Summer 2011

The Effects of System Reliability and Time Pressure on Unoccupied Aircraft Systems Operator Performance and Mental Workload

Rania Wageh Ghatas
Embry-Riddle Aeronautical University - Daytona Beach

THE EFFECTS OF SYSTEM RELIABILITY AND TIME PRESSURE ON UNOCCUPIED AIRCRAFT SYSTEMS OPERATOR PERFORMANCE AND MENTAL WORKLOAD

by

RANIA WAGEH GHATAS
B.S. in Molecular Biology and Microbiology, University of Central Florida, 2009
B.S. in Psychology, University of Central Florida, 2009

A Thesis Submitted to the
Department of Human Factors & Systems
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Human Factors and Systems Psychology.

Embry-Riddle Aeronautical University
Daytona Beach, Florida
Summer, 2011

The Effects of System Reliability and Time Pressure on Unoccupied Aircraft Systems Operator Performance and Mental Workload

By: Rania Wageh Ghatas

This thesis was prepared under the direction of the candidate's thesis committee chair, Dr. Dahai Liu, Ph.D., Department of Human Factors & Systems, and has been approved by members of the thesis committee. It was submitted to the Department of Human Factors & Systems and has been accepted in partial fulfillment of the requirements for the degree of Master of Science in Human Factors and Systems Psychology.

THESIS COMMITTEE:

Dahai Liu, Ph.D., Chair

Christina Frederick-Recascino, Ph.D., Member

John A. Wise, Ph.D., Member

Master of Science in Human Factors and Systems Program Coordinator

Department Chair, Department of Human Factors and Systems

Associate Vice President for Academics

Acknowledgements

I would like to gratefully acknowledge the enthusiastic supervision of Dr. Dahai Liu during this work. I thank him for always helping me through encouragement, feedback and comments, and taking time out of his busy schedule in order for me to finish my thesis and do the best work that I can. Thank you also to my committee members, Dr. John A. Wise and Dr. Christina Frederick-Recascino, who have both encouraged me with their support, comments, and feedback and who have taken the time to be part of my committee and helping me.

I would like to thank all my great friends, my friends/colleagues at the Next-Generation ERAU Advanced Research (NEAR) Lab, and the Human Factors & Systems faculty and staff who have always believed in me.

I am forever indebted to my amazing mother, Seham Metri, and my amazing brother, Remon Ghatas, whom God has tremendously blessed me with. I gratefully thank them both for everything they do for me; for all their continuous love and support, always believing in me, and encouraging me to follow my dreams. I would also like to thank my dad, Wageh Ghatas, who is in Heaven but I know is watching over me. Also, thank you to my family, who are all over the world, for encouraging me and believing in me.

Most importantly, I thank my Father in Heaven, my Lord and Savior Jesus Christ, Who through Him I can do all things.

"I can do all things through Christ Who strengthens me." – Philippians 4:13

Abstract

Author: Rania Wageh Ghatas

Title: The Effects of System Reliability and Time Pressure on Unoccupied Aircraft Systems Operator Performance and Mental Workload

Institution: Embry-Riddle Aeronautical University

Year: 2011

Unoccupied Aircraft Systems (UAS) are in the midst of aviation's next generation. UAS are being utilized at an increasing rate by military and security operations and are becoming widely popular in usage from search and rescue and weather research to homeland security and border patrol. The Federal Aviation Administration (FAA) is currently working to define acceptable UAS performance standards and procedures for routine access for their use in the National Airspace System (NAS). This study examined the effects of system reliability and time pressure on unoccupied aircraft systems operator performance and mental workload. Twenty-four undergraduate and graduate students, male and female, from Embry-Riddle Aeronautical University participated in this study on a voluntary basis. The primary tasks were image processing time and target acquisition accuracy; three secondary tasks were concerned with responding to events encountered in typical UAS operations. Mental workload, using the NASA-Task Load Index (TLX) form, and trust levels of Multi-Modal Immersive Intelligent Interface for Remote Operation (MIIIRO) system were also studied and analyzed. System reliability was found to produce a significant effect for image processing time, while time pressure produced a significant effect for target acquisition accuracy and mental workload. A significant effect was found for the interaction between system reliability and time pressure for pop-up threats re-routing processing time. The results were examined and recommendations for future research are discussed.

Table of Contents

List of Tables

List of Figures

Acronyms

ADS-B	Automatic Dependent Surveillance – Broadcast
ATC	Air Traffic Control
ATM	Air Traffic Management
ATR	Automatic Target Recognition
DNRS	Digitally Networked Radio System
DoD	Department of Defense
DV	Dependent Variable
FAA	Federal Aviation Administration
GCS	Ground Control Station
IA	Intruder Aircraft
IMD	Image Management Display
IV	Independent Variable
LS	Latin Square
MIIIRO	Multi-Modal Immersive Intelligent Interface for Remote Operation
MMI	Mission Mode Indicator
NAS	National Airspace System
NASA	National Air and Space Administration
NASA-TLX	NASA-Task Load Index
NVS	NAS Voice System
NextGen	Next Generation Air Transportation System
SMS	Safety Management System
SWAT	Subjective Workload Assessment Tool
TSD	Tactical Situation Display
UA	Unoccupied Aircraft
UAS	Unoccupied Aircraft Systems
UAV	Unmanned Aerial Vehicle
U.S.	United States
VoIP	Voice over Internet Protocol

Introduction

The next generation of flight is already underway and Unoccupied Aircraft Systems (UAS) are in the midst of aviation's next generation. Since their pre-aviation history, these systems have proven themselves to be versatile, efficient, and valuable, and as such approximately 50 companies, universities, and government organizations in the United States (U.S.) alone are developing and producing some 155 unoccupied aircraft designs (Dorr & Duquette, 2010). These same entities that are developing and producing these "spies in the sky" are also spending billions of dollars in research efforts to improve UAS in terms of hardware, software, and human-system interactions. Through years of research and development, UAS are flying toward a future with many possibilities.

Unoccupied Aircraft Systems (UAS)

UAS, or Unoccupied Aircraft Systems, is the term that is on the verge of replacing Unmanned Aircraft Systems, which was the official term coined by the Federal Aviation Administration (FAA) and introduced by the U.S. Department of Defense (DoD) to replace the term Unmanned Aerial Vehicle (UAV). A typical UAS consists of the Unoccupied Aircraft (UA), the Control System, the Datalink, and other related support equipment (Unmanned Aerial Vehicle, n.d.). UAV's come in a variety of shapes and sizes and have wingspans that could range as large as a Boeing 737 or even smaller than a radio-controlled model airplane. They serve diverse services and are becoming widely popular in usage from search and rescue and weather research to homeland security and border patrol. Until recently, military and security operations were mainly supported by UAS; however, due to their increasing popularity, unoccupied aircraft are growing to support "aerial photography, surveying land and crops, monitoring forest fires and environmental conditions, and protecting borders and ports against intruders" (Dorr &

Duquette, 2010) as they promise new innovations to increase efficiency, save money, and enhance safety and even save lives. The UAS consists of the UAV, the Ground Control Station (GCS), and other related support equipment. Since UAVs are unoccupied, a pilot on the ground is in charge of UAS operations at all times. Figure 1 shows a picture of a U.S. Department of Homeland Security Customs and Border Protection Predator UAV.

Figure 1: U.S. Department of Homeland Security Customs and Border Protection Guardian/Predator UAV (From ALEA E-Newspaper, 2010)

UAS History

Past (Pre-aviation to 1980's)

UAS history dates back to pre-aviation times; a time before manned aircraft even took flight. The primitive technology of UAVs used for surveillance and combat has existed well before manned aircraft first took flight on December 17, 1903. The first UAV took flight in the U.S. in the 1910's during World War I when the military took extra notice of their potential in combat after their success in test flights. For more than a decade after the end of World War I, pilotless aircraft development in the U.S. and abroad drastically declined; however, new UAVs emerged on the scene during the mid-to-late 1930's as an important combat training tool. In the

1940’s, in the occurrence of World War II, the U.S. laid groundwork for post-war UAV programs in light of Nazi Germany’s innovative V-1, which demonstrated the “formidable” threat that a UAV could impose in combat. The 1960’s and 1970’s brought on a new era of UAVs, which took on a new role during the Vietnam War: “stealth surveillance” and when the U.S. set its sights on other types of UAVs. An aggressive UAV developer, the Israeli Air Force, pioneered several new and important UAVs, which were integrated into the UAV fleet of the U.S. and many other countries during the late 1970’s and throughout the course of the 1980’s (Krock, 2002).

Present (1990’s to today)

UAVs in the 1990’s became a permanent and critical position in high-tech military arsenals ranging from the U.S. and Europe to Asia and the Middle East. UAVs are also playing key roles in keeping the peace of the Earth’s environment by commanding a monitoring role (Krock, 2002). Currently, military and government agencies represent the major players for the operation of UAVs; however, a large call for the expansion of UAVs into domestic and commercial operations is on the rise, such as usage in law enforcement settings. For this reason, the FAA, along with many private and educational agencies, is extensively researching the integration of UAS into the National Airspace System (NAS) with safety at the forefront of the research. Today’s UAS demands are paving the way for the future of the national airspace.

Future and the Next Generation Air Transportation System

As aviation’s portal to the future, the Next Generation Air Transportation System (NextGen) is transforming the way America flies using 21st century technology. NextGen will combine increased automation with new procedures to achieve economical, capacity, safety, environmental, and security benefits by the year 2025 (Prevot, Lee, Smith, & Palmer, 2005).

NextGen will be better for the environment and the economy (Federal Aviation Administration "Why NextGen Matters," 2011) by:

- Allowing travel to be predictable due to fewer delays.
- Reducing aviation's impact on the environment. Flying will be quieter, cleaner, and more fuel-efficient.
- Being more proactive in preventing incursions.
- Getting the necessary information to the right person at the right time.
- Improving the nation's economy.
- Making better use of airports by attracting new jobs and helping current employers expand their business.
- Meeting the needs of increasing national security and providing the highest levels of safety for travelers.

One of the strategic objectives of NextGen is to make "the National Airspace System (NAS) scalable and flexible enough to incorporate various and new types of aircraft," including UAVs. The FAA is currently working to "define acceptable UAS performance standards and procedures for routine access, all while maintaining safety" in order to alleviate existing restrictions associated with UAS operations. This will be done by improving UAS operations by using state-of-the-art technologies, such as Voice over Internet Protocol (VoIP); NAS Voice System (NVS); Digitally Networked Radio System (DNRS), which is used as an interim solution for UAS operations in the event of Loss of Voice Communication and may be utilized as a backup, or even primary means of communications network between Air Traffic Control (ATC), UAS GCS, or within Air Traffic Management (ATM); and Automatic Dependent Surveillance – Broadcast (ADS-B), which uses global positioning system satellites and on-board technology that are superior to radar and provides broader and more precise coverage of an aircraft's position, altitude, and velocity; among many other technologies (Federal Aviation Administration Task E: UAS Demo (3) Test Plan, 2010). These new technologies, some of which are currently being used in limited capacities, will greatly help with many integration issues, such as the UAV's ability to "sense-and-avoid," UAV Pilot situational awareness, and most importantly safety through the Safety Management System (SMS), which consists of four components: safety policy, safety risk management, safety assurance, and safety promotion (Figure 2) (Federal Aviation Administration Task E: UAS Demo (3) Test Plan, 2010).

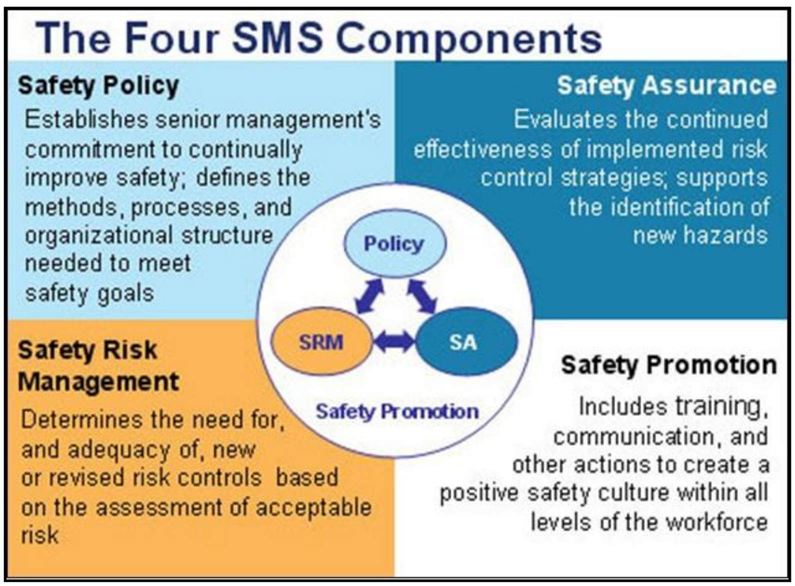

Figure 2: The Four Safety Management System (SMS) Components (From Federal Aviation Administration Task E: UAS Demo (3) Test Plan, 2010)

Human Factors in UAS

Interest in using UAS for an extensive range of purposes is quickly increasing within the aviation community, making UAS access to the NAS a priority (Federal Aviation Administration Task E: UAS Demo (3) Test Plan, 2010). The increasing popularity of UAS is creating a dilemma that is: how will they be integrated into the NAS? This question has brought the attention of research and development in military, government, and civil entities. Many critical issues have arisen in integrating UAVs into an airspace that is already populated with commercial airlines and private jets. Among these many issues are (Hottman & Sortland, 2006):

- Ground Control Stations: This issue relates to whether or not a single or multiple UAVs will be operated. Knowing this question can help in determining how many operators should be in the control station and what functions they should perform.
- UAV operator qualifications: Questions such as "what attributes and skills does the operator possess?" "How well does he or she know the system?" "What kind of background do they have?" "How much flying experience do they have?" All of these questions plus much more should be taken into consideration when screening and testing potential UAV operators. The degree to which they are qualified is important to keeping a safe environment.
- Validation of sense-and-avoid technologies: Making sure the sense-and-avoid technologies are all working as they should and doing what they are required to do is essential in the design and operating principles of UAVs.
- UAV call signs: Call signs are important to have when it comes to knowing which UAV is which. It also helps in knowing its location and what its intended mission is.
- UAV communication with ATC: Having a communication standard between the operator of the UAV and ATC is crucial when dealing with safety and security. This would allow the UAV operator to know what is going on in the surrounding airspace of the UAV and know whether or not it is entering any restricted space where a commercial flight or general flight may be in order to not cause any incursions.

UAS operations are multi-faceted and complex. Understanding what those operations entail and the automation behind them could help in determining how to use combinations of automation with UAS procedures in making the system a better one that not only decreases key human factors issues, such as workload on an operator, but also allows an operator to maintain full control at all times (Prevot, et al., 2005).

Humans and Automation

Today, humans tend to rely heavily on technology and automation. For many, going through a day without some sort of technology would be unheard of with the amount of dependency society has on technology. Automation is highly valuable, especially in dangerous and sensitive environments. Designers automate systems to replace or aid human performance for many reasons; however, these reasons could be generally placed into four categories (Wickens, Lee, Liu, & Becker, 2004):

1. Impossible or hazardous: This reason deals with automating process that would normally be either dangerous or impossible for humans to perform, such as in handling hazardous materials.
2. Difficult or unpleasant: While not impossible for a human to perform, some tasks may be very challenging or difficult for a human to properly perform, such as multiplying large numbers together. Although this may be done in the head or by multiplying on a piece of paper, it is usually easier and quicker to do using a calculator.
3. Extend human capability: This category deals with aiding, not replacing, the human operator to perform something in otherwise difficult circumstances.

4. Technically possible: Even though this category may provide little to no value to the human user, sometimes systems are automated simply because they can be; the technology is there and is inexpensive to do.

A fine balance between humans and automation needs to exist in order to reap the many benefits automation has to offer.

Automation Reliability

Throughout the years, automation has played significant roles in how the world operates and sometimes it may even seem as though automation is taking over how the world operates. Automation has a number of many real *benefits* in aiding human performance and alleviating humans from performing dangerous or tedious jobs (Wickens, et al., 2004), but like everything else in the world, automation is not perfect and thus comes with its own set of issues, literally and metaphorically. Researchers in the field of Human Factors have spent years researching different types of automation and developing statistics based off of a multitude of data gathered from experimental studies in order to integrate automation as user-friendly and user-centered in today's fast paced and highly technology dependent society. The problem with automation is not merely due to failures with hardware and software components, but rather with the system problems of automation that "are distinctly and inexorably linked to human issues of attention, perception, and cognition" in administering the automated system in its normally operating state (Wickens, et al., 2004). Although having systems that can be fully automated or semi-automated help ease the workload of the individual using it, these systems have a set of draw-backs. One issue with autonomy in UAS is the question of "how much automation should the system have" and "what part of the system should be controlled by the operator?" This issue leads to design challenges of the system for engineers in terms of how the system should be designed to where it helps the operator maintain productivity without work or cognitive overload but at the same time keep the operator active so as to not lose situational awareness. Another issue is sub-tasks with the question of "will a single operator or a team of operators need to conduct multiple tasks at once?" This issue is concerned with the causation of work overload, a decrease in situational awareness, and confusion, among other factors. It requires operators to use many cognitive skills that could be tiring and lead to a decrease in productivity (Weil, Freeman, MacMillan, Jackson, Mauer, Patterson, & Linegang, 2006). Other issues consist of displays and controls; automation and system failures; and crew composition, selection, and training (McCarley & Wickens, 2004). Additional issues, in relation to human-centered automation (Table 1), include workload; operational situation awareness and system-mode awareness; automation dependencies and skill retention, and interface alternatives (Wise, Hopkin, & Garland, 2010).

Table 1: Human-Centered Automation (From Wise, et al., 2010)

Workload	1. Too little workload in some phases of flight and parts of air-traffic control (ATC) operations to maintain adequate vigilance and awareness of systems status 2. Too much workload associated with reprogramming when flight plans or clearances change 3. Transitioning between different levels of workload, automation-induced complacency, lack of vigilance, and boredom on flight deck, ATC, and monitoring of system and service performance
Operational situation awareness and system-mode awareness	1. The ability of operators to revert to manual control when the advanced automation equipment fails 2. An inadequate "cognitive map," or "situational awareness" of what the system is doing 3. Problematic recovery from automation failures 4. The potential for substantially increased head-down time 5. Difficulty and errors in managing complex modes
Automation dependencies and skill retention	1. The potential for controllers, pilots, and others to over-rely on computer-generated solutions (e.g., in air-traffic management and flight decisions) 2. Hesitancy of humans to take over from an automated air-traffic and flight deck system 3. Difficulty in maintaining infrequently used basic and critical skills 4. Capitalizing on automation-generated alternatives and solutions 5. Monitoring and evaluating pilot and controller skills where computer-formulated solutions disguise skill weaknesses 6. Supporting diagnostic skills with the advent of systems that are more reliable and feature built-in self-diagnostics (e.g., those in "glass cockpit" systems and fully automated monitoring systems)
Interface alternatives	1. Major system-design issues that bridge all the aviation operations including selecting and presenting information for effective human–computer interface 2. Devising optimal human–machine interfaces for advanced ATC systems and for flight deck avionics 3. Devising strategies for transitioning to new automation technologies without degrading individual or contemporary system performance

Despite of the long list of human factors issues with automation of all types, including UAS automation, it is necessary to narrow research efforts to specific issues of automation in order to produce progress in making the automation better for the operators. As such, many military, government, and civil entities are tirelessly conducting research in hopes of improving UAS automation, such as the FAA's NextGen Flight Deck Human Factors Research and Development Program. In conjunction with other FAA research and development programs, this program is aiming to identify and resolve human factors issues through research activities (Federal Aviation Administration "NextGen Flight Deck," 2009).

Over Trust of the Automation

Over trust, also known as complacency or misuse, happens as a result of humans placing too much trust in automation, which could lead to severe negative consequences if the automation in question is not fully reliable. Over trust is likely to occur when an individual uses a particular system that rarely encounters failures; hence the skewed perception of the individual, which leads to the belief that the automation is perfect. The problem that arises from complacency, or the failure to monitor the system adequately and thus causing problems for the human to properly intervene when a system failure takes place, is due to the following three distinct categories:

(1) Detection: The complacent operator would take a longer time than what is necessary to detect a real failure in a system. It is said that "the more reliable the automation, the rarer the 'signal events' become, and the poorer is their detection" (Parasuraman et al., 1996 from Wickens, et al., 2004).

(2) Situation awareness: An individual who is actively participating or monitoring something has a greater dynamic state of awareness and of his or her surroundings in comparison to an individual that is passively monitoring something or someone (Wickens, et al., 2004). This phenomenon is known as the generation effect (Slamecka & Graf, 1978; Endsley & Kiris, 1995; Hopkin & Wise, 1996) and results in a less likelihood of that individual intervening correctly or appropriately as a result of being out of the loop and not fully understanding the momentary state of the system (Sarter & Woods, 2000). In addition to this, situation awareness becomes even more problematic if the system is poorly designed to give adequate feedback in regards to the ongoing state of the automated process.

(3) Skill loss: As automation increases, operator skills tend to decrease. Wiener (1988) describes a skill loss term known as *deskilling*, which refers to the "gradual loss of skills" than an operator may inherently experience due to not being actively participating in the control operation or decision making process of an automated system. Losing the ability to remember certain skills and operations may have two implications on the operator (Lee & Moray, 1994). These two implications on the operator include: (1) becoming less confident in his or her own performance and thus more likely to continue using the automation to do everything, and (2) degrading the operator's ability to appropriately intervene should a failure in the system occur. Additionally, another implication may arise depending on how far the skills of the operator degrades in which the automation may end up masking the incompetence of the operator.

In addition to the above mentioned implications, the automated system may sometimes fail and hand over responsibility to the operator, which is usually the most challenging of problems for the human. As a result, more problems may arise in overcoming the system's failure and dealing with the situation at hand due to the operator's complacency in the automated system and decrease in skill level (Hopkin & Wise, 1996). Figure 3 shows the elements in which automation reliability and human trust correspond.

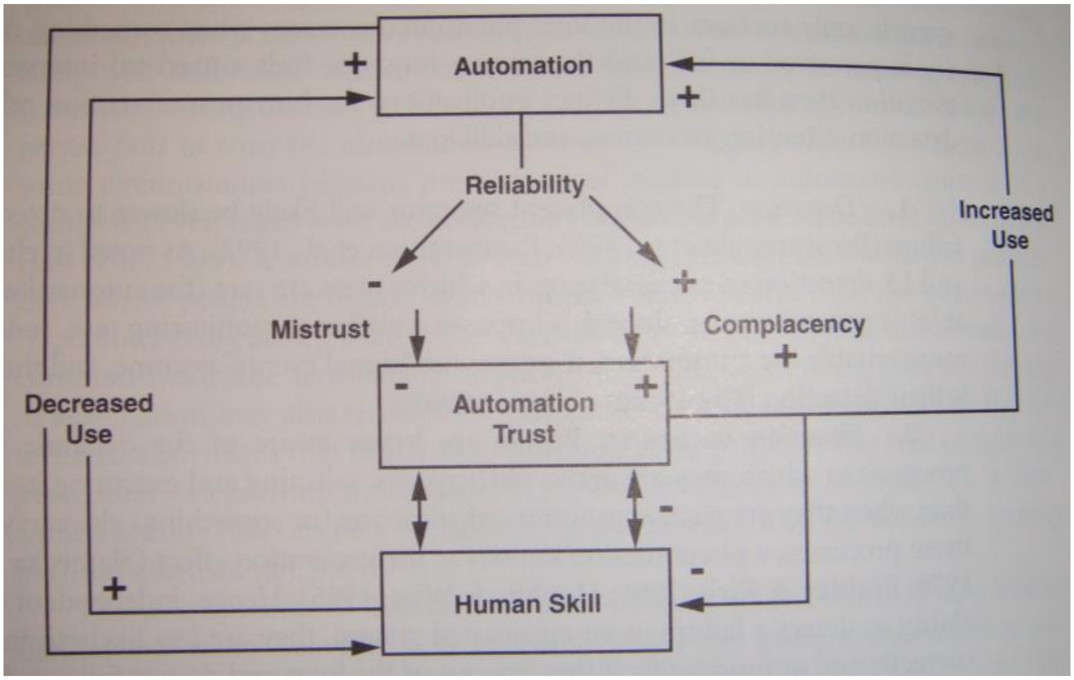

Figure 3: Elements of automation reliability and human trust (From Wickens, et al., 2004)

Human-Computer Interfaces

Precursors to modern computer technology can be traced back to ancient times with the age of electronic digital computers beginning roughly in the middle of the 20th century (Helander, Landauer, & Prabhu, 1997). Despite the fact that computers have been around for decades and have enormously impacted modern society; however, they are considered to be relatively new tools as they exist in a phase of rapid development and tend to be complex in nature. Many pieces of hardware and software that make up what is known as the modern computer are crucial to its function and so is its design and interface. General interface designs should incorporate eight principles in order to make the interface user-friendly. These eight principles (Wickens, et al., 2004) consist of:

1. Match between system and real world, such as speaking the user's language and using familiar conceptual models or metaphors.

2. Consistency of the interface internally, such as same information types located in same location throughout, and with respect to any existing standards.
3. Visibility of system status; make sure the user is informed about what is going on with the system.
4. Allowing the user to maintain control of the system and freedom to initiate actions
5. Error prevention, recognition, and recovery methods that help users recognize, diagnose, and recover from errors, if any arise, in addition to clear and explicit error messages.
6. Memory; providing lists of choices and picking from lists, using see-and-point instead or remember-and-type.
7. Flexibility and efficiency of use by providing shortcuts and the ability to initiate, reorder, and/or cancel tasks.
8. Simplicity and aesthetic integrity with all information appearing in a natural and logical order.

Although these eight principles should be applied to the design of human-computer interfaces, this is not always the case and as a result, many errors, whether from the human operator or from the computer system, arise. Many researchers have, and are currently, researching the issue of human-computer interactions and their interfaces, even though arguments have been brought up to the futile nature of these researches and have marked them unnecessary. However, these arguments bear little weight as research on human-computer interactions serve as a building stone for future technology "by revealing the aspects of human tasks and activities most in need of augmentation and discovering effective ways to provide it" (Helander, et al., 1997). Through these research studies, a multitude of information has shown what makes a good interface design and what doesn't, but more importantly which of these good designs are tailored around the end-user. Human factors issues that relate to interface designs are generally due to the lack of human-computer interactions. These interactions are necessary as many times the computer is either too difficult for the human to understand or the computer is simply too automated, leaving the human user unaware of the system's status.

UAS System Reliability and Trust

Technology has quickly immersed itself into society and the UAS is no exception. The relation between reliability and trust in automation is a critical one when discussing its importance in relation to human performance issues. Within the human factors context, the reliability of automation is said to be the extent to which it does what the human operator expects it to do, such as a copy machine that faithfully reproduces the number of pages requested or a car's cruise control that holds the car at a set speed. However, the reliability of automation is not the main concern, but rather its *perceived reliability*. At least four reasons exist as to why automation may be perceived as unreliable. First, considering that automated systems are complex and have more components than manually operated systems, the automation may be unreliable as a result of design flaws ending in component failure. Second, the automation's system design may not be suitable for certain tasks that the automation is performing as all automation have a limited operating range within which designers assume it will be used. Third, the automation may be incorrectly set up by the operator. Fourth, at times the logic behind the automation is too complex for the operator to understand whether or not it is performing correctly, whereas in actuality the automation is doing exactly what it is meant to do. An important aspect in automation reliability is realizing that often times the automation is asked to perform certain tasks that are themselves dynamic and uncertain in nature, such as weather forecasting or predicting enemy intent; therefore, it would be simply impossible for the automation to perform at a high level of reliability (Wickens, et al., 2004).

Trust "should be in direct proportion to its reliability," whether it is in another human or in a computer, and should be well "calibrated;" however, there is evidence that the trust a human puts into automation is not entirely well calibrated. Inappropriate calibration of trust, or "automation bias," (Mosier, Skitka, Heers, & Burdick, 1998) is defined as either too high or too low; too high a trust level in automation could lead to complacency, whereas too low a trust level could lead to distrust with the system going un-used. Distrust is a type of mistrust, which occurs when trust in something is not directly related to its reliability. Distrust in automation may occur due to a number of reasons, such as the failure to understand the nature of the automation and its algorithms, which may lead to inefficiency where the human may "reject the good assistance that automation can offer" (Wickens, et al., 2004) and prefer to do a certain task by hand, resulting in slower performance, less accuracy, and an increase in workload. Two terms that exemplify the inappropriate levels of trust are misuse and disuse. Misuse refers to the failures that crop up when people rely on automation inappropriately, while disuse refers to the failures that arise when people discard the capabilities of the automation (Parasuraman & Riley, 1997). Trust in automation by an operator needs to be appropriately calibrated as extremes could pose to be dangerous. Figure 4 shows the relationship between trust and automation reliability.

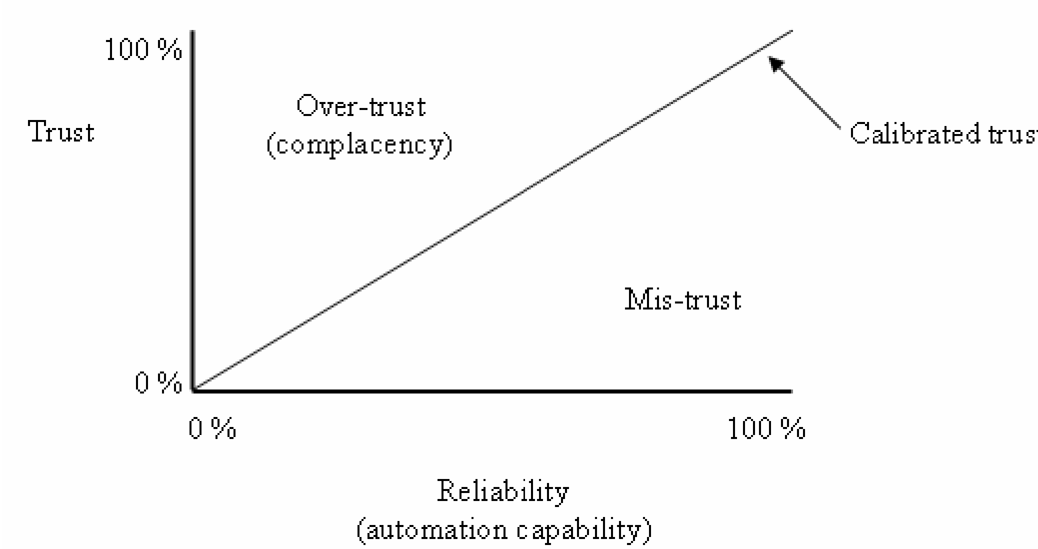

Figure 4: Relationship between trust and automation reliability (From Nisser & Westin, 2006)

Mental Workload

Mental workload is a primitive construct term that is fundamentally complex and multifaceted, which "'everyone knows,' but hardly anybody can define in precise, operationally useful terms" (Hancock & Meshkati, 1988). Sarno and Wickens (1995) refer to workload as the relationship between resource supply and task demand. It will be assumed that mental workload is representative of the cost that is incurred by the human operator to achieve a particular level in his or her performance and will thus be defined as human-centered, rather than task-centered. The subjective workload that an individual experiences consists of the influences of multiple and simultaneous factors in addition to objective demands imposed on the individual by a certain task. Workload can be summarized as an innate property that emerges from the "interaction between the requirements of a task, the circumstances under which it is performed, and the skills, behaviors, and perceptions of the operator" (Hart & Staveland, 1988). Figure 5 shows a conceptual framework relating human performance and workload.

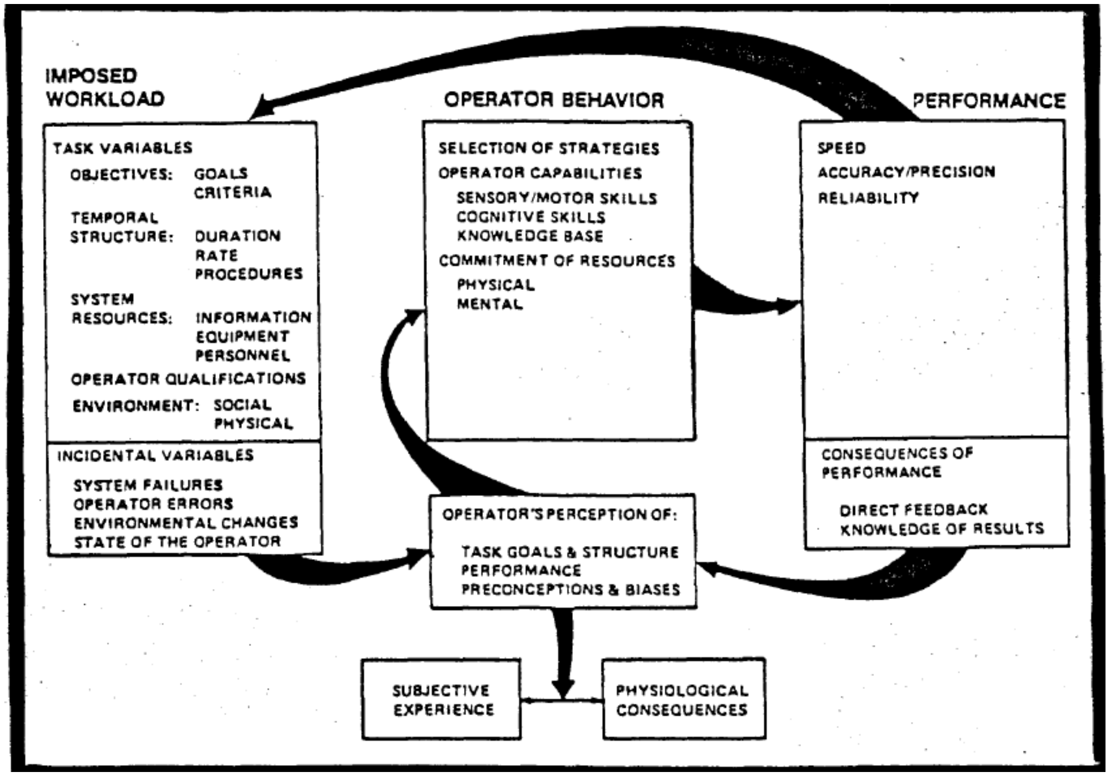

Figure 5: Conceptual Framework relating Human Performance and Workload (From Hart & Staveland, 1988)

In order to understand workload in a relatively easy manner, the concept could be described in terms of a ratio, TR/TA, which is representative of the time required (TR) to do a certain task/s to the time available (TA) to complete them. A workload timeline model depicting the different tasks that need to be performed and how long those tasks would typically be outlined if a researcher wishes to calculate the workload an operator would experience in a particular environment. This ratio calculation can be derived on the basis of a careful task analysis and is designed to accomplish two objectives, which include: (1) predicting how much workload a human would experience, and (2) predicting the extent to which performance would suffer due to workload overload. Thus, mental workload can be generally defined as the ratio of the resources required to the resources available (Wickens, et al., 2004).

Workload's innate property in the operation of UAS is crucial to an operator's performance and perceived mental workload. To alleviate the downside of workload, the implementation of automation can come in handy for a myriad of various tasks; making certain tasks plausible for the operator to perform more complex tasks for which they are better suited. High levels of automation, however, can affect the workload of a UAS operator in ways that require further scrutiny.

Research has found that allocation of flight control to automation led to higher performance on simultaneous target identification and system failure identification tasks, these, however, were attributed to the reduced level of workload (Dixon, Wickens, and Chang, 2003). High levels of workload can lead an operator performing detrimentally; however, levels of workload that are too low can cause the same effect (Crescenzio, Miranda, Periani, & Bombardi, 2007). Low workload levels often cause the operator to lose track of tasks the system is performing and lose situation awareness, which inevitably cause problems in the operator's performance levels. As a side effect of low workload levels, an increase in workload takes place when the operator needs to become involved in the system again as he or she attempts to regain the lost situation awareness, such as during an unforeseen event.

Measuring Mental Workload

As the modern world quickly immerses itself into the world of technology, the need to subjectively measure the mental workload of an individual performing certain tasks in particular environments is growing. The ability to understand and apply the findings of an individual's perceived mental workload is a sought after need and want by researchers in the field of human factors and ergonomics as searches for "higher levels of comfort, satisfaction, efficiency, and safety in the workplace" (Rubio, Diaz, Martin, & Puente, 2004) become more demanding. The evaluation of mental workload is dependent upon the following seven requirements (Rubio, et al., 2004):

1. Sensitivity: The ability for an assessment tool to detect changes in a task's difficulty or demands.
2. Diagnosticity: The ability for an assessment tool to identify changes in workload variation and the reason for such changes.
3. Selectivity/Validity: The ability for an assessment tool to differentiate sensitivity levels in changes to cognitive demands and not to changes in other variables, such as physical workload or emotional stress that are not essentially associated with mental workload.
4. Intrusiveness: The ability for an assessment tool to not interfere with the primary task performance, the load that is the actual object of evaluation.
5. Reliability: The ability for an assessment tool to consistently reflect the mental workload.
6. Implementation requirements: The inclusion of aspects related to such things as time, instruments, and software that are used for the collection and analysis of data.
7. Subject acceptability: The perception of the subject, or participant, of the validity and usefulness of the procedure.

With the ability to evaluate systems and assess operator mental workload, subjective workload measuring assessments have become increasingly important tools. Their growing popularity, ease of implementation, and frequent usage are due to their practical, non-intrusive, nature that aid in the collection of data and their ability to provide sensitive measures of operator load. As automation in human-machine systems is becoming more complex, the ability to evaluate an operator's performance is becoming more difficult, which, in turn, makes the need for subjective mental workload assessment even the more critical.

NASA-Task Load Index (TLX) Assessment Tool

The NASA-Task Load Index (TLX) Assessment Tool is one, among many, subjective mental workload assessment tools that is used to assess workload in a variety of human-machine environments. It is a multi-dimensional rating procedure that "derives an overall workload score based on a weighted average of ratings on six subscales" (NASA TLX: Task Load Index, n.d.) that includes: mental demand, physical demand, temporal demand, performance, effort, and frustration. It is assumed that some combination of these six subscales, or dimensions, is likely to represent an individual's experienced workload while performing certain tasks (Hart, 2006). Table 2 shows the descriptions of each of the above mentioned dimensions (Rubio, et al., 2004). The measurement of these dimensions in regards to a UAS operator is important in order to understand the operator's perceived mental workload. The weighting scheme that is used by this assessment tool was designed to increase assessment sensitivity, to relevant variables, while decreasing between-rater variability (Hart, 2006).

Table 2: NASA-Task Load Index (TLX) Dimensions Descriptions (From Rubio, et al., 2004)

TITLE	ENDPOINTS	DESCRIPTIONS
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

The NASA-TLX is available in either a traditional paper-pencil version or a computerized version. Since its development by the Human Performance Group at NASA's Ames Research Center, it has been subjected to a variety of independent studies in the evaluation of its assessing of reliability, sensitivity, and utility, in addition to being compared to other methods of measuring workload (Hart, 2006).

Time Pressure

Time pressure has become an increasingly prominent feature in work environments. Time pressure has been defined as "either subjectively perceived time pressure or the imposition of a deadline," which increases the rate of individual and group performance (Kelly & Karau, 1993, 1999 in Amabile, Mueller, Simpson, Hadley, Kramer, & Fleming, 2002). A "time famine" has been identified from both the business press and organizational literature in which individuals feel as though there are “never enough hours in the work day” (Perlow, 1999 in Amabile, et al., 2002). In an environment that is time sensitive, such as that of UAS operations, the effects of time pressure may be too great. Time pressure is a critical issue when it comes to performing a specific task or duty; having a constricted time limit to complete something adds on stress to an individual in addition to increased mental workload as cognitive processes work extra hard to meet the demands constricted upon it. It also leads to “faster motions in completing a task,” which may cause physiological stress, such as strain on the muscles (Hughes, 2004), on top of the stress already imposed on cognitive processes. Extensive research has been sought out on the issue of time pressure. Research has shown that the effects of time pressure, in relation to decision making, causes operators to submit to coping processes, which include acceleration, filtering, and omission. Acceleration denotes an increase in the rate of information processing; filtering refers to processing certain parts of information more than other parts; omission, which is also referred to as “shallower search for information,” implies completely ignoring particular parts of information. However, research has also shown that a process known as “regression to learnt behaviors” is a common cognitive strategy in which the operator has “a tendency to lock into one problem solving strategy under time pressure even if it suboptimal” (Boussemart, Donmez, Cummings, & Las Fargeas, 2009).

Time pressure in UAS operations is a critical factor when it comes to performance. Research demonstration has found that performance decreased as a result of an increase in workload due to time pressure, particularly in tasks, such as target acquisition, that already present high levels of stress to the operator (Hughes & Babski-Reeves, 2005). According to Burke, Oron-Gilad, Conway, and Hancock (2007), time pressure, during a target acquisition task, resulted in the degradation of operator ability in distinguishing friend from foe. Situations such as these pose serious threats to military operations. Although time pressure may make a somewhat tedious and boring task more interesting and enjoyable, it could also cause higher levels of stress and mental overload, which result in poorer performance levels, increased fatigue and mental workload, and poor decision making skills, among others.

Effects of Time Pressure

Stress

One of the effects caused by high time pressure is stress. Stress is a difficult and often confusing subject as it is a psychological concept that is not concrete; one cannot touch it or perceive it directly (Driskell & Salas, 1996). Stress can take effect from a multitude of factors; it can be imposed by having too much to do in too little time. According to Wickens, et al. (2004), the ratio TR/TA, which was described earlier in the section regarding mental workload, plays a significant role in causing stress. If there is not enough time available to perform a certain task, stress levels of an operator increases as he or she tries to finish the task quickly in the time allotted while at the same time do well in performance. Stress and performance go hand-in-hand. Figure 6 portrays a four-stage process model of stress and performance (Driskell & Salas, 1996) in which stimuli from the environment, the first stage of the process, is first activated by noise, time pressure, task load, threats, or group pressure. Once a threat is perceived, the process moves on to the second stage, appraisal, where the extent of that threat is evaluated. The appraisal then leads to the third stage, performance expectations. This stage is concerned with feelings of self-efficacy or mastery. If perceived resources are exceeded by demands, then negative performance expectations are formed; however, if it is perceived that the available resources exceed the perceived threat, then positive performance expectations are formed, which is a crucial factor in preparing personnel to operate under high-demand conditions. The addition of time constraints to completing tasks increases the demand of completing those tasks in a certain time frame, thus exceeding the perceived resources to complete them and resulting in negative performance expectations.

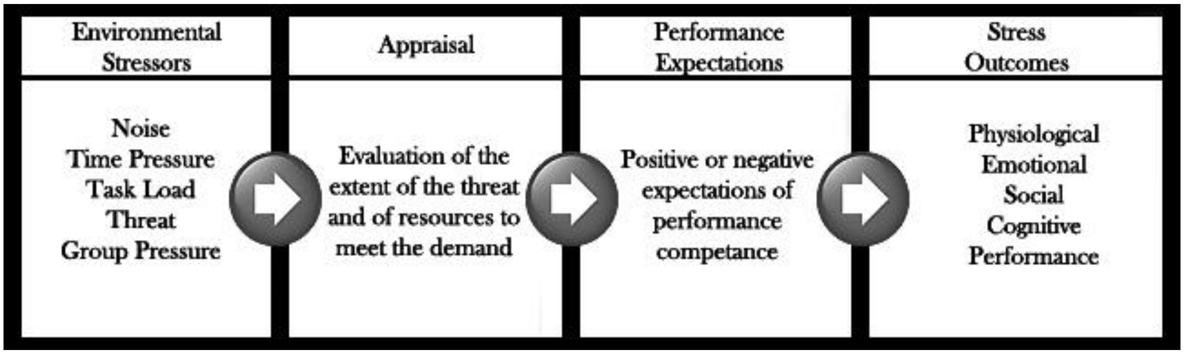

Figure 6: Four-stage model of Stress and Performance (Adapted from Driskell & Salas, 1996)

Lastly, stress outcomes, the fourth and final stage, consist of various types of stress, including:

- Physiological reactions that incorporate various measures relating to heart beat, pulse rate, blood pressure, blood glucose levels, eye blink duration, and respiratory rate, among many.
- Emotional reactions, which may include subjective feelings of anxiety or fear, frustration, annoyance, tension, and an increased concern of well-being for self and others.
- Social reactions that could lead to less cooperation in a team environment, neglect towards social or interpersonal cues, and increased interpersonal aggression with a decrease in tendency to provide a helping hand to others.
- Cognitive effects as a result of stress can cause narrowing of attention, distraction, tunnel vision, longer reaction time, increased errors, and memory deficits, among others.

- Performance in stress relates to performance accuracy (the number of errors incurred on a task), performance speed (the required time to perform a task), and performance variability (variability in accuracy and speed).

Completing a task, in general, may be stressful, especially in sensitive and dangerous environments. As a result, the issue of stress, and its outcomes, is an important factor to take into consideration, particularly when assigning tasks under high-pressure conditions.

Decision Making

Decision making is another effect caused by high time pressure. Each new day brings about its own challenges, one of which is the ability to make every day decisions; however, this ability can prove to be difficult at times. Decision making is generally a task in which (a) a person must choose one option from a number of alternatives, (b) there is a relative amount of information, with respect to the option, available, (c) the timeframe is relatively longer than a second, and (d) the choice chosen is not necessarily clear to be the best and thus is associated with uncertainty (Wickens, et al., 2004). Making decisions could either be riskless or risky. Riskless decisions are characterized by using mathematical models to identify decision strategies, such as the Elimination-by-Aspects Rule where an individual chooses between alternatives "by selecting the most important attribute and rejecting all alternatives that fail to meet the cutoff." Risky decisions, on the other hand, are characterized by "couplings between alternatives and outcomes that are probabilistic and thereby cannot be predicted with certainty" (Svenson & Maule, 1993). Through understanding these processes, good decision makers take the time to assess the costs and benefits of each decision, first; however, not all decisions come with the comfort of time. Limiting the time to make a decision decreases the ability to make a sound decision and increases stress as decision making is commonly represented by the following three phases or stages that often cycle and iterate in a single decision in working memory:

1. Acquiring, perceiving, and integrating information or cues relating to the decision,
2. Generating and selecting hypotheses or situation assessments about the meaning of the cues in regards to the current and future state of the decision, and
3. Planning and selecting choices to take based on the inferred state and the costs and benefits of the different possible outcomes.

These three stages impose cognitive limits that are crucial to conscious, effortful decision making. Decision making is so much more than just choosing one option out of many; it has a lot to do with how the human brain operates, from working memory to long-term memory, all have an important influence in the information-processing aspect of decision making. The information-processing model of decision making is depicted in Figure 7.

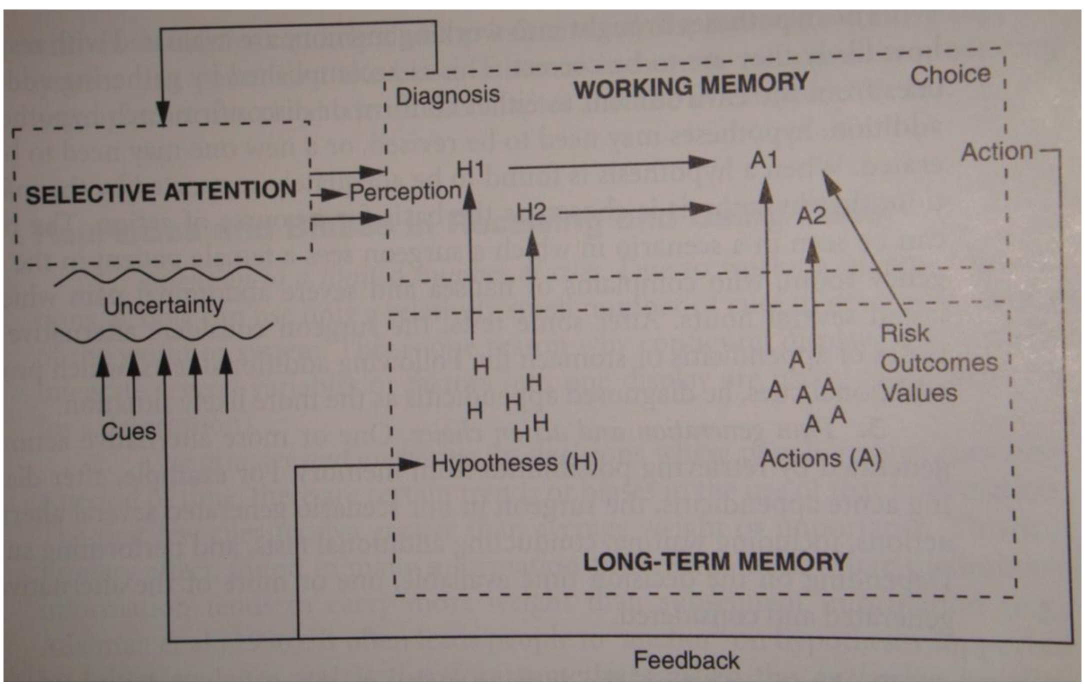

Figure 7: Information-processing model of decision making (From Wickens, et al., 2004)

Decision making is crucial to UAS operations as operators are expected to make important decisions frequently and rapidly, such as deciphering whether a tank or target is a friend or a foe. Research on decision making has been extensive, although, initially, the focus of the research had been on optimal, *rational* decision making. It was assumed that researchers would be able to specify the costs and benefits, or values, associated with making different choices and then apply mathematical models to those values in order to yield the optimal choice that would maximize their benefits and minimize their costs. This type of decision making is sometimes referred to as normative decision making, which revolves around the central concept, the overall value of the choice, and the "worth," or importance, each outcome has on the individual making the decision. In addition to this model, another model, known as descriptive decision making is where decision makers rely on rules-of-thumb, or simpler and less-complete means, to come to a decision (Wickens, et al., 2004). Although many models and heuristics exist to explain the process of decision making, it takes time and mental power to choose what is thought to be the optimal choice and limiting that time could lead to an increase in stress and making incorrect choices, which in turn could be disastrous in sensitive environments such as in national security.

Summary

In the midst of aviation's next generation, the UAS exists. The highly versatile, efficient, valuable, and autonomous nature of the UAS has been extensively debated by researchers. Since their pre-aviation history, the environments in which UAS operate and the operators themselves have been a major topic in the research field as they pose significant human factors questions. In order to attain the full potential of this technology, one must understand the relationship between the system and the human operator. This relationship exists through the interface in which the system and the human operator interact.

Since the dawn of their creation, the primitive technology of UAVs was used for surveillance and combat well before manned aircraft first took flight on December 17, 1903; now leading aviation's portal to the future with NextGen, transforming the way America flies using 21st century technology. Despite of the long list of human factors issues with UAS automation, a narrowing of research efforts is necessary in order to expand research knowledge in regards to human capabilities and limitations as it exists within UAS operations.

One area of research concern is UAS system reliability and the trust the operator places on that system. The relation between reliability and trust in automation is a crucial one when discussing its importance in relation to human performance issues. Reliability of a system is one in which the system does what it is meant to do or designed to do. However, nothing is perfect and the system is not 100% reliable at all times. This issue needs to be well researched in order to teach the operator what cues to look for when the system does not act accordingly. This is where the issue of trust in the system falls into place. Trust should be in direct proportion to its reliability, and should also be well calibrated; however, there is evidence that the trust a human puts into automation is not entirely well calibrated, being either too high or too low; too high trust in automation could lead to complacency, whereas too low trust could lead to distrust with the system going un-used.

Mental workload is one area of limitation that operators experience with their interaction with the UAS interface. Workload levels could range from too high, causing performance detriments, to too low, causing difficulty for the operator to maintain vigilance. To alleviate the downside of workload, the implementation of automation can make certain tasks plausible for the operator to perform more complex tasks; however, the implementation of this automation needs to maintain an appropriate balance so as to not dramatically lower workload but at the same time alleviate the overload imposed on the operator.

Time Pressure is another area of concern in the research field. In a time sensitive UAS environment, the effects of time pressure may be too great on the operator causing mental overload and resulting in diminished performance. Although time pressure may make a mundane task more interesting and enjoyable, having a small time frame to complete a task, such as target acquisition, could lead to an increase in the operator's stress level. This increase in stress levels could in turn lead to poor decision making skills that could be the difference in mistaking a friend for a foe.

Objective

The objective of this study was to investigate the effects of system reliability and time pressure on UAS operator performance and mental workload. Participants' perceived trust was also investigated for exploratory reasons.

Method

Participants

Twenty-four students, male and female, from Embry-Riddle Aeronautical University were recruited for this study on a voluntary basis. Participants received either extra credit in an undergraduate course or $10 for their participation and had the chance to win $50 for best overall performance. Participants were asked to sign an informed consent form (Appendix A) acknowledging their willingness to participate in this study. To mitigate any confounding variables, a biographical questionnaire (Appendix B), that elicited the participant's video/computer gaming experience, was given prior to the start of the study. Gender was not taken into consideration for recruitment purposes in this study.

Apparatus

The apparatus of this study consisted of a test bed built around the Multi-Modal Immersive Intelligent Interface for Remote Operation (MIIIRO) operator interface, which provided a 3D model of UAVs and their environment and was run on a standard computer. The MIIIRO software was designed by IA Tech, Inc., with support from the Air Force Research Laboratory, "to perform or simulate the operations of a number of autonomous UAVs" (Tso, Tharp, Tai, Draper, Calhoun, & Ruff, 2003). It is a synthetic task environment which allows for flexible emulation of "envisioned single operator supervision of multiple UAVs" (Galster, Nelson, & Bolia, n.d.) and consists of (1) a community of intelligent agents that are used to integrate, assimilate, and present data, as well as interact with the operator to plan and control the remote systems and mission payloads, (2) an immersive environment, and (3) multi-modal inputs that includes head tracker and joystick to allow for efficient interactions (IA Tech, Inc., n.d.).

The setup arrangement included two monitors, a standard computer mouse, and a QWERTY keyboard. The first monitor (Figure 8) portrayed the Tactical Situation Display (TSD), which provided a plan view of the mission environment, including waypoints, flight segments, targets, and threats, in addition to icons showing the positions and status for each of the UAVs. The Mission Mode Indicator (MMI), which displayed a series of lights (green, yellow, and red), was also displayed at the top of the TSD (Tso, et al., 2003). The second monitor (Figure 9) was used for image processing and showed the Image Management Display (IMD) that included an image cue and image display.

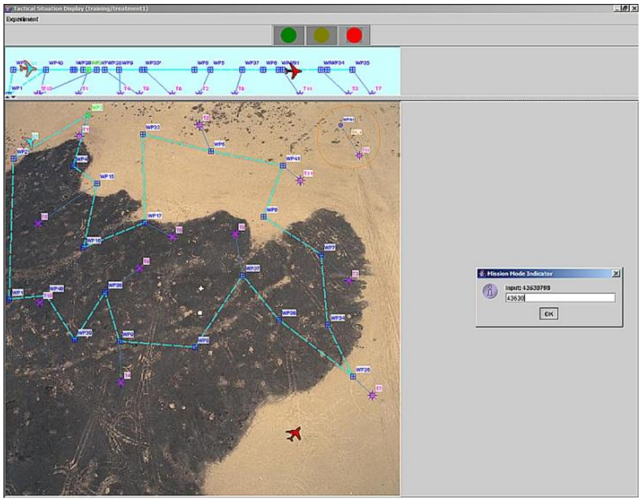

Figure 8: MIIIRO Tactile Situation Display (TSD)

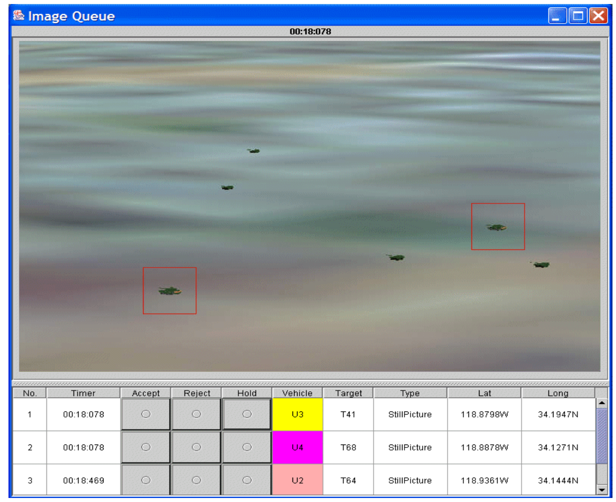

Figure 9: MIIIRO Image Management Display (IMD)

Design

A 2x2 within subjects, fully factorial design was used for this study. The study consisted of two IVs; system reliability and time pressure (Table 3). The first IV, system reliability, consisted of target images appearing with a high reliability measured at 80% or with a low reliability measured at 40%. The second IV, time pressure, consisted of either a 5 or a 10 second time limit during the target acquisition task in which the participant either had 5 seconds to acquire the target or 10 seconds to acquire the target. If the participant was unable to do so within the specific time limits, the MIIIRO software acquired the targets for them and moved on to the next target. These specific measurements for system reliability and time pressure were based on previous scholarly theses. A 4x4 Latin Square (LS) design (Table 4) was used to determine the order of which first and second IVs were to be presented to the participants; this was done in order to counter balance any learning effects. The LS design derived its name from "an ancient puzzle that was concerned with the number of different ways that Latin letters can be arranged in a square matrix so that each letter appears once in each row and each column"

(Weiner, Freedheim, Schinka, & Velicer, 2003). Table 5 shows the LS-4 design for this experiment, where the number 4, in LS-4, denotes the number of levels of the treatment. The first treatment scenario, 1, denotes 40% system reliability with 5 seconds time pressure. The second treatment scenario, 2, denotes 40% system reliability with 10 seconds time pressure. The third treatment scenario, 3, denotes 80% system reliability with 5 seconds time pressure. The fourth and last treatment scenario, 4, denotes 80% system reliability with 10 seconds time pressure. The order presented in Table 5 was repeated for every 6 participants, resulting in 4 repetitions for a total of 24 participants. The dependant variables (DVs) that were collected were operator performance and mental workload. Mental workload was subjectively reported by the participants using the NASA-TLX standardized subjective workload scale (Appendix C) after each of the four scenarios, while operator performance, in terms of image processing time, target acquisition accuracy, MMI processing time, pop-up threats re-route processing time, and Intruder Aircraft (IA) processing time, were objectively measured by the MIIIRO software.

Table 3: Experimental Design

		Time Pressure	
		5 seconds	10 seconds
System Reliability	40%	X	X
	80%	X	X

Table 4: 4x4 Latin Square Design

	1	2	3	4
1	---	1,2	1,3	1,4
2	2,1	---	2,3	2,4
3	3,1	3,2	---	3,4
4	4,1	4,2	4,3	---

Table 5: LS-4 Experimental Design

	Order of Treatment Scenarios			
Group 1 (Participants 1, 5, 9, 13, 17, and 21)	1	2	4	3
Group 2 (Participants 2, 6, 10, 14, 18, and 22)	2	3	1	4
Group 3 (Participants 3, 7, 11, 15, 19, and 23)	3	4	2	1
Group 4 (Participants 4, 8, 12, 16, 20,and 24)	4	1	3	2

Primary Task

The primary task in this study was that of target acquisition. Due to the high level of autonomy used for the UAS in the MIIIRO software, the participants were not required to directly control the flight of the UAV. To make up the flight path in which the UAV trailed, waypoints were preset in addition to 10 preset image capture locations. These 10 preset image capture locations were associated with certain waypoint locations along the UAV flight path. Participants were asked to view the images collected by the UAV and to decipher whether or not the Automatic Target Recognition (ATR) tool, that was part of the IMD on the second computer monitor and had the ability to recognize targets or objects based on data obtained from the MIIIRO software, had correctly selected the targets at the current waypoint the UAV was located. Each waypoint contained at least one terrain vehicle, which may or may not have been a target, in addition to distracters that were randomly present at certain waypoints. The ATR, which had two preset reliability percentages at 40% and 80%, placed a red box (Figure 10) around what it recognized as a target. Reliability percentage at 40% corresponded to the ATR being correct 40 percent of the time while 80% reliability corresponded to the ATR being correct 80 percent of the time. As a result, the ATR was not always correct and sometimes placed the red box around non-targets and/or distracters; in those cases, the participant was required to deselect the incorrect images, select the correct ones, and click on "accept" on the IMD using the standard computer mouse. In the cases where the ATR recognized all the correct targets, the participant should have clicked on "accept;" however, if the ATR recognized non-targets and/or distracters and no targets were present, the participant should have instead clicked on "reject." If no action was taken by the participant in the allocated time pressure, the automation processed and "accepted" the red boxed images, as is.

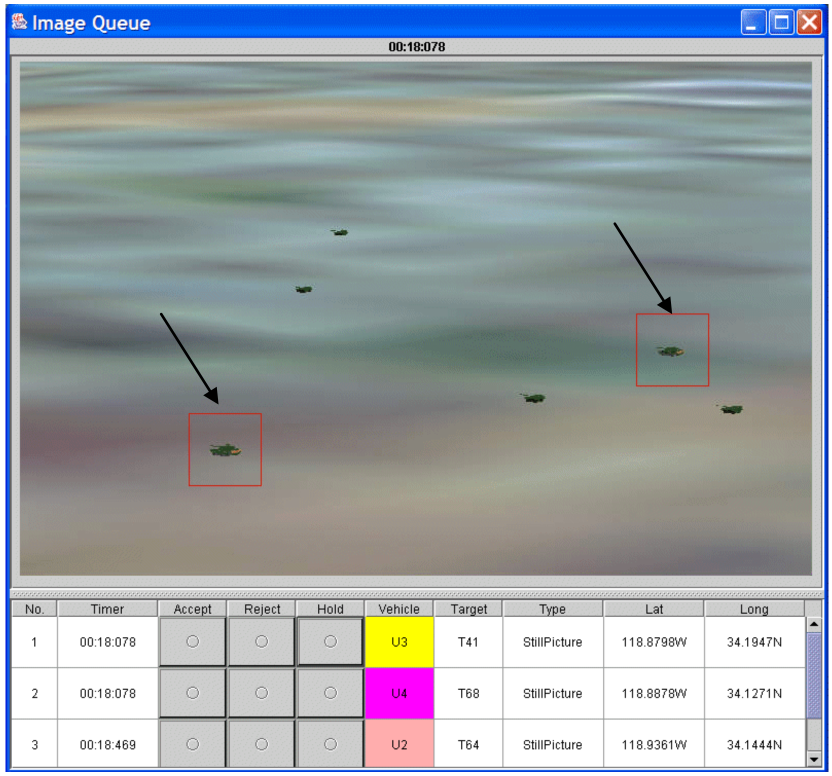

Figure 10: Image Management Display (IMD) with Automatic Target Recognition (ATR) placing red boxes around perceived targets

The MIIIRO software automatically measured the results of the primary task. These objective measures consisted of image processing time and target acquisition accuracy.

Secondary Task

There were three secondary tasks in this study that consisted of: (1) processing IA, (2) responding to automation made flight path change recommendations, and (3) monitoring the MMI.

The first secondary task consisted of processing IA that entered the operational airspace. The mission of this task was to imitate the occurrence of unexpected IA that may enter into the airspace, which required a quick and attentive response from the participant as it is considered to be a highly critical situation in typical UAS operations. In order to distinguish between the UAV in the study and the unexpected aircraft, the IA resembled a red aircraft (Figure 11) and was displayed three times at random intervals throughout the course of the simulation. Participants were required to click on the red aircraft using the standard computer mouse and then enter a predetermined code that was made available to them on a piece of paper located on the first computer monitor in order to overcome the situation.

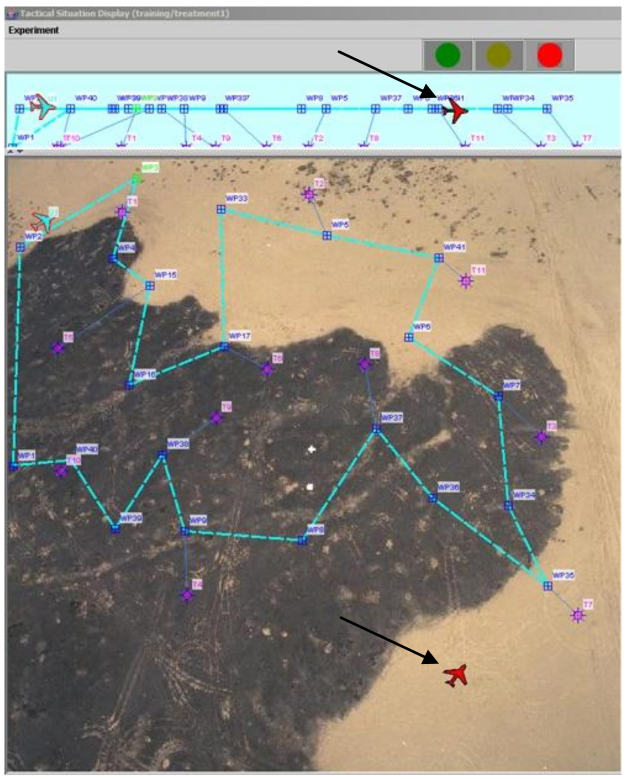

Figure 11: Intruder Aircraft (IA) displayed as a red aircraft

The second secondary task consisted of responding to recommendations, made by the automation, to change the UAV flight path (Figure 12). Participants were required to respond to these recommendations when "pop-up threats" were encountered by the UAV. These so-called "pop-up threats" were designed into the flight path; yet, they were made undetectable to the participant until the UAV encountered them at different waypoints. At that time, the automation made a recommendation to change the route of the flight path in order to avoid the “pop-up threat;” however, not all of the recommendations that were made by the automation were necessary. As a result, the participant was required to acknowledge the recommended change and either “accept” or “reject” the route change.

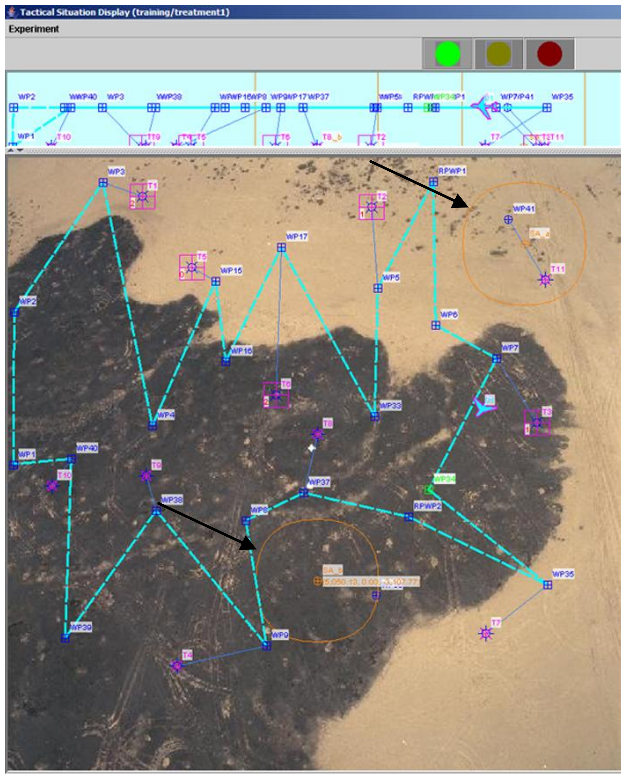

Figure 12: Flight path change recommendations for “pop-up threats”

The third and final secondary task involved the MMI (Figure 13) that was displayed at the top of the TSD. The MMI was represented by a series of three round lights (green, yellow, and red) organized in a horizontal line that is similar to a horizontal stoplight. These series of lights indicated the status of the UAV; green represented a state of good health, yellow indicated that action is needed, and red indicated that an urgent action is needed. If the status of the UAV was green, the participant did not need to take any action; on the contrary, if the status of the UAV was either yellow or red, the participant needed to take immediate action by clicking on the illuminated yellow or red light and correctly type in a string of text that appeared on the screen of the first computer monitor (Figure 14). Once the participant typed in the correct text string, the MMI returned to its original state, which was that of the color green, indicating that is had returned to a state of good health.

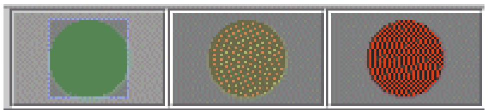

Figure 13: Mission Mode Indicator (MMI)

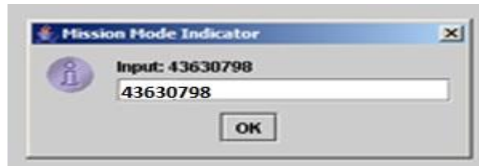

Figure 14: MMI pop-up "input code" screen

The MIIIRO software automatically, and objectively, measured the results of the three secondary tasks. These measures consisted of the number of events and response times for the IA, the "pop-up threats," and the MMI. In order to subjectively measure the participants' mental workload, the NASA-TLX standardized subjective workload scale (Appendix D) was used following the completion of the primary and secondary task in each of the four scenarios. The NASA-TLX measure provided an overall mental workload scale based on a weighted average of the following six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration.

Procedure

Once the participants arrived at the lab, they were asked to read, acknowledge, and sign an informed consent form (Appendix A) after the purpose of the study and the compensation of the study was explained to each one. Following this, each participant was given a biographical questionnaire. The participants were then introduced to the paper-pencil version of the NASA-TLX standardized subjective workload scale and the trust survey; the proper method of filling out the form was also explained at that time. In order to allow familiarization of the MIIIRO simulator, each participant went through a five-minute training session that included an instructional hands-on training session with the occurrence of all possible scenario events that were going to occur during the actual study session. However, 50% was used for system reliability and a 15 second time pressure was used to avoid any learning effects; reliability percentage at 50% corresponded to the ATR being correct 50 percent of the time while having a 15 second time limit to acquire the target. After the training process, the participant was given the opportunity to ask questions or comment on any concerns he or she may have had. Once any questions or concerns had been addressed, the participant began with the actual data collection phase of the study.

Each participant received four treatment scenarios using the LS-4 Design in order to avoid any learning effects. Those four treatment scenarios included, in no particular order: (1) 40% system reliability with 5 seconds time pressure, (2) 40% system reliability with 10 seconds time pressure, (3) 80% system reliability with 5 seconds time pressure, and (4) 80% system reliability with 10 seconds time pressure. Those treatments were selected on the basis of previous research studies that focused on other levels of treatment. Each of those treatment scenarios lasted about seven minutes with a five-minute break after the second treatment scenario. After the completion of each treatment scenario, participants were asked to fill out the NASA-TLX standardized subjective workload scale, resulting in four separate forms for the NASA-TLX workload scale for each participant. Each participant was also asked to complete a trust survey (Appendix D) of the MIIIRO software after each treatment scenario, which assessed perceived reliability, technical competence, perceived understandability, faith, and personal attachment; each participant was then verbally notified of their performance during the debriefing phase of the study and any further questions or concerns were addressed at that time. Once all participants had completed the study, the participant with the best overall performance was contacted to receive $50.

Statement of Hypotheses

The following hypotheses were tested:

Hypothesis 1: When participants are exposed to high system reliability, they will report lower mental workload than when they are exposed to low system reliability.

Hypothesis 2: When participants are exposed to high system reliability, they will report higher trust levels than when they are exposed to low system reliability.

Hypothesis 3: When participants are exposed to high system reliability, their processing time score will be lower (better) in the primary task than when they are exposed to low system reliability.

Hypothesis 4: When participants are exposed to high system reliability, their target acquisition accuracy score will be higher (better) in the primary task than when they are exposed to low system reliability.

Hypothesis 5: When participants are exposed to high system reliability, their processing time score for the Intruder Aircraft (IA) will be lower (better) in the secondary task than when they are exposed to low system reliability.

Hypothesis 6: When participants are exposed to high system reliability, their processing time score for the pop-up threats will be lower (better) in the secondary task than when they are exposed to low system reliability.

Hypothesis 7: When participants are exposed to high system reliability, their processing time score for the Mission Mode Indicator (MMI) will be lower (better) in the secondary task than when they are exposed to low system reliability.

Hypothesis 8: When participants are exposed to the lower time pressure condition, they will report lower mental workload than when they are exposed to the higher time pressure condition.

Hypothesis 9: When participants are exposed to the lower time pressure condition, they will report higher trust levels than when they are exposed to the higher time pressure condition.

Hypothesis 10: When participants are exposed to the lower time pressure condition, their processing time score will be lower (better) in the primary task than when they are exposed to the higher time pressure condition.

Hypothesis 11: When participants are exposed to the lower time pressure condition, their target acquisition accuracy score will be higher (better) in the primary task than when they are exposed to the higher time pressure condition.

Hypothesis 12: When participants are exposed to the lower time pressure condition, their processing time score for the Intruder Aircraft (IA) will be lower (better) in the secondary task than when they are exposed to the higher time pressure condition.

Hypothesis 13: When participants are exposed to the lower time pressure condition, their processing time score for the pop-up threats will be lower (better) in the secondary task than when they are exposed to the higher time pressure condition.

Hypothesis 14: When participants are exposed to the lower time pressure condition, their processing time score for the Mission Mode Indicator (MMI) will be lower (better) in the secondary task than when they are exposed to the higher time pressure condition.

Hypothesis 15: An interaction will exist between system reliability and time pressure for mental workload.

Hypothesis 16: An interaction will exist between system reliability and time pressure for trust levels.

Hypothesis 17: An interaction will exist between system reliability and time pressure for primary task processing time.

Hypothesis 18: An interaction will exist between system reliability and time pressure for primary task target acquisition accuracy.

Hypothesis 19: An interaction will exist between system reliability and time pressure for secondary task Intruder Aircraft (IA).

Hypothesis 20: An interaction will exist between system reliability and time pressure for secondary task pop-up threats.

Hypothesis 21: An interaction will exist between system reliability and time pressure for secondary task Mission Mode Indicator (MMI).

Results

The intention of this study was to analyze the effects of system reliability and time pressure on UAS operator performance and mental workload. Repeated measures analysis of variance (ANOVA) tests were used to analyze the effect of each independent variable on the following dependent variables: image processing accuracy, image processing time, MMI processing time, pop-up threats re-route processing time, IA processing time, and mental workload scores which were subjectively collected by the NASA-TLX form. Trust of the system was collected and analyzed through the use of a survey for exploratory purposes. It is included in the outcome of the results described in this section. A type I error alpha value (α) of 0.05 was used to determine significance.

Primary Task

There were two primary task performance measures collected during this study which included image processing time and target acquisition accuracy. Repeated measures ANOVAs were conducted to analyze the hypotheses made regarding each primary task performance measure.

Image Processing Time

Image processing time was the first primary task performance dependent measure to be tested. Hypothesis 3 stated that the processing time score of participants will be lower (better) in the primary task when they are exposed to high system reliability (80%) than when they are exposed to low system reliability (40%). Hypothesis 10 stated that when participants are exposed to the lower time pressure condition (10 seconds), their processing time score will be lower (better) in the primary task than when they are exposed to the higher time pressure condition (5 seconds). The means and standard deviations for image processing time are presented in Table 6. The results of the ANOVA for image processing time are shown in Table 7.

Table 6: Primary Task's Image Processing Time Means and Standard Deviations

	Mean	Standard Deviation	N
40% System Reliability / High Time Pressure	2384.500	449.459	24
40% System Reliability / Low Time Pressure	2506.125	757.080	24
80% System Reliability / High Time Pressure	2389.667	474.585	24
80% System Reliability / Low Time Pressure	2844.083	623.191	24

Table 7: ANOVA Source Table for Primary Task Image Processing Time

Source Greenhouse-Geisser	Sum of Squares (SS)	*df*	Mean Square	*F*	*p*	Observed Power
System Reliability	1990944.010	1	1990944.010	13.613	.001*	.942
Time Pressure	706408.594	1	706408.594	3.964	.059	.479
System Reliability*Time Pressure	664501.760	1	664501.760	3.223	.086	.405
Error (System Reliability)	3363760.740	23	146250.467			
Error (Time Pressure)	4099122.156	23	178222.702			
Error (System Reliability*Time Pressure)	4742710.990	23	206204.826			

* indicates p value < 0.05

The main effect of system reliability on image processing time was analyzed first and was found to be statistically significant with $F(1, 23) = 13.613$, $p = .001$. The significance of this effect indicates that the participants' processing time were significantly lower (better) in the primary task when they were exposed to the high system reliability (80%) than when they were exposed to the low system reliability (40%). As a result, hypothesis 3 was supported.

The main effect of time pressure on image processing time was also analyzed with $F(1,23) = 3.964$ and $p = .059$ and was shown to be statistically insignificant. The insignificance of this effect indicates that the participants' processing time scores were not impacted significantly by the time pressure. As a result, hypothesis 10 was not supported.

The results of these two main effects on image processing time are shown in Figure 15.

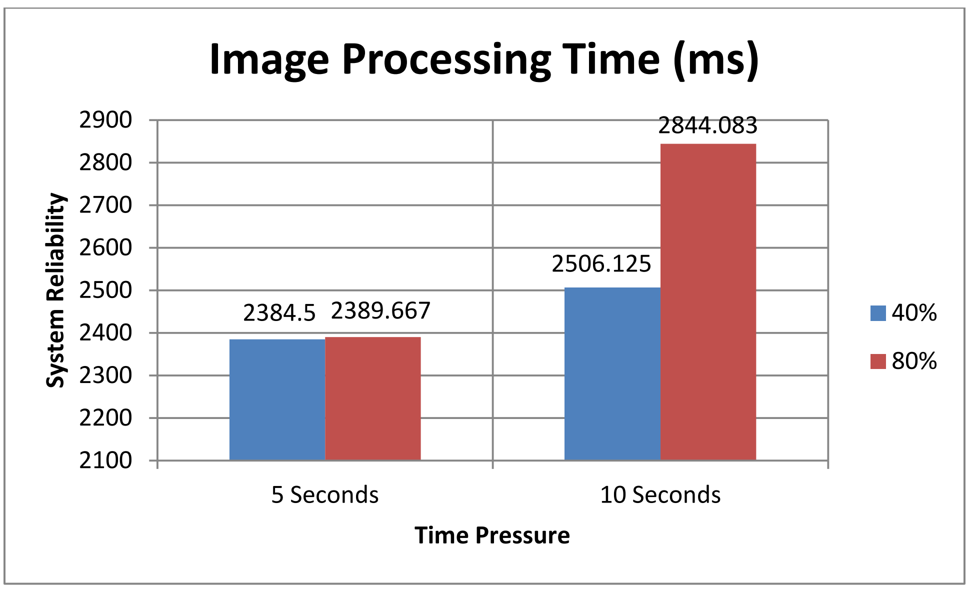

Figure 15: System Reliability and Time Pressure Effects on Image Processing Time

In addition to the main effects of image processing time reported previously, hypothesis 17 stated that an interaction will exist between system reliability and time pressure for primary task processing time. The means and standard deviations are presented in Table 8.

Table 8: Means and Standard Deviations for the Interaction of System Reliability and Time Pressure on Image Processing Time

				95% Confidence Interval	
System Reliability	Time Pressure	Mean	Standard Error	Lower Bound	Upper Bound
Low	High	2384.500	91.745	2194.710	2574.290
	Low	2389.667	96.874	2189.267	2590.067
High	High	2506.125	154.538	2186.438	2825.812
	Low	2844.083	127.208	2580.933	3107.234

The interaction effect between system reliability and time pressure was found to be statistically insignificant with $F(1, 23) = 3.223$ and $p = .086$. The insignificance of this effect indicates that there was no significant interaction between system reliability and time pressure on image processing time; as a result, hypothesis 17 was not supported.

Target Acquisition Accuracy

Target acquisition accuracy was the second primary task performance dependent measure to be tested. Hypothesis 4 stated that when participants are exposed to high system reliability (80%), their target acquisition accuracy score will be higher (better) in the primary task than when they are exposed to low system reliability (40%). Hypothesis 11 stated that when participants are exposed to the lower time pressure condition (10 seconds), their target acquisition accuracy score will be higher (better) in the primary task than when they are exposed to the higher time pressure condition (5 seconds). The means and standard deviations for target acquisition accuracy are presented in Table 9. The results of the ANOVA are presented in Table 10 and are meant to reflect percentages.

Table 9: Primary Task's Target Acquisition Accuracy Means and Standard Deviations

	Mean	Standard Deviation	N
40% System Reliability / High Time Pressure	84.875	16.201	24
40% System Reliability / Low Time Pressure	86.000	10.299	24
80% System Reliability / High Time Pressure	85.542	12.635	24
80% System Reliability / Low Time Pressure	77.792	19.580	24

Table 10: ANOVA Source Table for Primary Task Target Acquisition Accuracy

Source Greenhouse-Geisser	Sum of Squares (SS)	*df*	Mean Square	*F*	*p*	Observed Power
System Reliability	263.344	1	263.344	3.641	.069	.448
Time Pressure	341.260	1	341.260	6.717	.016*	.699
System Reliability*Time Pressure	472.594	1	472.594	3.622	.070	.446
Error (System Reliability)	1663.406	23	72.322			
Error (Time Pressure)	1168.490	23	50.804			
Error (System Reliability*Time Pressure)	3001.156	23	130.485			

* indicates p value < 0.05

The main effect of system reliability on target acquisition accuracy was examined first and was found to be statistically insignificant with $F(1, 23) = 3.641$ and $p = .069$. The insignificance of this effect indicates that the target acquisition accuracy of participants was not significantly higher (better) when they were exposed to the high system reliability (80%) than when they were exposed to the low system reliability (40%). As a result, hypothesis 4 was not supported.

The main effect of time pressure on target acquisition accuracy was also analyzed with $F(1,23) = 6.717$ and $p = .016$ and was shown to be statistically significant ($p < 0.05$). The significance of this effect indicates that the target acquisition accuracy of participants was higher (better) when they were exposed to the lower time pressure condition (10 seconds) than when they were exposed to the higher time pressure condition (5 seconds). As a result, hypothesis 11 was supported.

The results of these two main effects on target acquisition accuracy are shown in Figure 16.

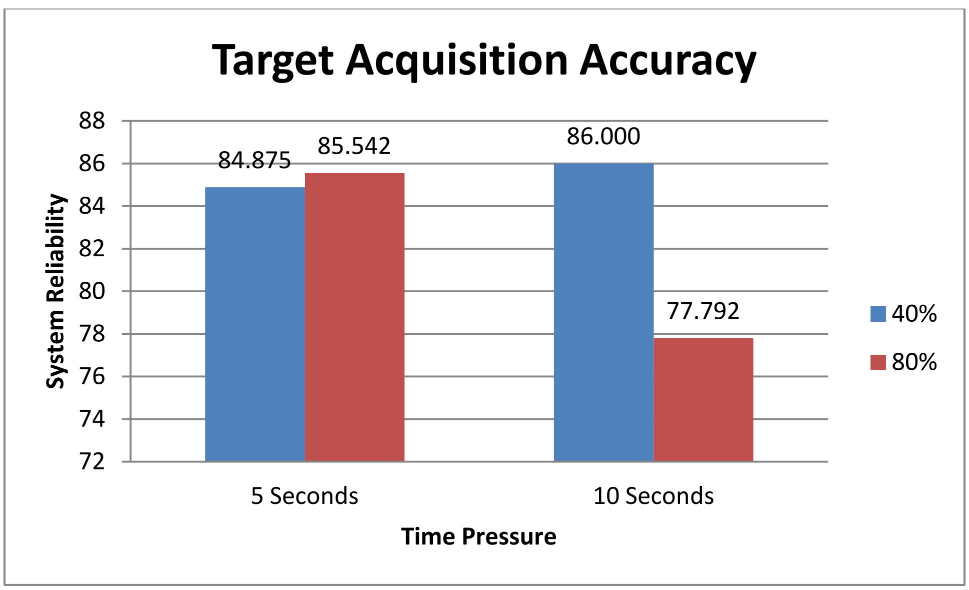

Figure 16: System Reliability and Time Pressure Effects on Target Acquisition Accuracy

In addition to the main effects for target acquisition accuracy reported previously, hypothesis 18 stated that an interaction will exist between system reliability and time pressure for primary task target acquisition accuracy. The means and standard deviations are presented in Table 11.

Table 11: Means and Standard Deviations for the Interaction of System Reliability and Time Pressure on Target Acquisition Accuracy

				95% Confidence Interval	
System Reliability	Time Pressure	Mean	Standard Error	Lower Bound	Upper Bound
Low	High	84.875	3.307	78.034	91.716
	Low	85.542	2.579	80.206	90.877
High	High	86.000	2.102	81.651	90.349
	Low	77.792	3.997	69.524	86.060

The interaction effect between system reliability and time pressure was found to be statistically insignificant with $F(1, 23) = 3.622$ and $p = .070$. The insignificance of this effect indicates that there was no significant interaction between system reliability and time pressure on target acquisition accuracy; as a result, hypothesis 18 was not supported.

Secondary Task

In addition to the primary tasks, there were three secondary task performance measures collected during this study, which included Intruder Aircraft (IA) processing time, pop-up threats re-routing processing time, and Mission Mode Indicator (MMI) processing time. Repeated measures ANOVAs were conducted to analyze the hypotheses made regarding each secondary task performance measure.

Intruder Aircraft (IA) Processing Time

IA processing time was the first of the three secondary task performance dependent measures to be tested. Hypothesis 5 stated that the processing time score for the IA will be lower (better) when participants are exposed to high system reliability (80%) than when they are exposed to low system reliability (40%). Hypothesis 12 stated that when participants are exposed to the lower time pressure condition (10 seconds), their processing time score for the IA will be lower (better) in the secondary task than when they are exposed to the higher time pressure condition (5 seconds). The means and standard deviations for IA processing time are presented in Table 12. The results of the ANOVA for IA processing time are shown in Table 13.

Table 12: Secondary Task's IA Processing Time Means and Standard Deviations

	Mean	Standard Deviation	N
40% System Reliability / High Time Pressure	6112.583	1965.309	24
40% System Reliability / Low Time Pressure	6405.458	3631.532	24
80% System Reliability / High Time Pressure	5890.542	2210.470	24
80% System Reliability / Low Time Pressure	5902.708	1741.681	24

Table 13: ANOVA Source Table for Secondary Task IA Processing Time

Source Greenhouse-Geisser	Sum of Squares (SS)	*df*	Mean Square	*F*	*p*	Observed Power
System Reliability	558302.510	1	558302.510	.393	.537	.092
Time Pressure	3151937.760	1	3151937.760	1.188	.287	.181
System Reliability*Time Pressure	472783.010	1	472783.010	.075	.787	.058
Error (System Reliability)	3.264E7	23	1419028.293			
Error (Time Pressure)	6.101E7	23	2652628.543			
Error (System Reliability*Time Pressure)	1.457E8	23	6333738.619			

The main effect of system reliability on IA processing time was examined first and was found to be statistically insignificant with $F(1, 23) = .393$ and $p = .537$. The insignificance of this effect indicates that the IA processing time of participants was not significantly lower (better) when they were exposed to the high system reliability (80%) than when they were exposed to the low system reliability (40%). As a result, hypothesis 5 was not supported.

The main effect of time pressure on IA processing time was also analyzed with $F(1,23) = 1.188$ and $p = .287$ and was also shown to be statistically insignificant ($p > 0.05$). The insignificance of this effect indicates that the IA processing time of participants was not significantly lower (better) when they were exposed to the lower time pressure condition (10 seconds) than when they were exposed to the higher time pressure condition (5 seconds). As a result, hypothesis 12 was not supported.

In addition to the main effects for IA processing time reported previously, hypothesis 19 stated that an interaction will exist between system reliability and time pressure for secondary task IA. The means and standard deviations are presented in Table 14.

Table 14: Means and Standard Deviations for the Interaction of System Reliability and Time Pressure on IA Processing Time

				95% Confidence Interval	
System Reliability	Time Pressure	Mean	Standard Error	Lower Bound	Upper Bound
Low	High	6112.583	401.167	5282.706	6942.461
	Low	5890.542	451.210	4957.142	6823.941
High	High	6405.458	741.283	4871.997	7938.920
	Low	5902.708	355.519	5167.261	6638.156

The interaction effect between system reliability and time pressure was found to be statistically insignificant with $F(1, 23) = .075$ and $p = .787$. The insignificance of this effect indicates that there was no significant interaction between system reliability and time pressure on IA processing time; as a result, hypothesis 19 was not supported.

Pop-up Threats Re-routing Processing Time

The second of the three secondary tasks performance dependent measures to be tested was the processing time for the re-routing of pop-up threats. Hypothesis 6 stated that when participants are exposed to high system reliability (80%), their processing time sore for the pop-up threats will be lower (better) in the secondary task than when they are exposed to low system reliability (40%). Hypothesis 13 stated that when participants are exposed to the lower time pressure condition (10 seconds), their processing time score for pop-up threats will be lower (better) in the secondary task than when they are exposed to the higher time pressure condition (5 seconds). Table 15 presents the means and standard deviations for the pop-up threats re-routing processing time and Table 16 shows the results of the ANOVA.

Table 15: Secondary Task's Pop-up Threats Re-routing Processing Time Means and Standard Deviations

	Mean	Standard Deviation	N
40% System Reliability / High Time Pressure	2568.333	646.298	24
40% System Reliability / Low Time Pressure	3078.333	924.850	24
80% System Reliability / High Time Pressure	3074.167	721.195	24
80% System Reliability / Low Time Pressure	2901.417	858.656	24

Table 16: ANOVA Source Table for Secondary Task Pop-up Threats Re-routing Processing Time

Source Greenhouse-Geisser	Sum of Squares (SS)	*df*	Mean Square	*F*	*p*	Observed Power
System Reliability	682425.375	1	682425.375	1.345	.258	.199
Time Pressure	649117.042	1	649117.042	1.833	.189	.254
System Reliability*Time Pressure	2796885.375	1	2796885.375	6.142	.021*	.661
Error (System Reliability)	1.167E7	23	507302.092			
Error (Time Pressure)	8147110.458	23	354222.194			
Error (System Reliability*Time Pressure)	1.047E7	23	455406.484			

* indicates p value < 0.05

The main effect of system reliability on pop-up threats processing time was examined first and was found to be statistically insignificant with $F(1, 23) = 1.345$ and $p = .258$; $p > 0.05$. The insignificance of this effect indicates that the participants' pop-up threats re-routing processing times were not significantly lower (better) when they were exposed to the high system reliability (80%) than when they were exposed to the low system reliability (40%). As a result, hypothesis 6 was not supported.

The main effect of time pressure on pop-up threats processing time was also analyzed with $F(1,23) = 1.833$ and $p = .189$ and was shown to be statistically insignificant ($p > 0.05$). The insignificance of this effect indicates that the participants' pop-up threats re-routing processing times were not significantly lower (better) when they were exposed to the lower time pressure condition (10 seconds) than when they were exposed to the higher time pressure condition (5 seconds). As a result, hypothesis 13 was not supported.

In addition to the main effects for pop-up threats re-routing processing time reported previously, hypothesis 20 stated than an interaction will exist between system reliability and time pressure for secondary task pop-up threats. The means and standard deviations are presented in Table 17.

Table 17: Means and Standard Deviations for the Interaction of System Reliability and Time Pressure on Pop-up Threats Re-routing Processing Time

				95% Confidence Interval	
System Reliability	Time Pressure	Mean	Standard Error	Lower Bound	Upper Bound
Low	High	2568.333	131.925	2295.426	2841.241
	Low	3074.167	147.213	2769.633	3378.700
High	High	3078.333	188.784	2687.804	3468.863
	Low	2901.417	175.272	2538.838	3263.995

The interaction effect between system reliability and time pressure was found to be statistically significant with $F(1, 23) = 6.142$ and $p = .021$; $p < 0.05$. The significance of this effect indicates that there was an interaction between system reliability and time pressure for pop-up threats re-routing processing time; as a result, hypothesis 20 was supported. The results of this interaction are shown in Figure 17.

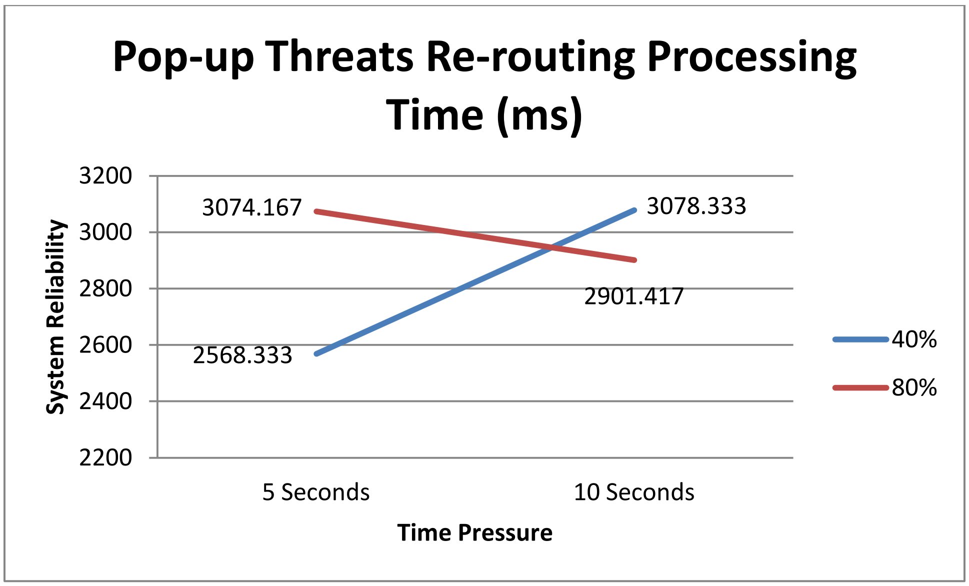

Figure 17: System Reliability and Time Pressure Interaction on Pop-up Threats Re-routing Processing Time

Mission Mode Indicator (MMI) Processing Time

The final secondary task performance measure, or DV, that was tested was the MMI processing time. Hypothesis 7 stated that when participants are exposed to high system reliability (80%), their processing time score for the MMI will be lower (better) in the secondary task than when they are exposed to low system reliability (40%). Hypothesis 14 stated that when participants are exposed to the lower time pressure condition (10 seconds), their processing time score for MMI will be lower (better) in the secondary task than when they are exposed to higher time pressure condition (5 seconds). Table 18 shows the means and standard deviations for the MMI processing time. The results of the ANOVA are presented in Table 19.

Table 18: Secondary Task's MMI Processing Time Means and Standard Deviations

	Mean	Standard Deviation	N
40% System Reliability / High Time Pressure	9934.667	3129.013	24
40% System Reliability / Low Time Pressure	9860.667	3820.227	24
80% System Reliability / High Time Pressure	8994.500	3148.219	24
80% System Reliability / Low Time Pressure	9229.500	4649.934	24

Table 19: ANOVA Source Table for Secondary Task MMI Processing Time

Source Greenhouse-Geisser	Sum of Squares (SS)	*df*	Mean Square	*F*	*p*	Observed Power
System Reliability	155526.000	1	155526.000	.024	.879	.052
Time Pressure	1.481E7	1	1.481E7	2.960	.099	.378
System Reliability*Time Pressure	572886.000	1	572886.000	.088	.769	.059
Error (System Reliability)	1.518E8	23	6600177.152			
Error (Time Pressure)	1.151E8	23	5005449.123			
Error (System Reliability*Time Pressure)	1.499E8	23	6515358.891			

The main effect of system reliability on MMI processing time was analyzed first and was found to be statistically insignificant with $F(1, 23) = .024$ and $p = .879$. The insignificance of this effect indicates that the participants' MMI processing times were not significantly lower (better) when they were exposed to the high system reliability (80%) than when they were exposed to the low system reliability (40%). As a result, hypothesis 7 was not supported.

The main effect of time pressure on MMI processing time was also analyzed with $F(1,23) = 2.960$ and $p = .099$ and was shown to be statistically insignificant ($p > 0.05$). The insignificance of this effect indicates that the participants' MMI processing times were not significantly lower (better) when they were exposed to the lower time pressure condition (10 seconds) than when they were exposed to the higher time pressure condition (5 seconds). As a result, hypothesis 14 was not supported.

In addition to the main effects for pop-up threats re-routing processing time reported previously, hypothesis 21 stated that an interaction will exist between system reliability and time pressure for secondary task MMI. The means and standard deviations are presented in Table 20.

Table 20: Means and Standard Deviations for the Interaction of System Reliability and Time Pressure on MMI Processing Time

				95% Confidence Interval	
System Reliability	Time Pressure	Mean	Standard Error	Lower Bound	Upper Bound
Low	High	9934.667	638.707	8613.400	11255.933
	Low	8994.500	642.628	7665.124	10323.876
High	High	9860.667	779.801	8247.526	11473.807
	Low	9229.500	949.164	7266.005	11192.995

The interaction effect between system reliability and time pressure was found to be statistically insignificant with $F(1, 23) = .088$ and $p = .769$. The insignificance of this effect indicates that there was no significant interaction between system reliability and time pressure for MMI processing time; as a result, hypothesis 21 was not supported.

Mental Workload

Mental workload was subjectively measured using the NASA-TLX after each trial. The subjective ratings were on a scale ranging from 0 to 100 on six different workload factors, with 100 being the highest level of workload and 0 being the lowest level and adjusted based on the pair-wise comparison among workload factors. Hypotheses 1, 8, and 15 referred to mental workload. Hypothesis 1 stated that when participants are exposed to high system reliability (80%), they will report lower mental workload than when they are exposed to low system reliability (40%). Hypothesis 8 stated that when participants are exposed to the lower time pressure (10 seconds), they will report lower mental workload than when they are exposed to the higher time pressure condition (5 seconds). A repeated measures ANOVA was conducted to test the hypotheses regarding mental workload. Table 21 shows the means and standard deviations for mental workload and the results of the ANOVA are presented in Table 22.

Table 21: Mental Workload Means and Standard Deviations

	Mean	Standard Deviation	N
40% System Reliability / High Time Pressure	39.139	19.910	24
40% System Reliability / Low Time Pressure	32.083	18.880	24
80% System Reliability / High Time Pressure	36.903	23.476	24
80% System Reliability / Low Time Pressure	35.862	20.305	24

Table 22: ANOVA Source Table for Mental Workload

Source Greenhouse-Geisser	Sum of Squares (SS)	*df*	Mean Square	*F*	*p*	Observed Power
System Reliability	14.281	1	14.281	.074	.789	.058
Time Pressure	393.470	1	393.470	4.292	.050	.510
System Reliability*Time Pressure	217.060	1	217.060	2.346	.139	.312
Error (System Reliability)	4467.827	23	194.253			
Error (Time Pressure)	2108.460	23	91.672			
Error (System Reliability*Time Pressure)	2128.183	23	92.530			

System reliability on mental workload was analyzed first and was found to be statistically insignificant with $F(1, 23) = .074$ and $p = .789$. The insignificance of this effect indicates that the participants' mental workload scores were not significantly lower (better) when they were exposed to the high system reliability (80%) than when they were exposed to the low system reliability (40%). As a result, hypothesis 1 was not supported.

The main effect of time pressure on mental workload was also analyzed with $F(1,23) = 4.292$ and $p = .050$ and was shown to be statistically significant ($p = 0.05$). The significance of this effect indicates that the participants' mental workload scores were significantly lower (better) when they were exposed to the lower time pressure condition (10 seconds) than when they were exposed to the higher time pressure condition (5 seconds). As a result, hypothesis 8 was supported.

The interaction between system reliability and time pressure for mental workload was also analyzed. Hypothesis 15 stated that an interaction will occur between the main effects. The means and standard deviations are presented in Table 23.

Table 23: Means and Standard Deviations for the Interaction of System Reliability and Time Pressure on Mental Workload

				95% Confidence Interval	
System Reliability	Time Pressure	Mean	Standard Error	Lower Bound	Upper Bound
Low	High	39.139	4.064	30.732	47.547
	Low	32.083	3.854	24.111	40.055
High	High	36.903	4.792	26.990	46.816
	Low	35.862	4.145	27.288	44.436

In addition to the main effects, the interaction effect between system reliability and time pressure was found to be statistically insignificant with $F(1, 23) = 2.346$ and $p = .139$; $p > 0.05$. The insignificance of this effect indicates that there was no significant interaction between system reliability and time pressure for mental workload; as a result, hypothesis 15 was not supported.

Trust

Trust was subjectively measured using a trust survey that was given after each treatment scenario. The trust survey assessed perceived reliability, technical competence, perceived understandability, faith, and personal attachment. It had a rating from 1 to 10 with 1 representing never happening or occurring and 10 representing always happening or occurring. Hypothesis 2 stated that the trust levels of participants will be higher (better) in the primary task when they are exposed to high system reliability (80%) than when they are exposed to low system reliability (40%). Hypothesis 9 stated that when participants are exposed to the lower time pressure condition (10 seconds), their trust levels will be higher (better) in the primary task than when they are exposed to the higher time pressure condition (5 seconds). Table 24 shows the means and standard deviations for trust for the four levels of treatment scenarios. Table 25 presents the results of the ANOVA and are meant to reflect percentages.

Table 24: Trust Means and Standard Deviations

	Mean	Standard Deviation	N
40% System Reliability / High Time Pressure	64.704	14.427	24
40% System Reliability / Low Time Pressure	66.529	12.305	24
80% System Reliability / High Time Pressure	68.529	12.339	24
80% System Reliability / Low Time Pressure	65.875	14.166	24

Table 25: ANOVA Source Table for Trust

Source Greenhouse-Geisser	Sum of Squares (SS)	*df*	Mean Square	*F*	*p*	Observed Power
System Reliability	60.325	1	60.325	1.784	.195	.249
Time Pressure	4.125	1	4.125	.105	.749	.061
System Reliability*Time Pressure	120.378	1	120.378	1.744	.200	.244
Error (System Reliability)	777.927	23	33.823			
Error (Time Pressure)	905.387	23	39.365			
Error (System Reliability*Time Pressure)	1587.645	23	69.028			

The main effect of system reliability on trust was analyzed first and was found to be statistically insignificant with $F(1, 23) = 1.784$ and $p = .195$. The insignificance of this effect indicates that the participants' trust levels were not significantly higher (better) when they were exposed to the high system reliability (80%) than when they were exposed to the low system reliability (40%). In this case, hypothesis 2 was not supported.

Time pressure on trust was also analyzed with $F(1,23) = .105$ and $p = .749$ and was shown to be statistically insignificant ($p > 0.05$). The insignificance of this effect indicates that the participants' trust levels were not significantly higher (better) when they were exposed to the lower time pressure condition (10 seconds) than when they were exposed to the higher time pressure condition (5 seconds). Hypothesis 9 was not supported in this case.

The interaction effect between system reliability and time pressure was also studied and was found to be statistically insignificant with $F(1, 23) = 1.744$ and $p = .200$. The insignificance of this effect indicates that there was no significant interaction between system reliability and time pressure for trust. As a result, hypothesis 16 was not supported. The means and standard deviations are presented in Table 26.

Table 26: Means and Standard Deviations for the Interaction of System Reliability and Time Pressure on Trust

				95% Confidence Interval	
System Reliability	Time Pressure	Mean	Standard Error	Lower Bound	Upper Bound
Low	High	64.704	2.945	58.612	70.796
	Low	66.529	2.512	61.333	71.725
High	High	68.529	2.519	63.319	73.739
	Low	65.875	2.892	59.893	71.857

The results previously described have been an account of the outcomes of the experimental findings. In the coming section, discussion, the perceived reasons of these results will be discussed.

Discussion

The objective of this study was to examine the effect of system reliability and time pressure on UAS operator task performance and mental workload when conducting certain tasks relating to UAV operation. The intention of this study was to broaden and advance current knowledge about the mental workload associated with operating a highly autonomous UAS, particularly with technology whose system is not 100% reliable during tasks and during extremely time sensitive operations. This study's aim was to provide knowledge, which can be applied to the design of future UAS systems, in order to improve operator performance through appropriate levels of mental workload. The knowledge provided is in regards to different levels of system reliability and time pressure and their effects on an operator trying to maintain a high level of performance in the midst of uncertainty and pressure. The results of this study are discussed here, organized into four main areas of interest: primary task performance measures, secondary task performance measures, mental workload, and trust for exploratory reasons.

Primary Task Performance Measures

There were two primary task performance measures collected during this study: image processing time and target acquisition accuracy.

Image Processing Time

From the results, image processing time showed significance for system reliability and no significant effect for time pressure. There was also found to be a no significant interaction of system reliability and time pressure with regards to image processing time. The significance for system reliability indicated that the participants' processing time scores were lower (better) when they were exposed to the high system reliability (80%) than when they were exposed to the low system reliability (40%). This coincided with predictions made about system reliability and image processing time. With high system reliability, participants did not need to make many clicks to fix the target recognizer, thus it was expected to save time, which was the one reason why such significance was believed to have occurred. An explanation for this is that when system reliability was high, it meant that most of the targets were pre-selected by the ATR from the MIIIRO software; as a result, participants only needed to click "accept" rather than select targets, or deselect distracters, and then click on "accept;" it saved time, thus having a lower processing time.

There was no significance found for time pressure on image processing time. This insignificant effect of time pressure on image processing time suggests that participants' processing time scores were not significantly lower (better) when they were exposed to the lower time pressure condition (10 seconds) than when they were exposed to the higher time pressure condition (5 seconds). This directly contradicted the predictions made about time pressure regarding image processing time. The fact that the participants were under more time pressure (5 seconds) to process the images possibly led them to exceed the 5 second time limit thus having the MIIIRO system automatically answering for them. This would also explain why there was a significant effect for system reliability and not for time pressure. High time pressure, such as 5 seconds, does not give participants enough time to properly think through a task in order to make a clear and concise decision; either a wrong decision will be taken or no decision at all, which in the latter case would result in having the MIIIRO system to time out and move on to the next target.

The interaction of system reliability and time pressure showed insignificant results for image processing time. The insignificance of this effect indicates that there was no significant interaction between system reliability and time pressure on image processing time. In other words, the effects of system reliability and time pressure did not impact the time it took for participants to process the images presented to them. This interaction was hypothesized to be significant, but a possible explanation for this occurrence could be related to the insignificance of time pressure on image processing time. If participants were not pressured for time, they may have been able to more carefully examine the images presented to them than under a time constraint; however, when under a more rigorous time pressure constraint, participants may have resorted to guessing or completely missing the opportunity to guess by having the MIIIRO software answer for them, resulting in insignificant processing times. One explanation for this could be from the side effects that occur as a result of time pressure, such as stress and poor decision making. Wickens, et al. (2004) described that an increase in stress levels, caused by time constraints, causes an operator to have lower performance ratings than if he or she had more time allotted and lower stress levels. High time constraints and stress also play a role in decision making, altering a person's ability to clearly think through a problem and make a precise and accurate decision. These reasons could explain why time pressure had so significant effect on image processing time.

Target Acquisition Accuracy

The second primary performance measure collected in this study was target acquisition accuracy. Results showed that there was no significant effect for system reliability on target acquisition accuracy. The insignificance of this effect indicates that the target acquisition accuracy of participants was not significantly higher (better) when they were exposed to the high system reliability (80%) than when they were exposed to the low system reliability (40%). In other words, the participants' accuracy did not matter on whether or not the reliability of the system was high or low. This could be due to a number of reasons, including the reality that participants did not know the reliability of the software prior to their experience. As a result, they may have treated each trial in the same manner and took their time to make each decision. This may have occurred in place of putting complete trust in the software's ability to choose the correct targets with the ATR tool, which had the ability to recognize targets or objects based on pre-set data obtained from the MIIIRO software.

There was a significant effect for time pressure on target acquisition accuracy. The significance of this effect indicates that the target acquisition accuracy of participants was higher (better) when they were exposed to the lower time pressure condition (10 seconds) than when they were exposed to the higher time pressure condition (5 seconds). A reason for this could be due to the fact that having more time, or having lower time pressure, allowed the participants' to more carefully examine the images presented to them and thus making better decisions as to which image was a target and which image was a distracter. A trade-off effect may have affected the results of target acquisition accuracy in which the quality of system reliability was lost in return for gaining better quality from the time pressure. This could be seen from the results that showed a significant effect for time pressure in contrast to system reliability, which did not have a significant effect on the accuracy of target acquisition.

There was no significant interaction found between system reliability and time pressure for target acquisition accuracy. This could be due to the insignificant effect that system reliability had on accurately acquiring targets. It could also be due to the manner in which participants treated each trial; the same manner. As a result, accuracy scores were compromised.

Secondary Task Performance Measures

Secondary task performance measures were often used as another measure of mental workload, attempting to determine how much excess capacity was available while performing the primary task. There were three secondary task performance measures involved in the current study: IA processing time, pop-up threats re-routing processing time, and MMI processing time.

Intruder Aircraft (IA) Processing Time

The secondary task of IA processing required the participants to respond to a red aircraft icon by clicking on the icon and typing in a given code, which appeared on the TSD. IA processing time yielded no significant differences for the effects of system reliability, time pressure, and their interaction. As a result, hypotheses 5, 12, and 19 were not supported. A possible explanation for this lack of significance is the number of IA events that took place during each treatment scenario; only two IA events occurred per treatment. With such few opportunities for the different tasks to conflict, the system reliability and time pressure effects placed on the primary task had little chance to affect performance on IA processing time. From the results, a conclusion can be made that the participants either had enough time to respond; consequently, if participants missed the response time, the IA would disappear and count against them. However, there were not enough IA events to make a significant difference.

Pop-up Threats Re-routing Processing Time

The secondary task measure for pop-up threats re-routing processing time required the participants to either accept or reject a recommended flight path change made by the automation in order to avoid a threat which had appeared during the simulation. Pop-up threats re-routing processing time yielded no significant differences for the effects of system reliability and time pressure. As a result, hypotheses 6 and 13 were not supported. A possible explanation for this lack of significance was the number of pop-up threats that took place during each treatment scenario; only two events occurred per treatment. In addition, the automation automatically "accepted" the flight path change if the participant missed his or her time frame to "accept" or "reject;" the time out period would impact the significant results. The interaction between system reliability and time pressure, however, was shown to be significant, thus hypothesis 20 was supported.

Mission Mode Indicator (MMI) Processing Time

The secondary task of MMI processing required the participants to respond to either a yellow light or a red light within the indicator. The participant would respond by clicking on whichever light came one and typed in a number string given to them by the MIIIRO software. The processing time for MMI showed no significant results for the main effects of system reliability and time pressure, as well as no significant results for their interaction. This lack of significance does not support hypotheses 7, 14, or 21. An explanation for this lack of significance could be due to the participants being more concerned with image processing than with secondary tasks; it is evident that when participants take longer to process the images, as in the case of the low system reliability images, secondary task performance is affected. As MMI is another secondary task, its priority is of less importance in comparison to primary tasks; in addition, changes in this task are not easily detected, unless they are looked at specifically. This may be an explanation which fits the data for the interaction.

Mental Workload

Mental workload was subjectively measured using the NASA-TLX after each treatment scenario, resulting in four NASA-TLX forms for each participant. The subjective ratings were on a scale ranging from 0 to 100, with 100 being the highest level of workload and 0 being the lowest level. The results for mental workload showed no significant results for the main effects of system reliability and time pressure, as well as no significant results for their interaction. This lack of significance does not support hypotheses 1, 8, and 15. As a result, the results do not support the argument made regarding an increase in mental workload as system reliability decreases and time pressure increases. One explanation for these results is the tasks involved in the study. Each individual participant had his or her own perceived thoughts on the mental workload they experienced during each task in a trial. This, in turn, leads to another explanation, which is the sensitive nature and validity of the NASA-TLX itself. The NASA-TLX is based around personal feelings and opinions, hence why it is subjective and not objective. This fact alone plays a major role in the outcomes of the results as each person perceives situations and tasks differently from the next person; as a result, mental workload is difficult to accurately measure. In addition, the results of mental workload are in correspondence with the results of the secondary task performance measures in terms of their insignificance. An explanation for this could be that due to the insignificance of the secondary tasks, mental workload was not at an increased level to the participants and thus resulted in insignificant results.

Trust

Trust was assessed for exploratory reasons and was not a main focus of the study. Participants were given a trust survey to fill out after each treatment scenario, resulting in four surveys per participant. The trust survey assessed perceived reliability, technical competence, perceived understandability, faith, and personal attachment. It had a rating from 1 (never) to 10 (always). The results for trust showed no significant results for the main effects of system reliability and time pressure, as well as no significant results for their interaction. The insignificance of these effects indicate that the participants' trust levels were not significantly higher (better) when they were exposed to the higher system reliability condition (80%) and lower time pressure condition (10 seconds). As an implication to the results, hypotheses 2, 9, and 16 were not supported. This could be due to the participants' not having enough knowledge about the MIIIRO software. It could also be due to not having enough trust in the reliability of the system, in addition to not having enough trials over an extended time period to build trust levels with the software.

Practical Implications

The next generation of UAS operation, both abroad and within the NAS, holds many opportunities for the future of flight. Many tasks can be performed more safely and efficiently by unoccupied aircraft; however, in order to implement the use of UAS for these tasks, more research and development needs to take place in order to assess the capabilities and limitations of the aircraft system. Researching factors that influence operator performance and mental workload need to be sought out in order to gain more knowledge and broaden understanding. This should be done so as to avoid possible hazards that could lead to UAV accidents related to poor performance levels due to too high or too low levels of mental workload.

System reliability plays a crucial role in determining how to train operators and how to evaluate which operators are better suited for performance, in cases where the reliability of the system is in jeopardy; however, time pressure also plays an integral role in determining how well a task can be performed and in determining its effect on an operator's mental workload. Time pressure is particularly important when dealing with time sensitive environments, such as UAS operation. Research focusing on trying to understand the factors that increase time pressure will be important to study as these types of research could help in future designs of systems that help alleviate some of those factors. Understanding how and when the automation should assist the UAS operator will be crucial to know in order to avoid the effects of time pressure, in addition to avoiding the effects of increased mental workload.

This study demonstrated a significant effect of system reliability on processing times for images. A significant effect of time pressure on target acquisition accuracy was also found in the primary task performance measures. The time pressure differences for image processing time, however, were small enough that they may not require any design changes. Although time pressure was found to be statistically insignificant for many of the secondary tasks, the time differences may be not negligible when dealing with real-life situations. The topic of system reliability and time pressure should be researched further and reviewed before implementing design strategies.

Recommendations for Future Research

The next generation of flight is already underway and UAS is in the midst of aviation's next generation. These highly autonomous vehicles involved in UAS can open many doors and broaden many horizons on how operators perform crucial and time sensitive tasks. Research and development on issues relating to UAS operator performance can open doors for understanding what is needed, or not needed, for one operator to supervise multiple vehicles at the same time. What kind of system is needed? What level of mental workload is expected if an operator is assigned to multiple vehicles at once? These questions can lead to new questions and more research efforts. The workload involved with multiple UAS supervision is not known. However, through further research efforts, more knowledge and understanding can be developed in applying what is already known for supervision of one UAS to that of multiple UAS. The inclusion of time pressure, in future research, as a factor when assessing performance would also help in understanding workload levels of multiple UAS supervision.

Future research on UAS operation can go in many directions, as many factors go into operating such highly autonomous systems. Another research effort that could be taken into consideration is gender in choosing operators. Research efforts on gender differences and how well each multi-task could help in gaining knowledge on what the needs are of designing systems from a psychological perspective.

Other research efforts could include the manipulation of participants' knowledge regarding the reliability of the system and the time constraint that is involved, in addition to providing feedback. The manipulation could go in many ways, including having two groups, one with prior knowledge of the effects and another group with no knowledge. Feedback could be given to the participants following each trial, as well, in order to give them the chance to improve in the remaining trials of the study.

In addition to the above mentioned factors, system reliability and the uncertainty surrounding the system are crucial factors to take into consideration when performing research and development studies for UAS design purposes. Due to the operator being separated from the UAV and the environment surrounding it, it is necessary to have reliable systems and designs. This is necessary in order to have an operator who is comfortable enough to rely on the system but, yet at the same time, knowledgeable enough to know when something is going wrong with the system and realizing that taking manual charge is necessary. Due to the higher level of uncertainty related to UAS operation in comparison to manned flight, uncertainty in system reliability is still a notion that can provide insight into operator performance and mental workload in the course of UAS flight. By further understanding system reliability and uncertainty when conducting UAS operations, UAS designs can be implemented to reduce the uncertainty that contributes to higher levels of mental workload and lower levels of performance.

Conclusion

UAS play a crucial role within military and security operations. Their use has grown considerably in the last decade and continues to grow, making UAS an important part of NextGen's and the NAS's future. UAS have a wide range of capabilities that allow them to provide a much safer and efficient method for performing a number of tasks while not putting an operator's life in jeopardy. Although a lot of research and development has taken place within the history of UAS operations, a number of concerns still exist with regard to UAS flight safety within the NAS and abroad. These concerns need to be addressed before the full potential of this technology is realized in order to take full advantage of all its capabilities. Designing this technology with the human in mind is necessary as understanding the human component of these systems would help resolve many of the human factors concerns, such as operator performance and mental workload. Once the human component is understood, issues of concern can be solved through design by providing the ability, for all the system's components, to perform at optimum levels. Studies such as this one will remain to be of importance to continue in order to improve UAS design so that the available technology could be used to its fullest potential to increase safety and efficiency.

The factors researched in this study, system reliability and time pressure on UAS operator performance and mental workload, has contributed to the knowledge and understanding of UAS operations. However, continuous research, applied to the findings of this study, is indispensable to future research and development involving UAS tasks and system designs that affect operator performance and mental workload, as well as other human factors concerns in order to create more safe and efficient UAS environments.

References

ALEA E-Newspaper (2010). In. FAA Approves Use of UAVs for Border Protection. Retrieved March 31, 20011 from http://www.alea.org/public/newsletters/10_07/index.htm

Amabile, T. M., Mueller, J. S., Simpson, W. B., Hadley, C. N., Kramer, S. J., & Fleming, L. (2002). "Time Pressure and Creativity in Organizations: A Longitudinal Field Study." Paper # 02-073, substantially revised version of HBS Working Paper #01-023, "The Influence of Time Pressure on Creative Thinking in Organizations."

Boussemart, Y., Donmez, B., Cummings, M. L., & Las Fargeas, J. (2009). Effects of time pressure on the use of an automated decision support system for strike planning. In *Proceedings of the 15th International Symposium on Aviation Psychology,* Dayton, OH.

Burke, K. A., Oron-Gilad, T., Conway, G., & Hancock, P. A. (2007). Friend/foe identification and shooting performance: Effects of prior task loading and time pressure. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 156-160.

Crescenzio F. D., Miranda G., Periani F., & Bombardi T. (2007). Advanced interface for UAV ground control station. In *Proceedings of the AIAA Modeling and simulation Technologies Conference and Exhibit*, Hilton Head, SC, AIAA 2007-6566.

Dixon, S. R., Wickens, C. D., & Chang, D. (2003). Comparing quantitative model predictions to experimental data in multiple-UAV flight control. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting,* 104-108.

Driskell, J. E., & Salas, E. (1996). *Stress and human performance*. Mahwah, NJ: Erlbaum.

Dorr, Jr., L., & Duquette, A. (2010). Fact Sheet – Unmanned Aircraft Systems (UAS). Retrieved March 18, 2011, from http://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=6287

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

Federal Aviation Administration Task E: UAS Demo (3) Test Plan. (2010).

Federal Aviation Administration. (2009). NextGen flight deck human factors research and development (R&D) program. Retrieved March 19, 2011 from https://www2.hf.faa.gov/HFPortalNew/NextGenRD.aspx?AspxAutoDetectCookieSupport=1

Federal Aviation Administration. (2011). Why NextGen matters. Retrieved March 19, 2011 from http://www.faa.gov/nextgen/why_nextgen_matters/

Galster, S., Nelson, W. T. & Bolia, R. (n.d.). *The Utilization of Synthetic Task Environments in Command and Control (C2) Domains*. [PowerPoint Slides]. Retrieved March 23, 2011, from http://www.dodccrp.org/events/10th_ICCRTS/CD/presentations/358.pdf

Hancock, P. A., & Meshkati, N. (1988). *Advances in psychology: Human mental workload*. Amsterdam, The Netherlands: Elsevier Science Publishers B.V.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, 904-908. Santa Monica: HFES.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: North Holland Press.

Helander, M. G., Landauer, T. K., & Prabhu, P. V. (1997). *Handbook of Human-Computer Interaction (2nd ed.)*. Amsterdam, The Netherlands: Elsevier Science Publishers B.V.

Hopkin, V. D., & Wise, J. A. (1996). Human factors in air traffic system automation. In R. Parasuraman & M. Mouloua (eds.), *Automation and human performance: Theory and applications* (pp.319-336). Mahwah, NJ: Erlbaum.

Hottman, S. B., & Sortland, K. (2006). UAV operators, other airspace users, and regulators: Critical components of an uninhabited system. In Nancy J. Cooke, Heather L. Pringle, Harry K. Pedersen, Olena Connor (ed.) *Human Factors of Remotely Operated Vehicles (Advances in Human Performance and Cognitive Engineering Research, Volume 7)*, Emerald Group Publishing Limited (pp.71-88). New York, NY: JAI Press.

Hughes, L. E. (2004). *Effects of time pressure and mental workload on physiological risk factors for upper extremity musculoskeletal disorders while typing*. (Master's thesis). Retrieved March 21, 2011, from http://scholar.lib.vt.edu/theses/available/etd-07262004-163847/unrestricted/LauraHughesThesis.pdf

Hughes, L., & Babski-Reeves, K. (2005). Time pressure and mental workload effects on perceived workload and key strike force during typing. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting.* 1390-1394.

IA Tech, Inc. (n.d.). MIIIRO: Multi-Modal User-Interface for Remote System Operation. Retrieved March 3, 2011, from http://www.ia-tech.com/

Jaramillo, M. (2011). *The Effects of System Reliability and Uncertainty under High Time Pressure on Operator Performance for Unmanned Aerial Vehicles*, thesis final report. Embry-Riddle Aeronautical University. Retrieved February 28th, 2011.

Krock, L. (2002). Spies that fly: Time line of UAVs. *NOVA Science Programming On Air and Online*. Retrieved March 10, 2011 from http://www.pbs.org/wgbh/nova/spiesfly/uavs.html

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

McCarley, J. S., & Wickens, C. D. (2004). *Human Factors Concerns in UAV Flight*. University of Illinois Institute of Aviation Technical Report. Urbana-Champaign, IL: Aviation Human Factors Division.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology, 8*, 47-63.

NASA TLX: Task Load Index (n.d.). Retrieved March 12, 2011, from http://human-factors.arc.nasa.gov/groups/TLX/

Nisser, T. & Westin, C. (2006). Human factors challenges in Unmanned Aerial Vehicles (UAVs): A literature review. Lund University School of Aviation. Retrieved March 13, 2011 from http://www.lu.se/upload/Trafikflyghogskolan/HumanFactorsChallengesUnmannedAerialVehicles.pdf

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

Peterson, T. (2010). *Effect of Time Pressure and Task Uncertainty on Human Operator Performance and Workload for Autonomous Aerial Vehicle Missions,* thesis final report. Embry-Riddle Aeronautical University. Retrieved February 28th, 2011.

Prevot, T., Lee, P., Smith, N., & Palmer, E. (2005). ATC technologies for controller-managed and autonomous flight operations. In *Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit (AIAA 2005-6043)*, San Francisco, CA.

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and Workload Profile methods. *Universidad Complutense de Madrid, Spain*. Applied Psychology: An International Review, 2004, 53 (1), 61–86.

Sarno, K. J., & Wickens, C. D. (1995). The role of multiple resources in predicting time-sharing efficiency. *International Journal of Aviation Psychology*, *5*(1), 107-130.

Sarter, N. B., & Woods, D. D. (2000). Teamplay with a powerful and independent agent: A full-mission simulation study. *Human Factors, 42*(3), 390-402.

Slamecka, N. J., & Graf, P. (1978). The generation effect: delineation of a phenomena. *Journal of Experimental Psychology: Human Learning & Memory, 4*, 592-604.

Svenson, O., & Maule, A. J. (1993). *Time pressure and stress in human judgment and decision making*. New York, NY: Plenum Press.

Tso, K. S., Tharp, G. K., Tai, A. T., Draper, M. H., Calhoun, G. L., & Ruff, H. A. (2003). A human factors testbed for command and control of unmanned air vehicles. *Digital Avionics Systems 22th Conference,* 8.C.1-1-12.

Unmanned Aerial Vehicle. (n.d.). Retrieved May 31, 2011, from http://en.wikipedia.org/wiki/Unmanned_Aircraft_System#UAS

Weil, S. A., Freeman, J., MacMillan, J., Jackson, C., Mauer, E., Patterson, M., & Linegang, M. P. (2006). Designing of a multi-vehicle control system: System design, team composition, and user interaction. In N. J., Cooke, et al. (Eds.), Advances ... (pp. 223-236) New York, NY: JAI Press.

Weiner, I. B., Freedheim, D. K., Schinka, J. A., Velicer, W. F. (2003). *Handbook of Psychology: Research methods in psychology (2nd volume)*. Hoboken, NJ: John Wiley & Sons, Inc.

Wickens, C. D., Lee, J., Liu, Y., & Becker, S. G. (2004). *An introduction to human factors engineering (2nd ed.)*. Upper Saddle River, NJ: Pearson Education, Inc.

Wiener, E. L. (1988). Cockpit automation. In E. L. Wiener & D. C. Nagel (eds.), *Human factors in aviation* (pp. 433-461). San Diego: Academic Press.

Wise, J. A., Hopkin, V. D., & Garland, D. J. (2010). *Handbook of aviation human factors (2nd ed.)*. Boca Raton, FL: CRC Press Taylor and Francis Group, LLC.

Appendix A

IRB Number: 11-130

Informed Consent Form
Unoccupied Aircraft Systems (UAS) System Reliability and Time Pressure Study

Conducted by Rania Wageh Ghatas
Advisor: Dr. Dahai Liu
Embry-Riddle Aeronautical University
600 South Clyde Morris Blvd, Daytona Beach, FL 32114

This is to certify that you hereby agree to participate in this research project, which is conducted by Rania Ghatas, a Human Factors and Systems Psychology graduate student at Embry-Riddle Aeronautical University, sponsored by Dr. Dahai Liu, (386) 226-6214. You acknowledge that you will receive a demographic survey and will participate in one session that will last approximately one and a half hours. During this session, you will be asked to complete four computer-based UAS simulation assessments and fill out two questionnaires after each assessment; one regarding your perceived feelings of mental workload and the other of trust.

You acknowledge that you have had the opportunity to ask questions about this study and feel comfortable with what will be expected of you. All information gathered from your participation will be completely confidential. Your name from the consent form will not be linked to the data collected from the experiment. You will be compensated for your participation by receiving either extra credit or $10 and will have the chance to win $50 for best overall performance. You may withdraw from the study at any time. Your assistance will help us better understand the effects of system reliability and time pressure on UAS operator performance and mental workload. If you wish to receive a copy of form or a final report of the study, you may do so when the results of the study are finalized.

There are no foreseen risks to participants. If at any time during this study you decide that you need to talk to a counselor, the ERAU counseling center can be reached at (386) 226-6035.

Thank you for your participation. If you have any questions, please ask during the experiment or feel free to call Rania Ghatas at (321) 276-9596 or Dr. Dahai Liu at (386) 226-6214.

Statement of Consent

I acknowledge that my participation in this experiment is entirely voluntary and I am free to withdraw at any time. I have been informed as to the general scientific purposes of the experiment. If I choose to withdraw from the experiment before its termination, I will not receive any compensation.

Participant's Name (Please Print): ____

Signature of Participant: ______________________________ Date: ________________

Experimenter: _____________________________________ Date: ________________

Appendix B

Biographical Questionnaire

ID#: ____________________ **Date: ____________________**

Please fill in the blanks or circle the appropriate response.

1. What is your age? ________ years

2. What is your gender? Male / Female

3. Do you have normal or corrected to 20/20 vision? Yes / No

4. Are you color blind? Yes / No

5. Are you: Right-handed / Left-handed

6. How many hours per week do you use computers: _____ hours

7. On a scale of 1 to 5, what is your confidence level in using computers:

Low confidence - 1 2 3 4 5 - High confidence

8. On average, how many hours per week do you spend playing computer/video games?
   0-5___ 6-10___ 11-15___ 16-20___ 21-25+___

9. What type of genre of gaming are you most accustomed to playing?
   Action___ Adventure___ Role-Playing___ Strategy___ Simulation___

10. Have you had any other experience participating in unoccupied aircraft simulations?
    Yes / No

11. Do you have any experience flying an unoccupied aircraft or remote controlled aircraft?
    Yes / No

If so, please explain: ____

Adapted from: Peterson, T. (2010). *Effect of Time Pressure and Task Uncertainty on Human Operator Performance and Workload for Autonomous Aerial Vehicle Missions*, thesis final report. Embry-Riddle Aeronautical University. Retrieved February 28th, 2011.

Appendix C

ID#: ____________ NASA Task Load Index (TLX) Form Date: ____________

We are interested in your subjective experience of workload ***for each test trial you completed***. Workload is a difficult concept to define precisely, but a simple one to understand generally. The factors that influence your experience of workload may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you felt.

One way to find out about workload is to ask people to describe the feelings they experienced. Because workload may be caused by many different factors, we would like you to evaluate several of them individually rather than lumping them into a single global evaluation of overall workload. This set of six rating scales was developed for you to use in evaluating your experiences during the test trial.

Please indicate the level of workload you experienced on each of the 6 scales by circling the line at the point which best reflects the level of workload you experienced. The ends of the scales are labeled to indicate very low and very high workload. Points in between those end points represent intermediate values of workload. ***Please note that the Performance scale goes from Good on the left to Bad on the right.*** *This order has been confusing for some people.*

EFFORT — How hard did you have to work (mentally and physically) to accomplish your level of performance?

| | | | | | | | | | | | | | | | | | | | |

Low **High**

PERFORMANCE — How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

| | | | | | | | | | | | | | | | | | | | |

Good **Poor**

FRUSTRATION LEVEL — How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

| | | | | | | | | | | | | | | | | | | | |

Low **High**

TEMPORAL DEMAND — How much time pressure did you feel due to the rate or pace at which the tasks or events occurred? Was the pace slow and leisurely, or rapid and frantic?

| | | | | | | | | | | | | | | | | | | | |

Low **High**

MENTAL DEMAND — How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching)? Was the task easy or demanding, simple or complex, forgiving or exacting?

| | | | | | | | | | | | | | | | | | | | |

Low **High**

PHYSICAL DEMAND — How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating)? Was the task physically easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

| | | | | | | | | | | | | | | | | | | | |

Low **High**

NASA Task Load Index (TLX) Weighting Form

The forms you filled out included six rating scale factors that can influence workload. We are interested in your assessment of the relative contribution of these factors to your experience of workload for ***each test trial you completed***.

People vary in their opinion of what contributes to workload. For example, some people feel that mental or temporal demands are the essential aspects of workload regardless of the effort they expended or the performance they achieved. Others feel that if they performed well, the workload must have been low and if they performed poorly, the workload must have been high. Yet others feel that effort or feelings of frustration are the most important factors in workload, and so on.

In addition, the factors that create levels of workload differ depending on the task. For example, some tasks might be difficult because they must be completed very quickly. Others may seem easy or hard because of the intensity of mental or physical effort required. Yet others feel difficult because they cannot be performed well, no matter how much effort is expended.

The evaluation you are about to perform is a technique developed by NASA to assess the relative importance of the six factors that were included in the workload rating scale in determining how much workload ***you*** experienced across all the test trials you just completed.

Below is a list of pairs of rating scale titles (for example Effort vs. Mental demand). For each pair, please circle the item that was **more important** to ***your*** experience of workload across all the test trials you just completed.

MENTAL DEMAND	VS	PHYSICAL DEMAND
TEMPORAL DEMAND	VS	MENTAL DEMAND
PHYSICAL DEMAND	VS	TEMPORAL DEMAND
EFFORT	VS	PERFORMANCE
PERFORMANCE	VS	FRUSTRATION
TEMPORAL DEMAND	VS	PERFORMANCE
MENTAL DEMAND	VS	PERFORMANCE
PERFORMANCE	VS	PHYSICAL DEMAND
EFFORT	VS	FRUSTRATION
TEMPORAL DEMAND	VS	EFFORT
EFFORT	VS	MENTAL DEMAND
PHYSICAL DEMAND	VS	EFFORT
FRUSTRATION	VS	TEMPORAL DEMAND
MENTAL DEMAND	VS	FRUSTRATION
FRUSTRATION	VS	PHYSICAL DEMAND

Appendix D

Trust Survey

The purpose of this questionnaire is to collect information about your working experience, your perception, and your trust of the Multi-modal Immersive Intelligent Interface for Remote Operations (MIIIRO) software and its automation.
Please base your assessment on your experience with the automation used in the simulation.

ID #: _______________ Treatment #: _______________ Date: __________________

<u>Perceived Reliability</u>

1. I could rely on the automation to function properly

1 2 3 4 5 6 7 8 9 10

Never Always

2. The automation performed reliably under a variety of conditions

1 2 3 4 5 6 7 8 9 10

Never Always

3. The automation provided an alert when I was required to make my decision

1 2 3 4 5 6 7 8 9 10

Never Always

4. The automation generates false-alerts

1 2 3 4 5 6 7 8 9 10

Never Always

5. The automation misses genuine conflicts/risks

1 2 3 4 5 6 7 8 9 10

Never Always

Technical Competence

6. The automation has appropriate operational understanding of what a conflict is (parameters used/classifications, etc.)

   1 2 3 4 5 6 7 8 9 10

   Never Always

7. The alert the automation produces is as good as that which a competent person could produce

   1 2 3 4 5 6 7 8 9 10

   Never Always

8. The automation correctly makes use of all the knowledge and information available to it to produce an alert

   1 2 3 4 5 6 7 8 9 10

   Never Always

Perceived Understandability

9. I understand well how the automation behaves

   1 2 3 4 5 6 7 8 9 10

   Never Always

10. I understand how the automation can assist me with decisions I have to make

    1 2 3 4 5 6 7 8 9 10

    Never Always

11. It is easy to interpret the automation output

    1 2 3 4 5 6 7 8 9 10

    Never Always

Faith

12. I believe the automation alerts even in uncertain situations

1 2 3 4 5 6 7 8 9 10

Never Always

13. When I am uncertain about a situation I believe the automation rather than myself

1 2 3 4 5 6 7 8 9 10

Never Always

Personal Attachment

19. I would feel a sense of loss if the automation were unavailable

1 2 3 4 5 6 7 8 9 10

Never Always

19. I find the automation suitable to my method of working

1 2 3 4 5 6 7 8 9 10

Never Always

16. I like using the automation for decision-making

1 2 3 4 5 6 7 8 9 10

Never Always

17. I have a personal preference of making decisions with automation rather than without automation

1 2 3 4 5 6 7 8 9 10

Never Always

18. At what level of reliability did you think the automation was working? (Please circle your answer)

40% 50% 70% 80% 90% 99%

19. Do you feel that the level of trust in the system in order to make decisions was affected by the reliability? (Please circle your answer)

Yes No

Adapted from: Jaramillo, M. (2011). *The Effects of System Reliability and Uncertainty under High Time Pressure on Operator Performance for Unmanned Aerial Vehicles*, thesis final report. Embry-Riddle Aeronautical University. Retrieved February 28th, 2011.