TYPE OF AUTOMATION FAILURE: THE EFFECTS ON TRUST AND RELIANCE IN AUTOMATION

A Thesis
Presented to
The Academic Faculty

by

Jason D. Johnson

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology
December 2004

TYPE OF AUTOMATION FAILURE: THE EFFECTS
ON TRUST AND RELIANCE IN AUTOMATION

Approved by:

Dr. Arthur D. Fisk, Advisor

Dr. Wendy A. Rogers

Dr. Gregory M. Corso

November 18. 2004

ACKNOWLEDGEMENTS

I want to extend a very special and heartfelt thank you to several people without whom this study would not have been possible. First and foremost, I thank my wife, Shayla, and my children, Aundreaniah, Isaiah, and Nathaniel. Your support and understanding during this past year and a half were paramount to the successful completion of this degree. It is as much yours as it is mine. Thank you and I love you all. Also, the members of the Human Factors and Aging Lab at Georgia Tech, your support has been invaluable. Specifically, I thank Dan and Wendy for taking a chance on someone who had to finish in 18 months. Most others were not willing to take that chance and, for you faith in me, I thank you very much. Neta and Julian, I cannot say enough about how integral your support was to completing the study on time. I am forever in your debt. Lastly, I thank the Air Force for giving me the opportunity to obtain my master's degree and Deere & Company for their financial contributions to this study.

TABLE OF CONTENTS

Acknowledgements iii

List of Tables vi

List of Figures vii

Summary iv

Introduction 1

  Type of Automation Error 4

  Aging and Automation 6

  Summary of Proposed Research 7

Method 9

  Participants 9

  Design 9

  Procedure 10

  Apparatus 12

    Engine Task 15

    Radar Task 20

    Questionnaires 20

    Point Scale 21

Results 22

  Task Performance 22

  Subjective Trust 26

  Perceived Reliability 28

Reliance 31

Perceived Workload 32

Discussion 34

Appendix A: Survey Instruments and Background Documents 40

References 62

LIST OF TABLES

Table 1 Description of Participants (e.g., Age, Abilities, and Education Level).... 9

Table 2 Location and Timing of Engine Events for each Experimental Condition........ 19

Table 3 Spearman Correlations (All Participants)........ 22

Table 4A Pairwise Comparisons of Task Performance by Experimental Condition (Younger Adults)........ 26

Table 4B Pairwise Comparisons of Task Performance by Experimental Condition (Older Adults)........ 26

Table 5 One Sample t-test Testing the Difference between Perceived and Actual Reliability........ 30

Table 6A Spearman Correlations (Younger Adults)........ 35

Table 6B Spearman Correlations (Older Adults)........ 35

LIST OF FIGURES

Figure 1 Types of Automation Malfunctions Based on Signal Detection Theory....4

Figure 2 Sanchez, Fisk, and Rogers (2004) Simulation Apparatus 13

Figure 3 Cockpit Simulator Apparatus for Present Study....14

Figure 4 Cockpit Simulator Engine Gauges....15

Figure 5A Pictorial Representation of Engine Events for Equal False Alarms and Misses Condition 18

Figure 5B Pictorial Representation of Engine Events for Majority False Alarms Condition 18

Figure 5C Pictorial Representation of Engine Events for Majority Misses Condition 18

Figure 6 Mean Task Performance for Younger and Older Adults by Experimental Condition....23

Figure 7A Mean Task Performance for Younger and Older Adults in the Equal False Alarms and Misses Condition by Experimental Block 24

Figure 7B Mean Task Performance for Younger and Older Adults in the Majority False Alarms Condition by Experimental Block 24

Figure 7C Mean Task Performance for Younger and Older Adults in the Majority Misses Condition by Experimental Block 25

Figure 8 Mean Subjective Trust Ratings for Younger and Older Adults by Experimental Condition....27

Figure 9 Mean Perceive Reliability (Scale) for Younger and Older Adults by Experimental Condition....29

Figure 10 Mean Perceive Reliability (Percentage) for Younger and Older Adults by Experimental Condition 31

Figure 11 Non-reliance on the Engine Status Indicator for Younger and Older Adults by experimental condition....32

Figure 12 The mean perceived workload (NASA-TLX) for younger and older adults by experimental condition..33

SUMMARY

Past automation research has focused primarily on machine-related factors (e.g., automation reliability) and human-related factors (e.g., accountability). Other machine-related factors such as type of automation errors, misses or false alarms, have been noticeably overlooked. These two automation errors correspond to potential operator errors, omission (misses) and commission (false alarms), which have proven to directly affect operator's trust in automation. This research examined how automation-error-type affects operator trust and reliance in and perceived reliability of automated decision aids. This present research confirmed that perceived reliability is often lower than actual system reliability and that false alarms significantly reduced operator trust in the automation more so than do misses. In addition, this study found that there does not appear to be an effect on the level of subjective trust within each experimental condition (i.e., type of automation error) based on age. There does, however, appear to be a significant difference in the reliance on automation between older and younger adult participants attributed to differences in perceived workload.

TYPE OF AUTOMATION FAILURE: THE EFFECTS
ON TRUST AND RELIANCE IN AUTOMATION

As the human race continues to progress, so does the sophistication of the systems we use. From military command and control systems and aircraft to nuclear power plants and automobile assembly lines, people encounter automated systems on a daily basis. Purely mechanical systems have been mostly replaced by computers and circuit boards allowing automated systems to perform tasks at which human beings are historically poor performers, such as monitoring during unengaging tasks (Parasuraman, Mouloua, Molloy, & Hilburn, 1996). Automation is defined as “technology that actively selects data, transforms information, makes decisions, or controls processes” (Lee & See, 2004). While these automated systems are designed to increase performance and decrease errors, they are not perfect. In addition, “When humans are involved, errors will [also] be made, regardless of the level of training, experience, or skill” (Park, 1997, p. 151). As a result, understanding constructs such as trust, in particular trust in automation, has becomes imperative in optimizing the overall relationship between the operator and the system.

Because of the potential for devastating accidents when automation is not allocated in appropriate levels (i.e., over trusting automation despite external cues of a potential malfunction), much research has been performed in the area of trust in automation concerning function allocation. For instance, Lee and Moray’s (1992) automated pasteurization plant experiments found that when overall system performance is low, based on efficiency and the occurrence of faults, the operators’ trust in automation was low (i.e., they generally opted for manual control). Very little, however, has been done to investigate how, if at all, the type of automation error, false alarm or miss, affects an operator’s trust in automation. Before discussing how this question can be investigated, this report will examine definitions of trust, trust's dynamics, how trust relates specifically to automation and reliability, the types of failures (human and automation), and the relationship each of these has with the operator's age.

In the social psychology literature, trust is traditionally defined on an interpersonal-level (e.g., "I do or do not trust a person or group"). According to Mayer, Davis, and Schoorman (1995), trust is "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (p. 712). In that regard, each person must decide if he or she will 1) give their trust to others and 2) act in a trust worthy manner as to gain other's trust (Harris & Provis, 2000). While the second scenario is not easily extrapolated to relationships between operators and machines, the first one very easily meets the analogy. To illustrate, from their research on trust in teams, Harris and Provis (2000) categorized trust into two levels: competence and intentions. In their view, the competence component of trust refers to a person's confidence that someone else (or group) has the ability to perform in the manner in which they are expected to or have advertised. Their competence component is easily mapped to automated systems. Operators have both preconceived and developed over time notions of an automated system's ability to perform, or its reliability.

During her field research of the modernization (i.e., computerization) of eight organizations including pulp and paper mills, insurance companies, and an international bank, Zuboff (1988) discussed three components of trust as reported by the participants in her study: understanding of the technology, trial-and-error experience, and faith. In her research, subjects reporting higher trust also reported greater understanding of the system (the amount to which they felt they understood the automated system, not necessarily the amount to which they actually understood the automation). As trial and error experience increased, those perceiving higher system reliability reported higher trust. Lastly, some subjects reported an almost "blind faith" that automation would perform as advertised and subsequently reported higher trust scores.

Much like Zuboff (1988), Lee and Moray (1992) provided a multi-component description of trust. They, however, described the development of trust as a progression across dimensions of trust: foundation, performance, process, and purpose. They derived their notion of foundation directly from work performed by Barber (1983) in which she stated that there are natural laws which form the foundation upon which all other dimensions of trust are formed. The second notion, performances, is characterized as an "expectation of consistent, stable, and desirable performance or behavior." Their third dimension, process, relies on the ability to understand the traits that motivate behaviors or actions. Lastly, they discussed the purpose component of trust or the "underlying motive or intents" (p. 1246). They interpreted this as representing the operator's responsibilities or the designer's intent for the system.

From the previously discussed definitions and characteristics, trust can be described as a subjective measure of one's confidence in something or someone else. In the context of automation, research has begun to distinguish this subjective rating from its closely related counterpart reliance. Trust refers to the subjective reports of the operators or their feelings about automation; whereas, reliance is the objective measure of performance, such as automation utilization or task efficiency (Wiegmann, Rich, & Zhang, 2001). Participants' subjective ratings of trust in automation may be typically lower than their usage of the automation (Sanchez et al., 2004; Wiegmann et al.). This was most notable when trust ratings were less than 100 percent despite utilization rates of 100 percent (Sanchez et al.; Wiegmann, et al.). This could quite possibly reflect the human feeling that nothing is ever perfect. Wiegmann et al. also showed that as the reliability of the automated system decreased the participants' subjective ratings of trust in the system in general decreased and vice versa. Although it is important to understand the effects of reliability in general, it is critical to understand how other changes in automated systems, such as type of automation error, affect an operator's subjective rating of trust.

Type of Automation Error

Much as described in classic signal detection theory, when an automated engine status indicator malfunctions, one of two types of errors will occur (Figure 1). First, if a system malfunctions and the automation does not indicate a malfunction (i.e., no signal detected), a miss has occurred. Second, if the automation erroneously indicates a malfunction when the system is working properly (a non-existent signal is detected), a false alarm has occurred. Research to date has primarily focused on how false alarms affect trust and reliance in the automation.

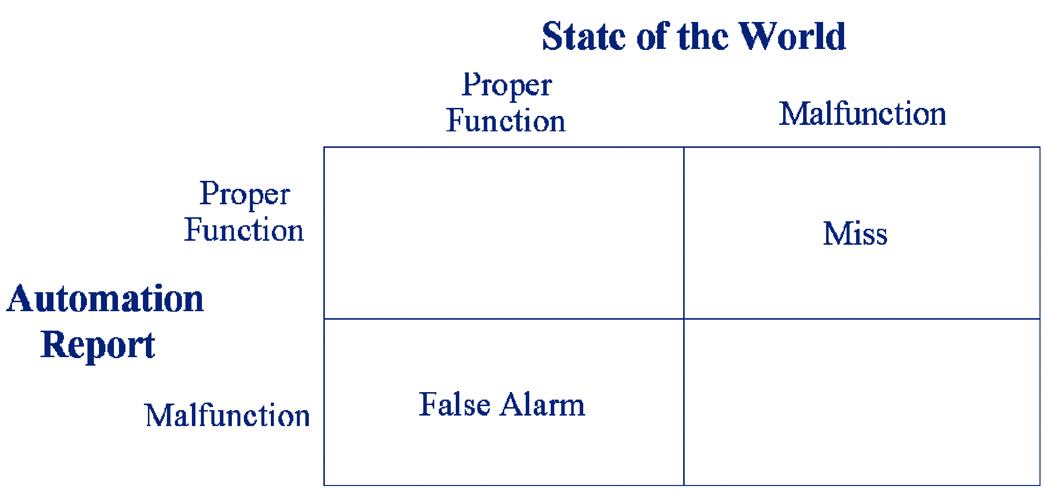

Figure 1. This is a representation of the types of automation failures based on Signal Detection Theory. For example, if the state of the world is that the system has malfunctioned but the automation reports proper function, then a miss has occurred

The effect automation error type has on trust appears to be situation dependent. For example, considering the state of the world today and the resulting heightened security, a bomb detection machine at an airport that falsely indicates a laptop computer contains explosive material (i.e., a false alarm) is much more acceptable than missing a laptop containing explosives and allowing it onto an airplane. Conversely, some may argue that it is more acceptable for blood analysis equipment to miss the fact that someone on trial for drug use actually used drugs rather than falsely indicating that an innocent person used drugs – resulting in their imprisonment for a crime they did not commit.

Much has been done to date to examine how false alarms affect trust. The following passage from Shlomo Breznitz's *Cry Wolf: The Psychology of False Alarms* (1984) delineates the various potential consequences of continued false alarms on an operator.

> Each false alarm reduces the credibility of a warning system. The credibility loss following a false alarm episode has serious ramifications to behavior in a variety of response channels. Thus, future similar alerts may receive less attention. They may elicit weaker fear reactions. The threat may be perceived as less intense or less probable. People may overestimate their ability to cope with the danger if and when it materializes. Or, most important, they may reduce their willingness to engage in protective behavior. (p.11)

Other researchers in addition to Breznitz's (e.g., Bliss & Dunn, 2000) have supported the notion that persistent or pervasive false alarms negatively affect operator trust in automated systems. Very little, however, has been done to examine how, if at all, misses affect trust. By inspection, one could imagine that automation misses might have a lesser effect on operator trust because the operator is experiencing less interaction with the system as a result of the automation not indicating the system malfunctions or because the operators are simply not realizing that misses have occurred (i.e., misses are less salient). While much research has been done to examine the effects of false alarms on operators' trust in and reliance on automation (Bliss, 1997; Bliss & Dunn, 2000; Breznitz, 1984; Xu, Wickens & Ratanen, 2004), the two examples above highlight the need to further examine if false alarms and misses differ in their effect on operators' trust and reliance in automation.

Based upon the two types of automation errors, operators can make two potential errors: omission and commission errors. For the present purpose, omission errors occur when an operator fails to respond to a system malfunction because the automation monitoring the system fails to register or detect a malfunction (i.e., miss occurs). Errors of commission occur when an operator inappropriately complies with the automation's directions concerning a system malfunction when no real malfunction exists (i.e., false alarm occurs). In many of these cases, information to verify or refute the malfunction is readily available, but commission errors are still prevalent (Skitka, Mosier, & Burdick, 2000).

Aging and Automation

Along with the growing amount of research examining trust and reliance in automation (e.g., Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Wiegmann, Rich, & Zhang, 2001; Sanchez, Fisk, & Rogers, 2004), more researchers are also interested in how these issues specifically relate to the effect of operator age. As "baby boomers" approach old age, the sheer number of older adults relying on automation seems to be increasing nearly exponentially. By determining age-related differences in one's approach to automation or formulation of trust in automation, systems can be designed and training programs developed with those differences in mind. Some research suggests that changes in trust developed based on automation reliability differ based on age (e.g., Sanchez et al.). In their study, older adults showed a significant loss of trust in automation when reliability degraded from 100 to 80 percent and then again from 80 to 60 percent. Younger adults, on the other hand, only reported a statistically significant loss in trust when reliability dropped from 100 to 80 percent. Understanding such age-related differences in trust and reliance in automation is important but underdeveloped. This present study adds to our understanding of aging and factors affecting use of automation.

Summary of Proposed Research

Recent research indicates that reliability and past experiences play an import role in developing or maintaining trust in automation (Dzindolet et al., 2003). It neglects, however, to explore the possibility that the type of error might also play a vital role in shaping a person's subjective trust in automation. Through the manipulation of a dual task flight simulator, this present study examined the role the type of automation failure, false alarm or miss, plays in shaping trust and reliance in automation. Both younger and older adult participants performed two experimental blocks in which they experienced one of three experimental conditions: equal false alarms and misses, majority false alarms, or majority misses. Prior to their training block and at the completion of each experimental block each participant completed a subjective trust questionnaire to determine 1) their initial trust in automated decision aids and 2) if the type of automation error over time affected their trust in the automated system. Based on previous research (e.g., Breznitz, 1984; Bliss & Dunn, 2000) it was expected that false alarms would adversely affect operator subjective trust ratings. Little, however, has been done to determine how automation misses affect operator trust and reliance.

This present study examined how both false alarms and misses relate to operator trust and built upon previous work completed by Sanchez et al. (2004) that examined age-related differences in trust and reliance in automation and attempted to replicate some of those results. Determining how such factors affect operator trust and reliance in automated systems is essential to furthering the understanding of how to design systems that older and younger adults alike will trust and upon which they will rely.

METHOD

Participants

A total of 60 older and younger adults participated in this study. The 30 younger adults consisted of 15 males and 15 females between the ages of 18 and 24 years of age, *M*=20.7, *SD*=1.49, from within the Georgia Institute of Technology's undergraduate psychology student population. The 30 older adults consisted of 14 male and 16 female volunteers between 65 and 74 years of age, *M*=69.67, *SD*=2.82, from within Atlanta, GA and the surrounding areas.

Table 1. Description of Study Participants

	Age		Education*		Shipley Vocabulary Test**		Digit-symbol Substitution		Reverse Digit Span	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Younger Adults										
Equal	20.30	1.34	2.70	0.48	31.00	3.80	75.60	6.29	10.90	2.92
Majority False Alarms	20.50	1.58	2.90	0.32	34.10	4.01	67.80	9.74	10.00	2.49
Majority Misses	21.30	1.49	3.00	0.00	29.90	2.77	72.90	9.01	10.20	2.20
Total	20.70	1.49	2.87	0.35	31.67	3.89	72.10	8.82	10.37	2.50
Older Adult										
Equal	69.10	2.85	4.00	1.56	34.90	4.15	65.90	18.98	7.90	2.08
Majority False Alarms	72.00	1.94	3.67	1.12	35.30	2.87	56.60	13.51	7.30	3.47
Majority Misses	67.90	1.97	3.80	2.04	28.30	9.63	47.60	16.02	6.60	3.50
Total	69.67	2.82	3.83	1.58	32.83	6.88	56.70	17.49	7.27	3.03

* 1 = Less than high school, 2 = High school graduate/GED, 3 = Some college, 4 = Bachelor's degree, 5 = Some graduate work, 6 = Master's degree, 7 = Ph.D., M.D., J.D., or other advanced degree
** Best possible score =
+ Best possible score =
++ Best possible score = 14

Design

The type of automation failure was a between groups variable with age employed as a grouping variable. Each age subgroup, younger adults and older adults, was separated into three groups. Each group received one of three experimental conditions: equal number of false alarms and misses, majority false alarms, or majority misses.

Procedure.

Upon beginning the study and providing informed consent, each participant completed a demographic questionnaire and an initial trust questionnaire (Appendix A) to ascertain their familiarity with decision aids, level of comfort with the use of decision aids, and their perceptions with regard to the different types of failure in automation. Participants then completed three abilities tests: the Digit Symbol Substitution test (Wechsler, 1997), Reverse Digit Span test (Wechsler, 1997), and The Shipley Institute of Living Scale vocabulary test (Shipley, 1986) (Table 1). They were then trained on both the engine task and the radar task through the use of scripted scenario and simulator descriptions and a computer tutorial. Upon the completion of a training block, the participants completed two experimental blocks. Each participant received the same experimental condition in both blocks (i.e., equal false alarms and misses, majority false alarms, or majority misses). As a way to measure task performance and provide performance feedback, each participant began each block of testing with 1000 points and lost points for each mistake they made or each time they verified the status of the engine gauges. In the engine task, five points were deducted for each failure to reset a malfunctioning engine, each reset of a working engine, or for resetting the wrong engine for the given scenario (e.g., resetting engine one when engine two is malfunctioning). One point was deducted for each time the participant pressed the view gauges button. These points were deducted regardless of whether the decision aid had instructed the participant to do the correct procedure, wrong procedure, or nothing at all. In the radar task, participants lost five points for each incorrect symbol they counted and for each symbol that was displayed and they failed to count.

Participants performed an orientation training block as a way of learning the tasks and learning strategies to simultaneously complete the engine task and the radar task. This training block consisted of three seven minute sections. If the participant completed the first section with at least 70 percent task proficiency (i.e., lost no more than 100 points during the seven minute period), they were advanced to the first experimental block. If they did not meet the criterion, they continued with section two of the training block. The training followed that same pattern until the participant's performance exceeded criterion or they reached the end of training section three. At the end of the third training section, each participant proceeded to the first experiment block. Therefore training varied from 7 to 21 minutes. Following the completion of their "orientation ride," participants performed 2 experimental blocks, lasting 20 minutes each. At the completion of each block, participants were given the subjective trust questionnaire and performance feedback.

Previous trust in automation research conducted by Sanchez et al. (2004) found that when automation reliability was high older adult performance on a secondary monitoring was approximately 90 percent and younger adult performance was approximately 100 percent (or near ceiling). As automation reliability decreased, older adult performance decreased substantially; whereas younger adults continued to perform at or near ceiling. Therefore, this study adjusted the difficulty of both the primary and secondary tasks (discussed in the apparatus section) to stabilize performance below ceiling. Also, Lee and Moray (1992) found that operators can recover from severe automation errors, but the recovery was not instantaneous when they compared the effects of errors in automation to trust ratings. Therefore, in an effort to control for this recency effect (e.g., a participant is notified of a fault two seconds before the end of the trial), the final minute of each trial consisted of only reliable automation (i.e., correct error notifications from the decision aid). Lastly, the timing of the faults within the trials, while randomly generated, occurred at same time across all the groups.

There were six dependent variables collected and analyzed: engine task and the radar task performance, objective trust (the number of times "view gauges" button was pushed), the subjective trust ratings, perceived reliability, and NASA-TLX subjective workload assessment (Hart & Staveland, 1988).

Apparatus

The experiment was conducted using a low-fidelity cockpit simulator. The simulation was displayed on a high-resolution, color monitor. Participants interacted with the simulation using a standard two button mouse. Only the left mouse button was operable during the experiment. Participants were required to monitor the status of the decision aid, two engine performance gauges, and the radar task.

As mentioned earlier, this present research evaluated and extended previous research conducted in the Human Factors and Aging Lab at the Georgia Institute of Technology. Therefore, the apparatus designed for this study, while aesthetically different from the Sanchez et al. (2004) apparatus (Figure 2), was designed to replicate its function (this report will highlight any significant differences).

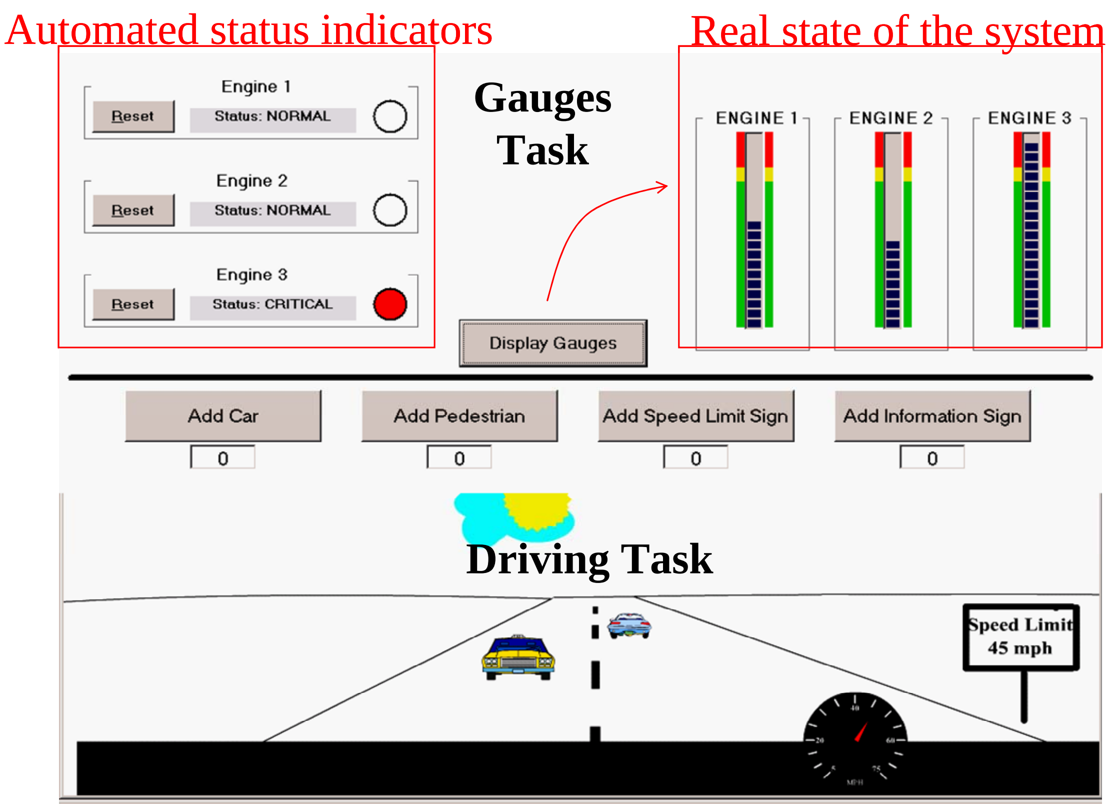

Figure 2. Driving simulator used to test relationships between automation reliability and operators' subjective trust (Sanchez et al., 2004).

Participants were trained to perform a dual task scenario in which they monitored a flight simulator cockpit's engine performance and responded to engine malfunctions, henceforth referred to as the "engine task," and monitor a radar scope and report the number and type of objects they observed on the scope, henceforth referred to as the "radar task." An automated engine status indicator (the "decision aid") was available to assist the participants in monitoring engine performance. Participants were instructed during training that the decision aid might not be 100 percent reliable and that it could be verified by selecting the "view gauges" button to confirm that the engine gauges indicated the appropriate settings for the condition displayed on the decision aid (e.g., gauges in the red indicated a malfunctioning engine) (Figure 3). When the decision aid worked properly, it accurately represented the true status of the engines (e.g., indicating an engine malfunction when a malfunction truly exists).

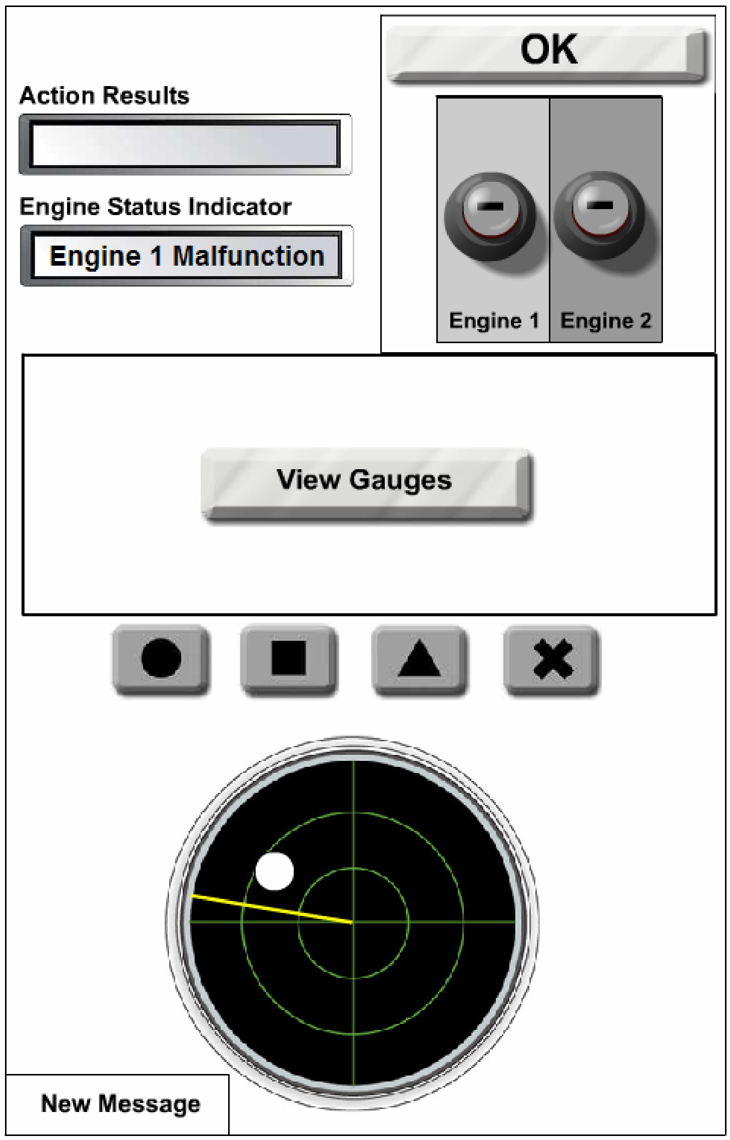

Figure 3. Cockpit simulator to be used in the present study. The top half is the engine task and the bottom half is the radar task.

Engine Task. The engine task, which mimicked the Sanchez et al. (2004) gauges task, consisted of the operator monitoring the decision aid to ascertain when the automation indicated an engine malfunction had occurred and resetting the engines when an actual malfunction occurred. The participants did this by following the decision aid's instructions and resetting the engine indicated or depressing the view gauges button (Figure 3) to determine the actual state of each engine (Figure 4).

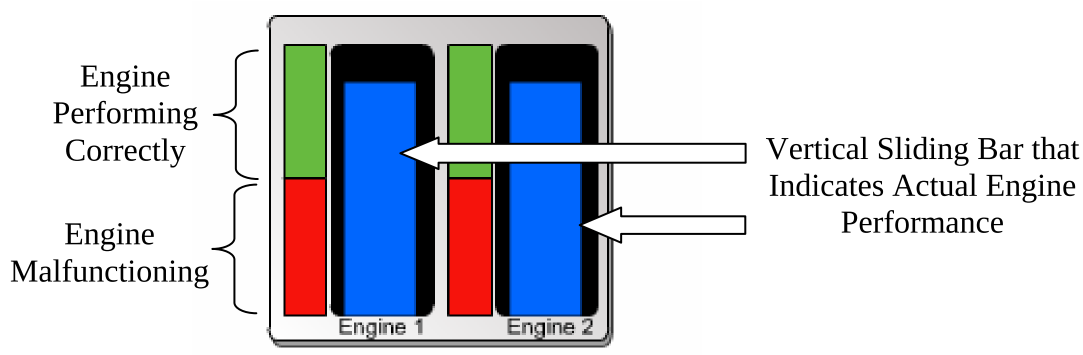

Figure 4. The gauges that become visible when the view gauges button is depressed. If the blue slider is in the green top portion, it represents proper engine performance (i.e., as depicted by both engines). If it is in the red lower portion of the scale that particular engine is malfunctioning.

The decision aid had a reliability of 80 percent (i.e., the decision aid indicated true engine malfunctions 80 percent of the time). The remaining 20 percent consisted of false alarms and misses in varying proportions that were condition specific: 6:6, 9:3, and 3:9 false alarms to misses respectively, for the equal false alarms and misses, majority false alarms, and majority misses conditions. The 80 percent reliability of the decision aid was chosen to replicate the condition by Sanchez et al. (2004) in which automation reliability was 80 percent with an equal number of false alarms and misses. When the participants depressed the view gauges button, they had the ability to confirm or refute the information the decision aid provided. The gauges indicated a malfunction (red) or proper function (green) and the corresponding engine (Figure 4).

The engine gauges only remained visible for 5 seconds when the view gauges button was depressed.

When a true engine malfunction occurred, the participants were required to reset that engine. Resetting was accomplished by depressing the engine button corresponding to the malfunctioning engine and then pressing the OK button (Figure 3). This is a slight change from the Sanchez et al. (2004) procedure in which participants needed only to depress a reset button located adjacent to the automated status indicators (Figure 2). This change slightly increased the cognitive demand on the participants (i.e., they remember three steps to reset the engine as opposed to one) and, in turn, slightly increased the difficulty of the engine task.

The "Engine Status Indicator" window, or the decision aid, located on the upper half of the interface displayed messages for the operator on its perceived status of the system (e.g., blank equates to proper functioning and engine one error means equates to a malfunction in engine one (according to the automation)).

The "Action Results" window was used to provide feedback to the user on whether they appropriately addressed a malfunctioning engine or inappropriately reset a properly working engine. When the operator appropriately reset a malfunctioning engine, the window read "Situation Improved." If the operator committed an error of commission, the window read "Incorrect Action" to indicate the operator complied with faulty advice from the decision aid. On the other hand, if the operator committed an error of omission, the window read "Engine Reset" to indicate that in fact the operator missed an engine malfunction.

The engine task consisted of 60 events (accurate malfunction reports, false alarms, and misses) during each twenty minute block. Because the system reliability was held constant at 80 percent, each block consisted of 48 accurate reports and a combination of 12 false alarms or misses. The three experimental conditions consisted of either six false alarms and six misses, nine false alarms and three misses, or three false alarms and nine misses. There was 1 event every 20 seconds. Each event lasted 10 seconds. Therefore, if a participant reset a malfunctioning engine 5 seconds after the fault occurred, the next event did not occur for another 15 seconds. The final minute consisted of three accurate reports to reduce the chance of recency effects on the subjective trust rating questionnaires. The heuristic followed for the dispersion of false alarms and misses throughout each block was that the automation did not commit two consecutive malfunctions. Once a pattern of false alarms and misses amongst the 60 events was established for the equal false alarms and misses condition, that pattern was used for all participants in that condition. The false alarm and miss place holders established for the equal false alarms and misses condition were also the basis for each of the other conditions. For the majority false alarms condition three misses were changed to false alarms. The same approach was used for the majority misses condition. Figures 5A – 5C and Table 2 show the distribution of automation failures throughout the 48 actual engine malfunctions. In conjunction with monitoring the gauge statuses in the top half of the screen, each participant also monitored a radar scope on the bottom half of the screen.

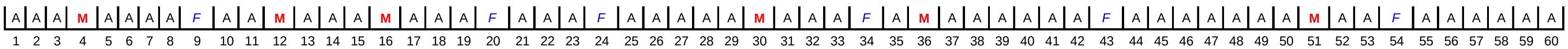

Figure 5a. Actual location of the automation failures in the equal false alarms and misses condition. Each bin represents an event (i.e., actual engine malfunction or automation perceived malfunction). The letter A represents actual engine malfunctions, the letter *F* represents the location of automation false alarms, and the **M** represents the location of the automation misses.

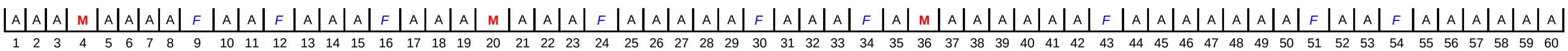

Figure 5b. Actual location of the automation failures in the majority false alarms condition. Each bin represents an event (i.e., actual engine malfunction or automation perceived malfunction). The letter A represents actual engine malfunctions, the letter *F* represents the location of automation false alarms, and the **M** represents the location of the automation misses. Note that all 12 instances of automation failures occur in the same locations as the equal false alarms and misses condition.

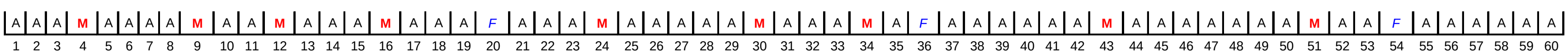

Figure 5c. Actual location of the automation failures in the majority misses alarms condition. Each bin represents an event (i.e., actual engine malfunction or automation perceived malfunction). The letter A represents actual engine malfunctions, the letter *F* represents the location of automation false alarms, and the **M** represents the location of the automation misses. Note that all 12 instances of automation failures occur in the same locations as the equal false alarms and misses condition.

Table 2. A Table representations of the time distributions of the engine events. The first engine event occurred at the five second point of the simulation and each subsequent event occurred on 20 second intervals. There were 60 total events (48 actual engine malfunctions and 12 automation failures).

Equal False Alarm and Misses Condition			Majority False Alarms Condition			Majority Misses Condition		
	Elapsed Time (sec)	Event		Elapsed Time (sec)	Event		Elapsed Time (sec)	Event
1	5	Actual	1	5	Actual	1	5	Actual
2	25	Actual	2	25	Actual	2	25	Actual
3	45	Actual	3	45	Actual	3	45	Actual
4	65	**Miss**	4	65	**Miss**	4	65	**Miss**
5	85	Actual	5	85	Actual	5	85	Actual
6	105	Actual	6	105	Actual	6	105	Actual
7	125	*False Alarm*	7	125	*False Alarm*	7	125	**Miss**
8	145	Actual	8	145	Actual	8	145	Actual
9	165	Actual	9	165	Actual	9	165	Actual
10	185	**Miss**	10	185	*False Alarm*	10	185	**Miss**
11	205	Actual	11	205	Actual	11	205	Actual
12	225	Actual	12	225	Actual	12	225	Actual
13	245	Actual	13	245	Actual	13	245	Actual
14	265	**Miss**	14	265	*False Alarm*	14	265	**Miss**
15	285	Actual	15	285	Actual	15	285	Actual
16	305	Actual	16	305	Actual	16	305	Actual
17	325	Actual	17	325	Actual	17	325	Actual
18	345	Actual	18	345	Actual	18	345	Actual
19	365	Actual	19	365	Actual	19	365	Actual
20	385	*False Alarm*	20	385	**Miss**	20	385	*False Alarm*
21	405	Actual	21	405	Actual	21	405	Actual
22	425	Actual	22	425	Actual	22	425	Actual
23	445	Actual	23	445	Actual	23	445	Actual
24	465	*False Alarm*	24	465	*False Alarm*	24	465	**Miss**
25	485	Actual	25	485	Actual	25	485	Actual
26	505	Actual	26	505	Actual	26	505	Actual
27	525	Actual	27	525	Actual	27	525	Actual
28	545	Actual	28	545	Actual	28	545	Actual
29	565	Actual	29	565	Actual	29	565	Actual
30	585	**Miss**	30	585	*False Alarm*	30	585	**Miss**
31	605	Actual	31	605	Actual	31	605	Actual
32	625	Actual	32	625	Actual	32	625	Actual
33	645	Actual	33	645	Actual	33	645	Actual
34	665	*False Alarm*	34	665	*False Alarm*	34	665	**Miss**
35	685	Actual	35	685	Actual	35	685	Actual
36	705	**Miss**	36	705	**Miss**	36	705	*False Alarm*
37	725	Actual	37	725	Actual	37	725	Actual
38	745	Actual	38	745	Actual	38	745	Actual
39	765	Actual	39	765	Actual	39	765	Actual

Table 2. (continued)

40	785	Actual
41	805	*False Alarm*
42	825	Actual
43	845	Actual
44	865	Actual
45	885	Actual
46	905	Actual
47	925	Actual
48	945	Actual
49	965	Actual
50	985	Actual
51	1005	**Miss**
52	1025	Actual
53	1045	Actual
54	1065	*False Alarm*
55	1085	Actual
56	1105	Actual
57	1125	Actual
58	1145	Actual
59	1165	Actual
60	1185	Actual

40	785	Actual
41	805	*False Alarm*
42	825	Actual
43	845	Actual
44	865	Actual
45	885	Actual
46	905	Actual
47	925	Actual
48	945	Actual
49	965	Actual
50	985	Actual
51	1005	*False Alarm*
52	1025	Actual
53	1045	Actual
54	1065	*False Alarm*
55	1085	Actual
56	1105	Actual
57	1125	Actual
58	1145	Actual
59	1165	Actual
60	1185	Actual

40	785	Actual
41	805	**Miss**
42	825	Actual
43	845	Actual
44	865	Actual
45	885	Actual
46	905	Actual
47	925	Actual
48	945	Actual
49	965	Actual
50	985	Actual
51	1005	**Miss**
52	1025	Actual
53	1045	Actual
54	1065	*False Alarm*
55	1085	Actual
56	1105	Actual
57	1125	Actual
58	1145	Actual
59	1165	Actual
60	1185	Actual

Radar Task. The radar task, which is analogous to the Sanchez et al. (2004) driving task as depicted in Figure 2, consisted of a radar scope that displayed four different symbols in varying amounts. The participants were instructed to depress a symbol's corresponding button when they first noticed it appear on the radar screen (Figure 3). Previous research by Sanchez et al. suggested that this task should be made more difficult to prevent the younger adults from performing at ceiling. In their experiment, each object in the driving task was displayed for approximately 12 seconds. To increase the difficulty of the radar task in this present research, the symbols remained on the screen for approximately seven seconds. Pilot tests were conducted to ensure seven seconds did not make the radar task overly burdensome.

Questionnaires. The experiment included a basic demographic questionnaire and subjective trust questionnaires (Appendix A). The trust questionnaires were administered after training, after block one and after block two. At the completion of the study, the trust questionnaires were analyzed to determine each participant's overall subjective trust in the decision aid and changes in their trust based on the type of automation errors they encounter. The subjective trust questionnaire asked similar but distinctly different questions. The first three questions ascertained the participants' trust in the system and the last two questions were used to determine each participants perceived reliability of the decision aid (i.e., their perception of how often the decision aid correctly identified engine malfunctions).

Point Scale. Participants were provided a scale to internally rank their performance. The scale consisted of point ranges associated with popular "fighter jock" labels. This also provided each participant with performance feedback at the end of each block and a goal should they not make the top category in their first block. Scores between 1,000 and 950 were classified as "Top Gun", 949 to 900 as "Ace", 899 to 850 as "Instructor Pilot", 849 to 800 as "Pilot Trainee", and less than 800 as "Cadet". During their study, Sanchez et al. found that providing this type of point scale seemed to increase participants' desire to optimize their performance.

RESULTS

Analyses of variance (ANOVAs) were performed to assess the main effects of type of automation error, experimental block, and age and any interaction effects. Specifically, this manuscript will discuss the results of the participants' task performance, perceived reliability, subjective trust, reliance (objective trust), and perceived workload. The a priori hypotheses of this study were that the majority false alarms condition would produce lower trust (both subjective and objective) and perceived reliability than the equal false alarms and misses condition and the majority misses condition. Prior to completing the ANOVAs, Spearman's correlations (Table 2) were computed to determine any relationships that exist between the variables of interest in this study.

Table 3. Spearman Correlations for All Participants

	Experimental Block	Task Performance	Reliance	Subjective Trust	Perceived Reliability	Perceived Workload
Age group	0.00	-0.60**	0.47**	0.13	-0.09	0.45**
Experimental Block		0.27*	0.02	0.07	0.06	0.000
Task Performance			-0.05	0.11	0.28**	-0.36**
Reliance				0.35**	0.22*	0.33**
Subjective Trust					0.66**	0.04
Perceived Reliability						-0.13

**Correlation is significant at the .01 level (2-tailed).
*Correlation is significant at the .05 level (2-tailed).

Task Performance. The graph depicting overall task performance is shown in Figure 6 as a function of age and experimental condition. Figures 7A, 7B, and 7C show task performance for each experimental condition by age group and experimental block. As is evident from Figures 6, 7A, 7B, 7C younger adult performance was higher regardless of experimental block or experimental condition. In general, younger adults' task performance scores were statistically higher than older adults, $F(1, 108) = 47.88$, $\eta^2 = .31$, $p < .05$. In addition, both younger and older adults' performance significantly increased from block 1 to block 2, $F(1, 108) = 12.24$, $\eta^2 = .10$, $p < .05$. The type of experimental condition (ratio of types of automation errors) significantly affected task performance scores, $F(2, 108) = 10.72$, $\eta^2 = .17$, $p < .05$; there was also an age x experimental condition interaction, $F(2, 108) = 3.53$, $\eta^2 = .06$, $p < .05$. The mean scores for both younger and older adults were lowest in the majority false alarm condition.

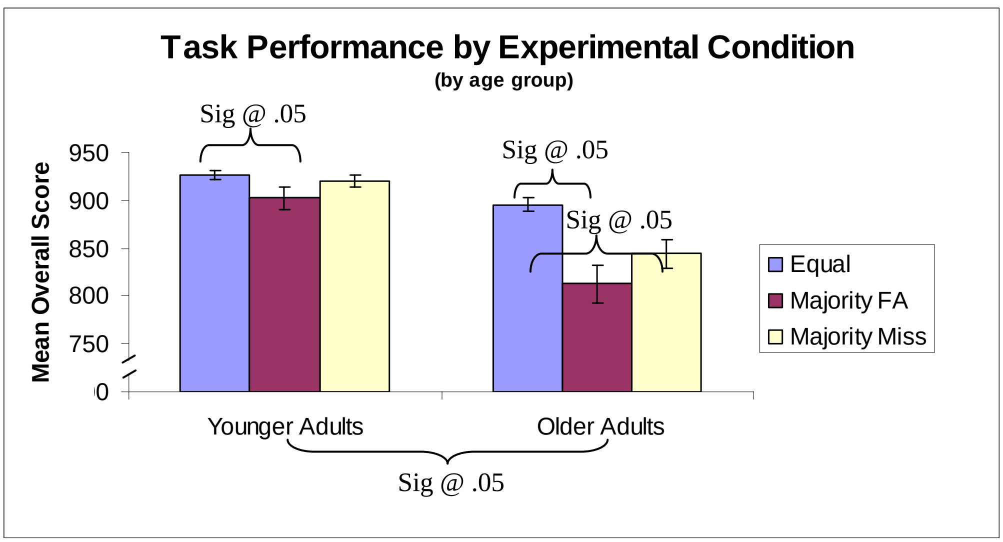

Figure 6. The mean task performance for younger and older adults by experimental condition. Bars represent standard error of the mean.

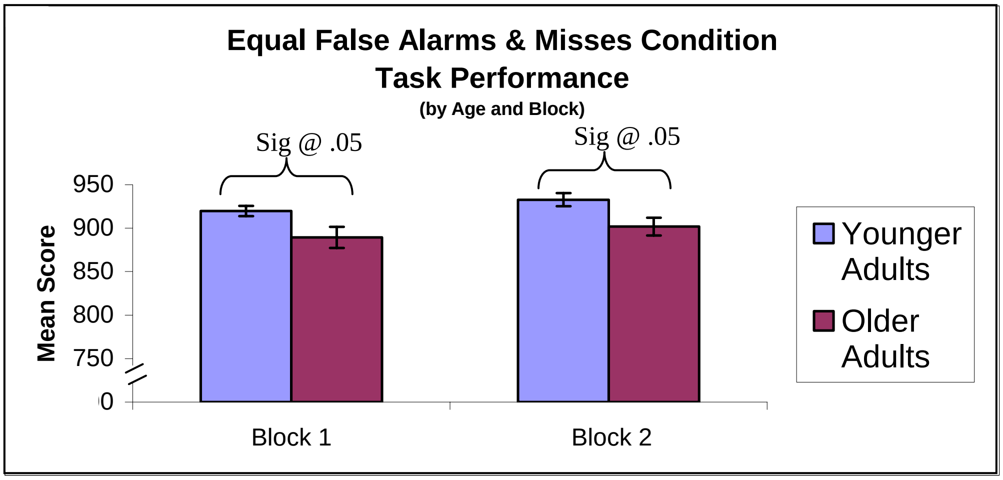

Figure 7A. The mean task performance for younger and older adult participants in the equal false alarms and misses condition by experimental block. Bars represent standard error of the mean.

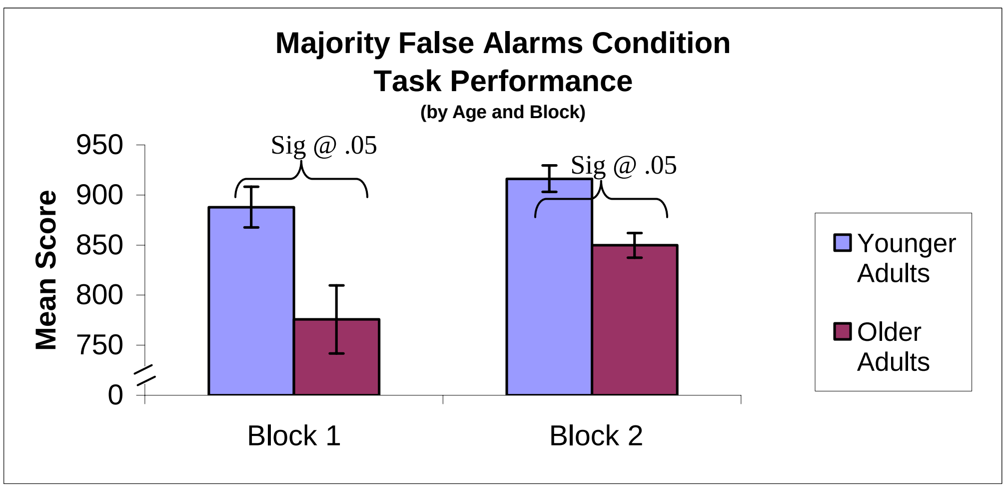

Figure 7B. The mean task performance for younger and older adult participants in the majority false alarms condition by experimental block. Bars represent standard error of the mean.

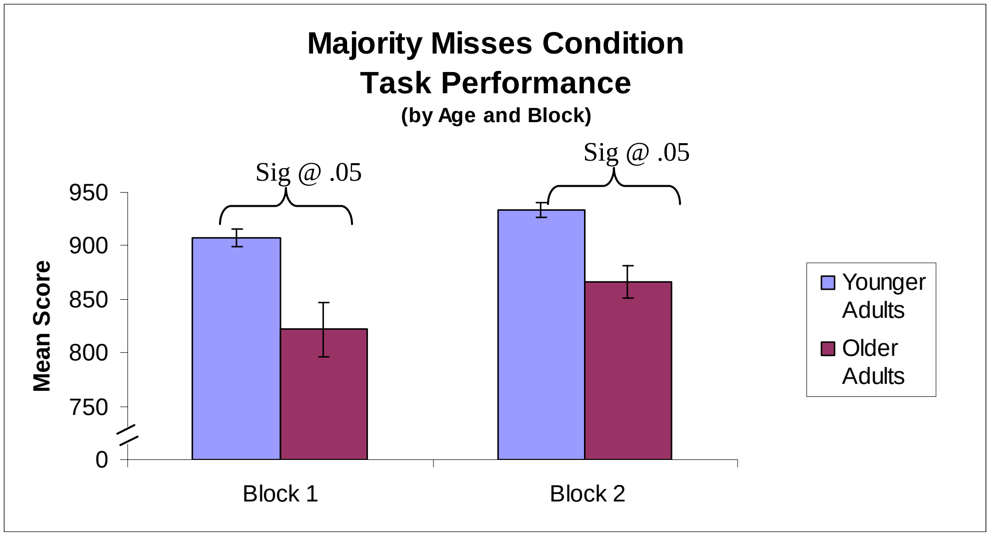

Figure 7C. The mean task performance for younger and older adult participants in the majority misses condition by experimental block. Bars represent standard error of the mean.

Pairwise comparisons for experimental condition (Tables 3A & 3B) revealed that participants in the majority false alarms condition scored significantly lower than participants in the equal false alarms and misses condition for both younger (24.3 points on average) and older adults (82.9 points on average) ($p < .05$). Older adults also displayed a significantly lower mean score in the majority misses condition compared to the equal false alarms and misses condition (51.5 points on average) ($p < .05$). While the younger adults' mean difference between the majority misses condition was numerically lower than that of the equal false alarms and misses condition, it was not statistically lower. Both younger and older adult mean task performance in the majority misses condition was numerically higher on average than the majority false alarms condition, but there was not a statistical difference.

Table 4A. Younger Adults' Pairwise Comparisons of Task Performance Scores of Experiment Condition

Experimental Condition	Experimental Condition	Mean Difference	Std. Error	Sig.	95% Confidence Interval for Difference Lower Bound	Upper Bound
Equal	Equal					
	False Alarm	24.3*	11.53	0.04	1.19	47.41
	Miss	5.8	11.53	0.62	-17.31	28.91
False Alarm	Equal	-24.3*	11.53	0.04	-47.41	-1.19
	False Alarm					
	Miss	-18.5	11.53	0.11	-41.61	4.61
Miss	Equal	-5.8	11.53	0.62	-28.91	17.31
	False Alarm	18.5	11.53	0.11	-4.61	41.61
	Miss					

Based on estimated marginal means
* The mean difference is significant at the .05 level.

Table 4B. Older Adults' Pairwise Comparisons of Task Performance Scores of Experiment Condition

Experimental Condition	Experimental Condition	Mean Difference	Std. Error	Sig.	95% Confidence Interval for Difference Lower Bound	Upper Bound
Equal	Equal					
	False Alarm	82.9*	20.10	0.00	42.59	123.21
	Miss	51.5*	20.10	0.01	11.19	91.81
False Alarm	Equal	-82.9*	20.10	0.00	-123.21	-42.59
	False Alarm					
	Miss	-31.4	20.10	0.12	-71.71	8.91
Miss	Equal	-51.5*	20.10	0.01	-91.81	-11.19
	False Alarm	31.4	20.10	0.12	-8.91	71.71
	Miss					

Based on estimated marginal means
* The mean difference is significant at the .050 level.

Subjective trust. The subjective trust questionnaire was administered at the end of the practice block, the end of block 1, and the end of block two; however, the analyses of subjective trust in this manuscript will be limited to the subjective trust ratings given at the end of each experimental block (i.e., not the practice block). This was mainly done because of the variable length of the practice block (i.e., between 7 and 21 minutes depending on performance). The graph depicting the mean subjective trust ratings for both younger and older adults by experimental condition can be seen in Figure 8. There was a significant main effect for condition, $F(2, 102) = 4.20$, $\eta^2 = .07$, $p < .05$, but not for age or experimental block. There were also no interaction effects. When examining more closely where the significant differences were found, Figure 8 shows that the younger adults exhibited no significant differences in their trust ratings based on experimental condition. Older adults, on the other hand, reported significant differences based on experimental condition, $F(2, 54) = 4.63$, $\eta^2 = .15$, $p < .05$. Figure 8 shows the significantly higher trust ratings in the mostly misses condition compared to both the equal false alarms and misses condition and the mostly false alarms condition, 1.05 and 1.5 points on average respectively, $p < .05$.

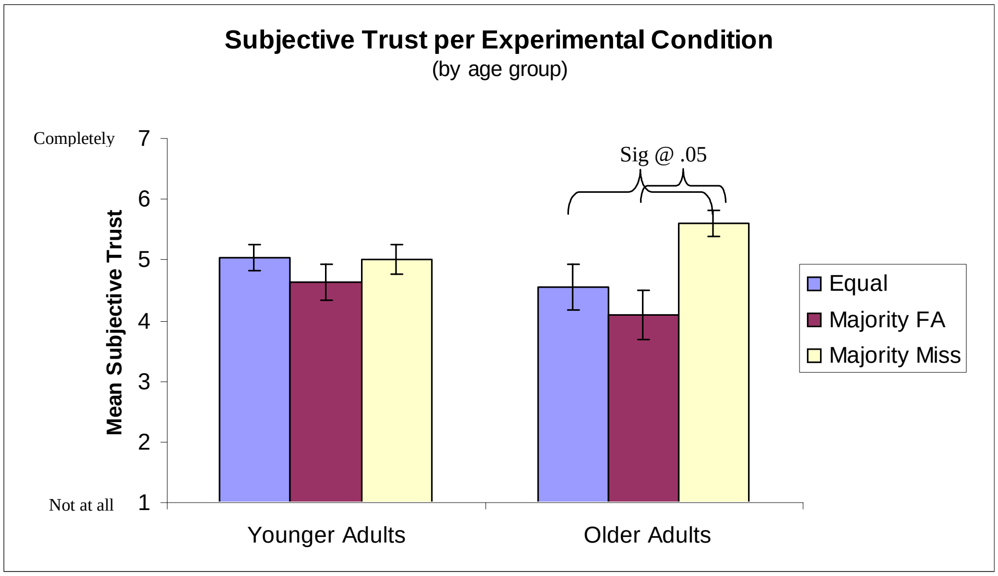

Figure 8. The mean subjective trust ratings for younger and older adults by experimental condition. Bars represent standard error of the mean.

Perceived Reliability. At the end of each subjective trust questionnaire, participants were asked to rate the reliability of the engine status indicator. The rating was accomplished with two questions. First, participants were asked to respond using the provided seven-point scale (1 = disagree, 7 = agree) to the statement "The Engine Status Indicator is reliable." The second question dedicated to understanding the participants' perceived reliability of the engine status indicator asked the participants to use a percentage to indicate the reliability of the engine status indicator (e.g., I think the engine status indicator was XX% reliable.). The correlation between the two questions was .59 was significant at $p < .01$. On the surface, this seems to be a very low correlation for two questions asking essentially the same thing. Further inspection revealed that one younger adult and one older adult failed to provide a percentage for their perceived reliability. In addition, four older adult participants appear to have misinterpreted the second perceived reliability question by providing inputs that were completely inconsistent with the first question (e.g. provided a 1 (disagree) that the engine status indicator was reliable and then provided 80 percent for their perceived reliability). When these six participants were excluded, the Spearman correlation increased to .74, $p < .01$.

For the purpose of the remainder of the perceived reliability analyses, the seven-point scale data are used (unless otherwise specified) as they are the less likely of the two sets of perceived reliability data to have been misinterpreted. Figure 9 shows that while there was no significant effects on perceived reliability due to age or experimental block, there was a significant effect due to experimental condition, $F(2, 108) = 8.20$, $\eta^2 = .13$, $p < .05$.

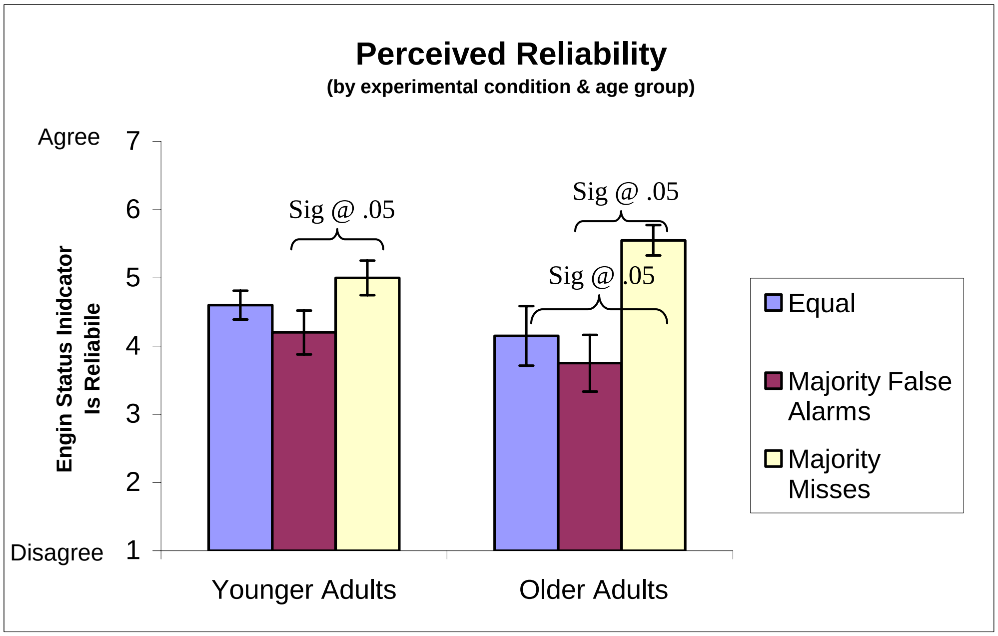

Figure 9. The mean perceived reliability for younger and older adults by experimental condition.

Pairwise comparisons of the means revealed that the mean perceived reliability rating in the majority false alarms condition, *M* = 3.97, *SD* = 1.66, was statistically lower than both the equal false alarms and misses condition, *M* = 4.38, *SD* = 1.53, and the majority misses condition, *M* = 5.28, *SD* = 1.09. The majority false alarms condition was rated .40 and 1.3 points lower on average respectively ($p < .05$). Further examination of these data showed that within each experimental condition only older adults accounted for a significant portion of the variability in perceived reliability ratings, $F(2, 54) = 6.24$, $\eta^2 = 12.47$, $p < .05$. Statically significant differences, $p < .05$, were found between the mean perceived reliability ratings in the majority misses condition, *M* = 5.55, *SD* = 1.00, and both the equal false alarms and misses condition, *M* = 4.15, *SD* = 1.95, and the majority false alarms condition, *M* = 3.75, *SD* = 1.86.

In addition to the ANOVA performed on the seven-point scale data, a two-tailed independent t-test was conducted on the percent reliable data to determine if the participants' perceived reliability was statically lower than the actual engine status indicator reliability (i.e., 80 percent). Table 5 and Figure 10 reveal that younger adults' perceived reliability was significantly lower than 80 percent in the majority false alarms condition, *M* = 69.45, *SD* = 22.18, *t*(19) = -2.13, *p* < .05 and approaching statistical significance in the equal false alarms and misses condition, *M* = 74.72, *SD* = 10.91, *t*(17) = -2.05, *p* < .06. Older adults' perceived reliability was only significantly lower than the actual reliability in the majority false alarms condition, *M* = 57.00, *SD* = 29.31, *t*(19) = -3.51, *p* < .05.

Table 5. One Sample t-test (2-tailed) Testing the Difference between Perceived and Actual Reliability of 80%

	Perceived Reliability			Test Value = 80			
	Mean	SD	n	t	df	Sig. (2-tailed)	Mean Difference
Younger Adults	75.03	15.74	58*	-2.38	57	0.02	-4.97
Equal False Alarms and Misses	74.72	10.91	18*	-2.05	17	0.06	-5.28
Majority False Alarms	69.45	22.18	20	-2.13	19	0.05	-10.55
Majority Misses	80.90	9.36	20	0.43	19	0.67	0.90
Older Adults	67.50	23.96	52*	-3.76	51	0.00	-12.50
Equal False Alarms and Misses	73.33	19.40	18*	-1.46	17	0.16	-6.67
Majority False Alarms	57.00	29.31	20	-3.51	19	0.00	-23.00
Majority Misses	75.00	15.06	14*	-1.24	13	0.24	-5.00

* Cases were excluded because reported perceived % reliability were inconsistent with participants' other reports of perceived reliability or missing

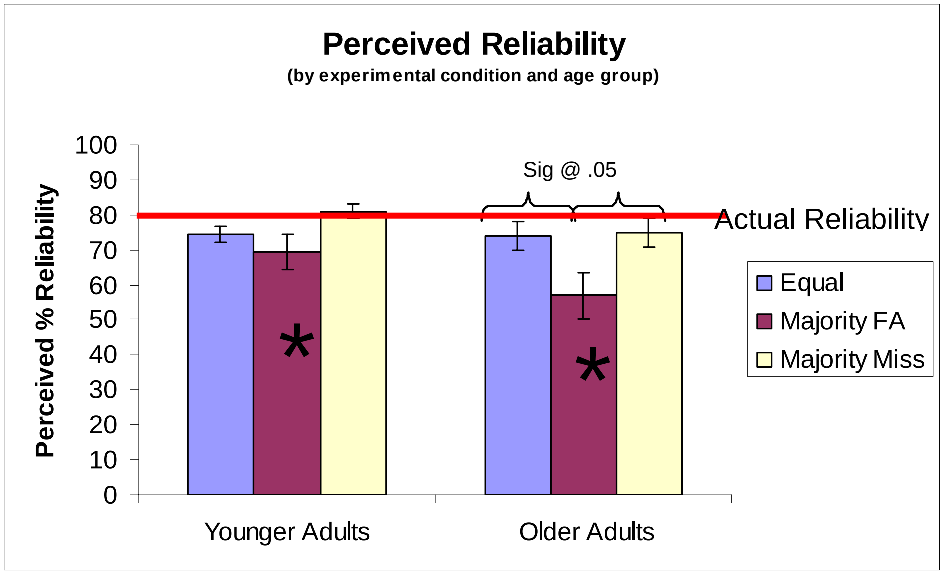

Figure 10. The mean perceived reliability for younger and older adults by experimental condition. The asterisks signify the conditions in which the reliability was perceived as significantly lower than the actual reliability.

Reliance. As depicted in Figure 11, there is a main effect for age only, $F(1, 108) = 19.703$, $\eta^2 = .15$, $p < .05$. While the effect of experimental condition was not statistically significant for reliance on the automation (i.e., choosing to *not* press the view gauges button), both younger and older adults pressed the view gauges button the least number of times on average in the majority misses condition (i.e., relied more on the automation or had higher trust that the automation would perform correctly). Within each experimental condition, older adults chose to rely on the automation significantly more ($p < .05$) than younger adults, $F(1, 36) = 8.09$, $\eta^2 = .18$; $F(1, 36) = 4.66$, $\eta^2 = .12$; $F(1, 36) = 7.26$, $\eta^2 = .17$, equal false alarms and misses, majority false alarms, majority misses respectively.

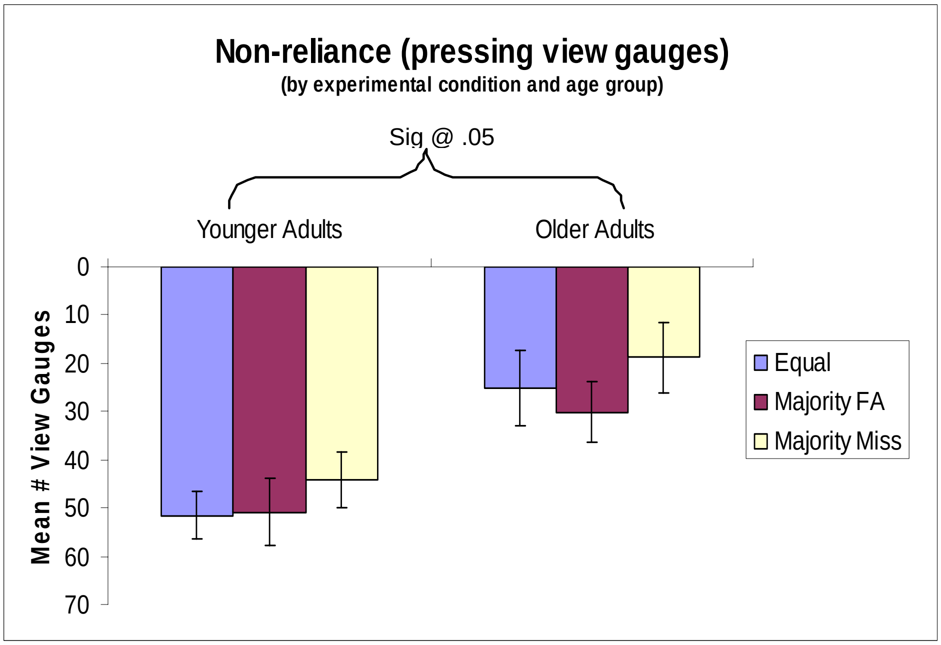

Figure 11. Non-reliance on the automation (i.e., number of times the view gauges button was pressed) for younger and older adults by experimental condition.

Perceived workload. Because experimental blocks 1 and 2 were exactly the same for each participant, mean task performance increased from block 1 to block 2 (i.e., participants were still learning their tasks), and the NASA-TLX workload assessment (Hart & Staveland, 1988) was collected only at the end of the block 2, comparisons of these data are limited to comparisons between the TLX data and the block 2 data. Figure 12 shows graphically the difference between older and younger adult perceptions of the overall workload for both the engine and radar tasks. There was no main effect for condition; however, older adults perceived a significantly higher workload overall than did younger adults, $F(1, 54) = 15.20$, $\eta^2 = .22$, $p < .05$. Pairwise comparisons of the means for older and younger adults within each experimental condition revealed that neither younger adults' not older adults' perceived workload assessments varied significantly by condition.

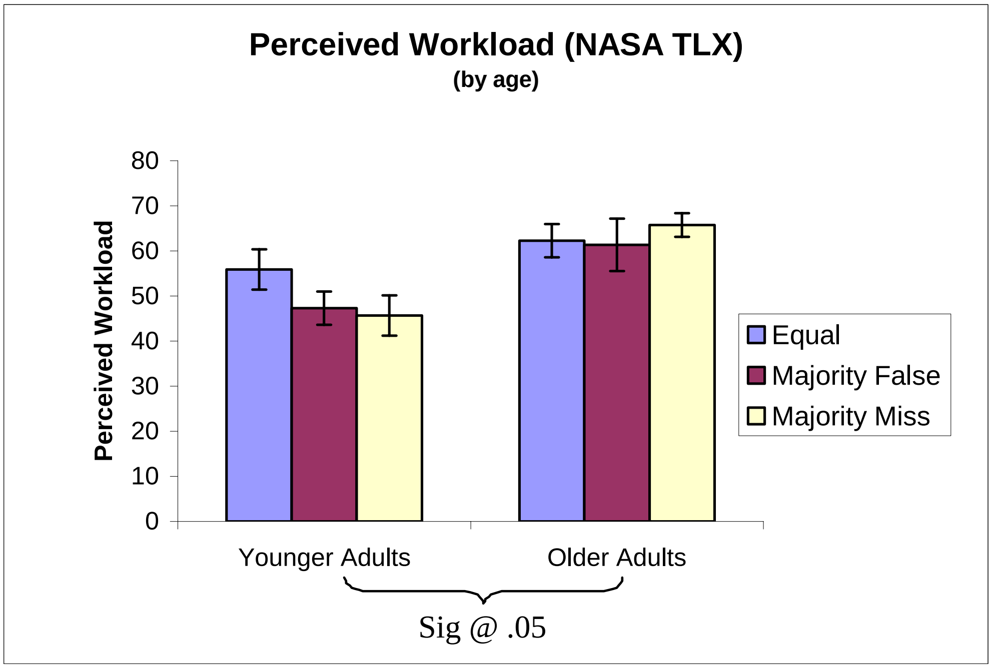

Figure 12. The mean perceived workload (NASA-TLX) for younger and older adults by experimental condition.

DISCUSSION

In general, automation is designed to work within a specific context. When used as intended, most automation performs fairly well. It is not, however, perfect. It can either false alarm or miss. This study was designed to examine how those false alarms and misses affected trust in the context of decision support aids. Since little has been done to specifically investigate the affect of automation error type on operator trust and reliance, this current study manipulated the type of automation error the participants were exposed to while holding constant other factors which can significantly affect trust in automation constant (e.g., reliability, scenario, and cost associated with committing omission errors and commission errors). Overall, younger and older adults showed no significant difference between their average subjective trust ratings (Figure 8) or between their perceived reliability ratings (Figures 9 & 10) despite the significant differences in their task performance scores. When these results are examined in conjunction with the correlations between task performance and subjective trust in Tables 1, 6A, and 6B, it appears that whether a participant decided to trust the automated decision aid was completely independent of how they performed while using the system. While there was no difference in level of perceived reliability between younger and older adults, it appears as though task performance may influence how perceived reliability is determined for older adults (i.e., there was a significant positive correlation between task performance and perceived reliability) (Tables 6A & 6B). This is a very interesting point and will be revisited after the discussion of how the participants’ perceived workload may have affected their perceived reliability and their decision to “rely on” the engine status indicator.

Table 6A. Spearman Correlations for Younger Adults

	Task Performance	Reliance	Subjective Trust	Perceived Reliability	Perceived Workload
Experimental Block	0.41**	0.16	0.07	0.06	--
Task Performance		0.56**	0.25	0.17	-0.10
Reliance			0.46**	0.44**	0.03
Subjective Trust				0.67**	-0.07
Perceived Reliability					-0.07

** Correlation is significant at the .01 level (2-tailed).

Table 6B. Spearman Correlations for Older Adults

	Task Performance	Reliance	Subjective Trust	Perceived Reliability	Perceived Workload
Experimental Block	0.21	0.05	0.09	0.08	--
Task Performance		0.09	0.15	0.32*	-0.04
Reliance			0.28*	0.21	0.31*
Subjective Trust				0.68**	0.08
Perceived Reliability					-0.09

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

As shown in Figure 9, both younger and older adult participants in majority false alarms condition reported statistically higher perceived reliability in the majority misses condition compared to the majority false alarms condition. In addition, both younger and older adults reported higher subjective trust ratings in the majority misses condition compared to the majority misses condition (statistically higher for older adults) (Figure 9). In fact, when comparing the older adults' data in Figures 8 and 9, the exact same pattern emerges.

A potential explanation of why the majority misses condition did not appear to affect subjective trust or perceived reliability to the same degree as the majority false alarms condition is the decreased saliency of the automation malfunction. In designing this simulator, attempts were made to ensure that the saliency of both false alarms and misses were the same. First, participants were required to verify when the engine status indicator made a false alarm. They did this by pressing the OK button without pressing either of the engine reset buttons. This procedure was added to ensure that participants did not “accidentally” ignore a false alarm condition if they simply didn’t see it. Secondly, a new message indicator was added to the bottom of the radar task (Figure 3) in an effort to prompt participants attending to the radar task that a new message was in either the engine status indicator or the action result window. Therefore both false alarms and misses should have been of similar saliency.

Upon further review, however, there is a potential difference that is inherent to the nature of false alarms and misses that could not be controlled for. Regardless of how salient the miss signal was made, participants in the majority misses condition received less information than those in the majority false alarms or equal false alarms and misses condition. For each false alarm a participant experienced they received two or three pieces of information from the simulator. First, they were informed that an engine had malfunctioned. Then, they either verified the potential malfunction (i.e., viewed the gauges), or directly accomplished the engine reset procedures. Lastly, they were informed by the simulator whether hey had performed the correct action. For misses, however, participants never received the initial engine malfunction notification. In addition, unless they were explicitly trying to “catch” the simulator in an automation miss condition, they also did not receive the malfunction verification. Typically, when the automation missed an engine malfunction, the participants also missed the malfunction. As a result, the participants received only the message that an engine had malfunctioned and the simulator reset it for them.

In initial design of this study, there was an explicit decision made to increase the cognitive complexity of this simulator over and above that of the simulator used in the Sanchez et al. (2004) study. This was done in an attempt to bring younger adult performance below ceiling. In addition, it was expected that by increasing task difficulty participants would make a decision to both trust in and rely on the provide automation or not trust in or rely on the automation and allow their overall score to suffer. In fact, the changes made to the current simulator appear to have accomplished the goal of lowering overall task performance. While older adults' performance was lowered along with younger adults' performance, it was still acceptable and the tasks did not appear to be overly difficult.

In general, older adults perceived a higher workload than younger adults (Figure 12) and the older adults were more reliant on the information provided by the engine status indicator (i.e., reliant on the automation through not pressing the view gauges button) (Figure 11) to decide whether to reset an engine. Those two facts taken in conjunction with no significant difference in subjective trust due to participant age (Figure 8) indicate that the decision to rely on the automation was heavily influenced by how difficult the task is perceived to be. This conclusion is further supported by the Spearman correlations (Table 6B) which indicate a significant positive relationship between perceived workload and reliance. As a result, it is unclear if the older adults believed the automation would perform better than they could because of the difficulty of the simulation or if they simply did not feel they had enough time to view the gauges and reset the engines. Lastly with regard to the relationships between workload, perceived reliability, trust, and reliance, it appears that as perceived workload increases reliance will potentially follow the same pattern shown in perceived reliability and trust. Admittedly, the differences in reliance for older adults are not significant, but reliance was highest in the majority misses condition, second in the equal false alarms and misses condition, and lowest in the majority false alarms condition. Further research with more difficult tasks should be conducted to determine if theses patterns and relationships between workload, trust, perceived reliability, and reliance can be replicated. If they can, designers of complex automated systems will have valuable piece of information. If their system predominantly misses rather than false alarms, assuming misses do not result in critical system failures, operators may be more willing to trust and use they system.

Trust and reliance in automation are dynamic traits that are significantly affected by many factors. Much research has been conducted to date to examine how some of these factors affect an operator's trust in increasingly automated systems, such as self-confidence, reliability, or prior knowledge of impending automation failure (e.g., Dzindolet et al., 2003; Moray, Inagaki, & Itoh, 2000; Sanchez et al., 2003; Wiegmann et al., 2001). Very little has been done to examine how the type of automation failure affects an operator's trust and reliance in automation. This study established the foundation to examine specifically how automation error type affected operator trust, reliance, and perceived reliability. This study attempted to control for variables that are know to or might affect trust, such as the scenario description, system reliability, and the cost of false alarms and misses. Like previous research, this present research confirmed that perceived reliability is often lower than actual reliability (Sanchez et al., 2004; Wiegmann et al., 2001) and that false alarms significantly reduce operator trust in the automation (Bliss & Dunn, 2000). In addition, this study found that there does not appear to be an effect on the level of subjective trust within each experimental condition based on age. There does, however, appear to be a significant difference in the reliance (i.e., objective trust) within each experimental condition between older and younger adult participants. This could very well be the result of older adults adapting to their perceived higher workload by actively choosing to not use the view gauges button in an effort to save time in traversing the display with the mouse.

Proposed follow-up research will need to be conducted to examine some the questions this study was not designed to answer and to control for perceived workload difference. Future work should investigate the effects of changing the scenario in which the participants find themselves, much like the bomb detection machine and blood analysis examples discussed earlier. Also of interest, and closely related to the scenario, is the degree to which the cost of the automation failure is varied. Lastly, a multifactor experiment should be conducted in which all of the aforementioned factors are systematically varied in order to determine their levels of interaction. Those results coupled with this single factor experiments will provide a more complete picture of the complex dynamics involved with determining and predicting operators' trust and reliance in automation.

APPENDIX A

Background Documents and Survey Instruments

Participant ID: ___-___-___

Please answer the following questions. All of your answers will be treated confidentially. Any published document regarding these answers will not identify individuals with their answers. If there is a question you do not wish to answer, please just leave it blank and go on to the next question. Thank you in advance

Demographics Questionnaire

Gender: ☐ 1 Male ☐ 2 Female

1. Education completed (check highest level)

☐ 1 Less than high school graduate (highest grade completed? _________)

☐ 2 High school graduate/G.E.D.

☐ 3 Some college, or trade, technical, or business school (how many years? _______)

☐ 4 Bachelor's degree

☐ 5 Some graduate work (how many years? _____)

☐ 6 Master's degree

☐ 7 M.D., J.D., Ph.D., other advanced degree

2. Current marital status (check one)

☐ 1 Single

☐ 2 Married

☐ 3 Separated

☐ 4 Divorced

☐ 5 Widowed

☐ 6 Other (please specify ___________)

3. Race/ethnicity

☐ 1 Black/African American

☐ 2 Asian American/Pacific Islander

☐ 3 White/Caucasian

☐ 4 Hispanic/Latino

☐ 5 American Indian/Alaskan Native

☐ 6 Multiracial (please specify ___________)

☐ 7 Other (please specify _______________)

4. In which type of housing do you live?

☐ 1 Residence hall/College dormitory

☐ 2 House/Apartment/Condominium

☐ 3 Senior housing (independent)

☐ 4 Assisted living

☐ 5 Nursing home

☐ 6 Relative's home

☐ 7 Other (please specify _______________)

5. Do you live alone a majority of the year?

- ☐ 1 Yes
- ☐ 2 No

6. What is your primary language?

- ☐ 1 English
- ☐ 2 Spanish
- ☐ 3 French
- ☐ 4 Creole
- ☐ 5 Portuguese
- ☐ 6 Other (please specify __________)

7. Occupational status (check <u>all</u> that apply)

- ☐ 1 Working full-time
- ☐ 2 Working part-time
- ☐ 3 Student
- ☐ 4 Homemaker
- ☐ 5 Retired
- ☐ 6 Volunteer worker
- ☐ 7 Seeking employment, laid off, etc.
- ☐ 8 Leave of absence
- ☐ 9 Other (please specify __________)

8. What is your current occupation?

____

<u>If retired:</u>

9. What was your primary occupation?

____

10. What year did you retire?

____

Health Information

1. In general would you say your health is:

❑ 1 Poor ❑ 2 Fair ❑ 3 Good ❑ 4 Very Good ❑ 5 Excellent

2. Compared to other people your own age, would you say your health is:

❑ 1 Poor ❑ 2 Fair ❑ 3 Good ❑ 4 Very Good ❑ 5 Excellent

3. How satisfied are you with your present health?

❑ 1 Not At All Satisfied ❑ 2 Not Very Satisfied ❑ 3 Neither Satisfied Nor Dissatisfied ❑ 4 Somewhat Satisfied ❑ 5 Extremely Satisfied

4. How often do health problems stand in the way of your doing the things you want to do?

❑ 1 Never ❑ 2 Seldom ❑ 3 Sometimes ❑ 4 Often ❑ 5 Always

5. Have you ever lost consciousness for more than 10 minutes because of a head injury?

❑ 1 Yes ❑ 2 No

6. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much? Select one box for each type of activity.

	Yes$_1$, **Limited a Lot**	Yes$_2$, **Limited a Little**	No$_3$, **Not Limited at all**
a. Vigorous activities, such as running, lifting heavy objects,or participating in strenuous sports (like swimming laps)			
b. Moderate activities, such as moving a table, pushing avacuum cleaner, bowling, or playing golf			
c. Lifting or carrying groceries			
d. Climbing **several** flights of stairs			
e. Climbing **one** flight of stairs			
f. Bending, kneeling, or stooping			
g. Walking **more than a mile**			
h. Walking **several blocks**			
i. Walking **one block**			

j. Bathing or dressing yourself			

7. Are you on post-menopausal estrogen replacement therapy?

❑1 Yes ❑ 2 No ❑ 3 Not Applicable

8. Do you take any medications (prescription or nonprescription) on a <u>regular</u> basis (at least once a week)?

❑1 Yes ❑2 No

9. **Please check which of following conditions you have now or have had in the past.**

Condition	**In Your Lifetime$_1$**	**Now$_2$**
a. Asthma or Bronchitis		
b. Cancer (other than skin cancer)		
c. Chronic liver disease or hepatitis		
d. Chronic migraine headaches		
e. Diabetes		
f. Emphysema		
g. Encephalitis or meningitis		
h. Epilepsy		
i. Heart attack or bypass surgery		
j. Heart problems		
k. High blood pressure		
l. Kidney disease		
m. Leukemia		
n. Multiple sclerosis		
o. Parkinson's disease		
p. Pneumonia		
q. Rheumatoid arthritis or other autoimmune disorders		
r. Stomach ulcers		
s. Stroke		
t. Other significant illnesses (please list)		

10. How many BONE FRACTURES have you had in the LAST FIVE YEARS?

❑ 1 None
❑ 2 1
❑ 3 2
❑ 4 3-5
❑ 5 More than 5

11. How many SURGERIES have you had in the LAST FIVE YEARS?

- ☐ 1 None
- ☐ 2 1
- ☐ 3 2
- ☐ 4 3-5
- ☐ 5 More than 5

12. How many times have you been HOSPITALIZED in the LAST FIVE YEARS?

- ☐ 1 None
- ☐ 2 1
- ☐ 3 2
- ☐ 4 3-5
- ☐ 5 6-10
- ☐ 6 More than 10

Medication Usage Details

Please list all medical products that you are currently taking. Include medicinal herbs, vitamins, aspirin, antacid, nasal spray, laxatives, etc., as well as prescription medications (copy names from label if possible). This information will be completely confidential.

EXAMPLE
Name of Medication: Zarontin_____

Reason for taking:____epilepsy___ Dosage (ea. time taken):______500 mg______

How often do you take the medication? (circle one)

daily (every other day) weekly as needed

On days that you take the medication, how many times per day do you take it? ___3_____

What time of day do you take the medication? morning, afternoon, evening______

How long you have been taking the medication? _________5 years

Does this medication cause any problems? _makes me sleepy_____

1. Name of Medication: ____

Reason for taking:_____________________ Dosage (ea. time taken):________

How often do you take the medication? (circle one)

daily every other day weekly as needed

On days that you take the medication, how many times per day do you take it? ___

What time of day do you take the medication? _____

How long you have been taking medication? ______________________________

Does this medication cause any problems? ________

2. Name of Medication: ____

Reason for taking:______________________ Dosage (ea. time taken):________

How often do you take the medication? (circle one)

daily every other day weekly as needed

On days that you take the medication, how many times per day do you take it? ___

What time of day do you take the medication? _____

How long you have been taking medication? ______________________________

Does this medication cause any problems? ________

3. Name of Medication: ____

Reason for taking:______________________ Dosage (ea. time taken):________

How often do you take the medication? (circle one)

daily every other day weekly as needed

On days that you take the medication, how many times per day do you take it? ___

What time of day do you take the medication? _____

How long you have been taking medication? ______________________________

Does this medication cause any problems? _

4. Name of Medication: ____

Reason for taking:______________________ Dosage (ea. time taken):________

How often do you take the medication? (circle one)

daily every other day weekly as needed

On days that you take the medication, how many times per day do you take it? ___

What time of day do you take the medication? _____

How long you have been taking medication? ____________________________

Does this medication cause any problems? ________

5. Name of Medication: ____

Reason for taking:______________________ Dosage (ea. time taken):________

How often do you take the medication? (circle one)

daily every other day weekly as needed

On days that you take the medication, how many times per day do you take it? ___

What time of day do you take the medication? _____

How long you have been taking medication? ____________________________

Does this medication cause any problems? ________

6. Name of Medication: ____

Reason for taking:______________________ Dosage (ea. time taken):________

How often do you take the medication? (circle one)

daily   every other day   weekly   as needed

On days that you take the medication, how many times per day do you take it? ___

What time of day do you take the medication? _____

How long you have been taking medication? ____________________________

Does this medication cause any problems? ________

7. Name of Medication: ____

Reason for taking:______________________ Dosage (ea. time taken):________

How often do you take the medication? (circle one)

daily   every other day   weekly   as needed

On days that you take the medication, how many times per day do you take it? ___

What time of day do you take the medication? _____

How long you have been taking medication? ____________________________

Does this medication cause any problems? ________

8. Name of Medication: ____

Reason for taking:_____________________ Dosage (ea. time taken):________

How often do you take the medication? (circle one)

daily   every other day   weekly   as needed

On days that you take the medication, how many times per day do you take it? ___

What time of day do you take the medication? _____

How long you have been taking medication? ______________________________

Does this medication cause any problems? ________

9. Name of Medication: ____

Reason for taking:_____________________ Dosage (ea. time taken):________

How often do you take the medication? (circle one)

daily   every other day   weekly   as needed

On days that you take the medication, how many times per day do you take it? _________

What time of day do you take the medication? _____

How long you have been taking medication? ______________________________

Does this medication cause any problems? ________

Automation Experience Questionnaire

The purpose of this questionnaire is to assess your familiarity and experience with automated devices. Please answer all questions.

According to Moray, Inagaki, and Itoh, automation is defined as any sensing, detection, information-processing, decision making or control action that could be performed by humans but is actually performed by a machine. A couple of everyday examples are the gas gauge in a car or the autopilot system in an airplane. Keep the above definition in mind when answering the following questions.

1. Please circle the pieces of automation (or automated devices) you use on a regular basis.

Computer	Personal Data Assistant (PDA)	Cruise control
Blood Glucose Meter	Washing and/or Drier	Remote Control
Game System (e.g., X-Box)	Thermostat	Copy Machine
Hearing aids	Alarm clock	Scale
Car seat adjustment	Hair drier	Vacuum
Voicemail	VCR	DVD/CD Player
Answering machine	Iron	Cell Phone
ATM machine	Other(s): ______________________	

2. Please circle the pieces of automation you use on an infrequent basis.

Computer	Personal Data Assistant (PDA)	Cruise control
Blood Glucose Meter	Washing and/or Drier	Remote Control
Game System (e.g., X-Box)	Thermostat	Copy Machine
Hearing aids	Alarm clock	Scale
Car seat adjustment	Hair drier	Vacuum
Voicemail	VCR	DVD/CD Player
Answering machine	Iron	Cell Phone
ATM machine	Other(s): ______________________	

3. Please list two automated devices/systems you find the MOST DIFFICULT to use.

4. For what reason or reasons do you have difficulty using these devices/systems?

5. Please list two automated devices/systems you find the EASIEST to use.

6. For what reason or reasons do you find it easy to use these devices/systems?

Using the scale below, select the answer in question 7 that best represents your opinion

7. Overall, how much do you trust automated devices/systems?

1	2	3	4	5	6	7
Not at all						Completely

Automation Attitude Questionnaire

Please mark the appropriate response

1. I feel comfortable using automated devices.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

2. Automation will never replace the need for working human beings.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

3. Learning to use automated devices is a worthwhile and necessary subject.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

4. Reading or hearing about automated devices would be (is) boring.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

5. Automated devices are making the jobs done by humans less important.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

6. Automated devices make me nervous.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

7. I don’t care to know more about automation.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

8. Automated devices would be (are) fun to use.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

9. I <u>don’t</u> feel confident about my ability to use automated devices.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

10. People are smarter than automated devices.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

11. Automated devices are too fast.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

12. Automation is confusing.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

13. Given a little time and training, I know I could learn to use most automated devices.

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly

Automation Questionnaire - 2

Please mark the appropriate response

1. Automation and/or automated devices do not scare me at all.

1	2	3	4
Strongly Agree	Slightly Agree	Slightly Disagree	Strongly Disagree

2. Working with automated devices would make me very nervous.

1	2	3	4
Strongly Agree	Slightly Agree	Slightly Disagree	Strongly Disagree

3. I do not feel threatened when others talk about automated devices.

1	2	3	4
Strongly Agree	Slightly Agree	Slightly Disagree	Strongly Disagree

4. I feel aggressive and hostile toward automated devices.

1	2	3	4
Strongly Agree	Slightly Agree	Slightly Disagree	Strongly Disagree

5. Automated devices make me feel uncomfortable.

1	2	3	4
Strongly Agree	Slightly Agree	Slightly Disagree	Strongly Disagree

6. I get a sinking feeling when I think of trying to use an automated device.

1	2	3	4
Strongly Agree	Slightly Agree	Slightly Disagree	Strongly Disagree

7. I would feel comfortable working with an automated device.

1	2	3	4
Strongly Agree	Slightly Agree	Slightly Disagree	Strongly Disagree

8. Automation makes me feel uneasy and confused.

1	2	3	4
Strongly Agree	Slightly Agree	Slightly Disagree	Strongly Disagree

Trust Questionnaire
(No Survey Title Used During the Study)

1. Overall, how much do you trust the Engine Status Indicator?

1	2	3	4	5	6	7
Not at all						Completely

2. To what extent can you trust the Engine Status Indicator to do its job?

1	2	3	4	5	6	7
Not at all						Completely

3. The Engine Status Indicator always notified you on what you where supposed to do.

1	2	3	4	5	6	7
Disagree						Agree

4. The Engine Status Indicator is reliable

1	2	3	4	5	6	7
Disagree						Agree

5. Please indicate, using a number, the reliability of the Engine Status Indicator
(Example: I think the Engine Status Indicator was XX % reliable)

Participant Study Instructions

1. Tutorial/Practice: Thank you very much for completing those preliminary steps. Now, let's talk about the simulation you will interact with today.

a) Scenario: the task you'll be doing is similar to driving in that you have to pay attention to outside influences (such as other cars on the road, pedestrians, information signs, etc) and you have to monitor your vehicle's performance (such as speed or engine malfunction warnings). In this scenario, you will be piloting a flight simulator. You will have to do two things 1) pay attention to a radar screen and 2) pay attention to the airplane's engine performance gauges. Both things are very important. Your goal is to keep as many points as possible by fixing malfunctioning engines and selecting targets each time you notice them appear on the radar screen. The tutorial to follow will provide you a step-by-step explanation on how to do this.

b) System explanation: in a moment you will participate in a tutorial that will fully explain the brief explanations I am about to give concerning the flight simulator tasks you will be interacting with today. The flight simulator has two parts: the radar task and the engine task

- a. The "Radar Task"
  - i. The radar screen on the bottom portion of the screen will randomly display four symbols. Each time you notice a symbol appear on the radar screen, you should use the mouse to "click" its corresponding button above the radar screen as quickly as possible. You will have the opportunity to practice this in just a few moments.
- b. The "Engine Task"
  - i. The engine task consists of the following parts
    1. The Engine Status Indicator: It is supposed to let you know what is happening in the gauges. When the display is empty, the engines are supposedly operating as they were designed to operate or in a non-critical state. When the display indicates that there is a malfunction, it means the respective engine has entered a critical state. However, it is possible that the reliability of the engine status indicator changes throughout each block. Basically, there may be times when it is wrong. For example, it could be telling you that there is an engine malfunction when there really is not one or it might not give you a message when in fact one of the engines is malfunctioning.
    2. View gauges button: the engine gauges, which are always 100% accurate will not be visible at all times, the only time you can see them is when you hit the view gauges button. When you do this the gauges will become visible for 5 seconds.
    3. Engine Reset Buttons: These buttons are used to reset malfunctioning engines. When you choose to reset an engine you must press the button and then the OK button.

4. Action Results Window: This window will indicate whether or not the action you took was correct or not. It will also indicate if you missed the opportunity to reset a malfunctioning engine.

c) Points system: Your goal in this task is to try to get as many points as possible. There are several ways you can lose points such as:
   a. -5 for not resetting a malfunctioning engine
   b. -5 for resetting a properly working engine
   c. -5 for resetting the incorrect engine
   d. -5 for missing a target
   e. -5 for selecting a target not on the radar screen
   f. -1 for clicking the "View Gauges" button.

Start with 1,000 points
Topgun: 950 – 1,000 points
Ace: 900 – 949 points
Instructor pilot: 850 – 899
Pilot trainee: 800 – 849
Cadet: <800

d) Run Tutorial
   a. We will now proceed with the tutorial. If you have any questions about something you see on the screen, please let me know. You can go back at any point to view previous screens. Upon completion of the tutorial, you will have the opportunity to practice what you just learned.

2. Block 1:

Now we are ready to start the real blocks. We are done with the practice. Do you have any other questions before we begin block one? Ok, now let's try out the block one. Remember, your goals are to try and keep as many points as you can. Would you like to see the list of things that will cause you to lose points, one more time before we start? The tasks you will perform in this block and the subsequent one are exactly the same as those you practiced in the tutorial. This block will take approximately 20 minutes to complete. When your score is displayed, that will signify the completion of block one. At this time, please wait until told to proceed. Please start block one.

3. Block 2:

You're almost done. This block will be similar to the one you just completed. When you finish, I will need you to fill out one more survey and then I will provide you a debriefing and conduct a short exit interview. Thank you again for participating in this study. Please start block two.

Trust in Automation (effects of error type)
Debriefing Information

Thank you very much for participating in this study.

This study was conducted to help us understand if the type of error that an automated system makes affects the trust people have in the automated system with which they are working. The system you were working with made two types of error, false alarms and misses. A false alarm occurs when the system says there is a failure and no failure really exists. A miss occurs when the system fails to tell you that a real failure has occurred. Previous research suggests that excessive or untimely false alarms can have a negative impact on an operator's trust in automation; this as been coined the "Cry wolf effect." Very little has been done, however, to examine if the same is true for instances when an automated system fails to notify its user of an actual malfunction or impending malfunction. We, therefore, set up conditions in which a participant interacted with an Engine Status Indicator that made an equal number of false alarms and misses, mostly false alarms, or mostly misses. Based on the type of errors experienced, task performance (engine and radar task), the number of times the view gauges button was pressed, and the answers the participants provided on the questionnaires, we'll be able to determine how trust changed, if at all.

We are also going to compare the answers of younger adults (aged 18-28) and older adults (aged 65-75) to assess whether trust differs across these age groups. Almost all of the research conducted to understand trust in automation has only tested younger adults. Because more and more systems encountered in daily life are automated systems, it is important to understand the factors that effect trust not only for younger adults but also for older adults. It may be that the younger adults, especially those at Georgia Tech, have had more experience with automated systems and have less tolerance for mistakes made by the automation (i.e., they expect near perfect performance from automated systems). Or, it may be that participants in both age groups' trust ratings differ similarly based on the type of automation error they experienced.

We will share a summary of the results with you by mailing you a newsletter. Because each individual's data and test scores are completely confidential, there will be no way for us to mail your individual results.

It is hoped that your data will provide valuable insight into how automation error type affects trust and reliance in automation. If you have any questions, please feel free to contact us.

Thank you for your time and cooperation.

Human Factors and Aging Lab

Georgia Institute of Technology (404) -894- 8344

Dr. Arthur D. Fisk & Dr. Wendy A. Rogers

REFERENCES

Barber, B. (1983). *Logic and the limits of trust*. Rutgers University Press: New Brunswick, NJ.

Bliss, J. P. (1997). Alarm reaction patterns by pilots as a function of reaction modality. *International Journal of Aviation Psychology, 7*(1), 1-14.

Bliss, J. P., & Dunn, M. C. (2000). Behavioral implications of alarm mistrust as a function of task workload. *Ergonomics*, *43*, 1283-1300.

Breznitz, S. (1984). *Cry wolf: The psychology of false alarms*. Lawrence Erlbaum Associates: Hillsdale, NJ.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*, 697-718.

Harris, H. & Provis, C. (2000). Teams, trust and norms: There's more to teams as a manufacturing management system than some might think. *Proceedings of 8th International Conference on Manufacturing Engineering. Sydney, Australia.*

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). North-Holland: Elsevier Science Publishers B.V.

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

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

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709-734.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*, *6*(1), 44-58.

Parasuraman, R., Mouloua, M., Molloy, R., & Hilburn, B. (1996). Monitoring of automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 91-115). Mahwah, NJ: Lawrence Erlbaum Associates.

Park, K. S. (1997). Human error. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed.) (pp. 150-173). New York: Wiley-Interscience.

Sanchez, J., Fisk, A. D., & Rogers W. A. (2004, March). *Age-related and reliability related effects on trust of a decision support aid.* Presented at the 2004 Human Performance, Situation Awareness and Automation Technology Conference, Daytona Beach, FL.

Shipley, W. C. (1986). *Shipley Institute of Living Scale*. Los Angeles: Western Psychological Services.

Skitka, L., Mosier, K., & Burdick, M. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies*, *52*, 701-717.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale III* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wiegmann, D., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: the effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science, 2*(4), 352-367.

Xu, X., Wickens, C. D., & Rantanen, E. (2004). *Imperfect conflict alerting systems for the cockpit display of traffic information* (Technical Report AHFD-04-8/NASA-04-02). Savoy, IL: University of Illinois at Urbana-Champaign, Institute of Aviation, Aviation Human Factors Division.

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. Basic Books: New York.