

U.S. Navy. (2004, January 15). *F14AAD-1.pdf.* Retrieved December 29, 2015, from publicintelligence.net: https://info.publicintelligence.net/F14AAD-1.pdf

U.S. Navy. (2005, January 7). *Navy & Marine Corps Mishap and Safety Investigation, Reporting, And Record Keeping Manual.* Retrieved October 17, 2020, from Navytribe.com: https://navytribe.com/wp-content/uploads/2015/11/opnavinst-5102-1d.pdf

U.S. Navy. (2008, March 15). *A1-AV8BB-NFM-000 Natops Flight Manual Navy Model AV-8B/TAV-8B 161573 And Up Aircraft.* Retrieved December 30, 2015, from publicintelligence.net: https://info.publicintelligence.net/AV-8B-000.pdf

U.S. Navy. (2008, September 15). *A1-F18EA-NFM-000 Natops Flight Manual Navy Model F/A-18E/F 165533 And Up Aircraft.* Retrieved December 29, 2015, from publicintelligence.net: https://info.publicintelligence.net/F18-EF-000.pdf

U.S. Navy. (2018, January 27). *F-14 Tomcat Pilot's Flight Operating Manual Vol. 1 - Unites States Navy - Google Books*. Retrieved from Google Books: https://books.google.com/books?id=J-8cAgAAQBAJ&lpg=PR4&ots=7XMcNHjUeN&dq=purpose%20of%20warnings%20notes%20and%20cautions%20in%20NATOPS&pg=PP4#v=onepage&q=purpose%20of%20warnings%20notes%20and%20cautions%20in%20NATOPS&f=false

Uhlarik, J., & Comerford, D. A. (2002). *A Review of Situational Awareness Literature Relevant to Pilot Surveillance Functions.* U.S. Department of Transportation, Office of Aerospace Medicine. Springfield, VA: National Technical Information Service.

Verdière, K. J., Roy, R. N., & Dehais, F. (2018, January 25). Detecting Pilot's Engagement Using fNIRS Connectivity Features in an Automated vs. Manual Landing Scenario. *Frontiers in Human Neuroscience*, 1-14. doi:10.3389/fnhum.2018.00006

Verschuur, W. L., & Hurts, K. (2008). Modeling Safe and Unsafe Driving Behaviour. *Accident Analysis & Prevention*, 644-656.

Victor, T. W., Tivesten, E., Gustavsson, P., Johansson, J., Sangberg, F., & Aust, M. L. (2018, December). Automation Expectation Mismatch: Incorrect Prediction Despite Eyes on threat and Hands on Wheel. *Human Factors, 60*(8), 1095-1116. doi:10.1177/0018720818788164

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician, 70*(2), 129-133. doi:10.1080/00031305.2016.1154108

Weaver, S. M., Roldan, S. M., Gonzalez, T. B., Balk, S. A., & Philips, B. H. (2022). The Effects of Vehicle Automation on Driver Engagement: The Case of Adaptive Cruise Control and Mind Wandering. *Human Factors, 64*(6), 1086-1098. doi:10. 1177/ 0018 7208 20974856

Weigmann, D., Faaborg, T., Boquet, A., Detwiler, C., Holcomb, K., & Shappell, S. (2005). *Human Error and General Aviation Accidents: A Comprehensive, Fine-Grained Analysis Using HFACS.* Federal Aviation Administration, Office of Aerospace Medicine. Washington, DC: Federal Aviation Administration.

Whitehouse, H. (2018). Dying For The Group: Towards A General Theory of Extreme Self-Sacrifice. *Behavioral and Brain Sciences*, 1-62. doi:10.1017/S0140525X18000249

Wickens, C. D., Clegg, B. A., Vieane, A. Z., & Sebok, A. L. (2015, August). Complacency and Automation Bias in the Use of Imperfect Automation. *Human Factors*, *57*(5), pp. 728-739.

Wickens, C. D., Hooey, B. L., Gore, B. F., Sebok, A., & Koenicke, C. S. (2009, October). Identifying Black Swans in NextGen: Predicting Human Performance in Off-Nominal Conditions. *Human Factors, 51*(5), 638-651. doi:10.1177/0018720809349709

Wickens, C. D., Sebok, A., Li, H., Sarter, N., & Gacy, A. M. (2015, September). Using Modeling and Simulation to Predict Operator Performance and Automation-Induced Complacency With Robotic Automation: A Case Study and Empirical Validation. *Human Factors, 57*(6), 959-975. doi:10.1177/0018720814566454

Wiener, E. L., & Curry, R. E. (1980). *Flight-Deck Automation: Promises and Problems.* National Aeronautic and Space Administration.

Wilson, J. R. (1998). *The Effect of Automation on the Frequency of Task Prioritization Errors on Commercial Aircraft Flight Decks: An ASRS Incident Report Study*.

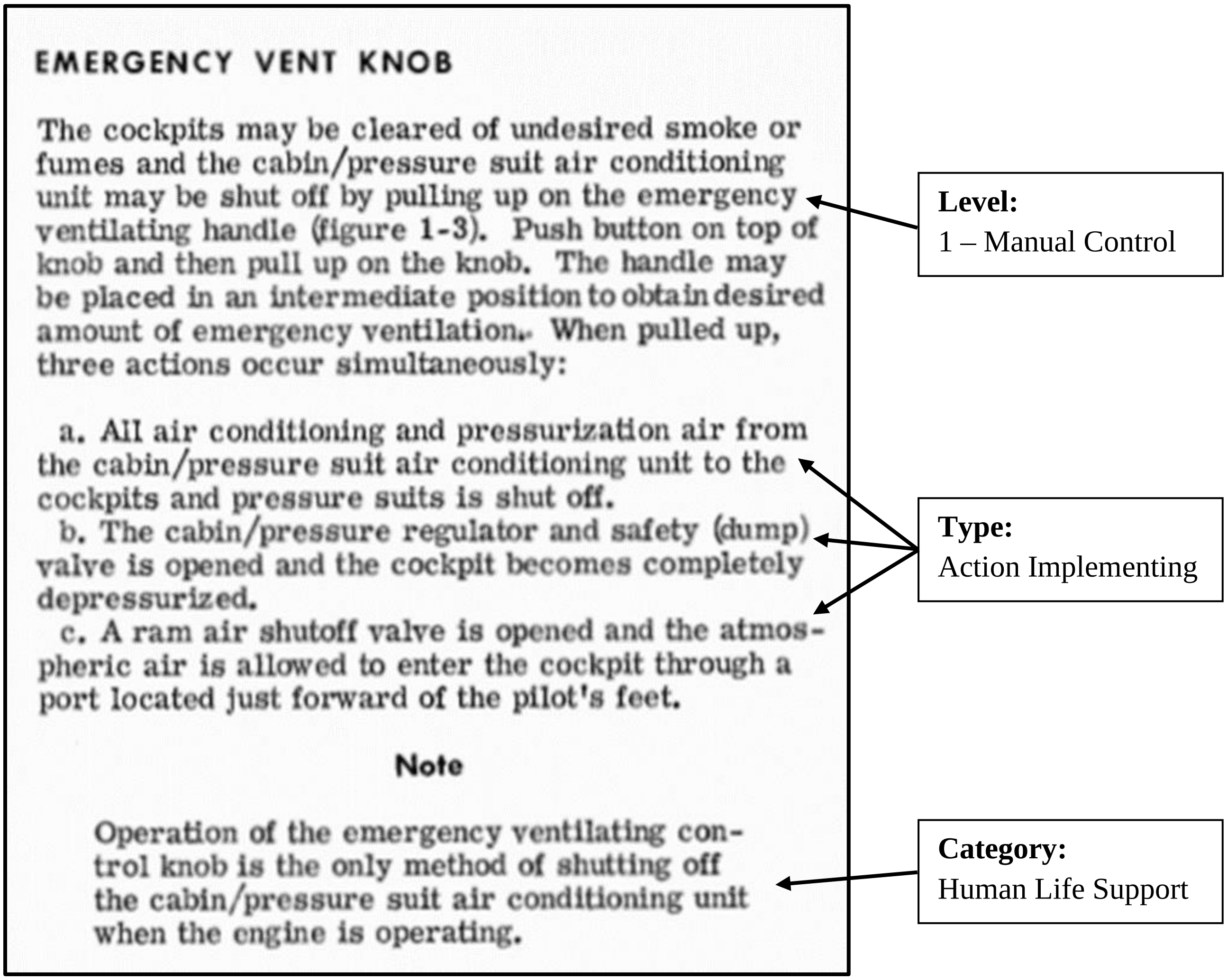
Wohleber, R. W., Matthews, G., Lin, J., Szalma, J. L., Calhoun, G. L., Funke, G. J., . . . Ruff, H. A. (2019, May). Vigilance and Automation Dependence in Operation of Multiple Unmanned Aerial Systems (UAS): A Simulation Study. *Human Factors, 61*(3), 488-505. doi:10.1177/0018720818799468

Wright, M. C., & Kaber, D. B. (2005, Spring). Effects of Automation of Information-Processing Functions on Teamwork. *Human Factors, 47*(1), 50-66.

Yamani, Y., & McCarley, J. S. (2016, May 3). Workload Capacity: A Response Time-Based Measure of Automation Dependence. *Human Factors, 58*(3), 462-471. doi:10.1177/0018720815621172

Yamani, Y., & McCarley, J. S. (2018, June). Effects of Task Difficulty and Display Format on Automation Usage Strategy: A Workload Capacity Analysis. *Human Factors, 60*(4), 527-537. doi:10.1177/0018720818759356

Appendix A. Example of Automation System Survey



EMERGENCY VENT KNOB

The cockpits may be cleared of undesired smoke or fumes and the cabin/pressure suit air conditioning unit may be shut off by pulling up on the emergency ventilating handle (figure 1-3). Push button on top of knob and then pull up on the knob. The handle may be placed in an intermediate position to obtain desired amount of emergency ventilation. When pulled up, three actions occur simultaneously:

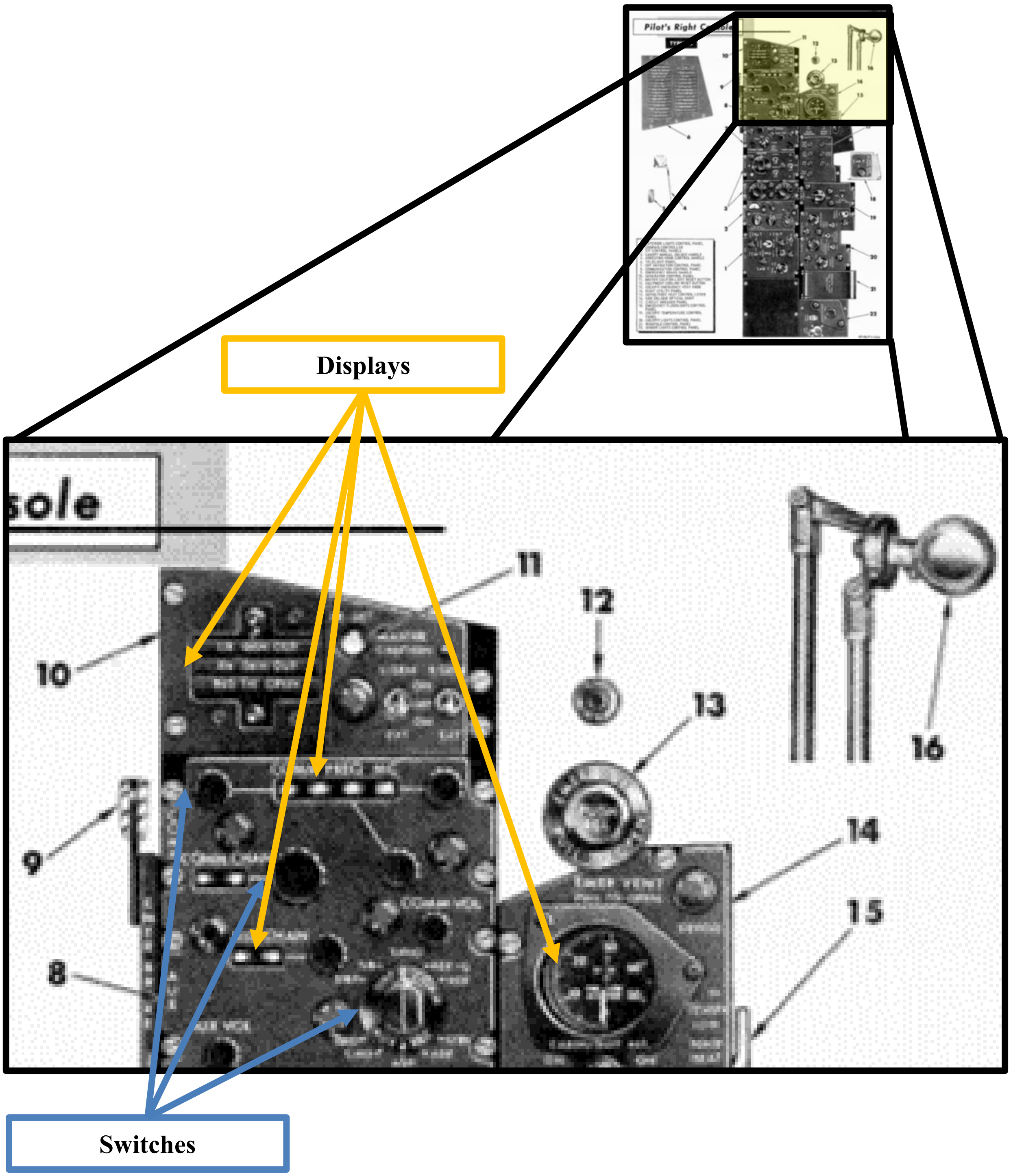
a. All air conditioning and pressurization air from the cabin/pressure suit air conditioning unit to the cockpits and pressure suits is shut off.
b. The cabin/pressure regulator and safety (dump) valve is opened and the cockpit becomes completely depressurized.
c. A ram air shutoff valve is opened and the atmospheric air is allowed to enter the cockpit through a port located just forward of the pilot's feet.

Note

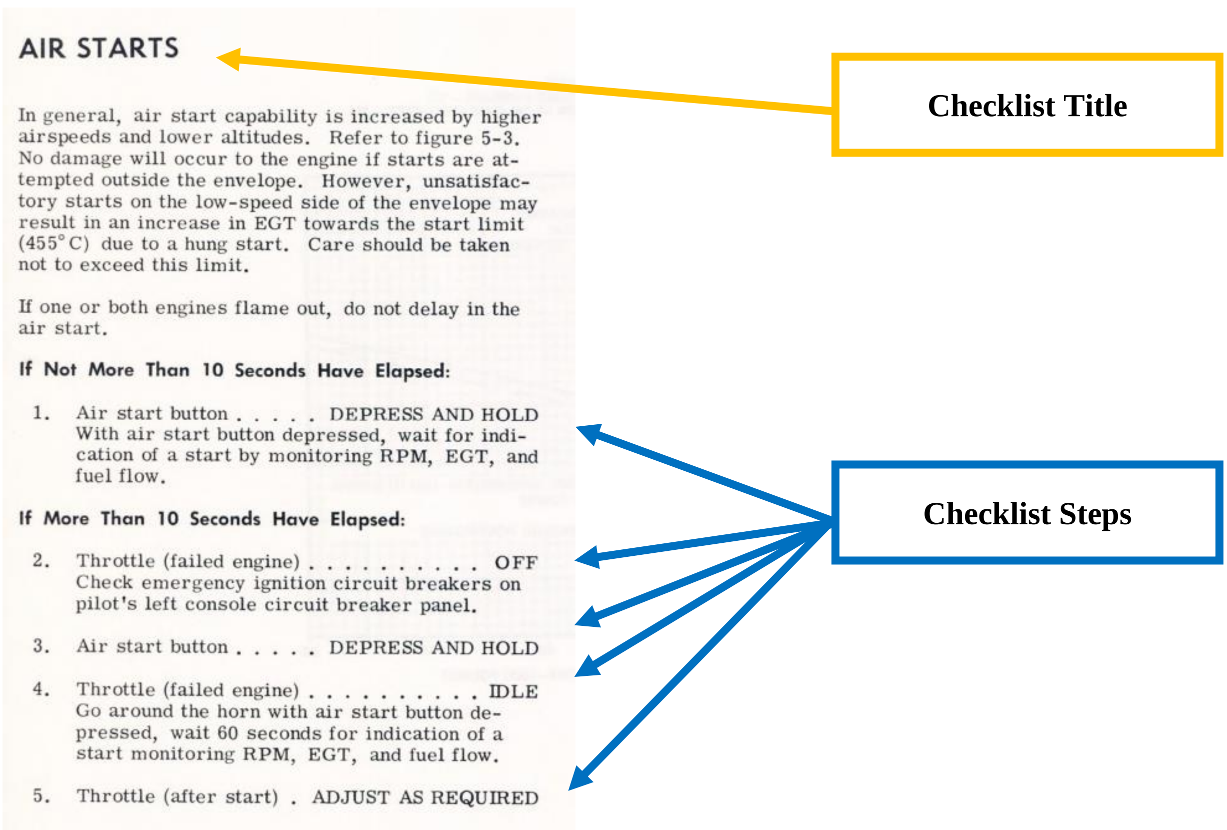
Operation of the emergency ventilating control knob is the only method of shutting off the cabin/pressure suit air conditioning unit when the engine is operating.

Figure A.1: F-4 Emergency Vent Knob Description.

Appendix B. Example of Cockpit Interface Survey



Appendix C. Example of Checklist Survey



AIR STARTS

In general, air start capability is increased by higher airspeeds and lower altitudes. Refer to figure 5-3. No damage will occur to the engine if starts are attempted outside the envelope. However, unsatisfactory starts on the low-speed side of the envelope may result in an increase in EGT towards the start limit (455° C) due to a hung start. Care should be taken not to exceed this limit.

If one or both engines flame out, do not delay in the air start.

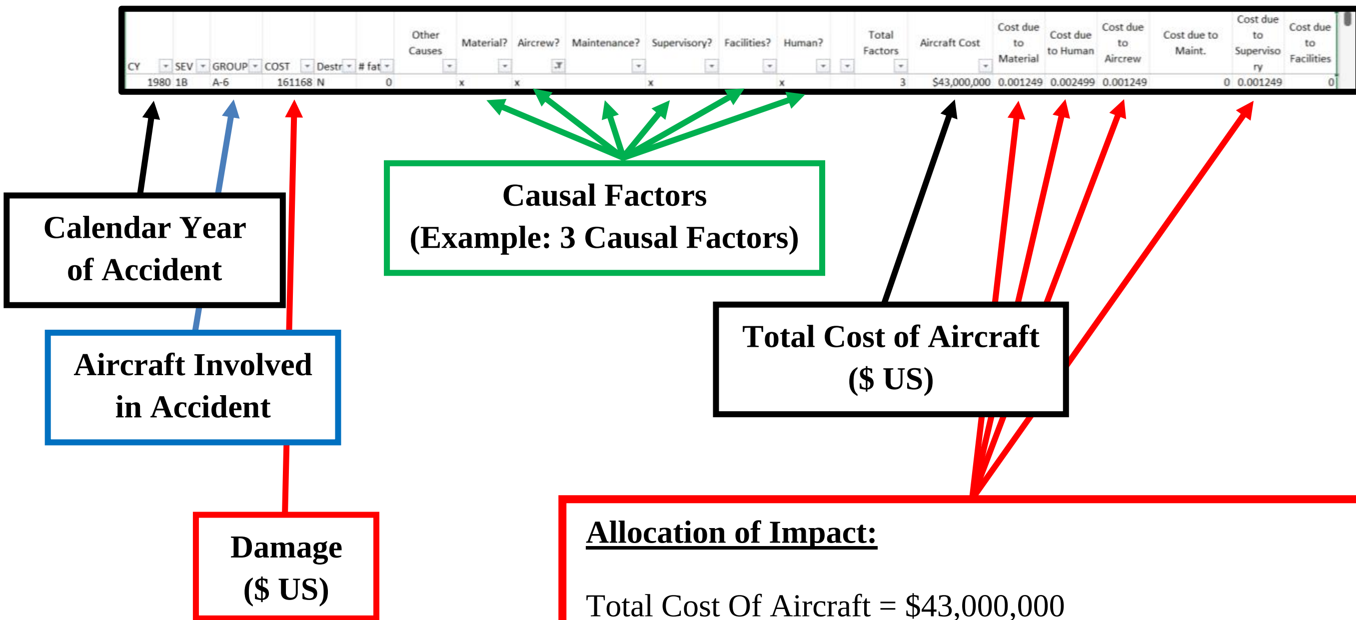
If Not More Than 10 Seconds Have Elapsed:

1. Air start button . . . . . DEPRESS AND HOLD
   With air start button depressed, wait for indication of a start by monitoring RPM, EGT, and fuel flow.

If More Than 10 Seconds Have Elapsed:

2. Throttle (failed engine) . . . . OFF
   Check emergency ignition circuit breakers on pilot's left console circuit breaker panel.
3. Air start button . . . . . DEPRESS AND HOLD
4. Throttle (failed engine) . . . . IDLE
   Go around the horn with air start button depressed, wait 60 seconds for indication of a start monitoring RPM, EGT, and fuel flow.
5. Throttle (after start) . ADJUST AS REQUIRED

Appendix D. Example of Allocation of Impact to Causal Factors



CY	SEV	GROUP	COST	Destr	# fat	Other Causes	Material?	Aircrew?	Maintenance?	Supervisory?	Facilities?	Human?		Total Factors	Aircraft Cost	Cost due to Material	Cost due to Human	Cost due to Aircrew	Cost due to Maint.	Cost due to Supervisory	Cost due to Facilities
1980	1B	A-6	161168	N	0		x	x		x		x		3	$43,000,000	0.001249	0.002499	0.001249	0	0.001249	0

Allocation of Impact:

Total Cost Of Aircraft = $43,000,000
Total Damage to Aircraft = $161,168
% Aircraft Damage = 0.3748%

Causal Factors = 3 (2 are subsets of Human Error)

- Human Error
  - Aircrew
  - Supervisory
- Material Failure

Allocation:

- Human Error = 0.2499% (Sum of Aircrew & Supervisory)
  - Aircrew: 0.1249%
  - Supervisory: 0.1249%
- Material Failure: 0.1249%

Appendix E. Measurement of Human Performance Tables

Table E.1: Rate of Accidents Caused By Human Error (HE-ALL).

A Fisher's Pairwise Comparison was used to evaluate the rate of accidents caused by any human error. Three statistically significant performance groups were identified with 5 of the 8 aircraft having performance that corresponded to more than one group. The AV-8, F-18, and E-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B	Group C
AV-8	11.6942	A		
A-6	5.2452	A	B	
F-14	3.8512	A	B	
EA-6	3.3114	A	B	
F-4	1.6777	A	B	C
F-18	1.0939		B	
C-2	-3.8763		B	C
E-2	-9.1641			C

Table E.2: Rate of Accidents Caused By Human Error with Material Failure (HE-MF).

A Fisher's Pairwise Comparison was used to evaluate the rate of accidents caused by any human error associated with material failure. Four statistically significant performance groups were identified with 6 of the 8 aircraft having performance that corresponded to more than one group. The AV-8 and E-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B	Group C	Group D
AV-8	3.30335	A			
C-2	1.46774	A	B		
F-14	1.07911		B	C	
F-4	0.96024	A	B	C	D
EA-6	0.39134		B	C	D
A-6	0.33122		B	C	D
F-18	-0.7522			C	D
E-2	-1.58003				D

Table E.3: Rate of Accidents Caused Only By Human Error (HE-O).

A Fisher's Pairwise Comparison was used to evaluate the rate of accidents caused only by human error. Three statistically significant performance groups were identified with 6 of the 8 aircraft having performance that corresponded to more than one group. The AV-8 and E-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B	Group C
AV-8	8.39083	A		
A-6	4.914	A	B	
EA-6	2.92008	A	B	
F-14	2.77214	A	B	
F-18	1.84607	A	B	
F-4	0.71744	A	B	C
C-2	-5.34403		B	C
E-2	-7.58408			C

Table E.4: Rate of Accidents Caused All or In Part By Aircrew Error.

A Fisher's Pairwise Comparison was used to evaluate the rate of accidents caused by any aircrew error. Three statistically significant performance groups were identified with 4 of the 8 aircraft having performance that corresponded to more than one group. The AV-8, F-14, C-2, and E-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B	Group C
AV-8	8.30695	A		
F-14	2.23498		B	
C-2	1.45016		B	
F-4	0.65407		B	C
F-18	0.43809		B	C
A-6	-1.16963		B	C
EA-6	-1.51881		B	C
E-2	-3.13371			C

Table E.5: Rate of Accidents Caused By Aircrew Error and Material Failure.

A Fisher's Pairwise Comparison was used to evaluate the rate of accidents caused by aircrew error associated with material failure. Three statistically significant performance groups were identified with 5 of the 8 aircraft having performance that corresponded to more than one group. The AV-8, F-14, and E-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B	Group C
F-14	1.07822	A		
AV-8	1.01903	A		
C-2	0.6232	A	B	
F-4	0.39458	A	B	C
A-6	0.06611	A	B	C
EA-6	-0.1292	A	B	C
F-18	-0.49328		B	C
E-2	-0.84557			C

Table E.6: Rate of Accidents Caused Only By Aircrew Error.

A Fisher's Pairwise Comparison was used to evaluate the rate of accidents caused only by aircrew error. Two statistically significant performance groups were identified. The AV-8 was the only aircraft that did not belong to Group B indicating a statistically significant different in accident rate compared to the other aircraft.

Aircraft	Relative Mean	Group A	Group B
AV-8	7.28791	A	
F-14	1.15676		B
F-18	0.93136		B
C-2	0.82696		B
F-4	0.25949		B
A-6	-1.23574		B
EA-6	-1.38961		B
E-2	-2.28814		B

Table E.7: Accident Rates Ranking Summary.

An ordinal ranking of accident rates by human error category indicates some consistency in relative position of the aircraft.

RANK	HE-ALL	HE-MF	HE-O	AE-ALL	AE-MF	AE-O
1	AV-8	AV-8	AV-8	AV-8	F-14	AV-8
2	A-6	C-2	A-6	F-14	AV-8	F-14
3	F-14	F-14	EA-6	C-2	C-2	F-18
4	EA-6	F-4	F-14	F-4	F-4	C-2
5	F-4	EA-6	F-18	F-18	A-6	F-4
6	F-18	A-6	F-4	A-6	EA-6	A-6
7	C-2	F-18	C-2	EA-6	F-18	EA-6
8	E-2	E-2	E-2	E-2	E-2	E-2

Table E.8: Aircraft Damage From Accidents Caused All or In Part By Human Error.

A Fisher's Pairwise Comparison was used to evaluate the aircraft damage from accidents caused by any human error. Five statistically significant performance groups were identified with 6 of the 8 aircraft having performance that corresponded to more than one group. The AV-8 and C-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B	Group C	Group D	Group E
F-4	9.55%	A	B			
AV-8	7.07%	A				
F-14	4.11%	A	B			
F-18	0.53%		B	C		
EA-6	-4.87%			C	D	
A-6	-6.06%			C	D	E
E-2	-10.88%				D	E
C-2	-11.61%					E

Table E.9: Aircraft Damage From Accidents Caused By Human Error and Material Failure.

A Fisher's Pairwise Comparison was used to evaluate the aircraft damage from accidents caused by human error associated with material failure. Three statistically significant performance groups were identified with 2 of the 8 aircraft having performance that corresponded to more than one group. The F-4, F-14, AV-8, F-18, E-2, and C-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B	Group C
F-4	6.81%	A		
F-14	6.19%	A		
AV-8	5.08%	A		
F-18	-4.06%		B	
A-6	-9.36%		B	C
EA-6	-10.66%		B	C
E-2	-13.32%			C
C-2	-14.37%			C

Table E.10: Aircraft Damage From Accidents Caused Only By Human Error.

A Fisher's Pairwise Comparison was used to evaluate the aircraft damage from accidents caused only by human error. Three statistically significant performance groups were identified with 2 of the 8 aircraft having performance that corresponded to more than one group. The F-4, AV-8, F-18, EA-6, C-2, and E-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B	Group C
F-4	6.03%	A		
AV-8	5.73%	A		
F-18	3.19%	A		
F-14	1.07%	A	B	
EA-6	-11.41%			C
A-6	-12.02%		B	C
C-2	-19.09%			C
E-2	-19.28%			C

Table E.11: Aircraft Damage From Accidents Caused All or In Part By Aircrew Error.

A Fisher's Pairwise Comparison was used to evaluate the aircraft damage from accidents caused by any aircrew error. Three statistically significant performance groups were identified with no aircraft having performance that corresponded to more than one group. Of note, the F-4 was the only aircraft in Group A indicating statistically significant difference in performance than the other aircraft.

Aircraft	Relative Mean	Group A	Group B	Group C
F-4	9.00%	A		
AV-8	1.25%		B	
F-14	1.11%		B	
F-18	0.81%		B	
A-6	-4.24%			C
EA-6	-4.70%			C
E-2	-5.60%			C
C-2	-5.64%			C

Table E.12: Aircraft Damage From Accidents Caused By Aircrew Error and Material Failure.

A Fisher's Pairwise Comparison was used to evaluate the aircraft damage from accidents caused by aircrew error associated with material failure. Four statistically significant performance groups were identified with 4 of the 8 aircraft having performance that corresponded to more than one group. The F-4, AV-8, F-18, E-2, and C-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B	Group C	Group D
F-4	6.56%	A			
F-14	2.49%	A	B		
F-18	-1.21%		B	C	
AV-8	-1.47%			C	
A-6	-4.04%			C	D
EA-6	-4.45%			C	D
E-2	-5.37%				D
C-2	-5.51%				D

Table E.13: Aircraft Damage From Accidents Caused Only By Aircrew Error.

A Fisher's Pairwise Comparison was used to evaluate the aircraft damage from accidents caused only by aircrew error. Three statistically significant performance groups were identified with only 1 of the 8 aircraft having performance that corresponded to more than one group. The F-18 belonged to both Group A and Group B.

Aircraft	Relative Mean	Group A	Group B	Group C
F-4	9.50%	A		
F-18	2.93%	A	B	
AV-8	0.81%		B	
F-14	-0.96%		B	
A-6	-9.31%			C
EA-6	-9.44%			C
E-2	-10.82%			C
C-2	-10.82%			C

Table E.14: Accident Costs Ranking Summary.

An ordinal ranking of accident costs by human error category indicated differences between the six combinations of human error, but relative ranking of the aircraft remained relatively consistent.

RANK	HE-ALL	HE-MF	HE-O	AE-ALL	AE-MF	AE-O
1	F-4	F-4	F-4	F-4	F-4	F-4
2	AV-8	F-14	AV-8	AV-8	F-14	F-18
3	F-14	AV-8	F-18	F-14	F-18	AV-8
4	F-18	F-18	F-14	F-18	AV-8	F-14
5	EA-6	A-6	EA-6	A-6	A-6	A-6
6	A-6	EA-6	A-6	EA-6	EA-6	EA-6
7	E-2	E-2	C-2	E-2	E-2	E-2
8	C-2	C-2	E-2	C-2	C-2	C-2

Table E.15: Crew Fatalities From Accidents Caused All or In Part By Human Error.

A Fisher's Pairwise Comparison was used to evaluate the crew fatalities in accidents caused by any human error. Two statistically significant performance groups were identified with 5 of the 8 aircraft having performance that corresponded to more than one group. The EA-6, E-2, and C-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B
EA-6	11.76%	A	
E-2	9.24%	A	
F-14	0.36%	A	B
F-18	-2.24%	A	B
AV-8	-2.58%	A	B
F-4	-2.90%	A	B
A-6	-3.20%	A	B
C-2	-8.38%		B

Table E.16: Crew Fatalities From Accidents Caused By Human Error and Material Failure.

A Fisher's Pairwise Comparison was used to evaluate the crew fatalities in accidents caused by human error associated with material failure. Two statistically significant performance groups were identified with 5 of the 8 aircraft having performance that corresponded to more than one group. The E-2, C-2, and EA-6 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B
E-2	9.36%	A	
F-18	0.45%	A	B
AV-8	-0.88%	A	B
F-4	-1.32%	A	B
A-6	-1.45%	A	B
F-14	-1.86%	A	B
C-2	-4.86%		B
EA-6	-4.86%		B

Table E.17: Crew Fatalities From Accidents Caused Only By Human Error.

A Fisher's Pairwise Comparison was used to evaluate the crew fatalities in accidents caused only by human error. Two statistically significant performance groups were identified with 4 of the 8 aircraft having performance that corresponded to more than one group. The EA-6, AV-8, E-2, and C-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B
EA-6	15.18%	A	
F-14	-2.78%	A	B
F-18	-5.04%	A	B
A-6	-7.48%	A	B
AV-8	-7.65%		B
E-2	-10.07%		B
F-4	-10.68%	A	B
C-2	-15.13%		B

Table E.18: Crew Fatalities From Accidents Caused All or In Part By Aircrew Error.

A Fisher's Pairwise Comparison was used to evaluate the crew fatalities in accidents caused by any aircrew error. Two statistically significant performance groups were identified with 6 of the 8 aircraft having performance that corresponded to more than one group. The E-2 and C-2 only belonged to one group each.

Aircraft	Relative Mean	Group A	Group B
E-2	2.27%	A	
EA-6	0.24%	A	B
F-18	-0.09%	A	B
F-14	-0.18%	A	B
AV-8	-0.31%	A	B
A-6	-0.96%	A	B
F-4	-1.35%	A	B
C-2	-4.32%		B

Table E.19: Accident Fatalities Ranking Summary.

An ordinal ranking of accident fatalities by human error category indicated significant shifts. Of note, the EA-6 shifted from having the highest occurrence of fatalities in accidents due to any human error, but the lowest occurrence when assessed only with accidents due to human error and material failure while the C-2 remained in the bottom two rankings for the least quantity of fatalities.

RANK	HE-ALL	HE-MF	HE-O	AE-ALL	AE-MF	AE-O
1	EA-6	E-2	EA-6	E-2	F-4	F-18
2	E-2	F-18	F-14	EA-6	F-18	E-2
3	F-14	AV-8	F-18	F-18	F-14	EA-6
4	F-18	F-4	A-6	F-14	E-2	AV-8
5	AV-8	A-6	AV-8	AV-8	A-6	F-14
6	F-4	F-14	E-2	A-6	AV-8	F-4
7	A-6	C-2	F-4	F-4	C-2	A-6
8	C-2	EA-6	C-2	C-2	EA-6	C-2

Appendix F. Measurement of Human Performance ANOVA Tables

Accident Rate – Human Error-All (HE-ALL)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	8846	1263.7	3.85	0.001
YEAR	33	9619	291.5	0.89	0.646
Error	185	60710	328.2		
Total	225	78980			

Fits and Diagnostics for Unusual Observations

Obs	HEA-R-T	Fit	Resid	Std Resid	
47	143.43	38.49	104.94	6.57	R
53	-30.46	9.11	-39.58	-2.37	R
54	-63.80	-20.86	-42.94	-2.57	R
77	54.08	4.10	49.98	3.07	R
87	-33.28	3.82	-37.10	-2.23	R
122	22.23	-13.68	35.91	2.15	R
153	42.71	-0.41	43.13	2.70	R
181	47.60	7.51	40.09	2.40	R
182	127.02	14.08	112.93	6.77	R

R Large residual

Accident Rate – Human Error with Material Failure (HE-MF)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	507.8	72.55	4.23	0.000
YEAR	33	335.3	10.16	0.59	0.962
Error	185	3172.9	17.15		
Total	225	4022.0			

Fits and Diagnostics for Unusual Observations

Obs	HEM-R-T	Fit	Resid	Std Resid	
25	11.45	3.66	7.79	2.04	R
32	11.02	2.64	8.37	2.22	R
43	16.23	5.45	10.77	2.89	R
49	-3.15	5.86	-9.00	-2.47	R
52	19.64	1.48	18.16	4.76	R
60	10.19	1.30	8.89	2.33	R
75	-6.13	3.53	-9.66	-2.59	R
78	13.63	2.44	11.19	3.00	R
83	17.63	4.02	13.61	3.73	R
177	16.39	3.14	13.25	3.57	R
179	-4.49	3.23	-7.72	-2.08	R
180	-7.36	2.06	-9.42	-2.54	R

R Large residual

Accident Rate – Human Error-Only (HE-O)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	6179	882.7	2.82	0.008
YEAR	33	9473	287.1	0.92	0.601
Error	185	57894	312.9		
Total	225	73392			

Fits and Diagnostics for Unusual Observations

Obs	HEO-R-T	Fit	Resid	Std Resid	
47	145.23	36.56	108.67	6.97	R
53	-27.47	9.78	-37.25	-2.29	R
54	-61.44	-20.52	-40.92	-2.51	R
77	47.68	0.48	47.21	2.97	R
87	-30.28	7.54	-37.82	-2.32	R
122	24.60	-12.25	36.85	2.26	R
149	-2.03	31.09	-33.12	-2.12	R
153	44.33	0.95	43.38	2.78	R
181	51.82	8.25	43.57	2.68	R
182	130.01	16.97	113.04	6.94	R
210	-1.36	30.02	-31.38	-2.01	R

R Large residual

Accident Rate – Aircrew Error-All (AE-ALL)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	2746	392.24	5.54	0.000
YEAR	33	2443	74.02	1.05	0.410
Error	185	13103	70.83		
Total	225	18283			

Fits and Diagnostics for Unusual Observations

Obs	AEA-R-T	Fit	Resid	Std Resid	
2	-7.06	9.87	-16.94	-2.22	R
19	39.63	19.35	20.28	2.62	R
21	27.73	8.00	19.73	2.55	R
47	71.87	21.35	50.52	6.81	R
52	32.25	2.81	29.44	3.80	R
81	-3.29	14.49	-17.78	-2.40	R
87	-10.08	7.91	-17.99	-2.32	R
121	-6.17	9.53	-15.70	-2.03	R
181	-15.47	1.80	-17.27	-2.23	R
182	61.86	11.48	50.38	6.50	R
225	19.53	2.46	17.07	2.26	R

R Large residual

Accident Rate – Aircrew Error with Material Failure (AE-MF)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	106.7	15.242	1.77	0.096
YEAR	33	132.5	4.014	0.47	0.995
Error	185	1596.7	8.631		
Total	225	1841.1			

Fits and Diagnostics for Unusual Observations

Obs	AEM-R-T	Fit	Resid	Std Resid	
27	6.91	1.27	5.64	2.08	R
43	9.36	1.97	7.39	2.80	R
50	6.85	0.79	6.06	2.34	R
52	20.49	1.50	18.99	7.02	R
72	8.24	1.82	6.41	2.43	R
73	6.83	0.99	5.84	2.21	R
75	-4.16	2.78	-6.94	-2.63	R
78	7.20	0.73	6.47	2.45	R
114	8.32	-0.07	8.38	3.24	R
143	8.42	2.03	6.38	2.42	R
172	8.59	1.61	6.99	2.66	R
225	11.00	1.96	9.04	3.43	R

R Large residual

Accident Rate – Aircrew Error-Only (AE-O)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	1988	283.96	4.59	0.000
YEAR	33	2562	77.62	1.26	0.175
Error	185	11438	61.83		
Total	225	15973			

Fits and Diagnostics for Unusual Observations

Obs	AEO-R-T	Fit	Resid	Std Resid	
2	-5.89	11.12	-17.01	-2.38	R
19	41.81	19.64	22.17	3.06	R
21	22.70	7.12	15.57	2.15	R
47	72.47	20.81	51.65	7.46	R
77	17.53	1.56	15.97	2.26	R
81	-2.69	14.35	-17.04	-2.46	R
87	-7.90	10.06	-17.96	-2.48	R
121	-4.00	10.96	-14.96	-2.07	R
149	-2.69	12.13	-14.83	-2.14	R
155	-1.92	13.51	-15.43	-2.14	R
182	64.04	13.28	50.76	7.01	R
210	-0.83	14.45	-15.28	-2.21	R

R Large residual

Accident Cost – Human Error-All (HE-ALL)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	1.1787	0.168382	9.55	0.000
YEAR	33	0.2793	0.008462	0.48	0.993
Error	185	3.2633	0.017640		
Total	225	4.7148			

Fits and Diagnostics for Unusual Observations

Obs	HEA-C-T	Fit	Resid	Std Resid	
37	0.4127	0.0824	0.3303	2.77	R
45	0.4636	0.1284	0.3352	2.86	R
46	0.4907	0.0741	0.4166	3.56	R
51	-0.1258	0.2119	-0.3377	-2.89	R
153	0.9842	0.0926	0.8917	7.62	R
167	0.3241	0.0508	0.2733	2.27	R
169	0.3687	0.0201	0.3486	2.89	R
181	0.2677	0.0215	0.2462	2.01	R
225	0.7594	0.1436	0.6158	5.18	R

R Large residual

Accident Cost – Human Error with Material Failure (HE-MF)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	1.500	0.21424	10.80	0.000
YEAR	33	2.581	0.07822	3.94	0.000
Error	185	3.669	0.01983		
Total	225	7.880			

Fits and Diagnostics for Unusual Observations

Obs	HEM-C-T	Fit	Resid	Std Resid	
24	0.3570	0.0824	0.2746	2.12	R
37	0.5015	0.1560	0.3455	2.73	R
45	0.5124	0.1424	0.3700	2.98	R
51	-0.9257	-0.6151	-0.3107	-2.50	R
178	0.8235	0.1138	0.7097	5.63	R
214	0.0000	-0.7065	0.7065	5.69	R
219	0.2898	0.0372	0.2526	2.00	R
225	0.7522	0.1636	0.5885	4.67	R

R Large residual

Accident Cost – Human Error-Only (HE-O)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	2.380	0.34006	7.21	0.000
YEAR	33	1.133	0.03433	0.73	0.859
Error	185	8.720	0.04714		
Total	225	12.260			

Fits and Diagnostics for Unusual Observations

Obs	HEO-C-T	Fit	Resid	Std Resid	
37	0.6603	0.0612	0.5991	3.07	R
41	0.5637	0.0933	0.4704	2.41	R
46	1.5708	0.2057	1.3651	7.14	R
148	-0.3838	0.0344	-0.4182	-2.19	R
153	1.0041	0.0800	0.9241	4.83	R
169	0.9704	-0.0434	1.0138	5.14	R
225	0.7245	0.1303	0.5942	3.06	R

R Large residual

Accident Cost – Aircrew Error-All (AE-ALL)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	0.34375	0.049108	11.13	0.000
YEAR	33	0.08863	0.002686	0.61	0.954
Error	185	0.81638	0.004413		
Total	225	1.26432			

Fits and Diagnostics for Unusual Observations

Obs	AEA-C-T	Fit	Resid	Std Resid	
35	0.1430	0.0221	0.1209	2.02	R
39	0.2492	0.0454	0.2038	3.41	R
41	0.1624	0.0358	0.1266	2.12	R
46	0.1518	-0.0058	0.1577	2.69	R
50	0.1627	0.0259	0.1368	2.34	R
169	0.2307	0.0086	0.2221	3.68	R
223	-0.0502	0.0767	-0.1269	-2.13	R
224	-0.0937	0.0673	-0.1610	-2.71	R
225	0.6542	0.1496	0.5045	8.48	R
226	-0.0654	0.0632	-0.1287	-2.16	R

R Large residual

Accident Cost – Aircrew Error with Material Failure (AE-MF)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	0.2390	0.034145	6.56	0.000
YEAR	33	0.6720	0.020363	3.91	0.000
Error	185	0.9636	0.005209		
Total	225	1.9029			

Fits and Diagnostics for Unusual Observations

Obs	AEM-C-T	Fit	Resid	Std Resid	
24	0.1875	0.0079	0.1796	2.70	R
50	0.1887	-0.0416	0.2303	3.62	R
126	0.1230	-0.0219	0.1449	2.18	R
169	0.2145	0.0486	0.1659	2.53	R
178	0.4582	0.0687	0.3895	6.02	R
214	0.0000	-0.3464	0.3464	5.45	R
219	0.2862	0.0322	0.2540	3.93	R
220	0.2460	0.0399	0.2061	3.19	R
221	-0.0513	0.0882	-0.1395	-2.16	R
225	0.3887	0.1144	0.2742	4.24	R

R Large residual

Accident Cost – Aircrew Error-Only (AE-O)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	0.9305	0.13293	9.04	0.000
YEAR	33	0.3381	0.01025	0.70	0.891
Error	185	2.7216	0.01471		
Total	225	4.0547			

Fits and Diagnostics for Unusual Observations

Obs	AEO-C-T	Fit	Resid	Std Resid	
39	0.2876	0.0198	0.2679	2.46	R
41	0.2853	0.0405	0.2448	2.24	R
46	0.7067	0.0380	0.6687	6.26	R
169	0.5325	-0.0392	0.5716	5.19	R
190	0.3465	-0.0006	0.3471	3.11	R
224	-0.1772	0.0651	-0.2423	-2.23	R
225	0.9040	0.1905	0.7134	6.57	R

R Large residual

Accident Fatalities – Human Error-All (HE-ALL)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	1.045	0.14924	1.37	0.222
YEAR	33	1.649	0.04998	0.46	0.995
Error	185	20.214	0.10926		
Total	225	22.897			

Fits and Diagnostics for Unusual Observations

Obs	HEA-F-T	Fit	Resid	Std Resid	
13	-0.251	0.350	-0.601	-2.02	R
98	3.082	0.474	2.608	8.65	R
99	1.444	0.212	1.231	4.09	R
113	0.930	0.213	0.717	2.46	R
129	0.852	0.155	0.697	2.29	R
138	2.052	0.117	1.936	6.51	R
153	1.375	0.277	1.098	3.77	R

R Large residual

Accident Fatalities – Human Error with Material Failure (HE-MF)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	0.4697	0.06710	1.12	0.350
YEAR	33	0.9718	0.02945	0.49	0.991
Error	185	11.0553	0.05976		
Total	225	12.4945			

Fits and Diagnostics for Unusual Observations

Obs	HEM-F-T	Fit	Resid	Std Resid	
4	0.433	-0.012	0.445	2.01	R
13	-0.198	0.307	-0.505	-2.30	R
37	0.611	0.051	0.560	2.55	R
64	-0.198	0.273	-0.471	-2.11	R
98	3.135	0.415	2.720	12.21	R
113	0.857	0.152	0.705	3.27	R
132	-0.198	0.273	-0.471	-2.11	R
166	-0.198	0.303	-0.501	-2.26	R

R Large residual

Accident Fatalities – Human Error-Only (HE-O)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	1.901	0.27161	1.50	0.168
YEAR	33	2.721	0.08246	0.46	0.995
Error	185	33.392	0.18050		
Total	225	37.962			

Fits and Diagnostics for Unusual Observations

Obs	HEO-F-T	Fit	Resid	Std Resid	
36	-1.438	-0.593	-0.845	-2.21	R
60	0.711	-0.105	0.815	2.08	R
70	-1.438	-0.668	-0.770	-2.02	R
99	2.118	0.085	2.033	5.25	R
104	-1.438	-0.617	-0.821	-2.15	R
136	1.076	0.191	0.884	2.28	R
138	3.563	-0.364	3.927	10.28	R
153	1.350	0.347	1.003	2.68	R
170	0.826	0.012	0.814	2.11	R
172	-1.438	-0.544	-0.893	-2.35	R

R Large residual

Accident Fatalities – Aircrew Error-All (AE-ALL)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	0.07876	0.01125	0.69	0.677
Error	218	3.53200	0.01620		
Total	225	3.61076			

Accident Fatalities – Aircrew Error with Material Failure (AE-MF)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	0.01355	0.001936	1.00	0.434
Error	218	0.42322	0.001941		
Total	225	0.43677			

Accident Fatalities – Aircrew Error-Only (AE-O)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Aircraft	7	0.1677	0.02395	0.67	0.697
Error	218	7.7903	0.03574		
Total	225	7.9580			