EVALUATION OF AN ADAPTIVE AUTOMATION TRIGGER BASED ON TASK PERFORMANCE, PRIORITY, AND FREQUENCY

THESIS

Crystal A. Miller, First Lieutenant, USAF

AFIT-ENV-13-J-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-13-J-01

EVALUATION OF AN ADAPTIVE AUTOMATION TRIGGER BASED ON TASK PERFORMANCE, PRIORITY, AND FREQUENCY

THESIS

Presented to the Faculty

Department of Systems and Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Research and Development Management

Crystal A. Miller, BS

First Lieutenant, USAF

June 2013

AFIT-ENV-13-J-01

EVALUATION OF AN ADAPTIVE AUTOMATION TRIGGER BASED ON TASK PERFORMANCE, PRIORITY, AND FREQUENCY

Crystal A. Miller, BS
First Lieutenant, USAF

Approved:

______________________________	__________
Michael E. Miller, PhD (Chairman)	Date
______________________________	__________
Gloria L. Calhoun, Senior Research Engineer Psychologist (Member)	Date
______________________________	__________
Joseph R. Wirthlin, Lt Col, USAF (Member)	Date

Abstract

As demand for the number of Unmanned Aerial Vehicle (UAV) sorties increases faster than the number of available operators, a significant Air Force research thrust includes the vision of a single operator supervising multiple UAVs; this involves increasing use of automation, creating the potential for the operators to become complacent and over-reliant on automation. To avoid operator complacency, adaptive automation has been proposed, where changes in automation are triggered based upon operator performance or other attributes. This research sought to understand the effect of a weighted method for triggering changes in automation within a multitasking environment as compared to a more traditional method in which performance on tasks is treated equally. In this work, the weighted method considered the priority of each task when computing a measure of operator performance on which to trigger changes in automation. Although overall system, consisting of both the operator and automation system, performance was not statistically different between the two trigger implementations, the participants with the priority based triggering scheme tended to rate the level of automation changes as more aligned with their actual performance and were significantly less surprised by the actions of the automation than those participants with the non-weighted approach. The results of this study, combined with participant preference for workload based adaptations, suggest a benefit to the implementation of a hybrid approach. Future research should focus on task weights based on priority and operator specific threshold criteria, where automation aides are triggered once the summation of current tasks exceeds the given threshold.

AFIT-ENV-13-J-01

I would like to dedicate this thesis to my amazing husband. Thank you so much for being my rock through it all!

Acknowledgments

I would like to thank my advisor, Dr. Miller, for all of his patience and guidance throughout this entire process. You were truly a major part of my success. I would also like to thank my team leader, Gloria Calhoun, and the entire 711 HPW/RHCI family for supporting my crazy schedule, help with data collection, and standing behind me the entire way.

Crystal A. Miller

Table of Contents

Page

Page

List of Figures

Page

List of Tables

EVALUATION OF AN ADAPTIVE AUTOMATION TRIGGER BASED ON TASK PERFORMANCE, PRIORITY, AND FREQUENCY

I. Introduction

General Issue

With demand for the number of Unmanned Aerial Vehicle (UAV) sorties increasing faster than the number of available operators, a significant Air Force initiative is to explore technologies that support increasing the effectiveness of UAV operations. An approach to this problem includes increasing automation to lessen manpower requirements per sortie. This approach has the potential to result in significant savings as current operations require more than one operator per UAV. As a result, UAVs are becoming increasingly automated with the goal of reducing operator workload and ideally inverting the ratio such that a single operator can manage multiple UAVs. While many segments of flight can be fully automated, it is not possible to anticipate all operational conditions and therefore, human judgment is required to respond to certain complex, rapidly evolving and time-sensitive events. These events are not predictable or necessarily even detectable by the automation. Therefore it is critical that the operator be aware of the status of the vehicles and be able to modify system behavior under circumstances that the automation is not responding correctly.

Unfortunately automation can have unintended, negative consequences on the human's ability to detect and respond to automation failures or lapses. Some negative impacts of automation on operator behavior are complacency, reduced situational awareness, decision biases, vigilance gaps, over- or under-reliance on automation due to trust issues, and workload problems (Endsley & Kaber, 1999b; Sheridan & Parasuraman, 2006). For example, complacency can happen when the human does not feel a vital part of the system. As the system becomes increasingly automated, the human may become less conscious of the status of the system and current processes. Additionally, he/she will have less opportunity to practice the skills that are necessary to recover from unexpected failures when they arise. Moreover, an operator that does not understand the decision processes and actions employed by the automation will likely not trust the actions of the system. In instances where the system is not viewed as accurate and trustworthy, the operator is unlikely to relinquish any control to the automation, annulling any anticipated gains in effectiveness.

To overcome these problems and achieve an optimal balance of operator involvement and application of automation, it is important to ensure that the appropriate level of automation (LOA) is used for each task. One possible approach is to employ adaptive automation (AA) in which the LOA applied to each task changes in response to the current needs of the mission and the operator (Feigh, Dorneich & Hayes, 2012). For instance, as operator performance on mission related tasks degrades under increased workload/cognitive demands, either higher LOAs can be applied for one or more tasks or the number of tasks that are automated increases.

Problem Statement

To implement adaptive automation, the system designer must select the functions to automate, the degree to which they must be automated, and the conditions under which each function should be automated (de Visser, LeGoullon, Freedy, Freedy, Weltman & Parasuraman, 2008). These design choices become more difficult for complex application environments where the operator must perform multiple tasks. For example, all tasks could have the system's global LOA, or each task could have an independently-determined LOA, or an LOA that is personalized to the operator.

Research by Szalma and Taylor (2011) indicates that individual differences should be taken into account to determine which functions to automate and the LOA. For adaptive automation applications, one method is to automatically monitor the operator's real-time task performance and select the LOA as a result of this performance. Recent research has examined alternative methods for adapting the LOA of an image analysis task in multi-task simulations (Calhoun, Ward & Ruff, 2011; Calhoun, Ruff, Spriggs & Murray, 2012). In these experiments, measures of the participant's individual performance on multiple task types were used in the adaptive scheme to determine *when* and *how* to adapt the image analysis task LOA. While both of these experiments demonstrated the potential value of adaptive automation, the results also highlighted how specific parameters of the performance-based algorithm can influence the frequency and appropriateness of LOA changes. For example, an asymmetrical adaptive scheme in which performance thresholds differed in respect to increasing versus decreasing LOA helped keep task LOA at a lower autonomy level where automation-induced problems are less likely (Calhoun, et al., 2012).

To date, these performance-based adaptive automation experiments conducted within a multi-UAV, multi-task simulation have employed algorithms that are based solely on task performance (Calhoun, et al., 2011 & 2012). Specifically, each time one of five criterion task types was completed by the test participant, the corresponding task completion time measure was submitted to the performance-based adaptive algorithm.

Only the time measure was considered in the algorithm. Examination of the participants' comments from these studies suggests that the algorithm employed should also consider other task characteristics (e.g., task type, frequency completed, or priority to the mission). For instance, one participant reported the strategy of quickly completing the health task because it was easier than the image task (Calhoun, et. al, 2012). This strategy enabled the participant to remain in a low LOA, providing the participant a false indication of good performance, at the detriment of the remaining tasks.

Research Objectives/Questions/Hypotheses

This research will develop and evaluate a new algorithm for triggering changes in LOA within a system employing adaptive automation. This algorithm will augment the measure of individual task performance through the application of a priori knowledge regarding the relative priority of the task within the mission. The evaluation will be accomplished by comparing system performance (consisting of both operator performance and the impact of automation aides) between trials when automation is triggered by the new algorithm and system performance when automation is triggered by task performance alone. It is hypothesized that implementing AA triggers based on task priority in addition to performance will improve overall system performance and operator perception of the adaptive algorithm.

Research Focus

This research focused on improving the triggering of adaptive automation by employing a more tailored algorithm, especially as applied to UAV operations where some types of tasks are higher priority than others for mission success. Specifically, the experiment evaluated the utility of a performance-based adaptive algorithm that also considers the priority of the task to the mission. Experimental protocols required the test participants to perform multiple task types, such as image analysis, chat response, task allocation, reroute task planning, and change detection. The LOA of one of these tasks, the image analysis task, changed based on the triggering algorithm in effect. Objective performance measures were recorded on all task types, as well as subjective opinion and personality measures.

Investigative Questions

All of the tasks within the simulated multi-UAV environment were important to the performance of the mission and influenced the workload imposed on the operator. However, the overall goal of the present research was to understand if considering task priority in a performance-based adaptive automation triggering algorithm improves task performance. Task performance can be applied in a weighted (magnitude of importance based on task priority) fashion to determine the appropriate LOA. This research addressed the following questions:

1) Does performance on the image task improve when the LOA adaptation takes into account task priority and frequency, in addition to task performance?
2) If adaptive automation helps image task performance and resources are freed up to help with other tasks, does performance across tasks improve when the LOA adaptation takes into account task priority and frequency, in addition to task performance?

3) What is a recommended method for triggering LOA changes to improve performance? and

4) Do the participants' perceive the LOA adaptation as more appropriate when the triggering algorithm considers task priority and frequency?

Methodology

Human participants completed multiple UAV mission related tasks in trials using the Adaptive Levels of Autonomy (ALOA), multi-UAV simulation. In all experimental trials, the LOA of the image analysis task was determined by the adaptive algorithm in effect for the trial. In some trials, the LOA was triggered by a performance-based algorithm that also considered task priority. On other trials, the image analysis LOA adapted based on an algorithm that only considered task performance, not task priority. Both performance and subjective data were recorded and analyzed.

Assumptions/Limitations

Test participants included a mix of young lieutenants and students from local colleges, not specifically UAV operators. This may limit direct application to the current war fighter due to training and mission differences. Air Force UAV operators have a much greater training basis to understand high fidelity systems. The test bed provided a simulation of pilot workload without requiring the specialized and extensive training necessary for a UAV pilot. This enabled efficient training while simulating the types of tasks that a pilot completes. However, another assumption is that the simulation emulates the tasking and workload of future missions. The degree to which it emulates future missions impacts the generalization of the research findings. In that single-operator, multi-UAV supervisory control stations are not in operation, the fidelity of this simulation of a potential future system is difficult to determine. Further this research assumes that the automation and levels of automation are appropriate within this application and that an improvement in the method for triggering automation changes will result in improvements in system performance.

Implications

An increased understanding of the effects of AA on task performance will help enable the creation of future single operator multi-UAV platforms. Each mission is different and a priority/performance based AA scheme may increase the benefit and flexibility of automation aides.

II. Literature Review

Application of Automation

Concept Discussion

When a system is said to be automated, one can envision images of fantastical spacecraft crossing the galaxy without the need for human intervention. In reality, automated systems include any system with programmed aids. As such, automated systems range from simple calculators which aid a human operator in performing complex calculations to nuclear reactor control systems which monitor and react to the rate of fusion and power demand to generate an appropriate level of power output, to intelligent robotic machines which are able to perform an array of less structured tasks. The differing stages of responsibility given to the system refer to the system's autonomy.

The amount of autonomy a system has is directly related to the level of automation used. "Automation is any sensing, detection, information processing, decision-making, or control action that could be performed by humans but is actually performed by machine" (Moray, Inagaki & Itoh, 2000). The balance of control between system and human is of great interest as increasing levels of automation typically reduces the physical or mental demand to the human operator while simultaneously moving the locus of control from a human operator who may be able to adapt to unexpected circumstances to an automated system which can only respond to the circumstances foreseen during system design.

Role for the Human Operator

The primary focus for system programmers, designers, and engineers is to create a "perfect" system. However, perfectly reliable systems are difficult, if not impossible to create. System programming can only be reliable to the degree a real time situation could be known or anticipated by the programmer (Draper, et al., 2007). Unfortunately, unless complete reliability is certain, a system imposing a high LOA might impose too great a risk to itself or other entities in its environment if its actions could involve survivability, habitability, or overall human safety (Wickens, Mavor, Parasuraman & McGee, 1998). If operators have a greater confidence in their own abilities or an unwillingness to accept system driven actions, then they will never trust or use the automation (Parasuraman & Wickens, 2008; Billings & Woods, 1994). In fact, the automation paradox questions the human's desire for truly autonomous systems (Draper, et al., 2007). If humans cannot accept automation as a credible or reliable source of aid, then automation is "forever constrained to be nothing but an assistant", and any additional efforts to improve automation beyond aiding human activity are futile (Draper, et al., 2007). The human's unique capacity to apply situational parameters to enable more robust decision making will always be necessary to guide the system (Draper, et al., 2007).

Automation Research Approach

In the 1960's the Air Force was faced with the problem of integrating the human pilot and autopilot; resulting designs forced the pilot to seamlessly transition between the two extreme levels of control (Reising, 2002). In these systems, the machine was viewed as a substitute for the human (Calefato, Montanari & Tesauri, 2008). Allocation of tasks was seen as binary, with either the human or the machine in complete control. Task allocation was technology focused, with programmers automating what they could and leaving the rest for the human (Endsley & Kaber, 1999a). Since that time, a more progressive automation strategy, featuring functional allocation has been adopted. In this paradigm, the operator and system are treated as "team members" with each accounting for the other's weaknesses (Reising 2002). Figure 1 illustrates this concept (Fitts, 1951). The left half lists the processes where the human surpasses the machine and the right half displays the processes machines are suited for. Though each team member has strengths, to be a true team the system must be such that the members not only augment each other but account for each other's lapses.

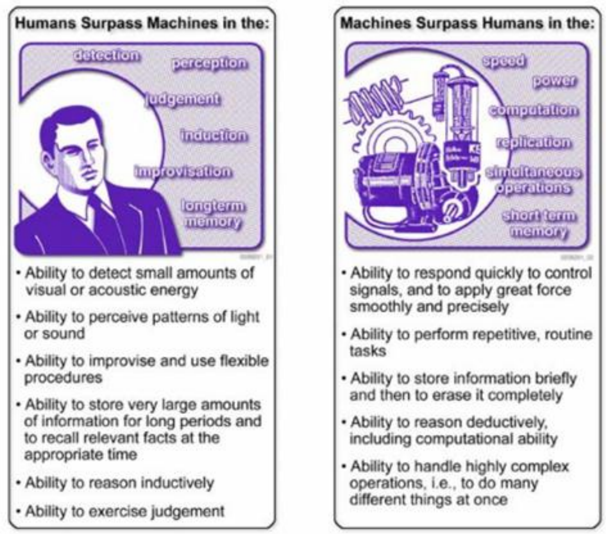


Figure 1: Capabilities of Humans and Machines (Fitts, 1951)

In practice, this alignment of tasks cannot be achieved as the programmer or system designer does not consider the changing needs of the operator (Reising, 2002). Therefore, when this division of tasks is made, automation is limited to serving as an aide to the operator, rather than a true teammate. Under differing sets of criteria, such as emergencies, the roles of the machine and human operator should change and interaction shift accordingly.

Supervisory Control

As automation technology improves, the idea that human activity will be replaced with automation leads to systems in which skill-based tasks are performed by the system and the operator is left only to monitor the actions of the system, assuming control or directing the system only with regard to knowledge-based decisions. These systems then require the operator to perform supervisory control (Moray, Inagaki, & Itoh, 2000). This type of control, referred to as human supervisory control (HSC), permits a shift in human interactions with the system from performing skill based tasks to knowledge based tasks (i.e., decision making). In these systems, the operator is not intended to practice skill-based tasks as these are to be performed by the system. Rather, the operator performs knowledge-based tasks only as required to direct or redirect the system. Supervisory control stems from the belief that "humans should always have ultimate decision-making authority in human-machine systems" (Moray, et al., 2000). Design of systems with HSC affects operator interactions with the automation, interpretation of feedback, and degree of command level (Cummings, Bruni, Mercier, & Mitchell, 2007). Figure 2 depicts Sheridan's HSC loop. This figure displays the mechanisms of control and not the level of operator control or machine automation. However, it demonstrates that the human interacts only with the computer, providing higher-level guidance, and the computer assumes all control of the actuators and sensors which enables the system to accomplish the task.

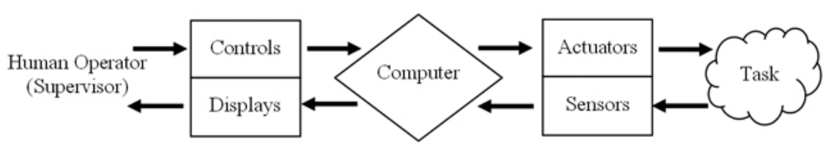


Figure 2: Human Supervisory Control (Sheridan, 1992)

Figure 3 illustrates the more hierarchical nature of UAV control. The inner dashed loop represents the basic guidance and flight control and the outer solid loop encompasses all of the more advanced tasks (Cummings, et al., 2007). The inner loop is the foundation for the complex mechanisms of the outer loop. Any failures with the inner loop trickle down and often produce failures in the more advanced tasks (Cummings, et al., 2007).

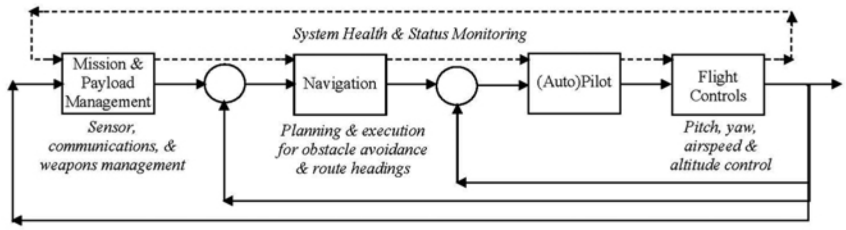


Figure 3: Hierarchical Control Loops for a Single UAV (Cummings, et al., 2007)

With the more advanced HSC envisioned for future UAV operators, the method of control will morph respectively. Figure 4 demonstrates the pull of the operator to be a supervisor of the higher level tasks and the resultant compensation of automation aids in the lower control loops (Cummings, et al., 2007). Such a system configuration permits one operator to potentially control multiple objects or processes, for instance multiple vehicles. However, a downside is that the time that each entity requires operator input is not scheduled and when the operator's responses are time critical, as is often the case for UAV tasks, it is entirely possible that the times the entities require attention can coincide, leading to periods of extreme, potentially unmanageable, workload followed by periods of boredom.

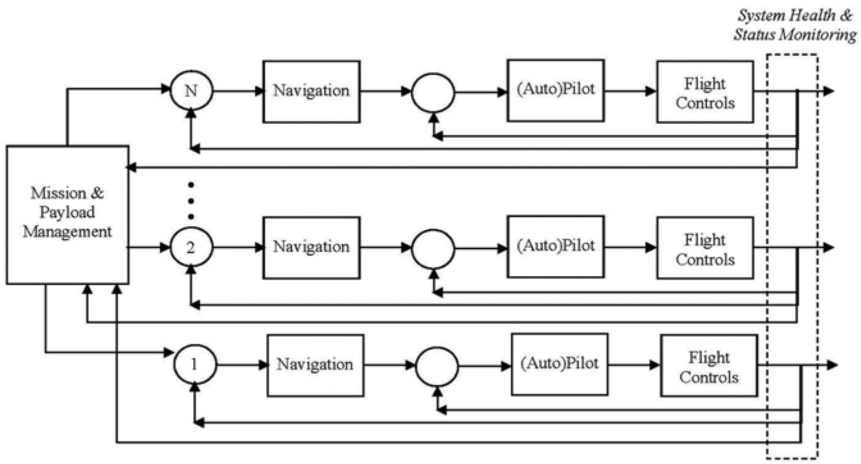


Figure 4: Hierarchical Control for Multiple Unmanned Vehicles (Cummings, et al., 2007)

This future control method will depend on the successful automation control of the inner control loops. Automation will need to reliably control the basic functions of the system, while keeping the supervisory operator aware of system status. This supervisory control concept is known as human-agent (H-A) teaming and is defined from the perspective of operator involvement, LOA, and the interaction between the operator and the control portion of the system (Chen, Barnes, & Harper-Sciarini, 2011). H-A teaming involves five operator tasks: planning, learning, monitoring, intervening, and teaching (Sheridan, 2002). The LOA used for each mission task is dependent on the capabilities of the human and automation (Chen, et al., 2011). This collaborative teaming enables the potential for greater effectiveness.

Advantages of Automation

This teaming concept allows the human and automation to augment each other and increase their efficiency (the whole is greater than the sum of its parts). Automation can aid the operator in a variety of situations and supervisory control environments. For

UAV applications, it can provide improvements in mission capability by freeing operators from the "dirty, dangerous, or dull" jobs, improve affordability through low operational costs, reduce chances of loss of operator life, and decrease workload (Reising, 2002; Draper, et al., 2007). Automation provides a trade space for improvements to safety, reliability, economy, and comfort (Billings, 1997). Taking the operator out of the cockpit improves safety and may reduce complexity, as the operator workstation does not need to be designed into the aircraft. The increase in the automation capabilities and expansion of environments allow for improvements in the reach of the system. "The key to success is to identify and apply the appropriate level of human skill/attention to each mission task and to provide operators powerful and flexible automation tools so they can focus their attention at the mission execution level" (Eggers & Draper, 2006, p. 1). It is only because of the advantages of automation that the concept of single operator control of multiple UAVs can even be considered.

Disadvantages of Automation

"Somewhat paradoxically, machines that can do more, and do it faster, provide the basis for systems that are increasingly demanding of the human operator, particularly in terms of cognitive requirements" (Howell, 1993, p. 235). If machines are exceedingly efficient, then what need is there for an operator? The short answer is that machines are not perfect and neither is the automation to control them. Irrespective of the fallibility of the automation, there are pros and cons to each LOA and they range from reduced situational awareness to complacency to trust issues (Endsley & Kaber, 1999b; Sheridan & Parasuraman, 2006).

Future UAV operators will need to be able to control multiple UAVs in a dynamic and constantly changing environment. Environments could be similar to current airspaces, with little air traffic and no fly zones, or civilian airspace, with commercial and civilian traffic and a large range of flying restrictions. This added complexity will have effects on situational awareness and operator workload (Chen, Barnes, & Harper-Sciarini, 2011). Like automation, situational awareness has different levels: perception of data points and elements in the environment, an understanding of the current status of tasks, and the ability to project current knowledge into the future (Endsley, 2005). Situational awareness can be negatively affected by switching tasks, error detection, and workload. Muthard and Wickens found that operators only detect 30 percent of experimenter induced automation errors (2002). Other research found an error detection rate of only 3 percent (Mumaw, Sarter, & Wickens, 2011). To better understand the impact of error detection, note that the National Transportation Safety Board found nearly 66 percent of aviation accidents caused by human error are due to operators failing to notice the error and revise their plans (Muthard & Wickens, 2002). Fischhoff, Slovic, and Lichtenstein found operators have extreme difficulty looking introspectively to evaluate their accuracy and tend to overestimate their capabilities (1977). This shows that humans are ill-equipped to know when they are in trouble. These problems with loss of situational awareness will only be exacerbated by the introduction of multi-UAV control.

The highly complex environment envisioned for UAVs will surely require multitasking on the part of the operator. Switching tasks during a mission may induce mode awareness issues (Cummings, 2004). Interrupting a primary task, such as supervisory control of Tomahawk missiles, with a secondary task, such as information requests in a chat box, can have a negative impact on one's mode awareness (Cummings, 2004). Mode awareness problems may be trivial, such as a mile to kilometer conversion in open airspace, or catastrophic, such as ignoring a ground warning indication because the airplane is supposed to be in autopilot.

Billings, Lauber, Funkhouser, Lyman, and Huff define complacency as "self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state" (1976). A complacency error is the result of overreliance in faulty automation (Parasuraman & Wickens, 2008). Some of the factors pertaining to an operator's potential for complacency are high levels of trust, reliance, and confidence in automation (Parasuraman, Molloy & Singh., 1993).

The topic of trust in automation is a double edged sword. Miller and Parasuraman state that "operators may not use well-designed, reliable automation if they believe it to be untrustworthy, or they may continue to rely on automation even when it malfunctions if they are overconfident in it" (Miller& Parasuraman, 2007). The end goal is to maintain involvement of operators without overwhelming them, degrading their situational awareness, or depleting their available resources.

Levels of Automation

When implementing any type of automation aid it is vital to determine the appropriate LOA. The LOA selection needs to balance the needs of the operator, overall system performance, and optimize the use of resources (Calefato, Montanari & Tesauri, 2008). This requires an understanding of how the human will need in interact with the automation on terms of safety, level of control required, and novelty of the environment. In one taxonomy, there are ten LOAs ranging from manual operation in level 1 to full automation in level 10 (Sheridan & Verplank, 1978). A detailed explanation of the levels is provided in Table 1.

Table 1: LOA Definitions (Sheridan & Verplank, 1978)

High	10	Full Autonomy: The automation makes all decisions, acts autonomously, and ignores the operator
	9	The automation informs operator after automatic execution, if it "decides" to
	8	The automation informs operator after automatic execution, if asked
	7	The automation informs operator after automatic execution
	6	The automation allows time for the operator to veto an alternative prior to automatic execution
	5	The automation asks on its suggestion with operator approval
	4	The automation recommends one option
	3	The automation narrows the set of alternatives
	2	The automation offers a complete set of alternatives for the operator to act on
Low	1	Manual operation: The automation offers no assistance, the operator must make all decisions

Different tasks may require a different optimal LOA. Higher LOAs might allow for multiple UAVs to be controlled by an individual operator, but they may result in the distancing of the operator from the mission and decreased system performance (Endsley & Kiris, 1994). Ruff, Narayanan, and Draper found “humans in the loop can provide the ability to make well-formed decisions in the absence of complete and correct information” (2002). The concept of keeping the human in the loop helps to mitigate the negative impacts stemming from the expansion of automation to novel and complex environments. The key is balancing the automation approaches to enable the benefits to safety, reliability, and economy while minimizing negative impacts. Miller suggests the use of intermediate LOAs to enable system flexibility while avoiding exclusive task control assignment to the operator of the automation (2007). Moray, Inagaki, and Itoh recommend intermediate levels, 5 through 7, as they contain “genuine collaboration between human and machine”, and generally a level 6 or higher should be used for safety (2000). With the advancements of technology, the tendency to use automation has pushed to an ever increasing capacity. This change will require a collaborative relationship between operator and automation and an intuitive interface to manage optimal LOA and control (Army Science Board, 2004).

When to Automate

AA is the dynamic assignment of control for mission tasks (Calefato, Montanari & Tesauri, 2008). AA involves a situation-dependant aide to a human operator resulting from the actions of the operator (Rouse, 1988; Scerbo, 1996). The counter view to AA is adaptable automation. In adaptable automation, the assignment of control and LOA is initiated by the operator (Scerbo, 1996). One can automate any number of tasks, including the decision of when to trigger a change in automation. Figure 5 depicts this automation decision process. The necessary LOA is dependent on the amount of control and the type of task being automated. The degree of control desired contrasted with the degrees of automation available for a given task will determine the appropriate LOA.

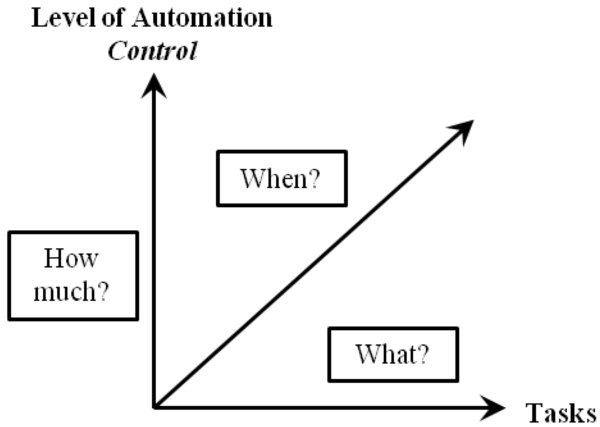


Figure 5: Automation Design Consideration (Endsley, 1996)

Adaptive Automation

De Greef, Arciszewski, and Neerincx define AA as "a mechanism that aids the human operator in real time by managing his or her workload, the latter fluctuating because of varying environmental conditions" (2010, p. 3). AA is known by many titles such as dynamic task allocation, dynamic function allocation, or adaptive aiding; each of these concepts tells the "real-time dynamic reallocation of work in order to optimize performance" (de Greef, Arciszewski & Neerincx, 2010, p. 3). The goal of AA is to determine when interjection of automation is necessary to optimize the task assignment process (Morrison, Cohen, & Gluckman, 1993). AA is the "optimal coupling" between operator workload and LOA (Parasuraman et al, 1992). Due to the varying nature envisioned for UAV missions, the coupling must fluctuate respectively. De Greef states "the automation should be regarded as a virtual partner, similar to a human actor" (2010, p. 3). As such, it should be capable to release or instill task load to maintain performance levels. A popular train of thought is to initiate automation aids to compensate for pilot issues, and return task control when under tasked (Prinzel, 2003). The purpose is not only maintaining operator performance, but "maintaining attentional focus on important tasks" (Chen, Barnes, & Harper-Sciarini, 2011, p. 13). The decision on when to initiate a control shift is determined by "invocation rules", and can be triggered by operator performance, models, physiological state, or some mixture (Parasuraman, Barnes & Cosenzo, 2007). Of these adaptive triggers, performance-based adaptive approaches should be considered for UAV applications since the missions will require dynamically changing cognitive demands. With performance-based adaptation, more automation can be applied during periods of decreased performance, presumably reflecting increased cognitive demands. (Note: other factors, such as operator skill, effort, time pressure, task component, and mission events can also influence workload level.) To apply more automation, either more tasks can be automated and/or higher LOAs are used for one or more tasks. If the cognitive demands are manageable, and performance is not degraded, task(s) LOAs can be kept lower so that the operator is more in-the-loop for task completion and less likely to be impacted by common automation-induced problems.

Review of Adaptive Automation Research

AA research has focused on the determining the process by which to trigger LOA changes (e.g., mission goals, critical events, operator performance, or a hybrid; de Visser, et. al, 2008). Here, the review will focus on AA research using performance-based triggers.

An early study on AA examined the effects of AA on monitoring tasks for the detection of failure with the automation (Parasuraman, Mouloua, & Molloy, 1996). This study compared a non-adaptive group (for which an engine task was automated for the first ten minutes, then allocated to the participant for ten minutes, and finally returned to the automation for the remaining ten minutes of the session) to an adaptive group (same as the non-adaptive group unless performance during the first ten minutes exceeded a threshold). The study found that AA can increase automation failure detection rates (Parasuraman, et al., 1996). This study claims to be the first experimental evaluation of AA and determined some of the key factors pertaining to AA: the "adaptive algorithm, the frequency of adaptive changes, automation reliability and consistency, the type of interface, and contextual factors specific to particular systems" (Parasuraman, et al., 1996). The next study utilized a simulated air traffic controller task to continue the thread aimed at determining what task types to automate. This study demonstrated that operators are better able to utilize AA applied to action tasks than to AA applied to cognitive decisions (Kaber, Wright, Prinzel, & Clapmann, 2005). Another study employed an AA scheme with three conditions: manual, fully automated, and experimenter induced adaptive (based on the experimenter's judgment of an operator's performance on a change detection task) (Cosenzo, Chen, Reinerman-Jones, Barnes & Nicholson, 2010). The results of this study demonstrated the effectiveness of an AA scheme that provides assistance when task load is high and decreasing automation when task load is low (Cosenzo, et al., 2010). This study also illustrated the need for a more time sensitive analysis of performance to trigger LOA changes. In addition to other research, these studies helped to lay the foundation for the effectiveness of AA.

The multi-UAV ALOA simulation test bed employed in the present experiment has been utilized in studies investigating the effects of adaptive automation on task performance. One study compared an adaptive condition, where performance on five task types initiated LOA changes for the image task with a static condition where the LOA in effect for the image task remained constant (Calhoun, Ward, & Ruff, 2011). The results showed that performance based AA improved performance on all task types; the participants also preferred the performance based AA due to a sense of reduced workload coupled with improved performance (Calhoun, et al., 2011). In this first experiment, participants' performance in respect to criteria tended to keep the LOA at a high level enabling problems such as complacency. The next experiment implemented an asymmetrical adaptive scheme where the criteria to decrease LOA was easier to achieve (the criteria to increase LOA went unchanged); the adaptation scheme was again compared to the static condition. The results demonstrated that the asymmetrical adaptive scheme helped to keep participants at a low LOA while still realizing performance benefits (speed and accuracy) for the image task (Calhoun, Ruff, Spriggs, & Murray, 2012). These studies provide support for importance of AA and its effects on task performance and neutralizing effect on automation induced problems.

Problem with Priority not Being Taken into Account

In the ALOA studies to date, the adaptive algorithm scheme has not taken into account the priority of one task verses the other. Given that operators are informed of an ordinal priority for each of the task types, the system should be such that adaptation aides are appropriately matched with the mission priorities. Otherwise, the system results in less optimal strategies. Many participants admitted to ignoring the image (highest priority) task due to the cognitive workload associated with the task and focusing attentional resources on simpler, more frequent, lower priority tasks (often only requiring one click). This strategy typically results in maintaining a low LOA at the further expense of the high priority tasks. For maximal mission effectiveness however, the AA scheme needs to support performance on all tasks, especially those that are high priority. Hence, research is needed to evaluate a performance-based AA scheme that also takes task priority into account.

III. Methodology

This study investigated a new method to trigger changes in task autonomy level for complex supervisory control applications. More specifically, the study was designed to examine a new performance-based adaptive control algorithm that takes into account the priority of tasks, in addition to the operator's performance on tasks. Participants completed multiple tasks while completing trials in a multi-UAV simulation. An adaptive automation scheme was used to drive the LOA of an image analysis task based on real-time performance on five task types. The impact of including task priority in the adaptive algorithm was determined by comparing task performance between trials in which the calculations used a weighting scheme that matched the priorities of the task types with trials in which there was no weighting scheme. Subjective data were also recorded.

Participants

Thirty-two volunteers served as participants (18 males and 14 females, mean age = 26.69 (SD = 6.50). All participants reported having normal hearing, normal color vision, and normal (or corrected) visual acuity to 20/20. Twenty-six were military employees and 6 were members of a paid ($15/hr) experimental participant pool.

Experimental Design

A between subjects design was utilized (Kirk, R.E., 1968). The between-subject variable was the algorithm used for the adaptive-automation control scheme. For one subject group, the autonomy level of an image analysis task was tied directly to an algorithm based on the individual participant's task performance, as well as a task priority weighting scheme. For the second subject group, a performance-based adaptive algorithm was also employed, but with a non-weighted scheme. All participants completed three experimental trials with their assigned performance-based adaptive-automation condition (either with weighted or non-weighted performance scheme).

Adaptive Automation Conditions

For each of the two performance-based algorithms evaluated in this study, a three-step calculation process was conducted. Step 1 involved determining if the participant's performance was better, worse, or within experimenter-specified thresholds. In Step 2, an integer value was derived that either reflected the priority of the task performed (the weighted scheme) or equaled 1 (the non-weighted scheme that does not consider task priority). Step 3 took the output from Steps 1 and 2 in relation to outputs from previous tasks and determined whether the LOA should change across the three LOAs available in the system for the image analysis task. The three LOAs ranged from LOA 1 (low) to LOA 3 (high). The following subsections described each calculation step in detail.

Step 1: Real-time Performance Compared to Experimenter-Specified Thresholds

As each participant performed tasks during the trial, the data were subjected to near real-time analysis. Performance on five criterion task types was considered by the algorithm: a red airplane task (change detection), allocation of image tasks to UAVs, rerouting of UAVs, image analysis, and health analysis (these tasks are described in more detail later in this chapter). For each of these task types, two threshold values were established prior to data collection, to define an "expected time window" in seconds. The thresholds and time windows for each task type (see Table 2) were determined from earlier pilot studies to be sensitive to workload. The mean reaction time plus or minus 1.5 seconds was used to determine the expected time window for each task (Calhoun, et. al, 2012).

Table 2: Task Expected Time Window

Task	Time range (s)
Red Airplane	6-9
Allocation	6-9
Rerouting	33-36
Image Analysis	10-13
Health	8-11

During the trials, each instance that one of the criterion tasks was completed, its recorded completion time was immediately compared to the expected time window. If the task completion time was less than the lower threshold (e.g., < 6 s for allocation; faster than expected) a '-1' was logged; if greater than the higher threshold (e.g., >9 s;

slower than expected), a ‘1’ was logged. If the time was within the defined (e.g., 3 s) range for that task, a ‘0’ was logged. The algorithm’s calculation continued to Step 2.

Step 2: Application of Weighted or Non-weighted Scheme

The priority of each task type to envisioned multi-UAV applications was determined based on pilot input from previous AA studies. This priority was represented as a percentage and ranged from 10% (health response task) to 45% (red plane task). These values are shown in the left-most column of Table 3, and are listed in the order of priority, with the highest priority task in the first row. (Since the allocation and rerouting task were completed in tandem, these tasks are represented in the scheme as a single “Mission Planning” task.) The third column from the left in Table 3 provides the frequency with which each task type occurred in each 15 min trial. These two values, task priority and task frequency, were used to estimate a “task importance factor”. Specifically, calculations involved: a) dividing the task priority by the frequency, b) multiplying the result by two, and c) recording the integer of the result (as the simulation code required an integer for the priority adaptation algorithm). For example, for the red airplane task, the calculation was 45 (priority) divided by 17 (frequency) = 2.647. This result was multiplied by 2, which equals 5.294. The corresponding Task Importance Factor (TIF) is recorded as ‘5’. This example describes the algorithm step for the weighted scheme, with the TIF value reflecting both the priority and frequency of the task. The health task which has a lesser priority has a lower TIF value than that for the red airplane, a higher priority task.

Table 3: Weighting Scheme TIF Adaptive Automation Calculations

Task	Relative Task Priority	Task Frequency per trial	Task Importance Factor (TIF)
Red Airplane	45	17	5
Mission Planning	30	10	6
Image Analysis	15	30	1
Health	10	17	1

The TIF for the non-weighted adaptive algorithm was the same for all tasks and was equal to 1 (see Table 4).

Table 4: Non-weighted Adaptive Automation TIF Calculations

Task	Task Frequency per trial	Task Importance Factor (TIF)
Red Airplane	17	1
Mission Planning	10	1
Image Analysis	30	1
Health	17	1

Step 3: Tally System

The value determined in Step 1 (+1, -1, or 0) and the TIF value computed in Step 2 were then employed in Step 3 for both adaptive automation algorithms. The value from Step 1 was multiplied by the TIF value to achieve a task count (TC). Figures 6 (for the weighted AA algorithm) and 7 (for the non-weighted AA algorithm) illustrate the method employed by the algorithms to tally the task counts and create a cumulative TC, known as system tally (ST) for an example series of operator performance changes.

For the weighted adaptive automation algorithm depicted in Figure 6, the LOA increases moving from left to right. Each LOA can be thought of as a ladder with defined values (steps) ranging from 1 to 7. Each ladder is defined by a pair of limits, 0 and 8, and the LOA increased or decreased (became more or less automated) once the ST reached one of these limits (increased to next higher LOA at 8 and decreased to next lower LOA at 0). Additionally, each LOA has a value in the middle of the ladder (4) known as the reset value. This value is where the initial ST begins and where the ST resets to after any LOA change. In Figure 6, the three columns show the ladders for each of the LOAs. The red letters represent different hypothetical tasks (for the weighted condition) in alphabetical order, with the pre task ST at the tail of the arrow and the resulting ST (pre task ST plus the TC) at the head.

The weighted example begins with a ST of 4 in LOA 1. A participant's performance on task A exceeded the task expected time window (logging a 1) and had a TIF of 3. The TC is the logged value (1) multiplied by the TIF (3), equaling 3. This results in moving the ST to step 7 of LOA 1. This did not result in an automation increase.

Since the ST after task A was just below the upper limit, any further increase in the ST would result in an increase in LOA and a ST reset to 4. This is precisely what happened with the following task. Performance on task B exceeded the task expected time window (logging a 1) and had a TIF of 1. The TC of 1 hit the LOA 1 limit resulting in a LOA increase and a ST reset value of 4. Figure 6 depicts task B hitting the limit of LOA 1 (step 8). The resulting LOA increase and ST reset are represented by the dashed line and gray B*.

Performance of task C was faster than the task expected time window (logging a -1) and had a TIF of 3. The TC equaled -3, the logged value (-1) multiplied by the TIF

(3). This TC moved the ST down to 1 in LOA 2. It is important to note that any additional negative TC at this point would result a LOA decrease and ST reset to 4 in LOA 1. However, task C did not result in a LOA change.

Performance on task D exceeded the task expected time window (logging a 1) and had a TIF of 4. The TC of 4 (logged value multiplied by the TIF) moved the ST to 5 in LOA 2. This task did not increase the LOA.

Performance on task E exceeded the task expected time window (logging a 1) and had a TIF of 6. This resulted in a TC of 6. Because the TC for task E caused the ST to exceed the LOA upper limit (8), the LOA increased and the ST reset to 4, not 7 such that any additional increase beyond the LOA reset is lost. The right part of Figure 6 illustrates task E hitting the LOA ladder limit and forcing a LOA change and ST reset, without adding to the ST.

This final task (F) illustrates the difference between limits on LOAs (white steps) and the end barriers on the outermost LOAs (dark gray steps). A TC causing a ST landing at, or exceeding, the white limits will result in a LOA change and ST reset. A TC causing a ST landing at, or exceeding, the dark gray end barriers cannot result in a LOA change (because this evaluation only utilized three LOAs). In this case, the ST remains at the barrier value until the participant's performance starts to improve.

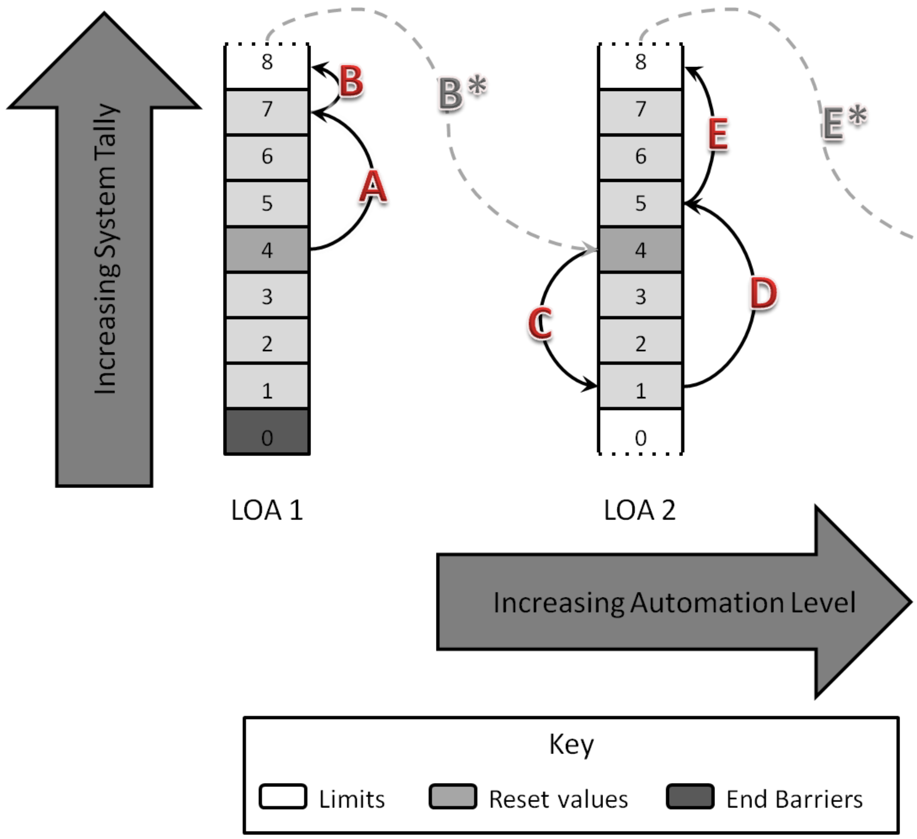


Figure 6: Example of Weighted Adaptive Automation Algorithm System Tally Logic

For the non-weighted adaptive automation algorithm depicted in Figure 7, the LOA increases moving from left to right. The previous ALOA studies also examined task completion time with respect to expected performance. In these studies, a 3 up and 2 down algorithm was employed, where (starting from the reset value) it took poor performance on three tasks to trigger an increase in LOA or good performance on two tasks to decrease in LOA. To match this method to the performance based process used in the non-weighted AA algorithm, each LOA can be thought of as ladders defined using values (steps) ranging from 1 to 4. Each ladder is defined by a pair of limits, 0 and 5, and the LOA increased or decreased (became more or less automated) once the ST reached one of these limits (increased LOA at 5 and decreased LOA at 0). The reset value for the non-weighted adaptive algorithm is 2. This value is where the initial ST begins and where the ST resets to in any LOA change. Like in the weighted scheme, the value determined in Step 1 (+1, -1, or 0) was multiplied by the TIF (1 for all tasks) to achieve the TC. The non-weighted algorithm tallies the task counts (+1, -1, or 0) to create a ST. Figure 7 uses the same symbolism employed in Figure 6 (the LOA limits are in white, the LOA barriers are dark gray, and the reset values are in medium gray).

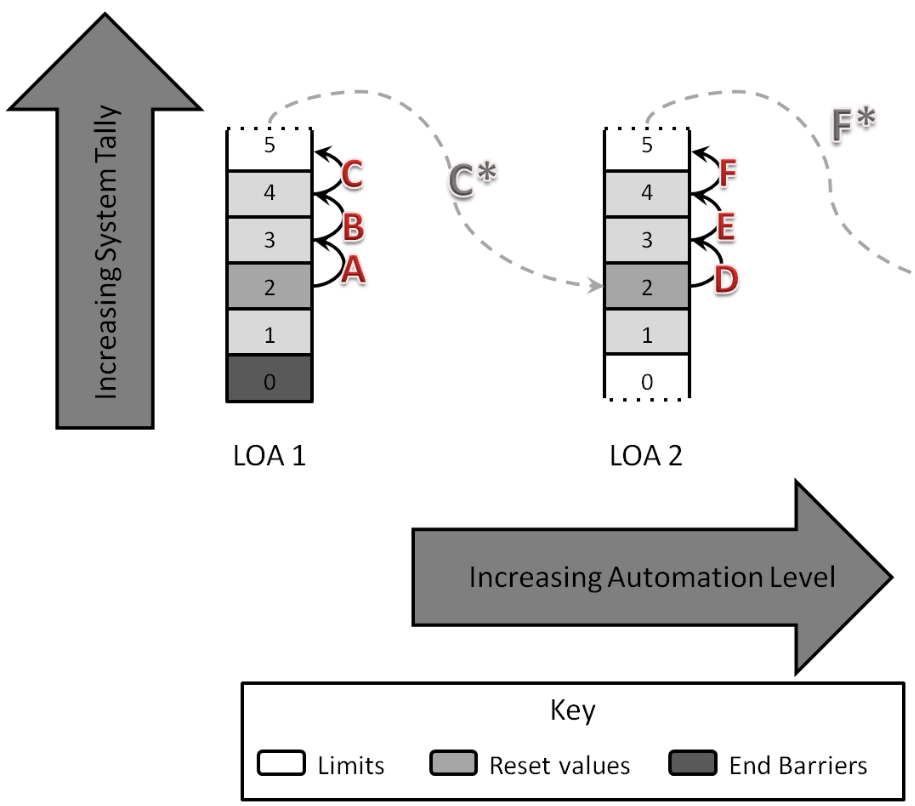


Figure 7: Example of Non-weighted Adaptive Automation Algorithm System Tally Logic

Differences between the Adaptive Algorithms

Table 5 demonstrates the differences between the two performance adaptive control schemes for a hypothetical trial. The rows represent the tasks in order of occurrence. The “Step1” column represents whether a task was completed within the normal time range (average) or outside of it (good or poor), the task count column is the performance score from Step 3, the system tally column is the cumulative performance score for a given LOA, and the LOA columns show the LOA at the end of the task, with highlighted values indicating the point of the trial where the LOA changes.

Table 5: Adaptive Automation Schemes

Automation Comparison							
Task	**Step 1** Task Completion Time Within Expected Window	Weighted Adaptive Automation Scheme			Non-weighted Adaptive Automation Scheme		
		Task Count	System Tally	LOA	Task Count	System Tally	LOA
Trial start	---	---	4	1	---	2	1
Red airplane	good	-5	0	1	-1	1	1
Health	good	-1	0	1	-1	0	1
Image	average	0	0	1	0	0	1
Health	poor	1	1	1	1	1	1
Health	poor	1	2	1	1	2	1
Health	poor	1	3	1	1	3	1
Red airplane	poor	5	4	2	1	4	1
Mission Planning	poor	6	4	3	1	2	2
Mission Planning	poor	6	8	3	1	3	2
Image	good	-1	7	3	-1	2	2
Health	good	-1	6	3	-1	1	2
Image	good	-1	5	3	-1	2	1
Image	good	-1	4	3	-1	1	1
Red airplane	good	-5	4	2	-1	0	1

In the weighted performance scheme the high priority tasks, red airplane and mission planning (allocation and rerouting), drive the LOA change. This is not the case in the non-weighted performance scheme. When the performance-based adaptive automation scheme does not take task priority into account the autonomy level change is more likely to be triggered by non mission essential tasks. Notice how the low priority health tasks in the non-weighted scheme (column 2) cause a decrease in automation level, while the high priority tasks, such as red airplane, have the same influence as the health task. When the algorithm does not consider task priority in its calculations, then changes in LOA are more likely to reflect which tasks the participant is strong in or devotes attention to (e.g., a frequent, low priority task such as health can be done quickly to artificially decrease the ST and resulting LOA).

Apparatus and Materials

A test bed developed by OR Concepts Applied was employed as it facilitates experimental manipulation of task LOA (ORCA; Johnson, Leen, & Goldberg, 2007). This Adaptive Levels of Automation (ALOA, Version 3.0) test bed also incorporates the ORCA commercially available mission planner to provide needed complexity and realism. The simulation's computer was a Dell Precision T7500 Workstation with dual Intel® Xeon® CPU x5550 processors @ 2.67 GHz each, 12.0 GB RAM, and a 1.5 GB PCIe nVidia Quadro FX 4800 graphics card (Microsoft© Windows 7 Ultimate 64-bit Operating System). Two Dell 24 inch widescreen monitors provided numerous windows which were required to support participants' completion of the multiple tasks. A keyboard and mouse were used for participants' inputs.

Experimental Tasks

Figure 8 depicts the entire ALOA test bed. Completion time and accuracy were recorded for most tasks. The following describes each task in turn.

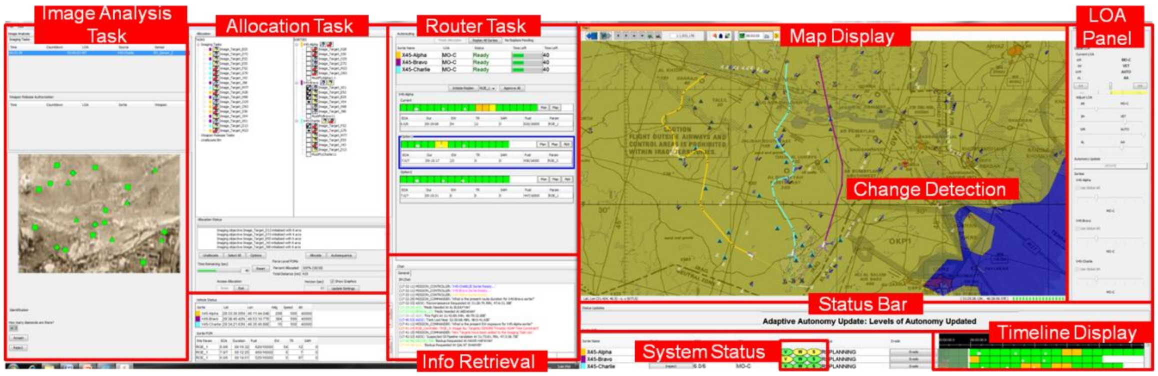


Figure 8: ALOA Control Station

<u>Image Analysis Task</u>

The image analysis task was the only experimental task in which the LOA adapted during the experimental trials based on the participant's performance. (The LOA was static for other experimental tasks.) There were 30 image analysis tasks per trial. Figure 9 shows the timeline used to identify the time an image arrived in the queue. The white plus symbols designated the image tasks, the white bar moved from left to right and represented the current time, and the colored blocks indicated the threat level for a given time interval based on the distance to threats. The threat colors were green (lowest threat), yellow, orange, and red (highest threat).

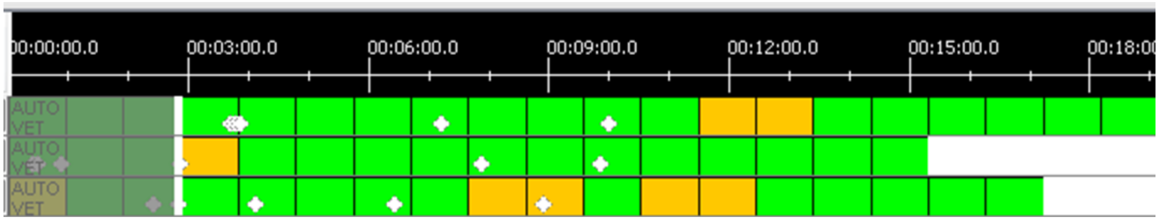


Figure 9: Image Task Timeline Display

Once the white bar passed over a plus symbol, images popped up in a queue in the image analysis panel shown in Figure 10. The images were listed in order of the time the image was taken. There were columns for the time the image was sent to the queue, the countdown time remaining, the LOA for the image, the aircraft that took the image, and the type of sensor used. Once participants clicked on an image row, the image analysis task popped up in the space below the image queue shown in Figure 11. Image task response time was measured from the time the image was sent to the queue until it was completed accurately. If completed inaccurately or not completed at all, the response time was not counted, to avoid creating a ceiling effect. The inaccuracy was reflected in the accuracy (percent correct) measure.

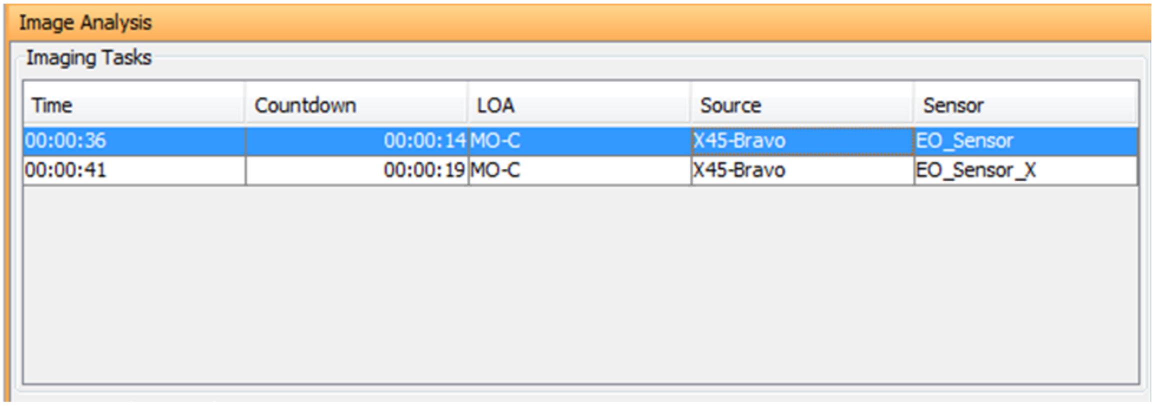


Image Analysis

Imaging Tasks

Time	Countdown	LOA	Source	Sensor
00:00:36	00:00:14	MO-C	X45-Bravo	EO_Sensor
00:00:41	00:00:19	MO-C	X45-Bravo	EO_Sensor_X

Figure 10: Image Queue

Figure 11: Image Analysis Task

The image analysis task required the operator to identify and count the number of green diamonds overlaid on the image. The green diamonds had to be distinguished from the remaining shapes (e.g., circles, triangles, and squares). The participant's next response depended on which of the three LOAs was in effect. In the "low" LOA (Figures 11 and 12), eight options were presented by the automation. To complete the task, participants clicked the bubble next to the correct count and pressed "enter". If no selection was made within 20 seconds, the image disappeared. With the "medium" LOA, the same eight options were presented, but one option was highlighted indicating which one the automation recommended. In the high "LOA", only the recommended option was presented. Participants had only two options: accept or reject the count recommended by the automation.

The image disappeared when participants clicked "Select" (low and medium LOA) and "Accept" or "Reject" (high LOA). In the low and medium levels, if an option was not clicked and selected within the 20 seconds, the task was counted as a miss and the image blanked. In the high LOA, the automation accepted the recommended option at the end of the 20 second window, if the participant didn't make a selection earlier. The 20 second countdown began once the image was taken and sent to the queue.

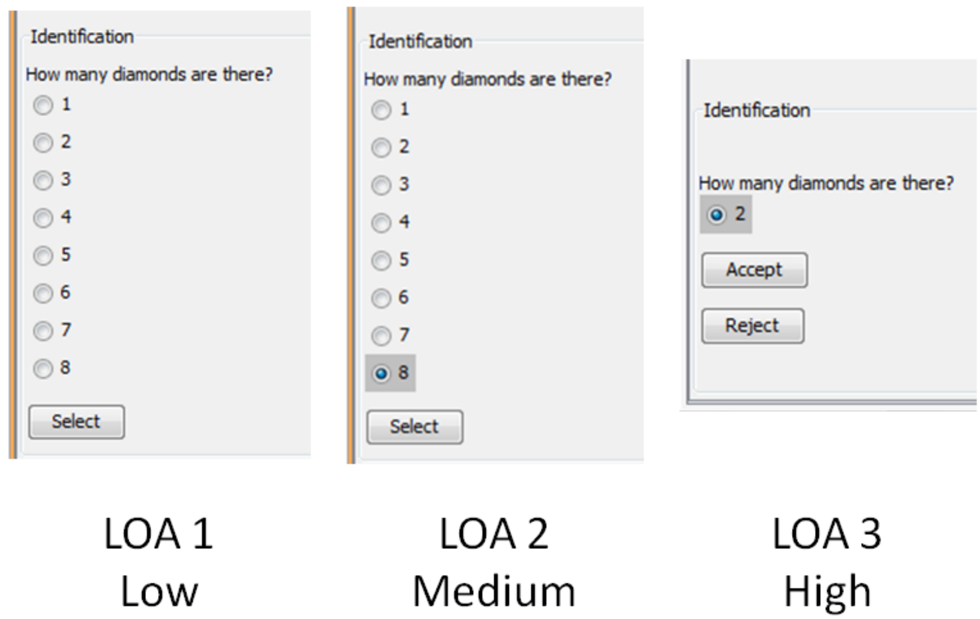


Figure 12: Image Analysis Task LOAs

Information on the LOA for the image task was located in a status bar (located below the map panel) and in the LOA panel (to the right of the map panel). Figure 13 presents an image of the LOA panel and status bar. The status bar displays "Adaptive Autonomy Update: Level of Autonomy Updated" to signal a LOA change (arrow 1). Arrow 2 on Figure 13 points to the location of the LOA for the image task (LOA 1). The remaining tasks maintained a static LOA.

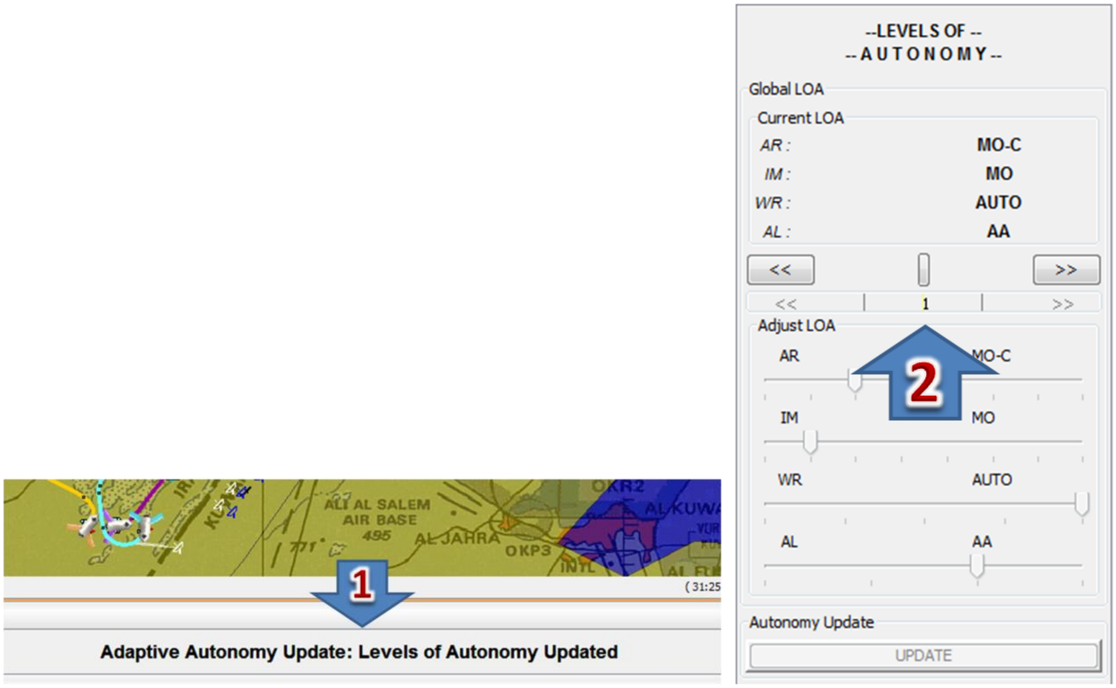


Figure 13: LOA Notifications in the Status Bar and LOA Panel

Allocation Task

Alerts for the assignment of new imaging targets were prompted by an auditory "ding", a system message "Theater Update: New Imaging Task" (Figure 14), and a chat notification from the Mission Commander "New Targets have been added to the Imaging Task List" (Figure 15). A new target assignment necessitated image assignment.

Figure 14: Notification of Image Analysis Task

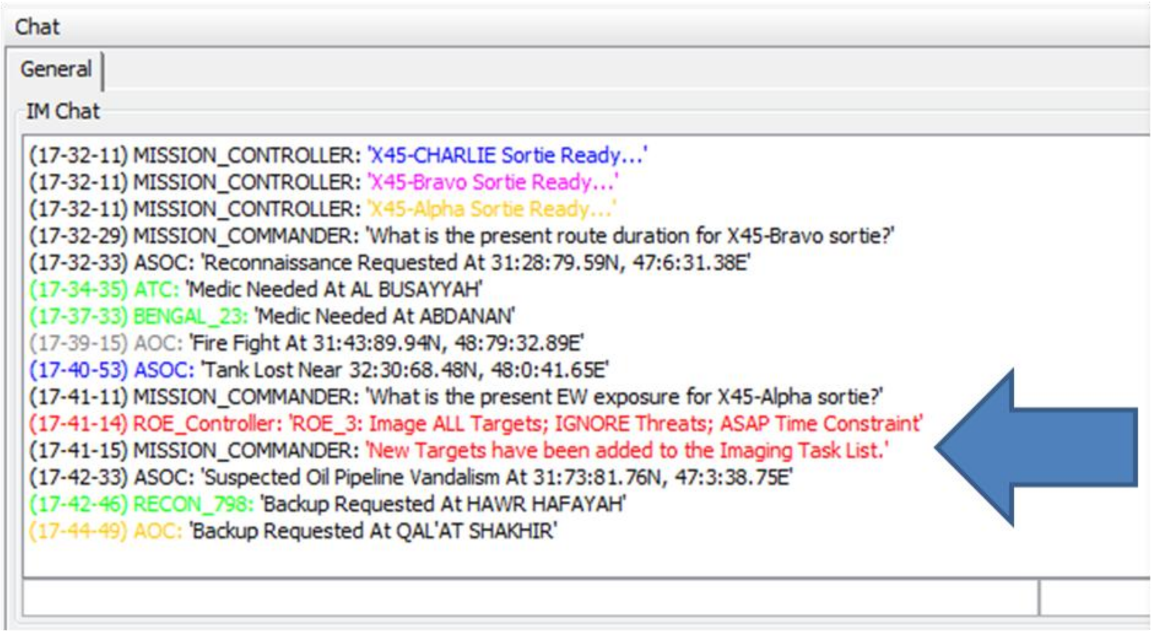


Figure 15: Chat Notification of New Image for Allocation Task

The two "mission planning" tasks (allocation and rerouting) were completed in tandem. Here the first task type in the sequence is described. Figure 16 shows the allocation task panel. The left part of the panel listed the existing imaging tasks. To the left of each task was an oval color coded to match current UAV assignment. Image target requests from the mission commander initiated a new target designation. New targets appeared in the allocation window as white unfilled circles (arrow 1 of Figure 16). They had to be allocated to the nearest aircraft with the needed sensor package, simplified by using color coded sensors. During this task (see Figure 16), the participant assigned the image targets by clicking the "Enter" (arrow 2), "Select All" (arrow 3), and "Allocate" buttons (arrow 4). Once the percent allocated was equal to 100% (arrow 5) the participant clicked the "Finish Allocation" button (arrow 6). If there was an allocation error (the percent allocated does not reach 100%) then the participant had to repeat steps illustrated by arrows 3, 4, and 5 prior to finishing the allocation (arrow 6). This task occurred 5 times during each trial. Allocation response time was measured from the moment the participants clicked “Enter” (arrow 2) until he/she clicked “Finish Allocation” (arrow 6). The allocation count was measured by the frequency of allocation plans completed, number of times the allocate button was pressed (arrow 4).

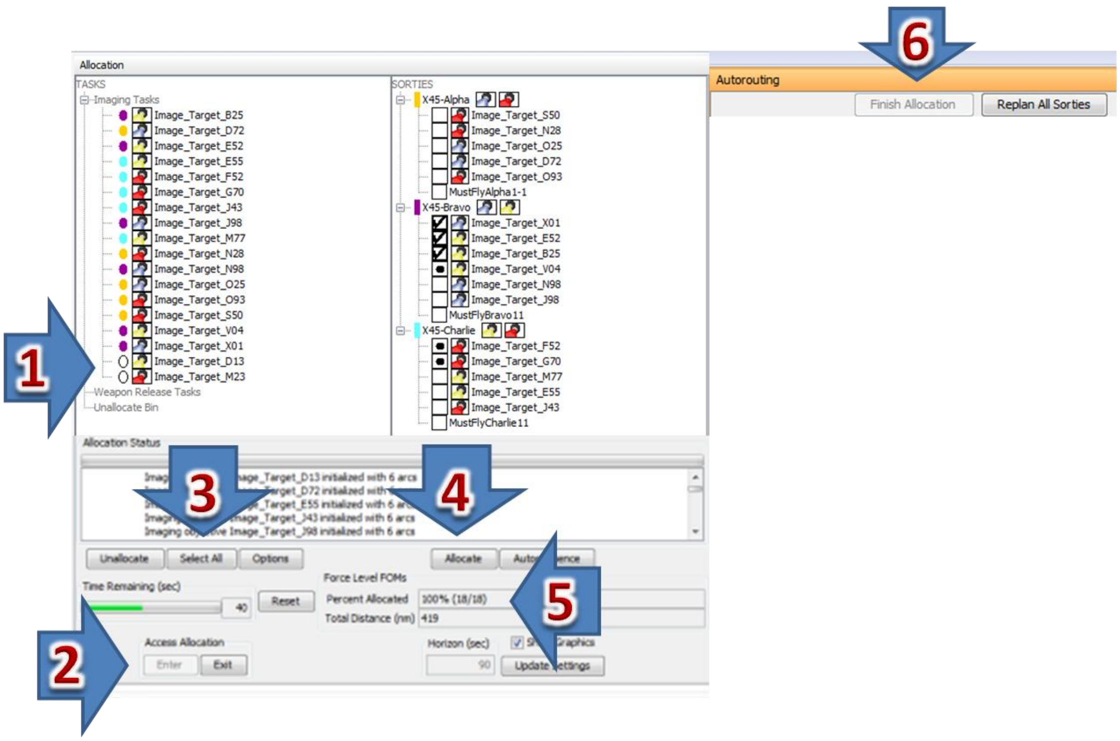


Figure 16: Allocation Panel and Task List

Reroute Task

The current routes for each UAV are displayed in the reroute task panel. Given the assignment of new targets in the allocation window, the UAVs had to be rerouted to match the current imaging task assignment. As such, the reroute task was accomplished a minimum of 5 times per trial. Figure 17 displays the reroute task accomplished for each

UAV individually. The participants had to enter the mission reroute phase by selecting the "Replan All Sorties" button (arrow 1). Once a route plan was ready, the participant clicked on a line with the word "Ready" (arrow 2) in green and three routes appeared (the top is the information for the UAV's current route before the allocation was changed, the second is the automations suggestion matching the current rules of engagement (ROE; e.g., ROE_3: Image ALL Targets; IGNORE Threats; ASAP Time Constraint) from the chat box, and the third is an option matching one of the other two ROEs). The participant approved routes for each of the following UAVs by clicking the "App" button (arrow 3). Once all routes were replanned, they appeared on the map panel. Participants then had to evaluate the new routes for errors (e.g., excessive threat levels or deviations from the general area of the targets). Any errors required the completion of an additional replan cycle (arrows 1, 2, and 3). Reroute frequency was measured as the number of replan cycles completed. Reroute response time was measured from the moment the participant clicked "Replan All Sorties" (arrow 1) until all three routes were approved.

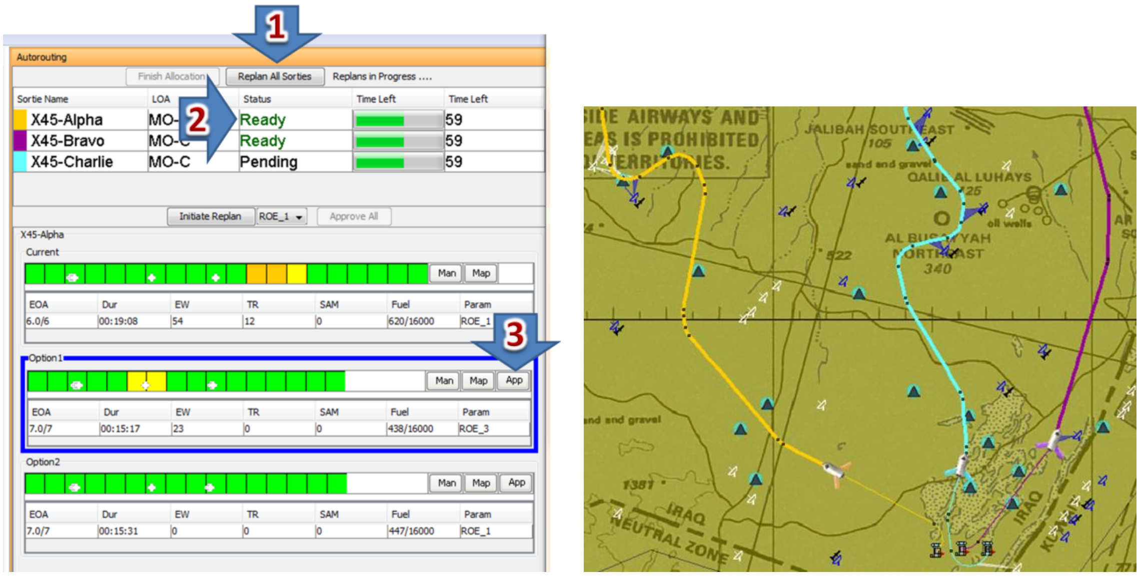


Figure 17: Reroute Task and Map Routes

<u>Red Airplane Task</u>

The system displayed the current routes for each UAV in the map panel. During the “Red Airplane” task, a red airplane symbol appeared on the map display at a random location and had to be noticed and selected within 10 seconds; otherwise it disappeared and was counted as a miss. Red airplane response time was measured from the moment the red airplane appeared until it was selected by the participant; this response time did not include the times for missed red airplanes. Accuracy for the red airplane task was measured as the percentage of red airplane selected within 10 seconds. This red airplane appeared 17 times per trial during the experiment. Figure 18 depicts the map panel and red plane task.

Figure 18: Red Airplane Task

Health Task

Figure 19 depicts the health task and its location in the test bed. To represent system failures, the warning lights changed from green to yellow 17 times per trial. Once warning lights turned yellow, they needed to be selected. Selection was completed with a single left mouse click. Lights not selected within 10 seconds remained yellow and were recorded as a miss. Health response time was measured from the moment a warning light turned yellow until it was selected by the participant.

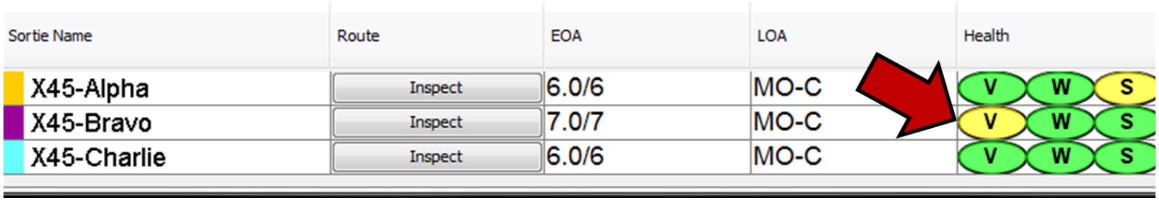


Figure 19: Health Task

Chat Task

The chat task entailed monitoring the chat panel for information requests from the mission commander (Figure 20). Participants had to left click on the chat bar, at the bottom of the panel, and respond to information requests such as "What is the present route duration for X45-Bravo sortie?" This task does not time out and participants were instructed to answer only those questions visible in the window without scrolling. For the present experiment, the AA schemes were not responsive to the chat task and the data was not analyzed.

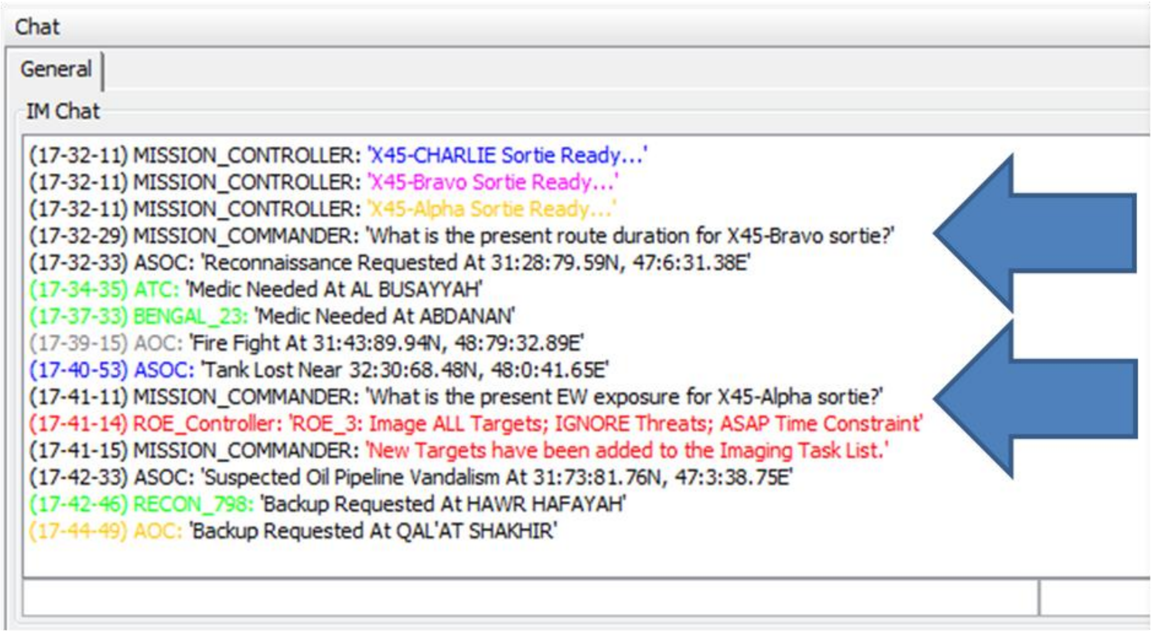


Figure 20: Chat Monitoring Task

Procedure

At the start of the session, the operators were given a written overview of the ALOA station to become familiar with the specifics of the tasks. Following the overview, participants read and signed an informed consent form. Background demographic information was collected. Prior to training, participants completed questionnaires on propensity to trust and personality (Questionnaires shown in Appendix A). Figure 21 provides the list of relative task priority given to all participants. All participants were given the same instructions. Relative task priority remained constant throughout the trials.

Mission Priorities	
Red Airplane	45%
Allocation Rerouting	30%
Image Analysis	15%
Health	10%

Figure 21: Task Priorities

Training was incremental and progressed through each of the six tasks in the following order: red unidentified aircraft, allocation, rerouting, image analysis, health, and chat. Operators had hands on training culminating in one or more practice trials. The practice trials simulated the task load and length of an experimental trial. A minimum accuracy on five task types had to be met prior to the conduct of the experimental trials, to avoid the impacts resulting from a common learning curve. Table 6 depicts the minimum task accuracy for each of the tasks.

Table 6: Training Thresholds

Task	Frequency	Minimum task accuracy
Red Airplane	17	12 of 17 correct
Allocation	5	4 of 5 correct
Rerouting	5	4 of 5 correct
Image Analysis	30	21 of 30 correct
Health	17	12 or 17 correct

The reliably of the automation was 80 percent. In six of the thirty images, the automation suggested an incorrect answer. In one of the five allocations, the automation failed to assign at least one image. In one of the five reroutes, the automation recommended one or more route plans failing to meet the current ROE. During training, participants were instructed on how to identify and correct errors in the image, allocation, and rerouting tasks. Participants were briefed “the automation is good but not perfect” and they were not informed of their performance during the trials.

Once participants completed at least one training trial that met the performance requirements, participants were asked to take a five minute rest break. Then three, fifteen minute experimental trials were completed. After each trial, participants completed an 11-item post-trial questionnaire and workload (NASA- TLX) questionnaire (located in Appendix A; Hart & Staveland, 1988). After the final trial, participants completed an additional post study questionnaire (Appendix A).

Data Analysis

SPSS 19 was employed to implement an Analysis of Variance (ANOVA) between-subjects model to analyze participants’ task performance and other LOA related parameters. Unless otherwise stated, all ANOVAs performed were one way with AA condition as the between subjects variable. Mission performance metrics (task completion time and task accuracy) were analyzed to assess if performance significantly varied between the two AA conditions. The frequency of LOA changes and time spent in each LOA were examined to evaluate the sensitivity of the two different AA algorithms. Subjective post-trial questionnaire data were also compared across AA schemes.

Questionnaire data comparisons were used to determine if perception of automation effectiveness varied due to automation condition. Additionally, questionnaire data on workload, personality, and perceived performance were assessed to determine variability between conditions. Data were pooled across trials unless otherwise stated. A chi-squared analysis was performed on the final questionnaire data.

IV. Results and Discussion

Chapter Overview

ANOVAs were performed on image task and LOA-related measures to gain insight into the effect of each adaptive algorithm. The participants in the weighted AA group were expected to perform better on the image task (which was the only task for which the LOA adapted) and remain in LOA 1 more than the participants in the non-weighted AA scheme. Next ANOVAs were performed focusing on the image task (task for which the LOA adapted) and red airplane task for each of the three LOAs. The red airplane task was chosen as it was the highest priority task in this experiment and the new weighted AA scheme takes task priority into account. Performance for both the image and red airplane tasks was expected to be better when the weighted AA scheme was in effect. To better understand the effect of each AA scheme on overall task performance, ANOVAs were also performed on the performance metrics for the tasks for which the LOA did not adapt. The participants with the weighted AA scheme were expected to have improved performance on all of the tasks for which the LOA did not adapt. ANOVAs were performed on the pre-session (personality, attention control, and desirability of control), NASA-TLX, and post-trial questionnaires to investigate the natural biases of the groups and the effect of AA scheme on workload and perceptions of the system. No differences were expected between the groups for the pre-session questionnaires. The participants with the weighted AA scheme were expected to perceive lower levels of workload on the NASA-TLX. For the post-trial questionnaire, the participants with the weighted AA scheme were expected to give the system better ratings than the participants with the non-weighted AA scheme. A chi-square analysis was performed to understand the distribution differences of the questionnaire data as a function of AA condition. The participants with the weighted AA scheme were expected to provide better ratings for the performance of the system.

Image Task and LOA

To understand the effectiveness of each adaptive scheme when balancing participant's workload through the prudent application of autonomy, it is first important to understand the effect of each scheme on image task performance and LOA status. Table 7 summarizes the ANOVA results of the image task response time and accuracy, the time spent in each LOA, and the frequency of LOA changes. Mean accuracy and response time did not differ significantly between the two AA schemes for the image task (F (1, 31) = 0.15, $p < .70$; F (1, 31) = 0.04, $p < .85$). Contrary to expectations, the mean time spent within each LOA across trials also did not differ significantly between the two AA schemes. However, the time spent in LOA 3 did approach significance with the weighted scheme; this resulted in a lower value for the time in the highest level of automation (F (1, 31) = 3.94, $p < .06$). Further, the mean frequency of LOA changes in each trial significantly differed as a function of AA condition: Trial 1 (F (1, 31) = 5.58, $p = .02$), Trial 2 (F (1, 31) = 14.06, $p < .001$), and Trial 3 (F (1, 31) = 5.85, $p = .02$).

Table 7: Image Task Performance and LOA Measures for the Non-weighted (NW) and Weighted (W) Adaptive Algorithm Schemes

	Mean (Standard Deviation)			*F*	*p*
	N-W	W	Total		
Image Accuracy	67.50 (10.66)	69.20 (13.98)	68.35 (12.26)	0.15	0.70
Image Response Time	11.95 (1.48)	12.04 (1.02)	12.00 (1.25)	0.04	0.85
Time Spent in LOA 1	577.76 (274.25)	686.86 (133.03)	632.31 (203.02)	2.42	0.13
Time Spent in LOA 2	159.34 (91.70)	143.69 (74.33)	151.51 (82.49)	0.28	0.60
Time Spent in LOA 3	163.17 (174.65)	69.71 (70.60)	116.44 (139.38)	3.94	0.06
LOA Change Frequency for Trial 1	4.38 (2.85)	7.50 (4.46)	5.94 (4.01)	5.58	0.02*
LOA Change Frequency for Trial 2	4.38 (3.18)	8.81 (3.51)	6.59 (3.99)	14.06	0.00**
LOA Change Frequency for Trial 3	3.88 (2.85)	7.06 (4.43)	5.47 (4.01)	5.85	0.02*

*$p < .05$, **$p < .01$

Figure 22 illustrates the fact that the LOA changed more frequently, for each of the three trials, with the weighted AA scheme compared to the non-weighted scheme. The weighted AA scheme employed both performance and task priority triggering mechanisms, allowing the system to be more responsive to declining performance on the high priority tasks (and be less responsive to the lower priority tasks). This weighted trigger mechanism was not expected to increase the frequency of LOA changes, but rather increase the reactiveness, or speed, of the change when performance for high priority tasks degraded. The increase in LOA change frequency for the participants with the weighted AA scheme may have been driven by the high priority tasks. For instance, a missed red airplane could cause an immediate LOA change, providing a signal to the participant that overall performance had decreased. This may have prompted the participant to refocus attention resources towards the high priority tasks that, in turn, would cause another LOA change, as a result of improved performance measures. Further research is needed to determine if this change is detrimental.

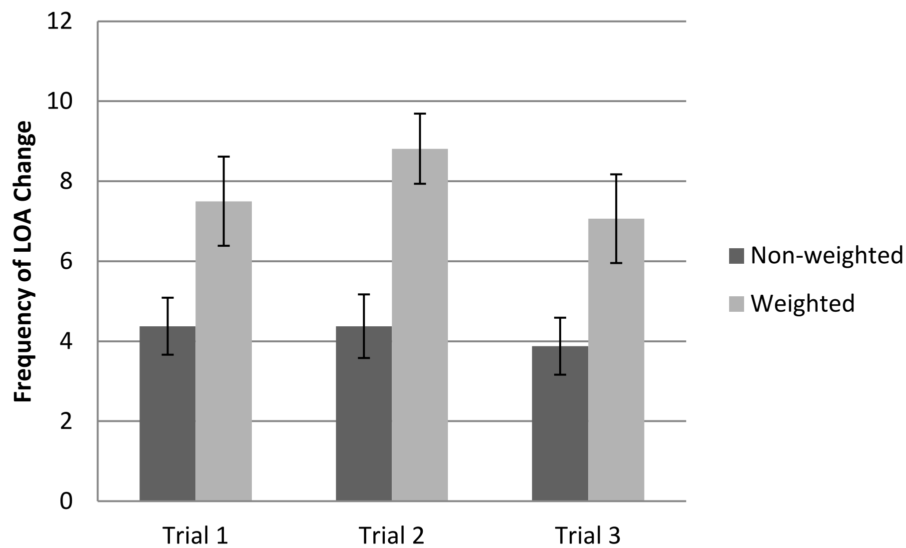


Figure 22: Frequency of LOA Changes by Trial for the Non-weighted and Weighted Adaptive Algorithm Schemes

Though the two participant groups differed significantly in terms of the frequency of LOA changes, the time spent in each LOA did not differ significantly as a function of AA. Figure 23 illustrates the mean time spent in each LOA for both groups. One goal of AA is to keep the operator involved in task completion without negatively affecting their performance. This aims to keep the operator involved in the decision making process as much as possible to avoid errors due to complacency, and other factors. To maintain operator involvement, it is optimal that the AA is such that more time is spent in the lowest LOA, if manageable. An increase in LOA would be needed if operator performance declines. For both AA schemes, the image task was at LOA 1 for the majority of the trial (Figure 23). Though the participants with the weighted AA scheme tended to spend more time with the image task in LOA 1, it was not significantly more than the time participants with the non-weighted AA scheme spent at the lowest autonomy level.

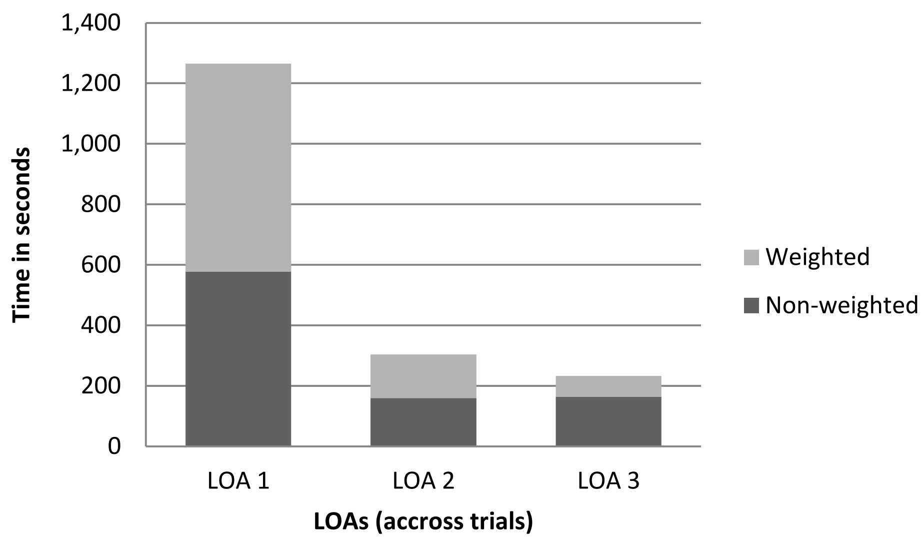


Figure 23: Mean Time Spent in each LOA Across Trials for the Weighted and Non-weighted Adaptive Algorithm Schemes

Figure 24 better illustrates how the percentage of time spent in each LOA with the non-weighted AA scheme increased as LOA increased. Though there was not a significant difference between the non-weighted and weighted AA schemes for time spent in LOA 1 or 2 and the time spent in LOA 3 only approached significance, the trend in Figure 24 suggests the weighted AA scheme tends be more effective at keeping participants in LOA 1 . As performance did not differ between AA schemes as discussed earlier, the weighted AA scheme kept the participants in lower LOAs without negatively affecting overall performance on the image task.

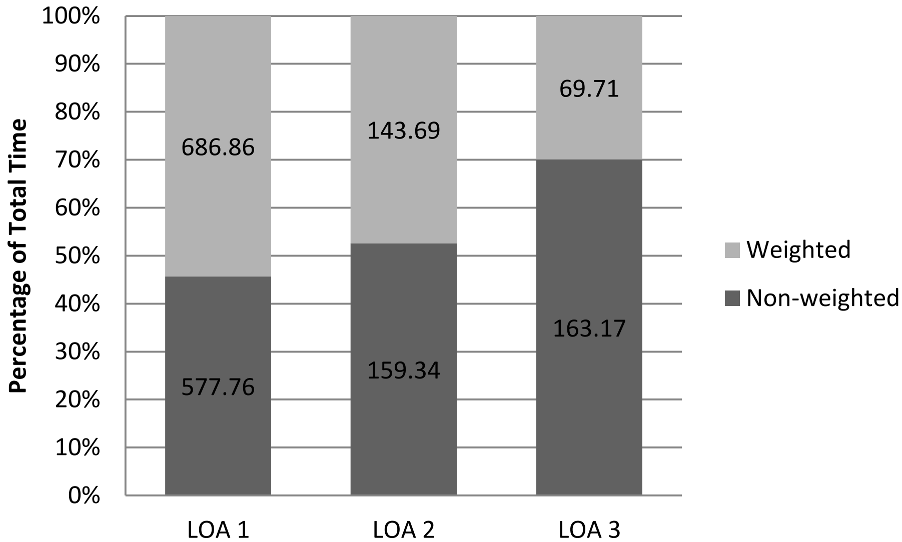


Figure 24: Percentage of Total Time Spent in each LOA for the Weighted and Non-weighted Adaptive Algorithm Schemes

Analysis of Tasks by LOA

To better understand why performance on the image task was similar regardless of AA in effect, it was decided to conduct a finer grain analysis examining performance separately with each of the three LOAs. In past studies utilizing this multi-UAV simulation, the image task was the highest priority and the only task for which the LOA adapted. In this study, the red airplane task was designated the highest priority task, but the image task remained the only task for which the LOA adapted. For this reason, it is important to look at the effects of the two AA schemes on performance of both the image and red airplane tasks. Table 8 summarizes the ANOVA results for accuracy and response time on these two tasks, within each of the LOAs. This was accomplished separately for each LOA and task measure. For example, the first row in Table 8 reports ANOVA results examining accuracy for image tasks across trials that were completed when the LOA was at the lowest autonomy level (LOA 1). As shown in the table, mean accuracy and response time did not differ significantly across the LOAs between the two AA schemes for either the image or red airplane tasks.

Table 8: Analysis of Tasks by LOA for the Non-weighted (NW) and Weighted (W) Adaptive Algorithm Schemes

	Mean (Standard Deviation)			*F*	*p*
	NW	W	Total		
Image Accuracy in LOA 1	56.76 (14.84)	66.93 (17.42)	61.85 (16.73)	3.16	0.09
Image Accuracy in LOA 2	70.13 (19.36)	73.34 (17.00)	71.78 (17.94)	0.24	0.63
Image Accuracy in LOA 3	86.88 (11.71)	85.19 (15.48)	85.93 (13.70)	0.09	0.77
Image Response Time in LOA 1	11.94 (1.47)	12.09 (1.27)	12.01 (1.35)	0.1	0.76
Image Response Time in LOA 2	12.00 (1.77)	11.78 (1.71)	11.88 (1.71)	0.12	0.73
Image Response Time in LOA 3	12.78 (2.59)	12.25 (2.91)	12.49 (2.73)	0.23	0.64
Red Airplane Accuracy in LOA 1	82.54 (9.69)	87.23 (7.15)	84.89 (8.71)	2.43	0.13
Red Airplane Accuracy in LOA 2	88.81 (11.06)	86.28 (11.05)	87.5 (10.95)	0.41	0.53
Red Airplane Accuracy in LOA 3	86.90 (14.67)	92.51 (9.00)	89.70 (12.24)	1.27	0.27
Red Airplane Response Time in LOA 1	3.63 (0.59)	3.79 (0.55)	3.71 (0.57)	0.66	0.42
Red Airplane Response Time in LOA 2	3.80 (1.45)	4.18 (1.18)	4.00 (1.31)	0.65	0.43
Red Airplane Response Time in LOA 3	3.36 (0.44)	4.31 (1.77)	3.83 (1.35)	3.27	0.08

Figure 25 illustrates the accuracy for the image task by LOA for both AA groups, as well as across groups. Though the two AA groups did not significantly differ on accuracy or response time, both performance measures tended to increase as LOA increased. This improvement could reflect the differences in the response steps and autonomy associated with each LOA. In LOA 1, the system provided eight options to choose from and there was no additional automation support. The system recommended an option in LOA 2 and the suggestion was correct 80 percent of the time. In LOA 3, the system was also 80 percent accurate, but only presented one option for the participant to accept or reject. In LOA 1 the combined average accuracy for the groups was only about 62 percent, while the addition of a suggested answer in LOA 2 increased the average to about 72 percent. This increase in accuracy could be attributed to the accuracy of the AA established by the experimenter. It is interesting to note that the participants' performance while using the automation aid remained lower than the performance of the automation aid itself. However, the accuracy of the AA alone cannot explain the accuracy increase in LOA 3, as the combined average is greater than the accuracy of the automation (e.g., 80 percent). The difference in the task created by a binary answer set allowed for a clearer understanding of the automation's recommendation (e.g., if the system recommended a 1, and a participant had already counted 2, he/she could reject the answer without finishing the task).

Though not significantly different, the trend of better image accuracy within LOA 1 for the weighted AA scheme is interesting. As stated earlier, the goal of AA is to keep operators involved in task completion without negatively affecting their performance. This result suggests that the weighted AA scheme is aligned with this goal: participants tended to spend more time in LOA 1 (Figure 24) and perform more accurately on the image task while in LOA 1 (F (1, 31) = 3.16, p < .1).

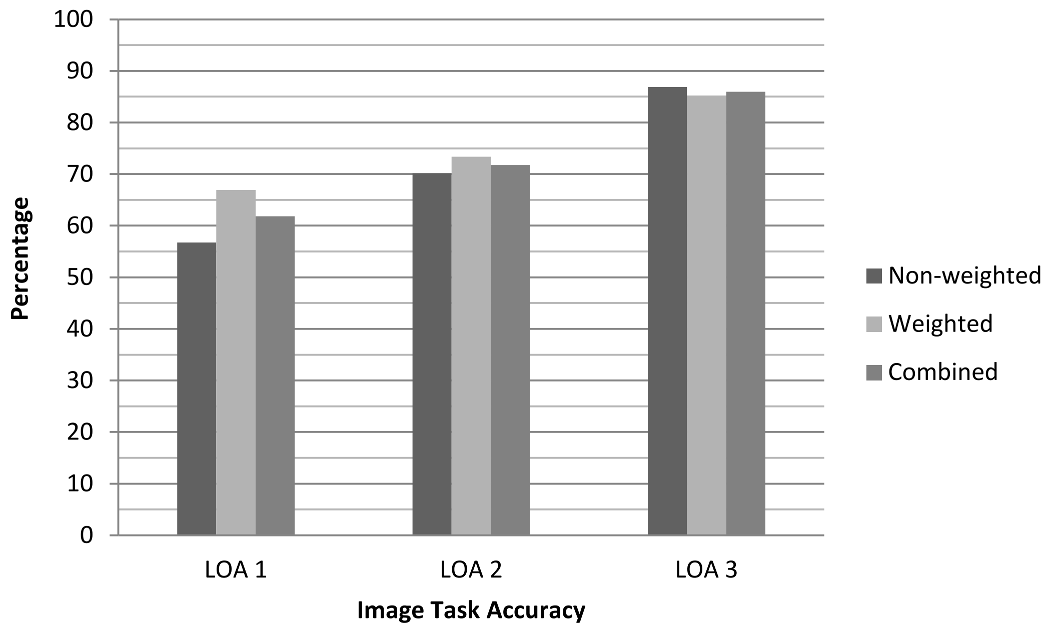


Figure 25: Image Task Accuracy by LOA for the Non-weighted and Weighted AA Schemes

Another implication of the difference for the image task between the three LOAs is the general trend of both groups to take longer to complete the image task as the LOA increased. This could be due to the fundamental differences in the steps to complete an answer selection for the image task. Many participants seem to approach the image the same way, regardless of LOA. In LOA 2, these participants seem startled when the automation recommendation did not match their answer; rather than trust the automation, they often took the time to double check the answer. Any unexpected mismatch of answers is magnified in LOA 3 due to the implementation of polarized answers. In LOA 2 participants seem to be more willing to accept the automation's answer recommendation when it was close to their own answer (e.g., "the automation recommended 8 and I only counted 7, so I must have missed one). In contrast, LOA 3's two answer choice involved a black or white answer. In fact, some participants voiced that it was easier to justify being off by one verses being completely wrong, in their reflections of their strategies with LOA 2 versus 3. This points towards the greater issue surrounding the fundamental difference in tasks due to the severity of the decision (e.g., risk to human life). For example, a weapon targeting decision is much more difficult to make when innocents are within the blast radius. This issue of decision severity may be responsible for real world tradeoffs between accuracy and response time, as the risk of failure overwhelms the importance of the target.

Figure 26 illustrates the response times for the red airplane task by LOA. Response times increased for each AA scheme as the LOA increased from LOA 1 to LOA 2; this does not match the expected result. If either AA is truly aiding the participant and decreasing workload, then one would expect to see a decrease in response time due to the increase in available resources. The response time for the non-weighted AA group decreased for LOA 3 as expected. However, the corresponding performance for the participants with the weighted AA scheme continued to decline (F (1, 23) = 3.27, $p < .1$). One result to note is the inconsistency between the red airplane task accuracy and response times. Generally, good performance on the red airplane task is denoted by high accuracy and low response time. The results from Table 8 do not support this expectation. Irrelevant of the reason, this discontinuity between response time and accuracy draws attention to the need for a clear determination what constitutes an increase in performance. As such, an overall system performance metric may need to be created prior to additional studies. This score would take into account task priority, task frequency, and importance of the different performance metrics.

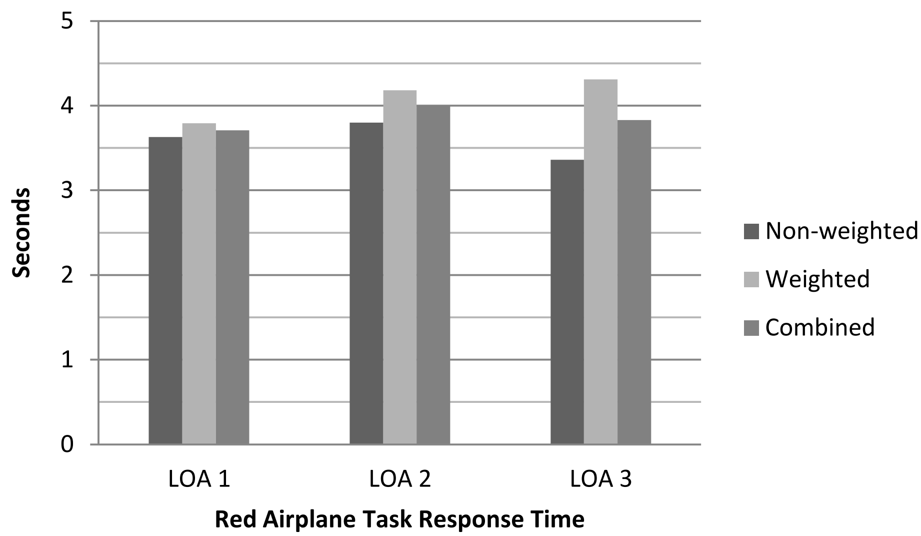


Figure 26: Red Airplane Task Response Time by LOA for the Non-weighted and Weighted AA Schemes

Analysis of Tasks for Which the LOA Did Not Adapt

While performance on the image analysis task did not differ significantly between the two groups, the AA scheme may have, in turn, had an effect on performance of tasks in which the LOA did not adapt during the trials. Table 9 summarizes the ANOVA results of the mean response time and accuracy across trials for the red airplane task, the response time and frequency for the allocation and reroute tasks, and the response time for the health task. As shown in Table 9, there was not a significant difference in any measure between the two AA groups. An exception is the reroute frequency measure, in which participants, on average, made more reroute interactions with the non-weighted AA than the weighted one (F (1, 31) = 5.59, p = .02).

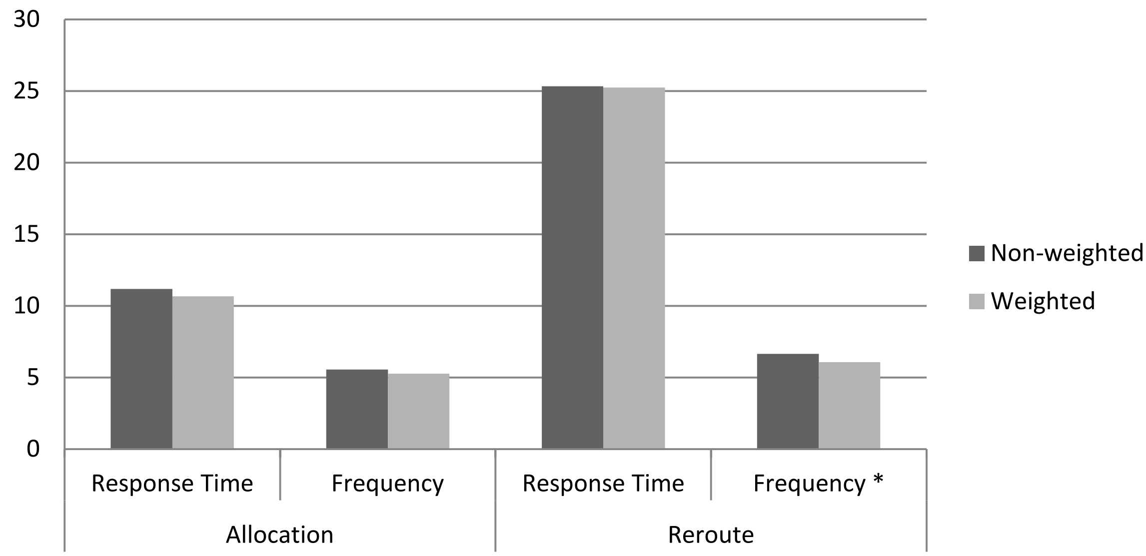
Table 9: Non-Adaptive Tasks for the Non-weighted (NW) and Weighted (W) Adaptive Algorithm Schemes

	Mean				
	NW	W	Total	*F*	*p*
Red Airplane Response Time	3.73 (0.43)	3.87 (0.39)	3.80 (0.41)	0.89	0.35
Red Airplane Accuracy	85.66 (6.82)	87.50 (7.16)	86.58 (6.94)	0.55	0.46
Allocation Response Time	11.18 (2.74)	10.67 (1.72)	10.93 (2.26)	0.4	0.53
Allocation Frequency	5.54 (0.78)	5.25 (0.35)	5.40 (0.61)	1.86	0.18
Reroute Response Time	25.32 (3.67)	25.24 (3.09)	25.28 (3.34)	0.01	0.94
Reroute Frequency	6.65 (0.75)	6.06 (0.64)	6.35 (0.75)	5.59	0.02*
Health Response Time	11.24 (3.34)	12.63 (3.57)	11.93 (3.47)	1.29	0.26

*$p < .05$

Figure 27 illustrates the mean response time and frequency of the allocation and reroute tasks for both AA schemes. The fact that reroute frequency is significantly higher for the weighted AA scheme is interesting because the frequency of allocation and reroutes are tied to the participants' trust in the automation. Lower replan frequency implies more trust in the automation. Trust is important because it is an essential component of human-automation teaming. The significant difference between the two AA schemes for the reroute frequency suggests the weighted AA scheme led to increased trust in the reroute automation. This result supports the hypothesis that LOA adaptations for one task can impact performance on a task for which the LOA did not adapt. This may reflect a freeing up of attention resources. Though the weighted AA scheme did not have an effect on the task for which the LOA adapted, these results suggest that it can improve performance on a different task. For real-world applications the transference of automation effects should be assessed when determining the effectiveness of the system.

Trust issues from one faulty subsystem may lead to an overall mistrust in the automated system and under reliance on automation.



*$p < .05$

Figure 27: Mean Response Time (mean seconds) and Frequency (mean number) of the Allocation and Reroute Tasks

Analysis of Pre-session Questionnaires

It is important to determine if group differences initially biased performance. Table 10 summarizes the ANOVA results of the pre-session questionnaires (personality, attention control, and desirability of control). The scores did not differ significantly between the two AA schemes for any of these instruments. This means the groups were considered homogeneous and there was no significant effect of personality or control factors.

Table 10: Pre-session Questionnaires for the Non-weighted (NW) and Weighted (W) Adaptive Algorithm Schemes

	Mean (Standard Deviation)				
	NW	W	Total	*F*	*p*
Personality Extraversion Score	5.50 (1.05)	6.28 (1.46)	5.89 (1.31)	3.01	0.09
Personality Agreeableness Score	7.12 (1.27)	7.20 (1.27)	7.16 (1.15)	0.04	0.85
Personality Conscientiousness Score	6.62 (1.30)	6.82 (1.23)	6.72 (1.25)	0.21	0.65
Personality Emotional Score	6.8 (1.12)	6.05 (1.39)	6.43 (1.30)	2.81	0.10
Personality Openness Score	6.77 (0.95)	6.52 (1.04)	6.64 (0.99)	0.50	0.48
Attention Control Score	57.31 (5.30)	56.06 (7.32)	56.69 (6.32)	0.31	0.58
Desirability of Control Score	100.63 (9.67)	98.63 (10.76)	99.63 (10.11)	0.31	0.58

Post-Trial Questionnaires

While one focus of AA is to ultimately improve performance, it does not relay the whole picture. It is arguably most imperative to assess if the AA schemes had an effect on the participants' perception of the system. Table 11 summarizes the ANOVA of the averaged results for the NASA-TLX and post-trial questionnaires. Note the results for the questions on task difficulty, workload, and surprise due to the actions of the AA were reverse coded such that higher results on Table 11 equate to better scores for all scales. NASA-TLX scores (based on a scale of 0 to 100), were averaged across the five measured subscales (effort, frustration, mental demand, temporal demand, and physical demand) and submitted to an ANOVA. The results showed that average workload value was less when the weighted AA condition was in effect (51.65) compared to when the non-weighted AA condition was used (53.10), but this difference was not statistically significant (F (1, 31) = .08, p < .77). The other post-trial scores consisted of a series of Likert-type ratings scales, and did not differ significantly between the two AA schemes for the questions on task difficulty, workload, and participant's perceived ability to complete the image task. Responses also did not significantly differ for questions addressing the AA in term of its ability to support the image task, trust in AA, detection of LOA changes, notification of LOA changes, conscious attention paid to LOA change, and the impact of LOA on the image task and non-adaptive tasks. However, the question "rate how often you were surprised by the actions of the automation" (shaded in Table 11) significantly differed as a function of AA condition: F (1, 31) = 6.43, p = .02.

Table 11: Workload and Post Trial Questionnaires for the Non-weighted (NW) and Weighted (W) Adaptive Algorithm Schemes

	Mean (Standard Deviation)				
	NW	W	Total	*F*	*p*
NASA-TLX Score	53.10 (9.69)	51.65 (17.41)	52.38 (13.88)	0.08	0.77
Task difficulty	2.25 (0.74)	2.33 (0.78)	3.71 (0.75)	0.10	0.76
Workload	2.42 (0.78)	2.63 (0.73)	3.48 (0.75)	0.61	0.44
Your ability to do the image task	2.75 (0.67)	3.13 (0.48)	2.94 (0.61)	3.27	0.08
Automation's ability to do the image task	2.58 (0.76)	3.06 (0.64)	2.82 (0.73)	3.77	0.06
Trust in automation	2.73 (0.83)	3.17 (0.50)	2.95 (0.71)	3.27	0.08
Less surprised by automation	3.33 (0.44)	3.83 (0.66)	2.42 (0.60)	6.43	0.02*
Notice LOA change	4.00 (1.42)	3.92 (1.31)	3.96 (1.34)	0.03	0.86
System LOA notification	3.26 (0.68)	3.07 (0.59)	3.16 (0.63)	0.69	0.41
LOA impact on image task	3.51 (0.66)	3.5 (0.63)	3.51 (0.63)	0.00	0.96
LOA impact on other tasks	3.52 (0.59)	3.59 (0.52)	3.56 (0.55)	0.10	0.76
Attention paid to LOA change	2.21 (0.78)	2.26 (1.01)	2.24 (0.89)	0.02	0.90

*p < .05

These same data are depicted in Figure 28, illustrating the trends of the post-trial response scores for the non-weighted and weighted AA schemes (data were recoded so that across questions, higher bars denoted a more favorable result). Most notably the group with the weighted AA scheme was significantly less surprised by the automation. This is important because it matches the goal of keeping operators involved in task completion without negatively affecting their performance. Being frequently surprised would indicate the automation's failure to maintain operator involvement. As such, the weighted AA scheme suggests better participant involvement. Other interesting trends supporting the above goal are the weighted group's better ratings for task difficulty/workload. The mean ratings ranged from difficult to neither difficult nor easy and busy to very busy. The participants with the weighted AA scheme had more trust in their abilities to complete the image task; the participants in the weighted group rated their confidence in their abilities as moderate to high, while the non-weighted group rated their abilities as low to moderate. Although not statistically significant, the mean values for the weighted group trended towards increased trust in the automation, the AA's ability to complete the image task and trust in the AA. The mean ratings ranged from moderate to high trust for the participants with the weighted AA scheme, compared to the low to moderate ratings from the participants with the non-weighted AA scheme. The participants with the weighted AA scheme seemed to perceive themselves as more aware and better able to perform all tasks. Though the system may not have been sensitive enough to detect a true difference between the two groups, it does point to the importance of participant perceptions. Perceptions of the utility and reliability of a system may make the automation more acceptable to operators.

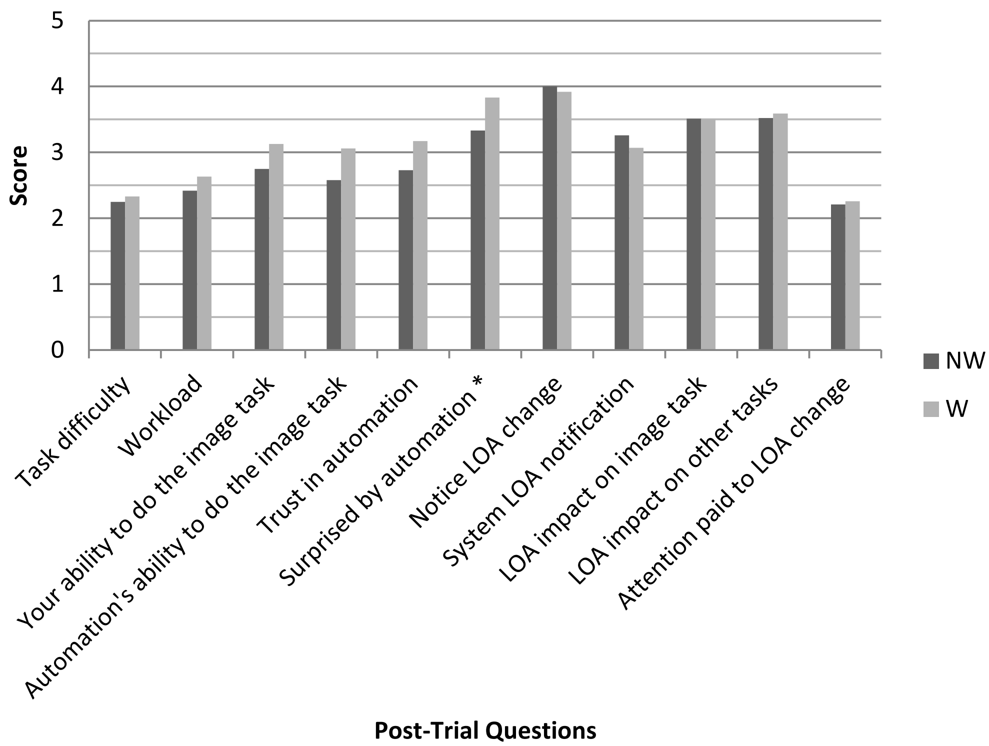


Figure 28: Post Trial Response Score for the Non-weighted (NW) and Weighted (W) Adaptive Algorithm Schemes

Final Questionnaire

A chi-square analysis was performed on the final questionnaire data to understand the difference in participant perceptions between the two AA conditions. Table 12 summarizes the chi-square results. (Note the response possibilities ranged from 1 to 5 except the last two questions were yes (1) or no (0)). Responses did not differ significantly between the two AA schemes for the questions on LOA frequency, LOA frequency adequacy, ability to complete image task, ability to complete non-adaptive tasks, situational awareness, mental workload, LOA preference, need for more responsive LOA change, and alignment of LOA change to actual performance.

Table 12: Chi-square Analysis for the Final Questionnaire for the Non-weighted and Weighted Adaptive Algorithm Schemes

Question	χ^2	df	*p*
LOA frequency	2.54	4	0.64
LOA frequency adequacy	2.33	3	0.51
Ability to complete image task	1.98	2	0.37
Ability to complete non-adaptive tasks	3.39	3	0.34
Situational awareness	5.05	3	0.17
Mental workload	0.42	3	0.94
LOA preference	2.28	4	0.68
Faster LOA change	0.13	1	0.72
Automation matched performance	1.13	1	0.29

Though the two AA schemes did not significantly differ on any response, the trends are as expected. The weighted AA group reported higher abilities to do all tasks, higher situational awareness, and a lower mental workload. Additionally, those in the weighted group felt the AA better matched their actual performance abilities. Although task accuracy and response time was not significantly better with the weighted AA as expected, this control scheme that took both performance and task priority into account tended to improve participants' perception in several important areas.

V. Conclusions and Recommendations

Answers to Investigative Questions

The overall goal of this study was to understand if including task priority in a performance-based adaptive automation algorithm improves task and system performance. With this new algorithm approach, task performance was used in a weighted fashion (based on the task's priority) to determine the appropriate LOA. Originally the design of the weighted AA scheme was to only account for task priority, but a frequency component was added to the algorithm to avoid diminishing the effects of a high priority/low frequency task with a low priority/ high frequency task (similar to the previous problems with weighting all task equally). The order of task priorities were changed from previous studies, such that the task for which the LOA changed was not the highest priority. This was altered to understand the effect of the AA schemes on improving performance on higher priority tasks. It is not enough to understand if the use of AA increased performance on the adaptive task, it is more interesting to investigate if the improvements due to the implementation of AA are transferable to other tasks.

Question 1

One focus of this research was to understand if performance on the image task improved for the participants with the weighted AA scheme. Though the groups did not significantly differ in terms of image task accuracy or response time, the mean values for the weighted group trended towards increased image task accuracy. These results lend support to the weighted AA scheme being an enhancement over the past non-weighted approach. In this experiment the both AA schemes were effective in supporting a balanced relationship between the operator and automation.

Question 2

The next expectation was that if AA was applied to improve image task performance and attention resources were freed up to help with other tasks, performance on other tasks should improve when the LOA adaptation takes into account task priority. Unfortunately, a statistically significant difference in performance between the two AA schemes was not present in the red airplane task performance data. The intent of the weighted AA scheme was to automate the image task to provide the operator more resources to perform the higher priority (e.g., red airplane) task. Given the lack of significant differences between the AA schemes for the image and red airplane tasks, it can be suggested that this system behavior did not occur as expected. Had the system freed up resources for the tasks in which the LOA did not adapt, then the results of the weighted scheme might had been significantly different.

Generally, the participants with the weighted AA scheme tended to perform the tasks more accurately than those with the non-weighted AA scheme and so it is possible that the sample size was not large enough to provide a statistically reliable trend. It was noted that AA switching occurred much more frequently with the weighted compared to the non-weighted AA scheme. Although one could argue that the trend of increased AA change frequencies by the weighted algorithm indicated that the algorithm was more sensitive to performance changes, one could also argue that the weighted AA scheme was perhaps oversensitive to relaxing the automation level. More research is required to understand when to trigger or relax automaton levels and identify the ideal algorithm that improves task accuracy, as well as reaction time. Given the significant difference in the number of LOA changes, the recommendation is to increase the size of the LOA "ladders" for the weighted AA scheme. For instance, the ladder size can be increased from eight to twelve and the reset value can be increased to an eight. This combined with a similar task load should decrease the "reactiveness" of the LOA trigger. However, since the frequency of the red airplane (highest priority) task was increased from the previous studies (4 to 17), the red airplane task affected the LOA trigger algorithm more than anticipated. To be more applicable to previous ALOA research, the frequency could be decreased to better effect the frequency of the LOA changes for the weighted AA scheme. The change in task frequency will not be applicable to a real-world system. For this reason, the primary focus needs to be on the alterations to the algorithm.

Question 3

The next question pertained to determining a recommended method for triggering LOA changes to improve performance. When it comes to AA, the weighted AA method implemented in this study seems to be better than the non-weighted AA scheme (based on the perceptions of the participants from the post-trial and final-questionnaires). However, many of the participants from both conditions specifically voiced concerns about the reactive nature of the system. These participants wanted the system to be more predictive of future performance (e.g., “I would rather the system see that I have five image tasks coming up and change LOA before I get the chance to perform poorly”). For this reason, a better recommended method for triggering LOA changes may incorporate workload based AA. It is worth noting that simply responding to the number of images in the queue for the image task may have produced a more responsive algorithm, without creating additional complexities.

A purely workload based AA scheme provides aide to the operators before they know they need it, however, this does not keep with the intent of keeping the operator involved. This scheme is more proactive to avoid overloading participants, but may be over reactive, leading to higher LOAs. This could induce issues related to mode awareness, complacency, and loss of situational awareness. A purely performance based AA scheme keeps the operator involved but may not be reactive enough to prevent the errors before task overload occurs (high image task load leading five missed tasks in a row). A hybrid approach that weighs the current task load with the performance limits of the participant may solve this dilemma.

Question 4

The final expectation was that participants would perceive the LOA adaptation taking into account task priority as more appropriate. The participants with the weighted AA scheme were less surprised by the actions of the AA (post-trial questionnaire data). This finding was statistically reliable. For the final questionnaire, the participants with the weighted AA scheme tended to rate the LOA changes as more aligned with their actual performance than those in the non-weighted AA group. Both results support the hypothesis that participants would find the weighted AA scheme more attuned to their performance.

Significance of Research

The literature review suggests that this is the first attempt at implementing a performance based AA scheme that also takes task priority into account when automating tasks within a multi-tasking environment. The present results lend support to its potential to provide improved system effectiveness. Though most of the performance metrics did not significantly differ as a function of whether priority was a factor in implementing the AA, data trends indicate this approach merits further consideration. One of the most interesting findings involves the participants' perception of the effectiveness and reliability in the system. Research evaluating candidate automation control schemes need to take into account the operator's perception of the automation, in addition to actual performance on tasks. For an effective scheme involving multiple highly autonomous systems, operators will need to understand and trust the automation in order to realize its benefits.

Recommendations for Future Research

Future studies should look at the effect of a workload/performance hybrid AA scheme on task performance. Participants are clearly open to a workload based system, so one would expect the perception of the appropriateness of the AA to improve. A system that proactively adapts to participant task load should improve task performance. Since people are not always good predictors of the moment when they will become overwhelmed, a workload based system should improve performance by providing support when it is needed (not after performance has started to decline or regardless of past performance). A weighted value could be applied to each task and an operator dependent threshold, a maximum and minimum ST, could be applied to the mission. A LOA change could be triggered once the value of the tasks exceeds the threshold, a ST less than the minimum threshold would trigger a decrease in LOA and a ST greater than the maximum threshold would trigger an increase in LOA. To explore this further, a study must first be conducted to understand the range of operator specific thresholds effective with the given task weights. Once acceptable range limits are determined, then training can be utilized to determine an individual’s baseline thresholds. This could be applied to future conditions where the operator dependent threshold can vary as a function of fatigue or experience.

The addition of a reversed AA scheme (where triggers do not align with task priority) could help to further understand the effects of priority based AA schemes. This would clarify if the performance increase is due to an effective AA scheme or merely because the LOA is changing.

Further understanding of the effect of operator perceptions on task performance is essential to determine the value of AA. Future studies should better assess the participants' perceptions on varying aspects of system interaction (e.g., the reactiveness of the AA, appropriateness of task load, accuracy of the system, and situational awareness). For an automation system to be effective, it must be both accurate and perceived as useful by the operators.

The results of this study suggest the need to create an overall performance score or ranking of metrics for overall performance (e.g., the tradeoff problems with the accuracy and response times for the red airplane tasks). Like in video games, the system needs a definitive set of guides for determining true goodness of an AA scheme. This score should be priority dependant while providing an overall score of performance. This would allow a more objective comparison of performance between AA conditions. It could also encourage greater involvement of the participants by offering incentives for maximizing overall performance.

One way to understand the effect of AA is through the analysis of each participant's attention resource allocation. An insight into the effect of AA conditions on the assignment of priorities and location of focus could be gained by tracking each participant's eye gaze and fixations. This will provide a better understanding of whether participants actually followed the assigned task priorities and insight into each participant's strategy.

Summary

Though performance measures, accuracy and response time, did not significantly differ with respect to AA scheme, the weighted AA method employed in this study seemed to be an improvement over the non-weighted AA scheme. The results of this study, combined with participant preference for workload based adaptations, suggest a benefit to the implementation of a workload/weighted performance hybrid approach. Future research should focus on task weights based on priority and operator specific threshold criteria, such that automation aides are triggered once the summation of current tasks exceeds a specified threshold.

Appendix A

Personality Questionnaire

Attention Control

Desirability of Control Questionnaire

NASA-TLX

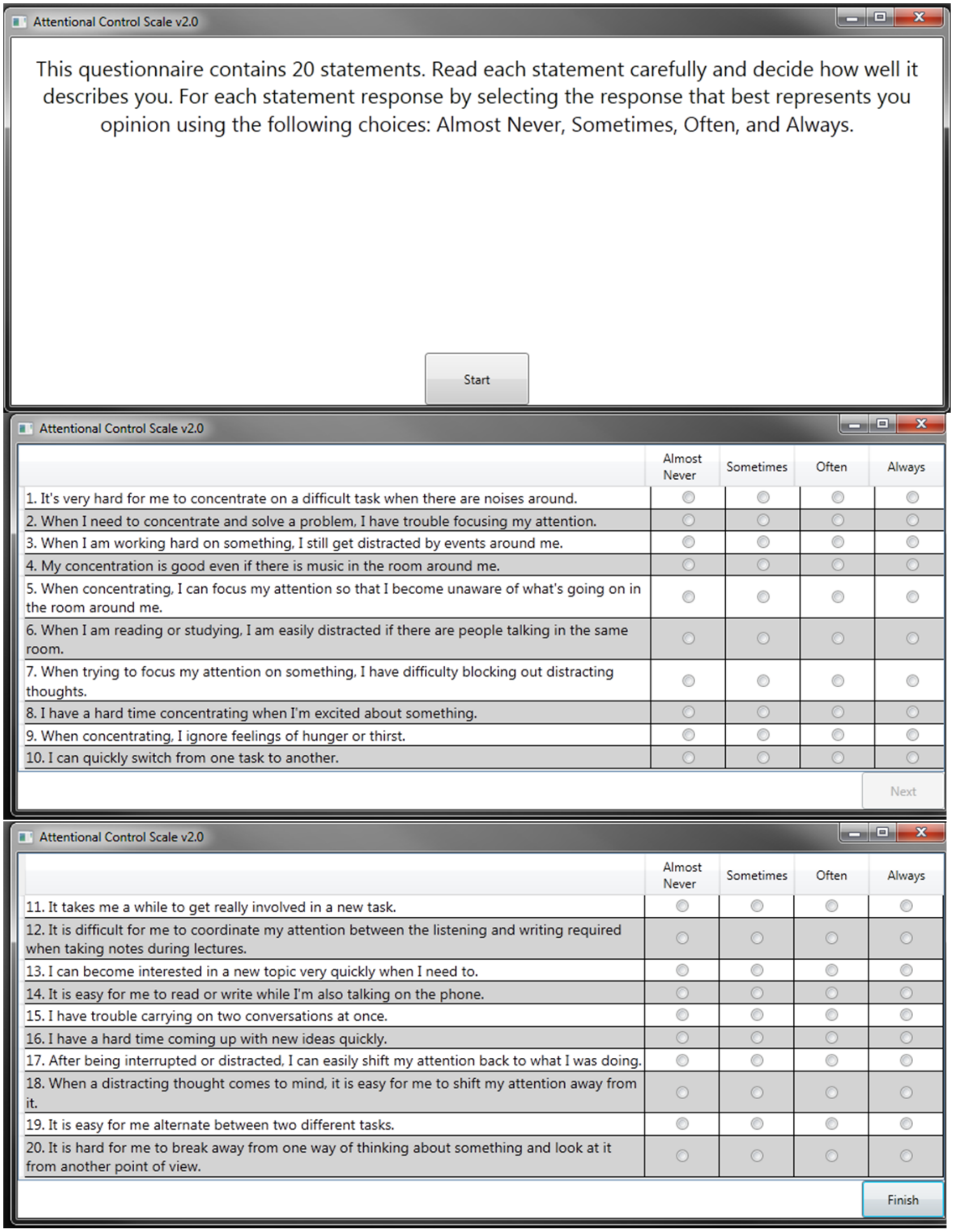
Post Trial Questionnaire

Final Questionnaire

Personality Questionnaire

	1 Extremely Inaccurate	2 Very Inaccurate	3 Moderately Inaccurate	4 Slightly Inaccurate	5 Neither Inaccurate Nor Accurate	6 Slightly Accurate	7 Moderately Accurate	8 Very Accurate	9 Extremely Accurate
Bashful	●	●	●	●	●	●	●	●	●
Bold	●	●	●	●	●	●	●	●	●
Careless	●	●	●	●	●	●	●	●	●
Cold	●	●	●	●	●	●	●	●	●
Complex	●	●	●	●	●	●	●	●	●
Cooperative	●	●	●	●	●	●	●	●	●
Creative	●	●	●	●	●	●	●	●	●
Deep	●	●	●	●	●	●	●	●	●
Disorganized	●	●	●	●	●	●	●	●	●
Efficient	●	●	●	●	●	●	●	●	●
Energetic	●	●	●	●	●	●	●	●	●
Envious	●	●	●	●	●	●	●	●	●
Extraverted	●	●	●	●	●	●	●	●	●
Fretful	●	●	●	●	●	●	●	●	●
Harsh	●	●	●	●	●	●	●	●	●
Imaginative	●	●	●	●	●	●	●	●	●
Inefficient	●	●	●	●	●	●	●	●	●
Intellectual	●	●	●	●	●	●	●	●	●
Jealous	●	●	●	●	●	●	●	●	●
Kind	●	●	●	●	●	●	●	●	●
Moody	●	●	●	●	●	●	●	●	●
Organized	●	●	●	●	●	●	●	●	●
Philosophical	●	●	●	●	●	●	●	●	●
Practical	●	●	●	●	●	●	●	●	●
Quiet	●	●	●	●	●	●	●	●	●
Relaxed	●	●	●	●	●	●	●	●	●
Rude	●	●	●	●	●	●	●	●	●
Shy	●	●	●	●	●	●	●	●	●
Sloppy	●	●	●	●	●	●	●	●	●
Sympathetic	●	●	●	●	●	●	●	●	●
Systematic	●	●	●	●	●	●	●	●	●
Talkative	●	●	●	●	●	●	●	●	●
Temperamental	●	●	●	●	●	●	●	●	●
Touchy	●	●	●	●	●	●	●	●	●
Uncreative	●	●	●	●	●	●	●	●	●
Unenvious	●	●	●	●	●	●	●	●	●
Unintellectual	●	●	●	●	●	●	●	●	●
Unsympathetic	●	●	●	●	●	●	●	●	●
Warm	●	●	●	●	●	●	●	●	●
Withdrawn	●	●	●	●	●	●	●	●	●

Attention Control Questionnaire


Attentional Control Scale v2.0

This questionnaire contains 20 statements. Read each statement carefully and decide how well it describes you. For each statement response by selecting the response that best represents you opinion using the following choices: Almost Never, Sometimes, Often, and Always.

Start

Attentional Control Scale v2.0

	Almost Never	Sometimes	Often	Always
1. It's very hard for me to concentrate on a difficult task when there are noises around.				
2. When I need to concentrate and solve a problem, I have trouble focusing my attention.				
3. When I am working hard on something, I still get distracted by events around me.				
4. My concentration is good even if there is music in the room around me.				
5. When concentrating, I can focus my attention so that I become unaware of what's going on in the room around me.				
6. When I am reading or studying, I am easily distracted if there are people talking in the same room.				
7. When trying to focus my attention on something, I have difficulty blocking out distracting thoughts.				
8. I have a hard time concentrating when I'm excited about something.				
9. When concentrating, I ignore feelings of hunger or thirst.				
10. I can quickly switch from one task to another.				

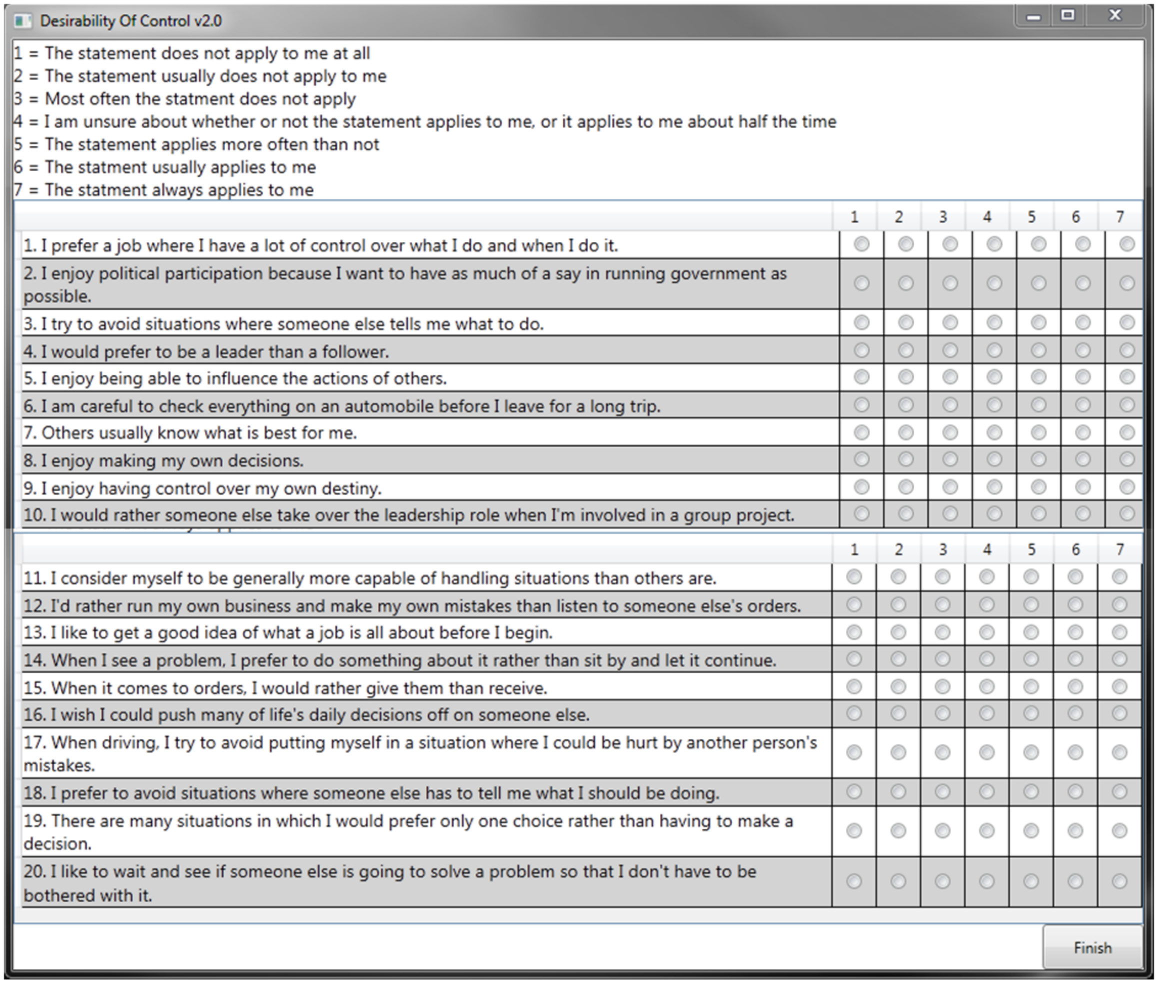
Next

Attentional Control Scale v2.0

	Almost Never	Sometimes	Often	Always
11. It takes me a while to get really involved in a new task.				
12. It is difficult for me to coordinate my attention between the listening and writing required when taking notes during lectures.				
13. I can become interested in a new topic very quickly when I need to.				
14. It is easy for me to read or write while I'm also talking on the phone.				
15. I have trouble carrying on two conversations at once.				
16. I have a hard time coming up with new ideas quickly.				
17. After being interrupted or distracted, I can easily shift my attention back to what I was doing.				
18. When a distracting thought comes to mind, it is easy for me to shift my attention away from it.				
19. It is easy for me alternate between two different tasks.				
20. It is hard for me to break away from one way of thinking about something and look at it from another point of view.				

Finish

Desirability of Control Questionnaire



Desirability Of Control v2.0

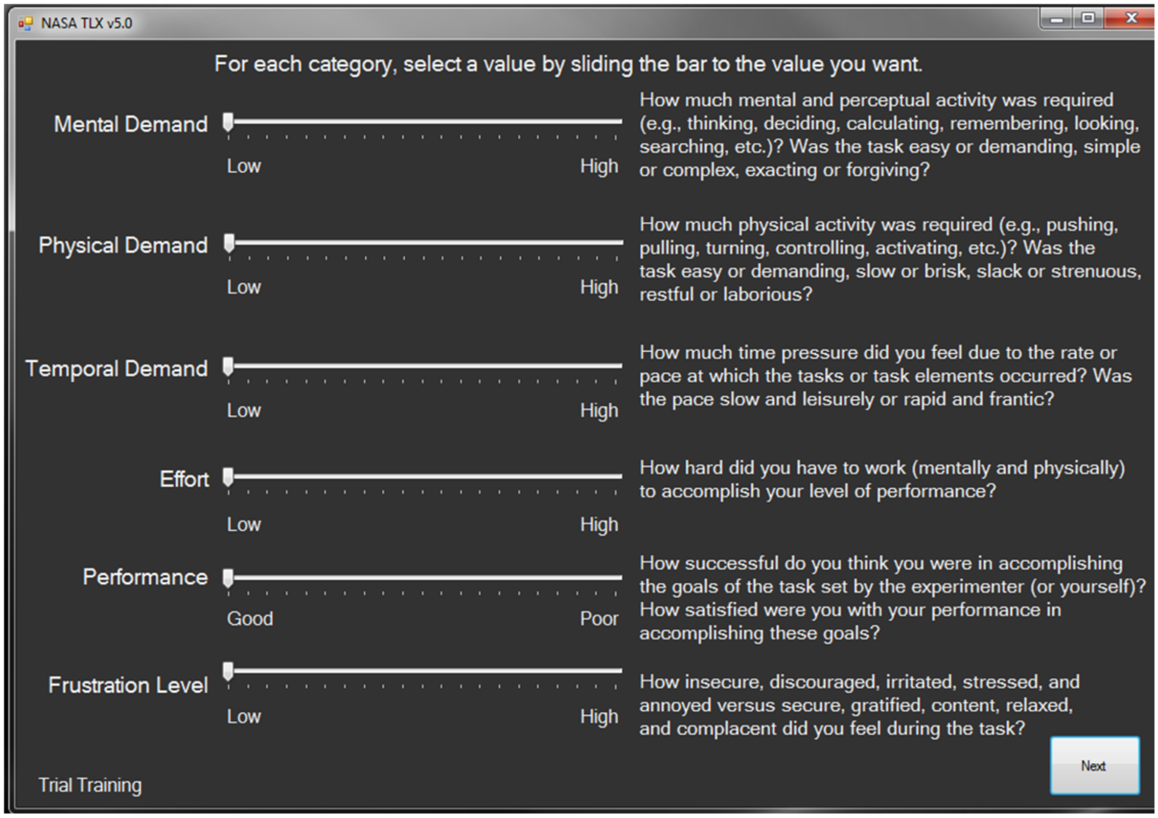
1 = The statement does not apply to me at all
2 = The statement usually does not apply to me
3 = Most often the statment does not apply
4 = I am unsure about whether or not the statement applies to me, or it applies to me about half the time
5 = The statement applies more often than not
6 = The statment usually applies to me
7 = The statment always applies to me

	1	2	3	4	5	6	7
1. I prefer a job where I have a lot of control over what I do and when I do it.							
2. I enjoy political participation because I want to have as much of a say in running government as possible.							
3. I try to avoid situations where someone else tells me what to do.							
4. I would prefer to be a leader than a follower.							
5. I enjoy being able to influence the actions of others.							
6. I am careful to check everything on an automobile before I leave for a long trip.							
7. Others usually know what is best for me.							
8. I enjoy making my own decisions.							
9. I enjoy having control over my own destiny.							
10. I would rather someone else take over the leadership role when I'm involved in a group project.							

	1	2	3	4	5	6	7
11. I consider myself to be generally more capable of handling situations than others are.							
12. I'd rather run my own business and make my own mistakes than listen to someone else's orders.							
13. I like to get a good idea of what a job is all about before I begin.							
14. When I see a problem, I prefer to do something about it rather than sit by and let it continue.							
15. When it comes to orders, I would rather give them than receive.							
16. I wish I could push many of life's daily decisions off on someone else.							
17. When driving, I try to avoid putting myself in a situation where I could be hurt by another person's mistakes.							
18. I prefer to avoid situations where someone else has to tell me what I should be doing.							
19. There are many situations in which I would prefer only one choice rather than having to make a decision.							
20. I like to wait and see if someone else is going to solve a problem so that I don't have to be bothered with it.							

Finish

NASA-TLX Questionnaire



ALOA-A5 POST TRIAL QUESTIONNAIRE ID_______ TRIAL_______ DATE_______

Please CIRCLE one answer to each of the following questions, giving your impression FOR ONLY THE TRIAL JUST COMPLETED

1	Rate ***how difficult*** it was to complete all tasks	Very Easy	Easy	Neither Easy nor Difficult	Difficult	Very Difficult
2	Provide a ***workload rating*** that represents your workload for this trial	Bored	Somewhat Busy	Busy	Very Busy	Overloaded
3	Rate your ***level of confidence*** in your decision making abilities for the ***image task***	Very Little Confidence	Low Confidence	Moderate Confidence	High Confidence	Very High Confidence
4	Rate your ***level of confidence*** in the automation's decision making abilities for the ***image task***	Very Little Confidence	Low Confidence	Moderate Confidence	High Confidence	Very High Confidence
5	To what extent did you ***trust*** the automation	Very Little Trust	Low trust	Some Trust	High Trust	Very High trust
6	Rate how often you were ***surprised*** by the actions of the automation	Never	Seldom	Occasionally	Often	Always
7	Did you notice the ***automation level change*** for the ***image analysis task***? If 'NO' stop here	NO				YES
8	Rate the adequacy of the ***system*** in giving you ***feedback*** on which automation level was currently in effect	Unacceptable	Bad	Satisfactory	Good	Optimum
9	How did having the ***automation level*** of the image analysis task ***tied to your performance*** impact performance on the ***image analysis task***?	Strongly Hurt Performance	Hurt Performance	No Impact	Aided Performance	Strongly Aided Performance
10	How did having the ***automation level*** of the image analysis task ***tied to your performance*** affect completion of ***all other tasks***?	Great Disadvantage	Slight Disadvantage	No Impact	Slight Advantage	Great Advantage
11	Rate how much ***attention*** you had to pay to the changing of the ***levels of automation***?	None	Very Little	Some	Quite A Bit	A Lot

COMMENTS:

ALOA-A5 FINAL QUESTIONNAIRE ID________ TRIAL________ DATE__________

In this experiment, the system tracked your performance on several tasks to determine if you were overloaded or not. If the system detected that you were overloaded then is would change from LOA 1 (options 1 through 8) to LOA 2 (options 1 through 8 with a highlighted suggestion). If the system still detected you were overloaded, the automation change to LOA 3 and only presented one answer for you to accept or reject. If the system detected that you were under loaded, the image analysis automation level changed to a lower automation level that enabled you to be more involved in the task.

Please CIRCLE one answer to each of the following questions, giving your impression FOR ALL TRIALS COMPLETED

1	Rate ***how frequently you observed*** the automation level change	Never (stop & tell experimenter)	Seldom	Occasionally	Often	A Lot
2	Rate your opinion of how ***frequently the automation level changed***	Insufficient/ Not Sensitive	Slightly Insufficient	About Right	Slightly Excessive	Excessive/Too Sensitive
3	Rate your ***ability to complete the image analysis task***	Unacceptable	Bad	Satisfactory	Good	Optimum
4	Rate your ability to ***accomplish all other tasks*** (red plane, mission planning, health, and chat)	Unacceptable	Bad	Satisfactory	Good	Optimum
5	Rate your ability to maintain ***situational awareness*** (degree you were aware of important elements in the environment)	Never	Seldom	Occasionally	Often	Always
6	Rate your ***overall mental workload***	Bored	Somewhat Busy	Busy	Very Busy	Overloaded
7	Of the three ***levels of automation*** (LOA 1, 2,3), indicate ***your preference*** to use for the majority of the trial	I don't like any of the levels of automation	No preference	Prefer LOA 1: 8 options shown	Prefer LOA 2: 8 options shown, one recommended	Prefer LOA 3: one option you could consent or veto

COMMENTS:

Final Questionnaire (page 1 of 2)

ALOA-A5 FINAL QUESTIONNAIRE ID__________ TRIAL__________ DATE__________

8 Did you ever wish the automation level changed sooner than it did?

Yes_______ No_______ *If YES, please explain below

9 Do you feel the change in automation level matched your performance?

Yes_______ No_______ *If NO, please explain how the automation differed

10 Please provide any additional comments concerning the experiment; training, tasks, and/or simulator you might have (include things you liked, things that were confusing, etc.)

Final Questionnaire (page 2 of 2)

Bibliography

Army Science Board (2004). Autonomy Levels for Unmanned Systems Workshop. Huntsville, AL.

Billings, C. E. (1997). Aviation automation: The search for a human centered approach. Mahwah, NJ: Erlbaum.

Billings, C. E., Lauber, J. K, Funkhouser, H., Lyman, G., & Huff, E M. (1976). NASA Aviation Safety Reporting System (Tech. Rep. TM-X-3445) Moffeit Field, CA: NASA Ames Research Center.

Billings, C. E. & Woods, D. D. (1994). Concerns about adaptive automation in aviation systems. In M. Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: Current research and trends (pp. 24–29). Hillsdale, NJ: Erlbaum.

Calhoun, G.L., Ruff, H.A., Spriggs, S., & Murray, C. (2012). Tailored performance-based adaptive levels of automation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 56, 1: pp. 413-417.

Calhoun, G.L., Ward, V.B.R., & Ruff, H.A. (2011). Performance-based adaptive automation for supervisory control. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 55, 1: pp. 2059-2063.

Calefato, C., Montanari, R., & Tesauri, F. (2008). The adaptive automation design. In K. Asai (Ed.), *Human computer interaction: New developments*. Available from: http://www.intechopen.com/books/human_computer_interaction_new_developments/the_adaptive_automation_design

Chen, J., Barnes, M., & Harper-Sciarini, M. (2011). Supervisory control of multiple robots: Human performance issues and user interface design. *IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews*, 41(4), 435-454.

Cosenzo, K., Chen, J., Reinerman-Jones, L., Barnes, M., & Nicholson, D. (2010). Adaptive automation effects on operator performance during a reconnaissance mission with an unmanned ground vehicle. *Proceedings of the Human Factors & Ergonomics Society*, pp. 2135-2139.

Cummings, M. (2004), “The need for command and control instant message adaptive interfaces: Lessons learned from Tactical Tomahawk human-in-the-loop simulations,” *CyberPsycholo. Behav*., vol. 7, no. 6, pp. 653-661.

Cummings, M., Bruni, S., Mercier, S., & Mitchell, P. (2007). Automation architecture for single operator-multiple UAV command and control. *International Command and Control Journal*, 1(2), 1–24.

de Greef, T., Arciszewski, H., & Neerincx, M. (2010). Adaptive automation based on an object-oriented task model: Implementation and evaluation in a realistic c2 environment. *Journal of Cognitive Engineering and Decision Making*, *4*(2), 152-182. doi: 10.1518/155534310X522851

de Visser, E.J., LeGoullon, M., Freedy, A., Freedy, E., Weltman, G. & Parasuraman, R. (2008). Designing an adaptive automation system for human supervision of unmanned vehicles: A bridge from theory to practice. In Proceedings of the *HFES* pp. 221-225.

Draper, M., Barry, T., Calhoun, G., Clark, J., Goodrich, M., Jansen, C., Kessens, J., Kooi, F., Lefebvre, A., Murray, S., Nelson, J., Nielsen, C., Osga, G., Oudenhuijzen, A., Quigley, M., Shively, R., Simpson, B., Stone, R., van Breda, L., van Delft, J., & van Erp, J. (2007). Advanced UMV operator interfaces in Uninhabitied military vehicles (UMVs): Human factors issue in augmenting the force (Technical Report TR-HFM-078), North Atlantic Treaty Organisation.

Eggers, J.W. & Draper, M.H.. (2006). Multi-UAV control for tactical reconnaissance and close air support missions: Operator perspectives and design challenges. Proceedings of the *NATO RTO Human Factors and Medicine Panel Symposium* HFM-3-135, NATO TRO.

Endsley, M. (1996), Automation and situation awareness, in R. Parasuraman and M. Mouloua (eds), Automation and human performance: Theory and applications (Mahwah, NJ: Lawrence Erlbaum Associates), 163–181.

Endsley, M. & Kaber D. (1999a). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492.]

Endsley, M. & Kaber, D. (1999b). Level of automation of a dynamic control task based on secondary task workload measurements. *International Journal of Cognitive Ergonamics*, 42(3), 462-492.

Endsley, M. & Kiris, E. (1994). The out-of-the-loop performance problem: Impact of level of automation and situation awareness. In M. Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: Current research and trends (pp. 50-56). Hillsdale, NJ: Lawrence Erlbaum Associates.

Feigh, K., Dorneich, M., & Hayes, C. (2012). "Towards a characterization of adaptive systems: A framework for researchers and system designers", *Journal of Human Factors and Ergonomics*. Vol. 54, No. 6, 1008-1024.

Fischhoff, B., Slovic, P., & Lichtenstein, S., (1977) “Knowing with certainty: The appropriateness of extreme confidence,” *Journal of Experimental Psychology: Human perception performance*, volume 3, 552-564.

Fitts, (1951). Human Engineering for an Effective Air Navigation in Traffic Control System. Washington D.C.: National Research Council.

Hart, S.G. & Staveland, L.E. (1988). Development of a NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & NH. Meshkati (Eds.), Human mental workload. Amsterdam: New Holland.

Johnson, R., Leen, M., & Goldberg, D. (2007). Testing adaptive levels of automation (ALOA) for UAV supervisory control (Technical Report AFRL-HE-WP-TR-2007-0068), Air Force Research Laboratory.

Kaber, D. B., Wright, M. C., Prinzel, L. J., Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human Factors*, 47(4), 730-741.

Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Brooks/Cole Publishing Company.

Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors*, 49, 57–75.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*, 6(1), 44-58. doi: 0.1037//0278-7393.6.1.44

Morrison, J. G., Cohen, D., & Gluckman, J. P. (1993). Prospective principles and guidelines for the design of adaptively automated crewstations. In J.G. Morrison (Ed.), The adaptive function allocation for intelligent cockpits (AFAIC) program: Interim research and guidelines for the application of adaptive automation (Tech. Report No. NAWCADWAR-93931-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.

Mumaw, R. J., Sarter, N. B., & C. D., Wickens, “Analysis of pilots’ monitoring and performance on an automated flight deck,” in Proc. 11th *Int. Symp. Aviation Psychol*. Columbus, OH: Ohio State University, 2011.

Muthard, E. K. & Wickens C. D., “Factors that mediate flight plan monitoring and errors in plan revision: An examination of planning under automated conditions,” University of Illinois, Urbana-Champaign, IL, Technical Report AFHD-02-11/NASA-02-8, 2002.

Parasuraman, R., Bahri, T., Deaton, J., Morrison, J., & Barnes, M. (1992). Theory and design of adaptive automation in aviation systems (Technical Report No. NAWCADWAR-92033-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.

Parasuraman, R., Barnes, M., & Cosenzo, K. (2007). Adaptive automation for human-robot teaming in future command and control systems. *The International C2 Journal*, 1(2), 43-68.

Parasuraman, R., Molloy, R., & Singh, I. (1993). Performance consequences of automation-induced "complacency. *The International Journal of Aviation Psychology*, *3*(1), 1-23.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, *38*(4), 665-679.

Parasuraman, R. & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors*, 50, 511–520.

Prinzel, K. (2003). Team-centered perspective for adaptive automation design (NASA-TM-2003-212154). Hanover, MD.: NASA Center for AeroSpace Information

Reising, J. (2002). Uninhabited military vehicles: What is the role of the operator (Technical Report RTO-MP-088), Air Force Research Lab.

Rouse, W. B. (1988). Adaptive aiding for human/computer control, in *Human Factors*, Vol. 30, 431-438.

Ruff, H., Narayanan, S., & Draper, M. (2002). *Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles*. (Master's thesis).

Scerbo, M. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), Automation and human performance: Theory and applications, 37-63. Mahwah, NJ: Erlbaum.

Sheridan, T.B. (2002). Humans and Automation: System Design and Research Issues. John Wiley &Sons, in cooperation with *HFES*: Santa Monica, CA.

Sheridan, T.B., & Parasuraman, R. (2006). Human-automation interaction. In R.S. Nickerson (Ed.), Reviews of human factors and ergonomics (Volume 1, Chapter 2, pp. 89-129)

Sheridan, T.B., & Verplank, W.L. (1978). *Human and Computer Control of Undersea Teleoperators* (Man- Machine Systems Laboratory Report). Cambridge: MIT.

Szalma, J.L., & Taylor, G.S. (2011). Individual differences in response to automation: The five factor model of personality. *Journal of Experimental Psychology: Applied*, 17(2), 71-96

Wickens, C. D., Mavor, A., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control: Human operators and automation*. Washington, DC: National Academy Press.

Vita

Name: Crystal A. Miller

Grade or Rank: First Lieutenant

Current Position Title: Lead Research Engineer

Location: 711 HPW / RHCI, Wright-Patterson Air Force Base OH

Education: B.S. Human Factors Psychology United States Air Force Academy, 2009

Publications:

Calhoun, G., Ruff, H., Miller, C., Murray, C., Hamell, J., Barry, T., & Draper, M. (2012, March). Flexible levels of execution-interface technologies (FLEX-IT) for future remotely piloted aircraft control applications. (Report No. AFRL-RH-WP-TR-2012-0077). Wright-Patterson Air Force Base, OH: Air Force Research Laboratory.

Calhoun, G.L., Ruff, H.A., Spriggs, S., & Murray, C. (2012). Tailored performance-based adaptive levels of automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 56, 1: pp. 413-417.

Ruff, H.A., Calhoun, G.L., Miller, C.A., & Breeden, C.R. (2013). Adaptive automation for multi-vehicle supervisory control: Impact of changing automation levels. *Proceedings of the 17th International Symposium on Aviation Psychology*. Dayton, OH: Wright State University.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*
13-06-2013

2. REPORT TYPE
Master's Thesis

3. DATES COVERED *(From – To)*
26 Mar 2012 - 13 Jun 2013

4. TITLE AND SUBTITLE
Evaluation of an Adaptive Automation Trigger Based on Task Performance, Priority, and Frequency

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Miller, Crystal A., First Lieutenant, USAF

5d. PROJECT NUMBER
7184

5e. TASK NUMBER
09

5f. WORK UNIT NUMBER
71840918

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way, Building 640
WPAFB OH 45433

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENV-13-J-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
711th HUMAN PERFORMANCE WING RHCI
Gloria Calhoun Gloria.calhoun@wpafb.af.mil
2210 8th Street, Building 146
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7334
Commercial: (937) 255-3856

10. SPONSOR/MONITOR'S ACRONYM(S)
711th HPW/RHCI

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT
A significant Air Force research thrust includes increasing use of automation, creating the potential for the operators to become complacent and over-reliant on automation. To avoid operator complacency, adaptive automation has been proposed, where changes in automation are triggered based upon operator performance or other attributes. This research sought to understand the effect of a weighted method for triggering changes in automation as compared method weighting all tasks equally. In this work, the weighted method considered operator performance, priority, and frequency of each task when computing a measure on which to trigger changes in automation. Although overall system performance was not statistically different between the two system implementations, the participants with the priority based triggering scheme tended to rate the LOA changes as more aligned with their actual performance and were significantly less surprised by the actions of the automation than those participants with the non-weighted approach. The results of this study, combined with participant preference for workload based adaptations, suggest a benefit to the implementation of a hybrid approach. Future research could focus on task weights based on priority and operator specific threshold criteria, where automation aides are triggered once the summation of current tasks exceeds the given threshold.

15. SUBJECT TERMS
Performance Based Adaptive Automation; Task Workload; LOA

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	102	Miller, Michael E., PhD (937) 255-6565, x 4651 (Michael.Miller@afit.edu)